LEARNING CURVES AND THEIR APPLICABILITY TO UNIT TRAINING LEVELS IN OPERATIONAL TESTING

A THESIS

Presented to

The Faculty of the Division of Graduate Studies

By

Jesse Lee Brokenburr

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Operations Research

Georgia Institute of Technology

June, 1978

LEARNING CURVES AND THEIR APPLICABILITY TO

UNIT TRAINING LEVELS IN OPERATIONAL TESTING

Russell G. Heikes, Chairman

Thomas L. Sadosky

Leslie G. Callahan, Jr.

Harrison M. Wadsworth

Date approved by Chairman 2 June 78

ACKNOWLEDGMENTS

A number of persons have assisted me in the conduct of this research. To all of them I express my appreciation. In particular, I would like to thank the following individuals:

Dr. Russell G. Heikes, my thesis advisor, has been the principal source of guidance and encouragement throughout this research.

Dr. Leslie G. Callahan, Dr. Thomas L. Sadosky, and Dr. Harrison Wadsworth, members of my reading committee, have provided many significant comments and useful suggestions.

The United States Army Operational Test and Evaluation Agency has provided both financial and technical support which helped greatly.

Finally, a special note of thanks goes to my family for their endless encouragement from the beginning to the completion of this effort.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

This research addresses the problem of determining the existence of a representative group/crew learning curve (or set of curves) and the development of a mathematical description of this curve applicable to training levels in operational testing. Emphasis is placed on the analysis of data from actual operational test reports.

An iterative procedure is developed to analyze sample data using regression techniques to screen data for suitability and to fit nonlinear learning models.

A representative learning curve for the data analyzed is selected by comparing the sum of squares regression and the lack of fit ratio for each model.

This comparison shows that the following models appeared to provide an adequate fit to the data analyzed.

(1) $\hat{Y} = at^{-b}$

(2) $\hat{Y} = a[\beta + (1-\beta)t^{-b}]$

(3) $\hat{Y} = at^{-b} + C$

(4) $\hat{Y} = ae^{bt}$

Since the variations of the power function, models (2) and (3) did not appear to provide a better fit to the data, model (1) was preferred from the standpoint of parsimony. It cannot be stated conclusively that model (1) provides a statistically better fit to the data than model (4). However, based on a survey of industrial applications of the power function model as reported in the literature, it was concluded

that the model $\hat{Y} = at^{-b}$ does adequately fit the empirical data analyzed and can be used as a representative group/crew learning model for this data.

CHAPTER I

INTRODUCTION

Background

The initial direction for this study was provided in a research task statement by the U.S. Army Operational Test and Evaluation Agency (OTEA).

> Conduct background research, including literature search covering both government publications and the general literature and field visits as appropriate to identify a general case learning curve (or set of curves, if necessary) existing in current test data; to describe this curve (or curves) mathematically in a manner such that the slope (first derivative) can be derived; to present evidence in support of the validity of such curves; and, to prepare a set of instructions explaining how to design a test to generate the needed data and then treat the data to record the curves.

OTEA is continually required to assess the impact of the training level of a crew or unit engaged in operational tests. This assessment is of particular importance because OTEA has the mission of assisting in the planning, directing, and evaluation of operational testing required during the materiel requisition process of all major systems and selected non major systems. Adequate and thorough operational testing is essential in determining an item or system's operational suitability and logistic support requirements (1,2).

Operational Testing (OT) is conducted in the most realistic test environment possible and utilizes the most representative configuration of the future operational system. Because operational testing is conducted throughout the development life cycle of materiel, it is usually begun using early prototypes and continues through the cycle by using production models.

To enhance the validity of generated test data, operational testing must be conducted by troop units, support personnel, and individuals who will actually be issued the materiel for use.

Through these tests a comparison is made between new materiel and existing equipment being operated under the same or similar mission profile. This testing concept greatly assists decision makers to accurately assess total operational suitability from a doctrinal, organizational and tactical viewpoint, and to collect performance and reliability, availability, and maintainability data that closely simulates that which would be experienced after the materiel is issued to the field. Results of testing are forwarded through channels to the Army Systems Acquisition Review Council (ASARC), with final decision of acceptance or rejection resting with the Secretary of Defense (3,4,5).

Essentially, the assessment of crew or unit training levels has traditionally been limited to qualitative techniques such as administering a proposed training program (with the assumption that the completed training equals a given training level) relying on ARMY TRAINING AND EVALUATION PROGRAM (ARTEP) results, or using military judgement. Training data is currently overwhelmingly qualitative, where as quantitative data is much to be preferred in operational test and evaluation.

It is generally agreed that a performance curve describing the progress of training is an asymptotic "learning curve". Assuming this, it should be possible to use the slope of a curve as a measure of how closely a unit has approached the asymptote. The slope of a curve may be expressed mathematically and can be treated rigorously. However, even though it is generally accepted that the individual "learning curve" follows this assumption and appears to be robust, it cannot be assumed that a representative "learning curve" for a crew or unit has these same properties.

Objective, Procedure, and Scope

Since operational testing usually involves the comparison of baseline systems to newly developed systems, participants are initially determined to be qualified or trained on the baseline system. Prior to the actual conduct of the test, refresher training and/or contractor training is provided on the new system. Through the use of randomization and test design the effect of learning during the test is generally expected to be lessened.

The objective of this study is to determine the existence of a representative learning curve (or set of curves) and develop a mathematical description of this curve applicable to training levels in operational testing.

This research involves an "after the fact" analysis of data from various test reports. Empirical data was collected, primarily from OTEA test reports and data made available through other training and analysis agencies. A more detailed description of the various data collected is provided in Chapter IV. The data obtained was plotted using consecutive trials versus a specified performance measure/measure of effectiveness (MOE) in order to determine if there were patterns which might suggest a demonstrable group "learning curve".

Linear regression models are used to screen sample data for suitability and further analysis, while nonlinear regression models are used to fit learning models to the sample data. Additionally, the fitted learning models will be tested for adequacy through a direct examination of residuals.

The scope of this research is concentrated on the analysis of data obtained from a military operational testing environment in which OTEA operates. A survey of the general literature is conducted to determine the existence of appropriate industrial studies of group or team learning which might support this study.

The initial background search involves the theory of learning along with the use and development of learning curves. This particular aspect is expanded to include group or team performance (learning models discussed in Chapter II).

The remainder of the study involves development of the methodology employed, a description of data collected, and a discussion of results including appropriate recommendations and conclusions.

CHAPTER II

REVIEW OF APPLICABLE LEARNING THEORY RESULTS

This chapter contains a review of general learning theory and the development of learning progress or performance improvement. It further summarizes the application of learning theory concepts to group/team learning.

Learning Theory

Learning is a fundamental process of life. Every individual learns and through learning develops modes of behavior by which he lives. Learning may occur intentionally, through organized or unorganized activity, and the variables which influence learning may be grouped under the three headings: (1) individual variables, such as capacity and motivation; (2) task variables, such as meaningfulness and difficulty; and (3) environmental variables, such as practice and knowledge of results (6).

The learning phenomenon has been studied by philosophers and psychologists for centuries. In fact Aristotle was the first to set forth laws in an attempt to explain the basis of learning (7).

In Mednick's book (7,8), learning has been defined in terms of four characteristics. These are:

1. Learning results in a behavioral change. This characteristic is the basic goal of any efforts at learning.

2. Learning is a result of practice. This eliminates behavioral changes due to illness, maturation, or motivation.

Although performance may be greatly altered by these variables, learning is not.

3. Learning is a relatively permanent change. A task which was learned sometime in the past can be easily resumed after a little practice.

4. Learning is not directly observable. Performance is affected by variables other than learning. Therefore, a record of successive performance is just that, and cannot be considered an exact representation of the learning process.

Mathematical Models

In order to measure learning or compute the rate of learning, mathematical models were developed. Experiments in learning phenomena are generally concerned with changes in some evidence of learning as a result of experiences on discrete trials. In most paired-associate learning paradigms (models) the subject's knowledge is tested after every exposure to the correct pairing (9). When a number (whether it be a probability value between 0 and 1, or some integer value) changes as a result of discrete opportunities, we are more likely to find more accurate mathematical analogies in difference equations than in differential equations. But difference equations were not known to psychologists until the late 1940's and early 1950's.

Clark L. Hull (10) is sometimes considered the first mathematical learning theorist, although there are other, earlier, quantitatively oriented theorists (9). The genesis of Hull's model was different from that of current models, and the difference is a critical one. The major mathematical technique used by Hull and his contemporaries was curve fitting. For Hull this meant a somewhat arbitrary selection of one from the many equations whose form would be compatible with previously obtained data. Theory dictated the selection of variables for his equations, but the precise forms of the equations were derived primarily out of attempts to fit past data. With the new quantitative techniques that have become available, it is now possible to permit the theory to imply the equation form directly, prior to data collection.

The capacity to derive equations from theory, and to see how these theoretically derived equations conform to data patterns, is what is meant by a true analogy between theory building in psychology and theory building in the physical sciences.

A further change from the past in learning theory that appears to be fairly general in more recent theory building is the abandonment of the belief in a general learning function that should cover all learning situations. More recent thinking recognizes that different theories, and therfore different mathematical functions, might be required for different learning situations. The earlier work assumed that a finding in one laboratory, stemming from one experimental paradigm, could contradict the theory of another experimenter using a different paradigm, with all assumed to be exploring a similar process.

Learning Curves

When several trials are given in an experiment and measures of learning or of retention are obtained, these measures may be plotted in the graphic form known as a learning curve, a graph which affords a comparison of the performance on each trial with a performance on other trials (6). It is customary to plot the independent variable on the horizontal axis, the abscissa, and the dependent variable on the vertical axis, the ordinate. The dependent variable changes as a result of the experimenter's manipulations. Scores on the dependent variable are dependent upon or are the function of the experimental factor and are usually some form of a learning score - error made, time consumed, and so on.

One of the things a learning curve reveals is the rate of improvement and the changes in this rate. A uniform rate of improvement is indicated by graphs of the type shown in Figure 2-1.

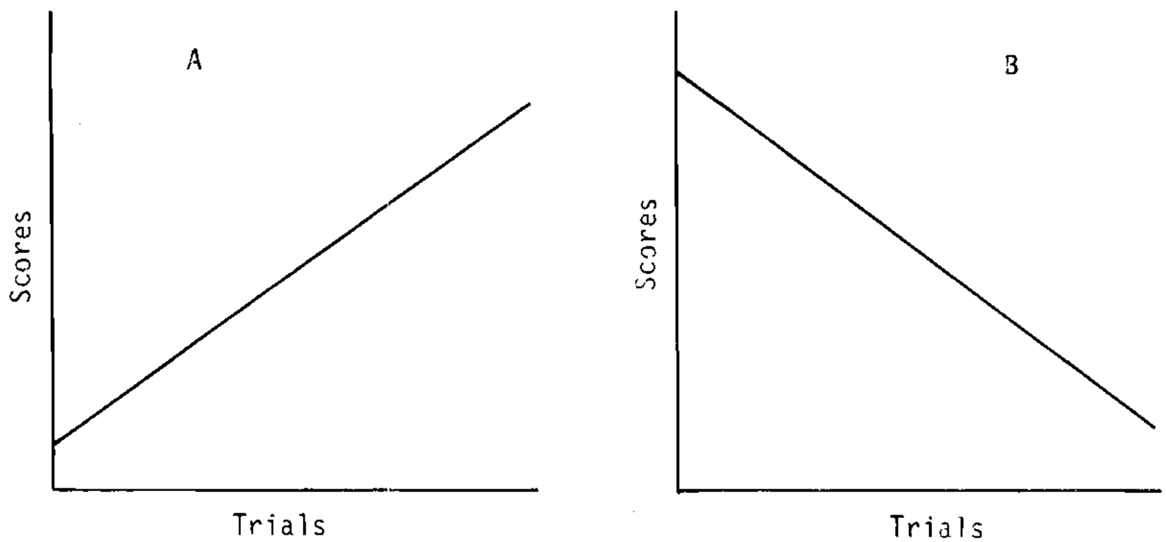


Figure 2-1. Theoretical learning curves showing zero acceleration, or a uniform rate of improvement. In A improvement is shown by an increase in scores. B depicts those learning situations wherein decreasing scores indicate improvement, such as fewer errors. (6)

Here progress is indicated by a straight line. Such a graph means that the increment of gain is the same for each successive trial. When the rate of improvement is constant, we have what is known as zero acceleration.

Most curves of learning show variations in the rate of improvement. Curves for motor learning usually show the fastest rate of gain at the beginning and a slowing up as practice continues. Such a change is called negative acceleration.

The authors, Garry and Kingsley, state that this should not be confused with a loss of skill. It refers to those cases wherein improvement is still being made, but the increment of gain is smaller on each successive trial. Theoretical curves for negative acceleration are presented in Figure 2-2.

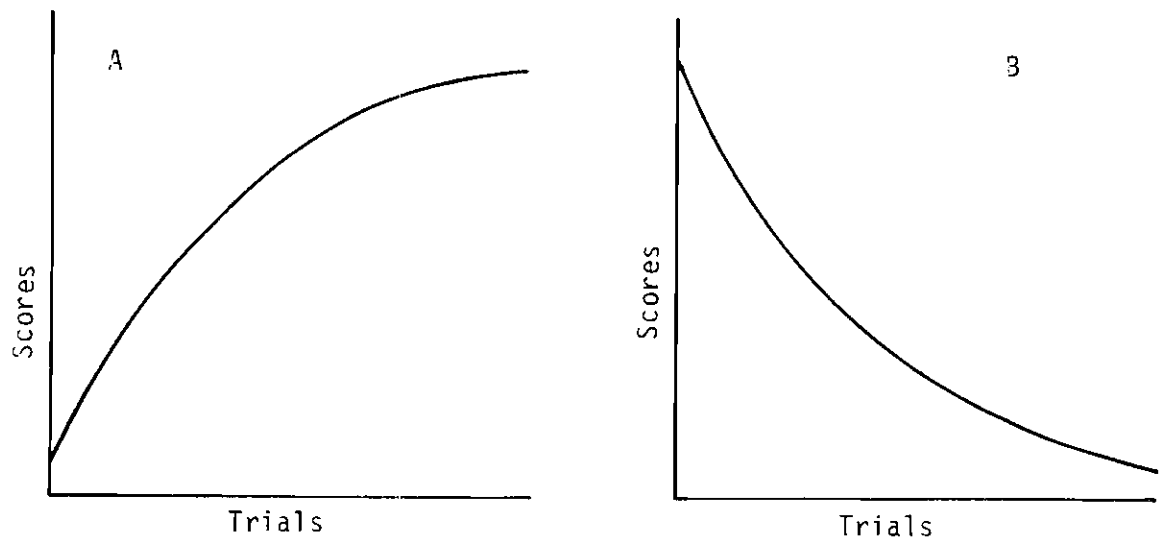


Figure 2-2. Theoretical curves of negative acceleration showing a decrease in the rate of gain. (6)

In the cases in which the scores grow smaller (time scores or error scores on successive trials) as performance improves, negative acceleration is indicated by a downward concave curve. Negatively accelerated curves are most frequently obtained in situations where

(1) the learning task is relatively simple,

(2) the subjects are of average or above ability (either practiced or bright),

(3) there is positive transfer from previous learning, or

(4) the tests are given toward the end of a series of trials.

Sometimes there is very slow progress at the start, with an increase in the increments of improvement as practice is continued. This increase in the rate of improvement is called positive acceleration, see Figure 2-3.

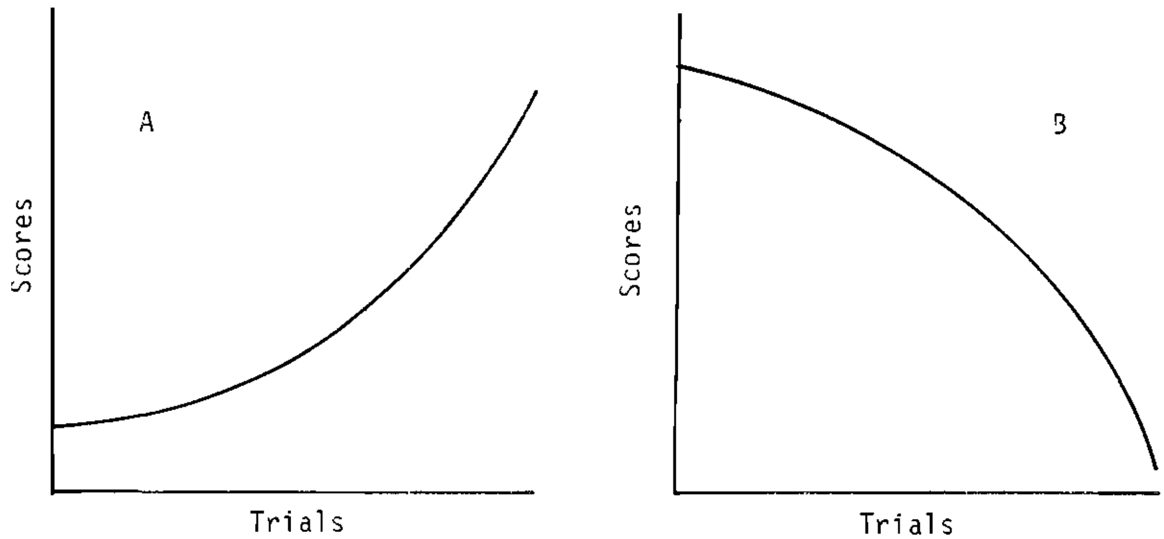


Figure 2-3. Two theoretical curves of positive acceleration. In both, the rate of improvement is faster in the second half of the learning period than in the first part (5)

Curves of positive acceleration are frequently found in motor learning or where previous learning interferes with the new learning. It is clear that positive acceleration cannot continue indefinitely, for sooner or later the learner reaches complete mastery or the curve levels off as he approaches the limit of his ability to improve (6).

It is likely that if we were able to plot a complete learning curve from zero to the absolute limit of improvement for any single performance, we should find the S-shaped curve with relatively slow progress at first followed by increasing increments of gain and leveling off with decreasing gains as the limit was approached (6).

It may be presumed that a very rapid initial rise in a learning curve is due to the fact that the learning task is not altogether new to the learner and that he therefore does not begin at a zero point.

The slowing down of the rate of improvement may be caused by several factors such as reaching the limit of improvement, fatigue, loss of interest, a sense of sufficiency, lack of desire for further advancement, and the needless repetition or overlearning of parts of the performance mastered in the early steps of learning.

The absolute limit of performance is rarely reached. In most instances, practical limits and motivational limits are the determinant factors.

Burns (7) reports that the first publication leading to the industrial application of the learning curve has been credited to T.P. Wright. Wright (11) showed that as the number of aircraft produced increases, the cumulative average per unit cost to procuce an aircraft decreases at a constant rate. The model employed was $Y = KX^c$, where

Y = the number of direct labor man hours required to produce the Xth unit

K = the number of direct labor man hours required to produce the first unit

X = the unit number

$C = \frac{\log B}{\log 2}$, where B equals the learning curve factor, a constant (.90, .85, .77, etc.)

The mathematical function is called an inverse variation and means that the dependent variable (Y) gets smaller as the independent variable (X) gets larger. This relationship is also referred to as an exponential (log-linear) equation. For a given learning curve, K and c are constants where K can assume any positive value and c is a constant between zero and minus one (12,13).

This has since become known as the cumulative average theory of the learning curve (14). Since this first publication, learning curve theory has been extended into many areas ranging from the setting of contract prices to production planning and control (15). In situations where the learning curve principles can be applied, the government is also using it in evaluating contract proposals.

In a related article (16), J.D. Patton states that the manufacturing progress curve is often referred to as a learning curve. He asserts that improvements usually come from tool design, methods, materials, procedures, as well as the employee's learning. This concept is also useful in the areas of training, maintenance, and other logistics concerns. He further states that the manufacturing progress function is assumed to describe a constant percentage improvement as the production quantities double and that all progress functions will have the same shape, even though they may differ in the percentage improvements between doubled production quantities and the direct labor hours required to complete the first unit. This progress learning curve utilizes the power function, $Y = KX^{c}$ developed by Wright (11).

An alternative model, $Y = \alpha a^{X-1} \beta$ was presented by Pegels (17). He states that:

> The startup or learning curve literature has in the past concentrated mainly on the algebraic power function or on versions based on this function. This concentration is not unusual because the power function has proven, in numerous studies, to fit empirical data quite well. However, other easy-to-apply algebraic functions should also be analyzed and considered. One such function, an exponential function, is shown to provide a better fit to several sets of empirical data than the traditional power function.

The other alternative models to which Pegels refers were usually intended for specific applications or contained restrictive assumptions. He specifically mentioned: (1) An S-type function proposed by Carr (18) which was based on the assumption of a gradual startup. An S-type function has the shape of the cumulative normal distribution function for the startup curve and the shape of an operating characteristics function for the learning curve, (2) Guibert (19) proposed a complicated multiparameter function with several restrictive assumptions, (3) De Jong (20) proposed a version of the power function which generates two components, a fixed component which is set equal to the irreducible portion of the task, and a variable component, which is subject to learning.

$$Y = a[\beta + (1 - \beta)X^{-b}]$$

De Jong calls this fixed component, the "factor of incompressibility". He explains that this factor is dependent not only on the nature of the work but also upon the commencing combination of skill and familiarity with the work in hand. The times for manual operations per cycle will fall gradually, but not to zero as porposed by the standard power function (Wright) at infinity. They will tend to approach a certain limiting value. (4) Levy (21) presented a learning function which reaches a plateau and does not continue to decrease or increase as does the power function.

An overriding point expressed was that there are no specific learning curves which have universal application.

Thus far, the discussion of learning and learning curves has been focused on the general theory, aspects of individual learning curves and some industrial applications of learning curve theory. This background will now be used to expand into the area of group/team training and performance.

Group/Team Training and Performance

Several studies and laboratory experiments have been conducted in the area of group/team training and performance. Some of these take the form of a literature survey on publications relevant to team training and evaluation, while others report on actual laboratory cases or experiments concerning team function, structure and performance.

A distinction was drawn between the terms team and small group. Glaser, Klaus and Egerman (22,23) state that although both refer to collections of individuals acting in consort, a team is usually well organized, highly structured, and has relatively formal operating procedures...as exemplified by a baseball team, an aircraft crew, or a ship control team. Teams generally display the following characteristics:

1. relatively rigid in structure, orgainization, and communication networks,

2. have well defined positions or member assignments so that the participation in a given task by each individual can be anticipated to a given extent,

3. depend on the cooperative or coordinated participation of several specialized individuals whose activities contain little overlap and who must each perform their task at least at some minimum level of proficiency,

4. are often involved with equipment or tasks requiring perceptual-motor activities.

5. can be given specific guidance on job performance based on a task-analysis of the team's equipment, mission, or situation (23).

A small group, on the other hand, rarely is so formal or has well-defined, specialized tasks --- as exemplified by a jury, a board of trustees, or a personnel evaluation board (23). As contrasted with a team, small groups generally have the following characteristics:

1. have an indefinite structure, organization, and communication network,

2. have assumed rather than designated positions or assignments so that each individual's contribution to the accomplishment of the task is largely dependent on his own personal characteristics,

3. depend mainly on the quality of independent, individual contributions and can frequently function well even when one or several members are not contributing at all,

4. are often involved with complex decision-making activities,

5. cannot be given much specific guidance beforehand since the quality and quantity of participation by individual members is not known.

In a review of team training and evaluation by the Human Resources Research Organization (HUMRRO) (24), the authors state that the review was undertaken in order to provide an information base that the Defense

Advanced Research Projects Agency could use as a foundation to facilitate decisions regarding future research program support.

The technical report (24) reported the following findings and implications:

As an aid toward organizing and analyzing the team training information obtained, a classification scheme was used to categorize the training techniques and situations discussed in this review along two dimensions. On one dimension, training focus, a distinction was made between "team" training and "multi-individual" training. Multi-individual training occurs in a group context but focuses on the development of individual skills. Team training, on the other hand, is focused on developing team skills such as coordination and cooperation. The type of task situation was the second dimension used to classify the training techniques reviewed. Task situations were categorized as either "established" or "emergent." Established situations are those in which the tasks and the activities required to perform these tasks can be almost completely specified. Emergent situations are those in which all tasks and activities cannot be specified and the probable consequences of certain actions cannot be predicted. This type of situation allows for unanticipated behaviors to emerge.

Team training studies and practices were categorized according to the classification scheme described. These studies followed two conceptual models of team behavior-response (S-R) and organismic. The S-R model adherents tended to study team training in laboratory settings derived from established task situations. More realistic environments were used by other researchers who attended to emergent factors in the job situation (the organismic approach). It was this latter group of investigators who demonstrated the need for training in team skills, even though individual skill proficiency was found to be a prerequisite for effective team training and performance, other conclusions which were drawn from the literature are:

1. The team context is not the proper location for initial individual skill acquisition.
2. Performance feedback is critical to the learning of team skills, as well as individual skills.

Several examples of team training techniques currently in use in the military services are also presented in the report; for example, ARMY TRAINING AND EVALUATION PROGRAM (ARTEP), REALTRAIN, Naval Training Device Center (NAVTRADEVCEN) program, etc.

In the Final Summary Report by Klaus, Glaser, and others (23), a brief description of the seven studies undertaken are briefly described along with their purpose and major results.

Report 1 described the approach being examined in the Team Training Laboratory, one which considered the team and its output or product rather than the performance of its individual members as the focus of investigation (25).

Report 2 reported on the acquisition and extinction of a team response, a demonstration that basic principles of individual learning could be applied to the team considered as a single entity (26).

Report 3 presented an experiment on the inclusion of parallel or "redundant" members in a team which confirmed an hypothesis derived from the underlying approach that redundancy could result in eventual decrements in team performance (27).

Report 4 further analyzed the effects of internal team structure on the development and maintenance of a team response based upon the degree of correspondence between individual performance and feedback supplied to the team (28).

Report 5 identified the relationships among team member characteristics, the conditions of team training and the speed and thoroughness with which teams developed proficiency that could be demonstrated empirically (29).

Report 6 explained the value of more gradually introducing the low ratios of reinforcement typical of early team performance providing supplemental, supervisory-furnished feedback to team members (30).

Report 7 presents three studies on the simulation of team environment which considered the degree to which the approach facilitated the replication of team learning phenomenon based on the performance of a single individual (31).

The studies enabled the researchers to derive a learning theory model of team performance from among those psychological models of individual behavior which have proved most useful in understanding the conditions likely to affect training practice.

The underlying model has three essential features (24). First a team is a functioning entity having an output which depends on a defined input from its members. Second, a team itself can be considered as the module of investigation and its responses as amenable to manipulation without necessary reference to the performance of individual team members. Third, team performance can and will vary as a function of the consequences of responses much the same as the performance of an individual learner.

In Technical Report 1 (25), the first team acquisition curve obtained in the Team Training Laboratory is shown in the bottom half of Figure 2-4.

The curve is a plot of the number of correct team responses per experimental period. It appears from the correspondence between the two curves that the team response shows acquisition characteristics similar to an individual response. The authors state that the apparent improvement in team performance leading to an asymptote, can tentatively be explained on the basis of a temporary reduction in individual proficiency upon entering a team reinforcement situation. Thus, the fact that the team changes in proficiency as a result of training does not require assumptions as to characteristics of a team which are over and above the learning characteristics possessed by its individual members.

This study is concerned with group or team models, where the data was obtained from operational tests. The type of tasks involved are those which depict learning situations wherein decreasing scores indicate improvement, such as fewer errors or decreasing performance times on successive trials. Therefore, the learning curves are expected to follow some form of the negative acceleration theoretical curve model.

Since the team/crews are organized into two or more members (tank crew, mortar crew) their organization is characteristic of those described by Glaser, Klaus, and Egerman (23). In that context the basic principles of individual learning curve robustness will be assumed and analysis of the empirical data will procede along that line.

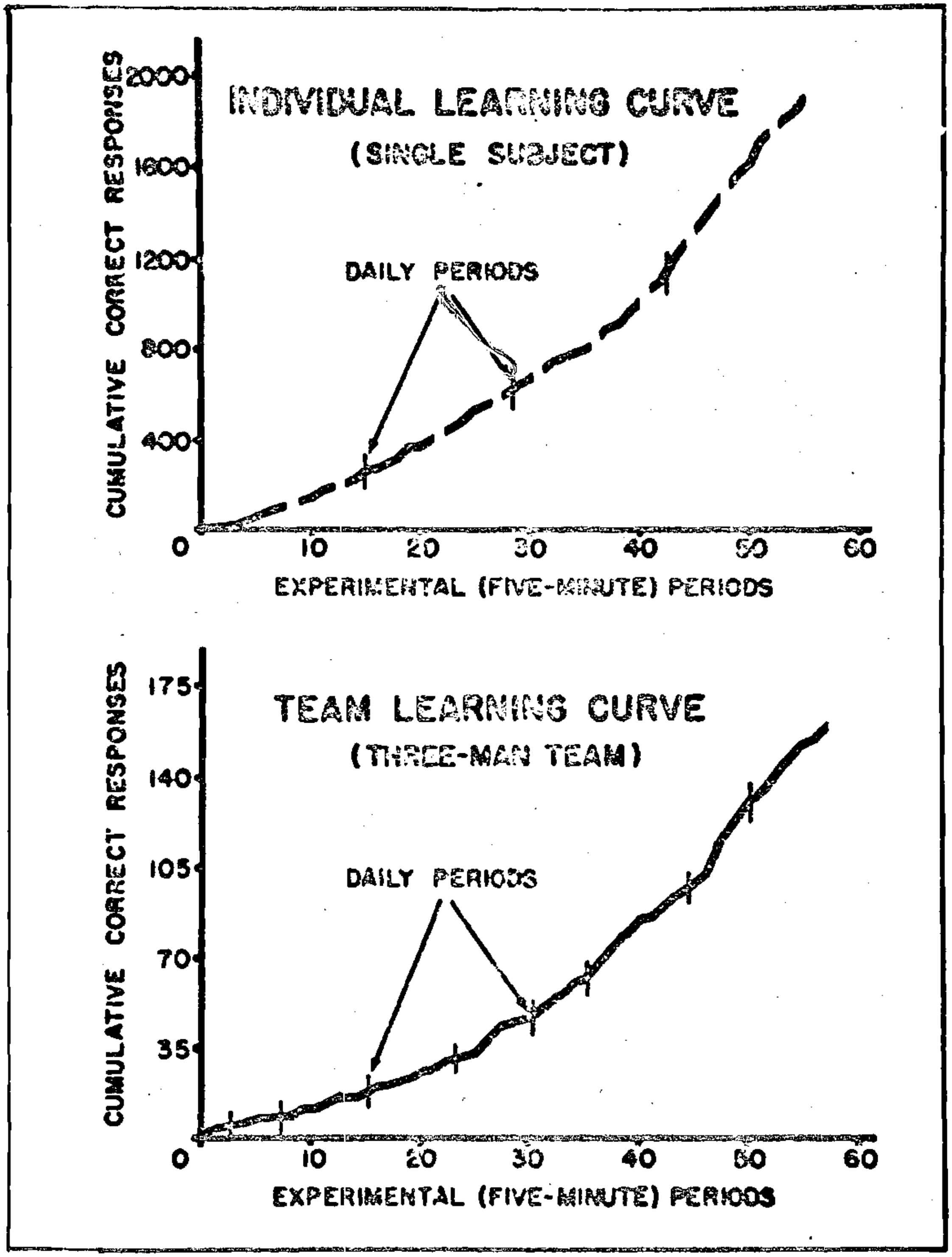


Figure 2-4. Comparison of Individual and Team Learning Curves (25)

Various models described previously, such as the power function with variations and exponential models, will be used to fit the empirical data and then analyzed for model adequacy. The methodology used to tie empirical data and analyze results will be discussed in Chapter III.

It was made clear through contacts with other sources of data that considerable interest is presently being generated in the area of group/team learning. Several proposed tests are being considered to analyze group learning. As discussed earlier, the analogy between individual learning and group learning suggests the substitution of the organization for the individual when using the classical learning model.

The Training and Doctrine Command (TRADOC) has conducted an extensive study into training cost procedures and the utilization of learning curve theory in the assessment of training proficiency. These studies include the assessment of both individual and group learning models along with validated performance measures. The Army Research Institute (ARI) has also planned tests which will attempt to make an assessment of group training.

CHAPTER III

METHODOLOGY

One of the principle objectives of this research is to determine the existence of a representative learning curve (or set of curves) and to develop a mathematical description of this curve applicable to training levels in operational testing. The existence of a representative learning curve could be used to develop improved operational test and evaluation methodology for training effectiveness. To determine whether there is a demonstrable learning curve for team/crew performance, it was necessary to collect and analyze data from operational test reports. Each data set will be analyzed iteratively utilizing the following procedures.

1. Determine graphically if learning patterns exist. Sample data will be plotted to determine if there are patterns in the empirical data which might suggest that learning can be detected. The performance measure is plotted against consecutive trials.

2. Fit Linear Model.

Simple linear regression is used to fit the linear model to empirical data and the null hypothesis, that the slope of the regression line is equal to zero, will be tested. In data sets where the time component or measurement of error is used as a performance measure, the slope of the regression line is expected to be negative and should not include zero in the confidence interval constructed around the slope. This condition reflects that there is an indication of learning in the data. If no learning is detected the data is not subjected to further analysis.

3. Fit Nonlinear Model.

Upon determining the suitability of the data, that is, graphically detecting discernible patterns and rejecting the null hypothesis that the slope of the regression line is zero, nonlinear models are used to fit the data. These include learning models suggested in the literature and/or variations based on the graphical patterns of the raw data (see Table 3-1). The selection of models is restricted to functional relationships between two variables whereby, the performance measure (Y) can be separated from the trials (t) in such a way that Y = f(t). Using this relationship, the performance measure is considered to be the dependent variable and the consecutive trial is the independent variable. Parameter estimates and a residual sum of squares are obtained by fitting the nonlinear model.

4. Test for Model Adequacy.

The assumption is made that the learning model fit in Step 3 is adequate. A test for "goodness of fit" of the model is used to verify that assumption utilizing the analysis of variance conducted for the significance of regression. A lack of fit test is performed when repeat observations in the data are available. This is done by constructing a lack of fit ratio which will be discussed later. Additionally, the statistical inferences on the model are checked through a direct examination of residuals. Model adjustments are made based on this examination of residuals and a careful examination of outliers

Table 3-1. Learning Models

Model	Origin
$\hat{Y} = at^{-b}$	T.P. Wright (11)
$\hat{Y} = a[\beta+(1-\beta)t^{-b}]$	De Jong (20)
$\hat{Y} = \alpha[a^{t-1}] + \beta$	Pegels (17)
$\hat{Y} = ae^{bt}$ *	*models suggested by graphical patterns in the data (32)
$\hat{Y} = ae^{b/t}$ *	
$\hat{Y} = at^{-b} + c$ *	
$\hat{Y} = \frac{a}{t+b} + c$ *	

if any. When adjustments are made, the iterative procedure returns to step 3 and the model is refit and tested for adequacy.

At this point another learning model or adjusted model is fit to the sample data and checked for model adequacy.

After fitting all selected models for a particular data sample, a comparison of models is conducted in step 5 and a new data set is introduced at step 1.

5. Selection of "Best" Model.

The criterion for evaluating the fitted learning models and selecting the model that provides the "best" fit to the empirical data will be based on the comparison of (1) the lack of fit ratio, and (2) the sum of squares for regression (SSR, the amount of variation in the model explained by regression). This criterion is used because it is a systematic and quantitative basis for selecting the "best" model.

The general procedures used in fitting the selected mathematical models to the empirical data and analyzing the models for adequacy involve regression techniques. These techniques provide:

(1) Parameter estimates for a given model.

(2) A measure of the error involved in estimating the parameters and the error variance around the fitted model. The sum of squares due to error is the amount of noise left in the data after the regression line has been fit. Where applicable, repeat observations are used to partition the error component into two parts, sum of squares due to pure error (random component) and sum of squares due to lack of fit (bias component). Normally, the data collected during operational

tests do not contain repeat observations over trials, therefore, an estimate of the sum of squares due to pure error is computed using different crew observations over a specific trial. This actually represents a measure of the random error between subjects (crews).

The regression procedures used are discussed in the following sections.

Fitting Linear Models

As stated previously, linear regression will be used to fit the linear model

$$Y_i = \beta_0 + \beta_1 t_i + \varepsilon_i, \; i = 1, 2, 3, \ldots, n \qquad (3\text{-}1)$$

where t is the i^{th} consecutive trial of the empirical data from various test reports. For a given trial t, a corresponding observation Y consists of the value $\beta_0 + \beta_1 t$ plus an amount ε, the increment by which any individual Y may fall off the regression line. β_0 and β_1 are the linear parameters in the model and are unknown as well as ε, the error or noise component which changes for each observation Y. The objectives of this model are

(1) Estimate β_0, β_1

(2) Screen data for suitability

The least-squares method is used to estimate the parameters β_0 and β_1. This method minimizes the sum of squares of deviations from the true line and is written (33)

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 t_i)^2 \qquad (3\text{-}2)$$

Estimates are chosen for β_0 and β_1 which produce the least possible value of S.

The usual basic assumptions for this model were made

(1) ε_1 is a random variable with mean zero and variance σ^2 (unknown), that is, $E(\varepsilon_i) = 0$, $V(\varepsilon_i) = \sigma^2$

(2) ε_i and ε_j are uncorrelated, $i \neq j$, so that COV $(\varepsilon_1, \varepsilon_j) = 0$. Thus, $E(Y_i) = \beta_0 + \beta_1$, $V(Y_i) = \sigma^2$ and Y_i and Y_j, $i \neq j$ are uncorrelated.

Recall that the linear model is fit to develop some idea of the relationship of the performance measure over consecutive trials. When estimates of the parameters β_0 and β_1 are obtained, a screening process is conducted to look at the slope (β_1) of the fitted model. This screening process is used to determine if there is an indication of learning over consecutive trials. We use the value from the t-distribution table (with the appropriate degrees of freedom) to obtain an estimate at a given level. We compare this value with the ratio given by

$$\frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{MS_{E/S_{xx}}}}$$

where MS_E is an estimate of the variance and S_{xx} is the corrected sum of squares of the trials. From this we would get some approximate idea of whether or not the slope is negative.

Since the performance measures in the data collected are time components and measurements of error over consecutive trials, a negative slope for the regression line would indicate that learning is taking place over consecutive trials. The hypothesis test on the slope can be modified since $\beta_{10} = 0$ to test for the significance of Regression and an Analysis of Variance can be conducted. For a further discussion of this procedure see Draper and Smith (33).

Fitting Nonlinear Model

When hypothesis testing conducted after fitting the linear model indicates that learning can be detected in the data, the nonlinear learning models mentioned earlier are fit to the data. Parameter estimates are obtained along with the residual sum of squares for use in the model adequacy test.

The SPSS (Statistical Package for the Social Sciences) Subprogram NONLINEAR (34) is used to apply nonlinear regression analysis to estimate parameters that appear in the regression model in a nonlinear fashion. The form of the learning models in Table 3-1 are known explicitly or come from an interpretation of the graphical patterns in the data. The SPSS NONLINEAR program utilizes the Least Squares Estimation function to estimate the unknown parameters by minimizing the error sum of squares. For each case, the performance measure (dependent variable) is defined:

$$Y_i = f_i(t,\theta) + \varepsilon_i, \; i = 1, 2, \ldots, n \qquad (3\text{-}3)$$

where $f_i(t,\theta)$ stands for the model function chosen, ε_i is the error term, and θ is a vector of parameter estimates.

The assumptions made are $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma^2$. The error sum of squares function can be written as

$$S(\Theta) = \sum_{j=1}^{n} [Y_j - f_j(t_j,\Theta)]^2 \tag{3-4}$$

The program minimizes the sum of squares for the model $f_i(t,\theta)$ by choosing suitable values for the unknown parameters (θ) in the model. This in turn will describe as close as possible the behavior of the dependent variable Y.

Marquardt's nonlinear minimization technique is used to estimate the unknown parameters. It is a compromise between the linearization (or Taylor series) method and the steepest descent method and appears to combine the best features of both while avoiding their most serious limitations. It almost always converges and does not slow down as it approaches the solution.

The idea of Marquardt's method can be explained briefly as follows (33,34). We start from a certain point in the parameter space, θ. The method of steepest descent is applied and a certain vector direction, δg where g stands for gradient, is obtained for movement away from the initial point. Because of attenuation in the $S(\theta)$ but may not be the best overall direction. However, the best direction must be within 90° of δg or else $S(\theta)$ will get larger locally. The linearization (or Taylor series) method truncated after the second term leads to another correction vector δ given by the linear model

$$\beta_0 = (Z_0^1 Z_0)^{-1} Z^1 (Y-f_0) \qquad (3\text{-}5)$$

where β_0 is the parameter estimate vector, Z_0 is an nxp matrix containing the first partial derivatives and Z_0^1 is its transpose matrix, and $(Y - f_0)$ is a vector containing the residuals (actual observation - predicted value).

However, instead of using the linear model to solve for the parameter estimates, Marquardt's method uses the following equation:

$$\beta_0 = (Z_0^1 Z_0 + \lambda I)^{-1} Z_0 (Y-f_0) \qquad (3\text{-}6)$$

where I is the identity matrix and λ is a correction factor. For the first iteration λ is set to zero and it remains zero for all subsequent iterations as long as the sum of squares function is reduced. If at some iteration, say iteration r, the sum of squares function is increased, then λ is replaced with the following expressions:

$$\lambda + \frac{\beta_r^1 \beta_r}{\beta_r^1 (Z_r^1 Z + \lambda I)^{-1} \beta_r}$$

and the solution in (3-6) is tried again. (This correction tends to reduce the Euclidean norm of β_r to one-half its previous value). The value of λ is corrected repeatedly until the sum of squares function is reduced (or until) the members in β_r are too small to be meaningful, i.e., the norm of β_r has been reduced beyond a tolerance level (34).

Since the program requires initial estimates of the unknown parameters, a computer program was used to provide them using data from the test reports and is listed in Appendix B.

After the nonlinear model is fit, a direct examination of residuals is conducted and a lack of fit ratio is computed for comparison with other models.

If the original observations of a sample data set do not conform to the model assumptions made, then a log transform of the model may possibly correct the problem. When a direct examination of the residuals for a model indicates that the error component is multiplicative instead of additive, then the log transform of the model should be computed and fitted to the sample data. For example, the model $\hat{Y} = at^{-b}$ has multiplicative error when expressed $\hat{Y} - at^{-b}\varepsilon$ and additive error when expressed as $\hat{Y} = at^{-b} + \varepsilon$. In the former case the log transform can be specified as $\ln\hat{Y} = \ln a - b\ln t + \ln\varepsilon$ but in the latter case the log transform cannot be specified. The multiplicative error is exemplified when variability becomes a function of the magnitude of the responses such as cases where large errors are linked with large responses.

When the log transform model is linear it is fit using step 2, when otherwise specified step 3 is used, and then tested for model adequacy. When comparisons are made between the log transform models and nonlinear models in step 5 of the iterative process, the parameter estimates must be converted in order to compare sum of squares.

Model Adequacy

As stated previously, the learning models chosen to fit to sample data from the various test reports are assumed to be tentatively correct. Under certain conditions we can check whether or not the models are correct. This will be done by testing for model adequacy using a "goodness of fit" test and through a direct examination of residuals. The residual at each trial is defined as the amount by which the actual observed value Y_i differs from the fitted value $\hat{Y}_i$ and can be written as $e_i = Y_i - \hat{Y}_i$. If the learning model chosen is not correct, then the residuals contain both random (variance error) and systematic (bias error) components.

Recall that during operational tests, repeat observations are not taken for each crew across trials. However, all crews are observed at each consecutive trial and are assumed to be similar in structure and training level. Therefore, several crew observations at the same trial t_i are considered repeat points in the data. These "repeats" are used to obtain an estimate of σ^2 and represents a measure of the random error between crews. As a consequence, we can test for the "goodness of fit" of our learning model. The hypothesis tested (33,35) can be stated:

H_0: The model adequately fits the data

H_1: The model does not fit the data

The test involves partitioning the error or residual sum of squares into the following two components:

$$SS_E = SS_{PE} + SS_{LOF} \qquad (3\text{-}7)$$

where SS_{PE} is the sum of squares attributable to random error between crews and SS_{LOF} is the sum of squares attributable to the lack of fit of the model. The pure error estimate of σ^2 is found by computing the contribution to the pure error sum of squares from the i^{th} consecutive trial when there are at least two observations, such that

$Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ are n_1 repeat observations at t_1

$Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ are n_2 repeat observations at t_2

$Y_{k1}, Y_{k2}, \ldots, T_{kn_k}$ are n_k repeat observations at t_k

The total sum of squares for pure error is calculated as follows:

$$SS_{PE} = \sum_{i=1}^{m} \sum_{\mu=1}^{n_i} (Y_{i\mu} - \overline{Y})^2 \qquad (3\text{-}8)$$

where m is the number of distinct levels of t,

n_i is the number of observations at trial i,

$Y_{i\mu}$ is a single observation, and

$\overline{Y}$ is the sample mean across a particular trial.

The total degrees of freedom associated with the total sum of squares pure error is computed as follows:

$$\text{total degrees of freedom} = \sum_{i=1}^{K} (n_i - 1) = \sum_{i=1}^{K} n_i - K = n_e$$

The sum of squares for lack of fit is computed by subtraction

$$SS_{LOF} = SS_E - SS_{PE}$$

with $n - 2 - n_e$ degrees of freedom, where n is the total number of observations (35). The mean square for pure error is

$$MS_{PE} = \frac{SS_{PE}}{ne} = \frac{\sum_{i=1}^{m} \sum_{\mu=1}^{n_i} (Y_{i\mu} - \overline{Y})^2}{\sum_{i=1}^{K} n_i - K}$$

and is an estimate of σ^2.

The pure error sum of squares is introduced into the analysis of variance procedure and the F-ratio is computed. This ratio, $F = \frac{MS_{LOF}}{MS_{PE}}$ is compared with the $100(1-\alpha)\%$ point of an F-distribution with $(n-n_e)$ and n_e degrees of freedom if the normality assumption is satisfied. If the ratio is

(1) Significant, this indicates that the model appears to be inadequate. Attempts would be made to discover where and how the inadequacy occurs.

(2) Not significant, this indicates that there appears to be no reason to doubt the adequacy of the model and both pure error and lack of fit mean squares can be pooled and used as estimates of σ^2 (33).

The usual tests which are appropriate in the linear model case are, in general, not appropriate when the model is nonlinear (33). As a practical procedure we can compare the unexplained variation with an estimate of $V(Y_\mu) = \sigma^2$ but cannot use the F-statistic to obtain conclusions at any stated level. In the absence of exact results for the nonlinear models, we can regard this sum of squares as being based on the total degrees of freedom for residuals/error. In the nonlinear case this does not in general, lead to an unbiased estimate of σ^2 as in the linear case, even when the model is correct.

A pure error estimate of σ^2 can be obtained from the repeat observations as discussed earlier. This provides a sum of squares (SS_{PE}) with n_e degrees of freedom. An approximate idea of possible lack of fit can be obtained by evaluating $SS_E - SS_{PE} = SS_{LOF}$ and comparing mean squares.

$$MS_{LOF} = \frac{SS_{LOF}}{n-n_e} \quad \text{and} \quad MS_{PE} = \frac{SS_{PE}}{n_e}$$

Draper and Smith state that an F-test is not applicable here but that we can use the value from the table (with the appropriate degrees of freedom) as a measure of comparison. From this we would get some approximate idea of how well the learning model fits. Measures of non-linearity suggested by E.M.L. Beale (36,37) can be used to help decide when linearized results provide acceptable approximations, but they are not used for this study.

Since residuals are measures of the error component, the assumptions made concerning the selected model and an assessment of model adequacy can be evaluated through a direct examination of residuals. Recall that residuals e_i, $i = 1, 2, \ldots, n$ represent the deviation of the observations after the regression line has been fit and can be expressed $e_i = Y_i - \hat{Y}_i$ where Y_i is an observation and $\hat{Y}_i$ is the corresponding fitted value obtained by use of the fitted regression equation

(33). From this definition, the residuals e_i are the differences between what is actually observed, and what is predicted by the regression equation. That is, the amount which the regression equation has not been able to explain or the observed errors if the model is correct.

The usual assumptions are that the errors are independent (uncorrelated), have zero mean, and a constant variance, σ^2. If in fact, the errors in the sample data follow a normal distribution, the F-test can be made. Through a direct examination of the residuals we can conclude either (1) the assumptions appear to be violated or (2) the assumptions do not appear to be violated. This direct examination will be done by plotting the residuals (1) overall, (2) in time sequence, and (3) constructing histograms of the residuals. If the learning model is correct the residuals should resemble observations from a normal distribution with zero mean. The patterns of the plotted residuals will also give indications about homogeneity of variances, abnormality, and an indication of possible outliers - unusual points in the data that are far greater than the rest in absolute value, and perhaps lies three or four standard deviations or further from the mean of the residuals. The errors may be linked to equipment failures or errors in recording the observations and should be obtained from background information concerning the various test reports.

To determine if the residuals are independent, an estimate of their autocorrelation function is obtained and examined. An estimate of autocorrelation coefficient at a particular lag is computed using the following expression:

$$\hat{\rho}(\ell) = \frac{1}{N-\rho-1} \frac{\sum_{i=1}^{N-\ell} (Y_t - \overline{Y})(Y_{t+1} - \overline{Y})}{S^2}$$

where N equals number of residuals, Y_t is the computed residual at trial t, ℓ equals lag, $\overline{Y}$ is the sample mean and S^2 is an estimate of the variance.

CHAPTER IV

DATA ANALYSIS

The first major task in this research study was that of data collection. Although OTEA was the primary source of data, other Army agencies in the training analysis area were also contacted. These include, the Army Research Institute (ARI), Training Development Division/System Analysis Branch of the Infantry School, The Infantry Board (USAIB), and the TRADOC Combined Arms Training Agency (TCATA). OTEA provided operational test reports or extracts concerning data relating to performance/learning in past tests, and made available, knowledgeable personnel to provide background information where possible.

Due to the nature of the study, there were limitations placed on the characteristics of the data required. The limitations are listed below:

1. Data had to come from an operational testing environment.
2. Tests conducted should involve team/crew tasks and performance objectives.
3. Criterion or measures of effectiveness must be applicable to team/crew tasks within the context of group or team definitions as discussed in Chapter II.
4. Test reports must provide a means of tracking a team/crew from start to finish. That is, performance measured over time or consecutive trials.
5. When applicable, test reports should provide some insight into the background infromation concerning the data relevant to this study, such as measurement error and conditions that may have affected the test results ("noise" in the data).

It became apparent from the outset that little empirical data was available in the context mentioned above. Factors affecting the availability of data were:

1. The cost is prohibitive or infeasible to conduct more than one or two trials in some data collection efforts.

2. Crew or group membership changes rapidly and significantly affects the results.

3. In some cases where test reports were selected, adequate information was not available to trace a particular crew from start ot finish. Therefore, changes in performance could not be adequately established or inferred.

Descriptions of the data collected and their analysis will be discussed in the following sections. Table 4-1 lists each sample data set and its origin.

Table 4-1. Data Base

Title	Origin
Improved Tow Vehicle (ITV) (38)	OTEA
Dragon (39)	OTEA
REALTRAIN Validation with Combat Units in Europe (40)	ARI
REALTRAIN Validation for Rifle Squads (41)	ARI
Project Stalk (42)	OTEA
Lightweight Company Mortar System (OTI) (43)	OTEA
Team Training (Experiment VIII) (44)	NAVTRADEVCEN

Improved Tow Vehicle (ITV)

The ITV operational test was conducted to compare four systems with each having six dedicated gun crews with alternates. The gunners tracked targets over four range bands which included two target profiles. All gunners were trained and ranked on a baseline system prior to allocation to separate systems. Additionally, contractor training was conducted for gunners assigned to the new system. A summarized description is provided below:

1. Performance measure - Root mean Square Error (RMS)
2. Characteristics

   (a) Four systems
   (b) 24 primary gunners
   (c) 5 gunners
   (d) Approximately 12 to 16 trials per gun crew with a total of 1760 observations
   (e) Type of activity - tracking

It should be noted that in the context of the definition of group/team learning tasks, the performance measure (RMS) analyzed does not reflect a team measure of effectiveness. However, since this was the initial data sample received and thought to contain detectable learning, an analysis was still performed.

In the initial analysis of the ITV data sample it was felt that there might be some effect on the data due to specific combinations of range and target profile (evasive maneuvers). Therefore, an analysis was conducted to determine if some adjustment was required for these effects. All possible combinations (8) of range and target profile were computed and a linear regression procedure was performed to estimate which combination should be adjusted. The results of the regression procedure indicated that while the overall regression appeared to be significant at the 5 percent level, the confidence intervals around the parameter estimates included zero and it was concluded that no specific combination of range and target profile had a significant effect. Therefore, no adjustment procedure was employed and the iterative analysis procedure was initiated.

Twenty-four (24) individual gun crew data polts were made to determine if a discernible pattern indicated learning over consecutive trials. The majority of the plots do not indicate such a pattern and there were only a few rare cases in which some slight indication of learning could be detected. Representative plots are shown in Figures A-1 through A-6. In addition 24 plots of the linear regression line with a 95 percent confidence interval were made and they depicted similar results.

An aggregate data sample for each system was developed using the average response for the crews at each trial. Fitting the linear model in step 2 of the iterative procedure shows the following results for the four systems analyzed.

System A

Source	d.f.	Sum of Squares	Mean Square
Regression	1	.00157	.00157
Residual	60	.04771	.0007952

$$\text{F-ratio} = \frac{.00157}{.00080} = 1.974$$

When compared to the F-distribution value for 1 and 60 degrees of freedom at the 5 percent level, there is no evidence to reject that $\beta_1 = 0$. The confidence interval around β_1 includes zero and it appears that learning cannot be detected.

System B

Source	d.f.	Sum of Squares	Mean Square
Regression	1	.00926	.00926
Residual	55	.03370	.00061

F-ratio = $\frac{.00926}{.00061}$ = 15.11089
*Significant at the 5 percent level

For the System B, the confidence interval around β_1 doe not include zero and $\hat{\beta}_1$ = .004388 which indicates that there is detectable learning.

System C

Source	d.f.	Sum of Squares	Mean Square
Regression	1	.00029	.00029
Residual	75	.10931	.00146

F-ratio = $\frac{.00029}{.00146}$ = .19907
*Not significant at 5 percent level

System D

Source	d.f.	Sum of Squares	Mean Square
Regression	1	.00163	.00163
Residual	89	.02448	.00028

$$\text{F-ratio} = \frac{.00163}{.00028} = 5.9146$$

*Significant at the 5 percent level

Systems B and D appear to have detectable learning while systems A and C did not. Since system B appears to have the largest F-ratio and slope estimate, the aggregate data sample was modified to use the individual crew response at each trial. This was done to provide an estimate of the lack of fit when the nonlinear models were fit in step 3 of the iterative procedure. The results of fitting the nonlinear models are shown in Table 4-2. The exponential model $\hat{Y} = ae^{bt}$ where $a = .040708$, $b = -.009424$ and the power function $\hat{Y} = at^{-b}$ where $a = .047369$ and $b = .13539$, appear to provide an adequate fit to the sample data.

Since the performance measure actually represents an individual measure of effectiveness further anlaysis was not undertaken.

Dragon

An operational test on the dragon weapon's system was conducted by OTEA using 32 gun crews. Gun crews tracked and fired on targets at various range bands. Each crew was observed over 15-20 consecutive

Table 4-2. Comparative Results for Fitted Models (System B(ITV))

Model	SS_E	SS_{LOF}	SS_R	Lack of Fit Ratio
$\hat{Y} = at^{-b}$	.29665006	.04937006	.336350	.78967
$\hat{Y} = ae^{bt}$	.2953304	.04775	.3375696	.7637638
$\hat{Y} = ae^{b/t}$	.3015603	.0542803	.3313397	.8683320
$\hat{Y} = a[\beta+(1-\beta)t^b]$	.30302805	.05574805	.32987195	.89169
$\hat{Y} = \alpha(a^{t-1}) + \beta$	.30302805	.05574805	.32987195	.89169
$\hat{Y} = at^{-b} + c$	.29606377	.04878377	.336836	.78029
$\hat{Y} = \frac{a}{t+b} + c$	.29557716	.0482972	.33732284	.77251

$SS_{PE} = .24728$

trials. A summarized description is provided below.

1. Performance Measure - Time components (seconds)

   (a) Identification of target to launch (T2)
   (b) Time between target hit and disposal of used round (T4)

2. Characteristics

   (a) 32 gun crews
   (b) Type of activity - tracking

The two time components, T2 and T4, were both plotted against consecutive tirals. The graphical representations show no discernible learning patterns in the data. Representative plots are shown in Figures A-7 through A-9. Furthermore, the linear regression shows that the slope (β_1) of the regression line is essentially zero.

T2 Aggregate

Source	d.f.	Sum of Squares	Mean Squares
Regression	1	35.29688	35.29688
Residuals	166	263026.55431	1584.449732

$$\text{F-ratio} = \frac{35.29688}{1584.49732} = .02228$$

*not significant at 5 percent level

T4 Aggregate

Source	d.f.	Sum of Squares	Mean Squares
Regression	1	3.83857	3.83857
Residuals	166	9285.15548	55.93467

$$\text{F-ratio} = \frac{3.83857}{55.93467} = .06863$$

*not significant at 5 percent level

Since the Dragon sample data fails to meet the suitability criteria during the screening process, no further analysis is performed.

REALTRAIN Validation with Combat Units in Europe

The REALTRAIN exercise provided a two-sided, free-play situation for infantry and armor units in a simulated tactical environment. It provided for a sequential record of events during each engagement which included an assessment of casualties. A summarized description is provided below.

1. Performance measure - Casualty rate
2. Characteristics
   - (a) Two teams (conventional training vs REALTRAIN methods)
   - (b) Each team consisted of
     - (1) Tank Platoon
     - (2) Two Infantry Squads
     - (3) Tow Section

This sample was deemed inappropriate because it contained consolidated data over two trials. That is, the exercise was run over two or three phases and all observations were averaged together and displayed in graphical form. Raw data for each unit was not available. Since our learning models contain at least two unknown parameters, further analysis would be misleading.

REALTRAIN Validation for Rifle Squads

This REALTRAIN exercise provided a two-sided, free-play situation for 18 rifle squads. Nine squads were trained using REALTRAIN techniques and the other nine squads were trained using conventional techniques. The rifle squads were pitted against each other (REALTRAIN vs

Conventional) in a simulated tactical environment. An assessment of the casualty rate (sustained vs inflicted) was recorded during each engagement. A summarized description is provided below.

1. Performance Measure - Casualty rate (sustained vs inflicted)
2. Characteristics

   (a) Two training methods - Conventional vs REALTRAIN
   (b) 18 rifle squads
   (c) 9 squads/training method
   (d) Type of Test - Tactical Exercise

Observations for all squads were averaged and displayed graphically. Only two phases (trials) of the exercise were conducted. Therefore, it was also concluded that this data sample was inappropriate for analysis.

Project Stalk

Twenty-five tank crews operating under conditions of competitive stress and rigidly uniform training were timed in their performance at hitting a stationary target which appeared suddenly as a result of the travel of their tank. Eleven different conditions of tank and fire control conditions were run by each of the twenty-five crews participating in the test. Crews were given instructions to obtain a target hit in a minimum time. Crews were timed in their speed at recognizing the target, loading the round, laying the gun, etc., until a hit was obtained. Two typed of test courses were used. On the first type, range and characteristics of the target and tank positions were repeatedly observed by the crews. On the second course none of these factors were known by the crews. The experimental design was such that factors related to differences in training, testing conditions, and crew proficiency could be accounted for when comparing the performance of the five tanks. A summarized description is shown below.

1. Performance Measure - Time of detection to hit on target
2. Characteristics
    (a) Twenty-five crews
    (b) Five types of tanks used
    (c) Each crew was trained on a tank immediately prior to firing it.
    (d) Type of activity - Tank gunnery

Data for sixteen of the twenty-five crews were used because it was felt that this provided an adequate number of degrees of freedom and the addition of the others would only marginally affect the results. In addition, because of the time required to extract the data from the test reports, it appeared that the sixteen crews selected adequately represented the data sample. Background information indicated no rank-order performance in assigning tank crews to the five platoons. Therefore, the selection of the 16 crews did not appear to perpetuate any bias effect in the analysis. Each crew was trained under rigidly uniform conditions and given the same instructions during the conduct of the test. Background information also reveals that

> The crew differences in recognition time are similar to crew differences observed for other operations and exhibit the normal spread of proficiency attainment of human beings. It has been observed that, whatever the ultimate cause of crew differences in recognition time, they were appreciable and reasonably constant.... The correlation coefficient between the average recognition time of each of the individual crews on the Test Course targets and the average recognition time of the corresponding crews on Training Test Courses targets is indicative of the crew consistency. (43)

Data was plotted for the sixteen crews and the patterns of the plots showed significant learning (see Figures A-10 through A-17).

Background information revealed that the recognition to hit time reflected the reduced times to perform the individual operations with training by decreasing from an average for the four non-transfer targets on the Test Course of 66.4 seconds for Phase I to 33.1 seconds in the final phase (43). Only observations for non-transfer targets were used because target 4 in the Test Traininc Course (TTC) and target 5 in the Test Course (TC) required the unloading and reloading of another round in the gun. For example, in the former case, target 3 required AP (antipersonnel) ammunition and the gun is immediately reloaded upon firing a round at any target in anticipation of another being required. After getting a hit on target 3, the loader had to unload the AP round and store it, then load the proper HE (high explosive) round for target 4. This procedure resulted in a longer first round load time by about 20 seconds more than was required at other targets (43).

The times to achieve a target hit were found to decrease markedly with crew training. Although the hitting probabilities were found not to increase with training, the time to load the rounds and lay the gun decreased greatly with the training given the crews during the test.

Two aggregate data sets for both the Test Training Course (TTC) and the Test Course (TC) were developed by combining the data for the 16 crews across the four non-transfer targets and the eleven conditions for each target. This provided a method of tracking the crew performances throughout the test according to the Greco-Latin test design used. The TTC data consisted of 678 observations and the TC data consisted of

674 observations over 44 trials. When the linear model was fit to both data sets in step 2 of the screening process, the following results were indicated.

TTC

Source	d.f.	Sum of Squares	Mean Square
Regression	1	87726.475	87726.475
Residuals	676	2827995.42068	4183.425

$$\text{F-ratio} = \frac{87726.475}{4183.425} = 20.97$$

TC

Source	d.f.	Sum of Squares	Mean Square
Regression	1	82522.39281	82522.39281
Residuals	672	2440878.25556	3632.259308

$$\text{F-ratio} = \frac{82522.39281}{3632.259308} = 22.719$$

When compared to the F-distribution value for the appropriate degrees of freedom at the 5 percent level, there was evidence to reject that $\beta_1 = 0$. The confidence intervals around β_1 for both data sets did not include zero. Since the estimates of β_1 were both negative, there was an indication that learning was occurring.

Both data sets satisfied the suitability criteria specified in the screening process; therefore, the nonlinear learning models listed in Table 3-1 were fit to the data.

Initially three models were fit.

(1) $\hat{Y} = at^{-b}$

(2) $\hat{Y} = ae^{bt}$

(3) $\hat{Y} = ae^{b/t}$

First analyze the Test Training Course data. Parameter estimates and a residual sum of squares were obtained by using the SPSS Nonlinear Subprogram.

(1) $\hat{Y} = at^{-b}$ where $a = 86.13708$ $b = -.173043$ $SS_E = 2851060.4$

(2) $\hat{Y} = ae^{bt}$ where $a = 77.2504$ $b = -.01792$ $SS_E = 2822300.5$

(3) $\hat{Y} = ae^{b/t}$ where $a = 51.61$ $b = .31028$ $SS_E = 2906957.3$

To obtain an approximate idea of the lack of fit of the models, a pure error estimate of σ^2 was computed as discussed in Chapter III by using the 16 crew observations over each trial.

$$SS_{PE} = \sum_{i=1}^{44} \sum_{\mu=1}^{n_i} (Y_{i\mu} - \overline{Y})^2 = 2339080.18552$$

Since $SS_E = SS_{PE} + SS_{LOF}$, the sum of squares for lack of fit was obtained by subtraction. Using the model $\hat{Y} = at^{-b}$,

$$SS_{LOF} = SS_E - SS_{PE} = 20851060.4 - 2339080.18552$$
$$= 511980.214$$

A lack of fit ratio was obtained by comparing the mean squares.

$$MS_{LOF} = \frac{SS_{LOF}}{n-n_e} = \frac{511980.214}{42} = 12190.00512$$

$$MS_{PE} = \frac{SS_{PE}}{n_e} = \frac{2339080.18552}{634} = 3689.4009$$

$$\text{Lack of Fit ratio} = \frac{12190.00512}{3689.4009} = 3.304$$

The lack of fit ratios for (2) and (3) are shown in Table 4-3. To further test the model for adequacy, a direct examination at residuals was conducted. Figure A-18 shows an overall plot of the average residuals across the 44 trials for the 16 crews. By visual inspection it appeared that the average residuals at trials 1, 4, and 42 were atypical of the others. The majority of the individual residuals appeared to be ± 3 standard deviations from the mean of the residuals at those trials. Even though there were one or two residuals which did not exceed the criteria, it was concluded that the removal of the entire set of observations would not adversely affect the analysis. The model $\hat{Y} = at^{-b}$ appears to fit the data and is selected as the "best" model. Even though De Jong's model and $\hat{Y} = at^{-b} + c$ appear to have a somewhat smaller lack of fit ratio with corresponding larger SS regression, the power function ($\hat{Y} = at^{-b}$) is selected due to parsimony. That is, it has fewer parameters and does not appear to be significantly different from the model $\hat{Y} = at^{-b}$ where a = 104.595 and b = -.26492.

After fitting and selecting the "best" model we must further examine its adequacy. We compute the residuals $e_j = Y_j - \hat{Y}_j$ and then

Table 4-3. Comparative Results for Fitted Models (TTC)

Model	SS_E	SS_{LOF}	SS_R	Lack of Fit Ratio
$\hat{Y} = at^{-b}$	2851060.4	511980.214	1991541.85	3.304
$\hat{Y} = ae^{bt}$	2822300.5	483220.314	2020301.75	3.119
$\hat{Y} = ae^{b/t}$	2906057.3	567877.114	1935644.95	3.665
$\ln\hat{Y} = \ln a - b\ln t$	374.1706	73.8112	12.50802	3.710
$\ln\hat{Y} = \ln a + bt$	369.6397	69.2803	17.03892	3.48186
$\ln\hat{Y} = \ln a + b/t$	384.5419	84.18246	2.13672	4.23078

SS_{PE} = 2339080.1855 (Nonlinear models)

SS_{PE} = 300.3594 (log transform models)

Table 4-4. Comparative Results for Fitted Models (TTC)

(Adjusted Data)

Model	SS_E	SS_{LOF}	SS_R	Lack of Fit Ratio
$\hat{Y} = at^{-b}$	1527619.0	166437.76	1626287.0	1.856
$\hat{Y} = ae^{bt}$	1529402.9	168221.66	1624503.1	1.876
$\hat{Y} = ae^{b/t}$	1545537.8	184356.374	1608368.2	2.06
$\hat{Y} = \frac{a}{t+b} + c$	1534004.0	172822.575	1519902.0	1.927
$\hat{Y} = a[\beta+(1-\beta)t^{-b}]$	1525856.8	164675.375	1628049.2	1.836
$\hat{Y} = at^{-b} + c$	1526337.5	165156.025	1627568.5	1.842
$\hat{Y} = \alpha(a^{t-1}) + \beta$	1597436.3	266255.06	1556469.7	2.635
$\ln\hat{Y} = \ln a - b\ln t$	310.0763	47.767	19.97871	2.765
$\ln\hat{Y} = \ln a + bt$	308.7863	46.4774	21.269	2.690
$\ln\hat{Y} = \ln a + b/t$	314.2337	51.925	15.82134	3.005
$\ln\hat{Y} = a' + b't$	.32069	.04804	.76351	2.675

SS_{PE} = 1361181.24 (Nonlinear models)

SS_{PE} = 262.30890 (Log transform models)

SS_{PE} = .27265 (other)

NOTE: Atypical points at trials 1, 6, 42 removed.

estimate and examine their autocorrelation function. The sample autocorrelation function of the residuals is denoted by $\{\hat{\rho}_k(e)\}$ (46). Again, the average residual across each trial is used. Rather than consider the $\hat{\rho}_k$ (e)'s individually, we obtained an indication of whether the first 11 residual autocorrelations considered together indicate adequacy of the model. As a general rule k lag coefficients are examined where $k \leq N/4$. This estimate is obtained through an approximate Chi-square test for model adequacy.

$$\hat{\rho}_1(e) = .02758 \qquad \hat{\rho}_6(e) = -.38102$$
$$\hat{\rho}_2(e) = -.38909 \qquad \hat{\rho}_7(e) = -.03358$$
$$\hat{\rho}_3(e) = -.02111 \qquad \hat{\rho}_8(e) = .37201$$
$$\hat{\rho}_4(e) = .38570 \qquad \hat{\rho}_9(e) = -.16558$$
$$\hat{\rho}_5(e) = -.34704 \qquad \hat{\rho}_{10}(e) = -.22597$$
$$\hat{\rho}_{11} = .02670$$

Approximate Chi-square statistic

$$Q = (N) \sum_{k=1}^{k} \hat{\rho}_k^2(e)$$

$k = 11$ lags

Test Statistic $Q = 34.57047$

Comparing Q with a 5 percent value chi-square variable w/43 degrees of freedom, we find $\chi^2_{0.05,43} \approx 59.34$. We conclude that there is no strong evidence to reject the model.

For the model $\hat{Y} = 104.595\ t^{-.26492}$ Figure A-19 shows a plot of the residuals for each observation and they appear to come from an approximate "peaked-normal" distribution. Figure A-20 shows a plot of the estimates of σ^2 at each trial (MS_{E_i}) and they tend to level off after the 16th trial.

The nonlinear models fit to the Test Course data provided the results shown in Table 4-5 for 674 observations over 44 trials. An overall plot of the average residuals indicated that there were some atypical points in the data sample. Atypical points were determined by background data which indicated that factors extraneous to the test considerations had exerted undue influence. Additionally residuals were judged to be atypical if they were ± 3 standard deviations from the mean of the residuals at a specific trial. A total of 82 observations were removed from the original aggregate data set. An adjusted data set was refit after removing atypical points at a specific trial. The results shown in Table 4-6 indicate that the estimate of the lack of fit improved slightly for the exponential model $\hat{Y} = ae^{bt}$ while the fit for the others appeared to get worse with the exception of De Jong's model, $\hat{Y} = a[\beta + (1-\beta)t^{-b}]$. It is also noted that the lack of fit ratios were twice as large in the adjusted TC data as compared to the TTC data. It appears that while learning was occurring, the "noise" or extraneous factors prevent the fitting of a smooth curve to the data. Those factors can be attributable to circumstances such as multiple misfires, mechanical or firing system failures, and where ammunition had to be drawn from storage wells. It is noted that a multi-parameter polynomial model may have fit the data but it was intuitive that a learning curve would be a smooth curve rather than a "zig-zag" curve in the case of a polynomial.

The parameter estimates for the two test courses are shown

Table 4-5. Comparative Results for Fitted Models (TC)

Model	SS_E	SS_{LOF}	SS_R	Lack of Fit Ratio
$\hat{Y} = at^{-b}$	2468607.8	493855.022	2704131.2	3.7513
$\hat{Y} = ae^{bt}$	2440991.9	466239.122	2731747.1	3.5415
$\hat{Y} = ae^{b/t}$	2514968.3	540215.522	2657770.7	4.103
$\ln\hat{Y} = \ln a - b\ln t$	389.61005	105.07188	16.90632	5.5391
$\ln\hat{Y} = \ln a + bt$	381.08081	96.54264	25.43555	5.0895
$\ln\hat{Y} = \ln a + b/t$	404.16331	119.625	2.35305	6.3063

SS_{PE} = 1974752.77787 (Nonlinear models)

SS_{PE} = 284.53817 (Log transform models)

Table 4-6. Comparative Results for Fitted Models (TC)

(Adjusted Data)

Model	SS_E	SS_{LOF}	SS_R	Lack of Fit Ratio
$\hat{Y} = at^{-b}$	513771.03	123782.0216	1321367.97	4.141
$\hat{Y} = ae^{bt}$	496887.26	106898.2516	1333251.74	3.576
$\hat{Y} = ae^{b/t}$	551335.23	161346.222	1283803.77	5.398
$\hat{Y} = \frac{a}{t+b} + c$	547924.68	157935.6716	1287214.32	5.284
$\hat{Y} = a[\beta+(1-\beta)t^{-b}]$	506392.74	116943.7316	1328206.26	3.913
$\hat{Y} = \alpha(a^{t-1}) + \beta$	562010.21	172021.2016	1273128.79	5.755

SS_{PE} = 389989.00838

NOTE: Atypical points removed from data.

below for both the power function and the exponential models.

TTC

$Y = at^{-b}$

$a = 104.595 \quad b = .26492$

$Y = ae^{bt}$

$a = 74.1207 \quad b = -.019076$

TC

$Y = at^{-b}$

$a = 76.3596 \quad b = .180306$

$Y = ae^{bt}$

$a = 67.5696 \quad b = -.017967$

A comparison indicates that the TTC model parameters are relatively larger than those for the TC. In addition, the learning factor which is represented by the parameter b, appears to be larger for the Test Training Course.

Lightweight Company Mortar System

The 81 mm Gunner's examination was conducted to establish baseline data to use in comparing the 81 mm mortar with the XM 224E1 Lightweight Company Mortar System. The purpose of the test was to establish the time it takes to set up and perform a mortar fire mission and to refamiliarize the test crews with the 81 mm mortar so that they may be better able to compare it with the XM 224E1. A summarized description is given below.

1. Performance Measure - Gunner's Examination Scores

2. Characteristics

(a) Two systems tested
(b) 3 mortar squads
(c) Number of observations 4 - 81 mm mortar
3 - XM224E1 mortar

(d) Type of activity - Performance Test

Seven complete gunner's examination were performed during OTI; four for the 81 mm mortar and three for the XM 224E1/LWCMS. The latter was not analyzed, even though there appeared to be learning patterns in the data, because there were only three distinct trials and since our learning models contain at least two unknown parameters, further analysis would be misleading. However, the four trials for the 81 mm mortar data were analyzed. At each trial or phase, there were six tasks performed:

(1) Mounting the mortar
(2) Small deflection and elevation change
(3) Referring the sight
(4) Large deflection and elevation change
(5) Reciprocal laging
(6) Manipulation for traversing

A plot of the data is shown in Figure A-21. The background information indicates that the initial times required to perform the phases of the gunner's examination were high due to the fact that the test platoon had not worked with mortars for several weeks and their level of training was low. Upon completion of the training program, times to perform the phases of the gunner's examination were minimized. (34)

The plot of the scores over consecutive trials (phases) show a discernible pattern which indicates learning. In addition when the linear model was fit in step 2, the following results were indicated.

Source	d.f.	Sum of Squares	Mean Square
Regression	1	15732.300	15732.300
Residuals	22	12880.200	585.46364

$$\text{F-ratio} = \frac{15732.30}{585.46364} = 26.87152$$

When compared to the F-distribution value for 1 and 22 degrees of freedom at the 5 percent level, there is evidence to reject that $\beta_1 = 0$. Additionally, the estimate of the negative slope ($\beta_1 = -22.9$) and the confidence interval around β_1 did not include zero, therefore the sample data was concluded to be suitable for further analysis.

The nonlinear model $\hat{Y} = at^{-b}$ was fit and the results are shown below.

$$\hat{Y} = 115.139\ t^{-.5837}$$

Source	d.f.	Sum of Squares	Mean Square
Regression	2	154831.0	
Residuals	22	13549.849	
(Lack of Fit)	2	2054.849	1027.425
(Pure Error)	20	11495.0	574.75

$$\text{Lack of Fit ratio} = \frac{1027.425}{574.75} = 1.788$$

Team Training

An air traffic control task was used in which each of two teammates portrayed a "pattern feeder" whose responsibility it was to guide aircraft into an approach gate by issuing verbal instructions via a simulated radio linked to the aircraft pilots. Two variables were manipulated in Experiment VIII: work load (for time stress) and team arrangement. Stress is defined in terms of the required approach rate (system criterion): one approach every 2 minutes for low stress, and one every minute for high stress. Team arrangement was defined in terms of the manner in which the two teammates coordinated, in order to satisfy the system criterion. The two team arrangements used were termed reciprocal and nonreciprocal. In the nonreciprocal arrangement the team was instructed to satisfy the low-stress criterion on each approach, independently of any time error incurred on previous approaches. In

The reciprocal arrangement, on the other hand, each radar controller (RC) was instructed to compensate for any time error which may have accrued over the previous approaches. A summarized description is presented below.

1. Performance Measure - Flight Errors by all groups of Experiment VIII

2. Characteristics

   (a) 4 groups
   (b) Two groups used reciprocal arrangement under both high and low stress conditions
   (c) Two groups used nonreciprocal arrangement under both high and low stress conditions
   (d) Four sessions (trials) for each group

A plot of data from Experiment VIII of the test report shows the performance measure, mean number of flight errors vs sessions (consecutive trials). The graph shows patterns which appear to indicate learning (see Figure A-22). The linear model was fit in step 2 of the iterative analysis process with the folloiwng results.

Source	d.f.	Sum of Squares	Mean Square
Regression	1	784.37812	784.37812
Residuals	14	689.48125	49.24866

$$\text{F-ratio} = \frac{784.37812}{49.24866} = 15.92689$$

When compared to the F-distribution value for 1 and 14 degrees of freedom at the 5 percent level, there is evidence to reject the hypothesis that $\beta_1 = 0$. Additionally, the estimate of the slope was negative ($\beta_1 = -6.2625$) and the confidence interval around β_1 did not include zero, therefore the sample data was concluded to be suitable for further analysis.

The nonlinear model $\hat{Y} = at^{-b}$ was fit and the results are shown below.

$$Y = 25.3582\ t^{-1.00391}$$

Source	d.f.	Sum of Squares	Mean Square
Regression	2	3954.11	
Residuals	14	589.140	
(lack of fit)	2	1.6875	0.72625
(Pure Error)	12	587.6875	48.974
TOTAL	16	4243.25	

$$\text{Lack of Fit ratio} = \frac{0.72625}{48.974} = .01483$$

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

This research has addressed the problem of determining the existence of a representative group/crew learning curve (or set of curves) and the development of a mathematical description of this curve applicable to training levels in operational testing. Data from OTEA test reports and data made available through other training and training analysis agencies was analyzed using an iterative procedure to determine if learning patterns could be detected.

A screening process was used to determine the suitability of data for further analysis, after which learning models suggested in the literature were fit to the screened data using nonlinear regression techniques. A comparison of the fitted models was conducted by comparing the Lack of Fit ratios and the sum of squares for regression computed for each model.

This comparison shows that the following models appear to provide an adequate fit to the data analyzed.

(1) $\hat{Y} = at^{-b}$ The power function

(2) $\hat{Y} = a[\beta + (1-\beta)t^{-b}]$ De Jong's model

(3) $\hat{Y} = at^{-b} + c$

(4) $\hat{Y} = ae^{bt}$

Since the variations of the power function, models (2) and (3) did not appear to provide a better fit to the data, model (1) was selected from the standpoint of parsimony or least parameters. In addition, it cannot be stated conclusively that model (1) provides a better fit than model (4). However, based on a survey of the industrial applications of the power function model as reported in the literature, it was concluded that the model $\hat{Y} = at^{-b}$ does adequately fit the empirical data analyzed and can be used as a representative group/crew learning model for this data.

Limitations of the Research

This research has been limited by the availability of adequate data representing several different crew and group learning situations. The lack of a larger data base limited the analysis to a small number of performance measures. These included tracking, tank gunnery and mortar examination scores. Since the analysis of a large number of data sets involving a variety of crew tasks and performance measures was not possible, this study concentrated on the analysis of suggested learning models for the limited data available.

Considerations for Test Design

Even though there is a limited amount of data available in the group/team context as discussed previously, future data may be analyzed using the iterative procedures developed in Chapter III. However, a review of the literature indicates that the following considerations should be made when providing input for the design of operational tests.

1. Insure that individual skill competencies are acquired prior to engaging in team training or testing. A consistent finding was that individual proficiency has been shown to be a significant factor in determining team performance (24).
2. Address the problem involved in the production of standardized

replicable test conditions, and the establishment of accepted group/team performance criteria by defining the tasks characteristics needed to identify realistic training objectives (24). These particular aspects are not clearly defined in current literature but objectives are outlined in these references (24,47,48,49, 50).

3. Distinguish between organizational type tasks and mission type tasks.

4. The detection, measurement, and recording of the value of an observable event at each occurrence (24). Current tests which use blocking and randomized test design should provide a vehicle for recording these consecutive occurrences in addition to recording the cell totals.

5. Assessment of learning effects. Procedures developed by Yealy (51) could be used to determine rate of learning at a specific trial during an operational test. These procedures could be employed in two ways: (a) Conduct initial stages of test in a sequential fashion, say, for the first three trials to determine rate of learning if any. If the rate of learning leveled off, then the participants are assumed to be at or approaching a fully learned state and the test could continue with learning effects considered negligible. On the other hand, if the rate of learning has not leveled off, then the test should be continued in a sequential fashion until learning effects become negligible. However, this approach appears to be too costly in terms of manpower and resources. An

alternative approach would be, (b) conduct a pretest and determine rate of learning at each trial. When a satisfactory level of learning is reached then the operational test could begin.

6. Avoid where possible the inclusion of order effects in the test design in which the participants, for example, learn where to look (learning the problem) rather than learning how to operate the equipment being evaluated.

Recommendations

The following recommendations for future research are made as a result of this study. One recommendation is the acquisition and analysis of more data using procedures outlined in Chapter III. Since this study was limited by the nonavailability of a large number of adequate data sets, further analysis of other sample data could be used to verify results obtained in the study. This would include the study of the adequacy of the power function, $\hat{Y} = at^{-b}$ vs the exponential model $\hat{Y} = ae^{bt}$ since both models appeared to fit sample data analyzed in this study. However, it could not be determined that the two models were statistically different.

Another recommendation involves the development of group/crew learning curves (or set of curves) for specific crews or units, i.e., Artillery battery, rifle squad, etc. Models should be developed on the basic research level to consider the interaction among crew members and a possible comparison of the performance by individuals and by the crew. This should be done because it appears that there is no single overall true model for all group learning. It is felt that since military teams or units are structured differently and have inherent mission capabilities, then the concept of an overall true model would not adequately reflect these differences.

APPENDIX A

This appendix contains representative plots used in the analysis of sample data in Chapter IV.

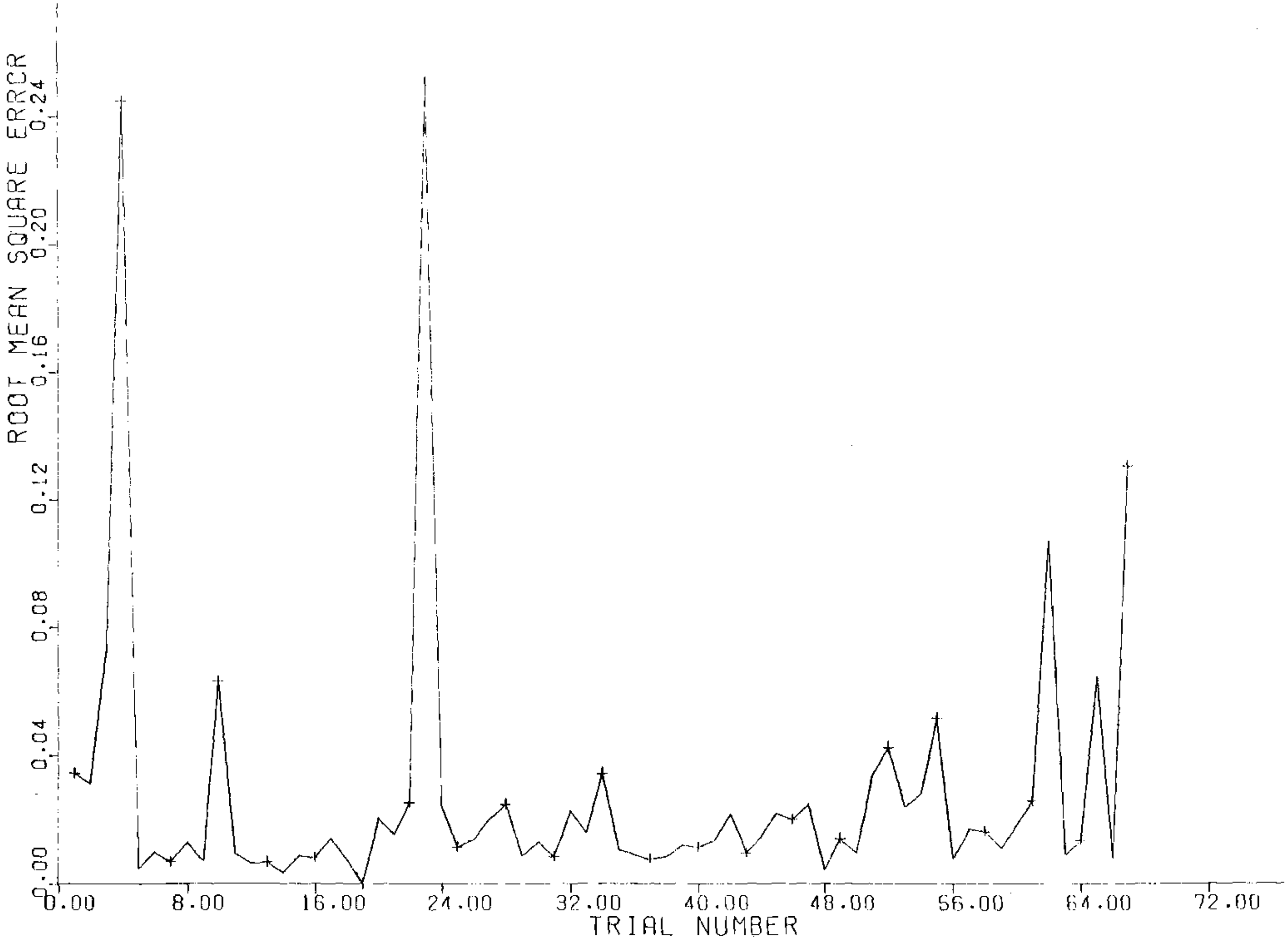


Figure A-1. Plot of Data for ITV Gun Crew (A-20)

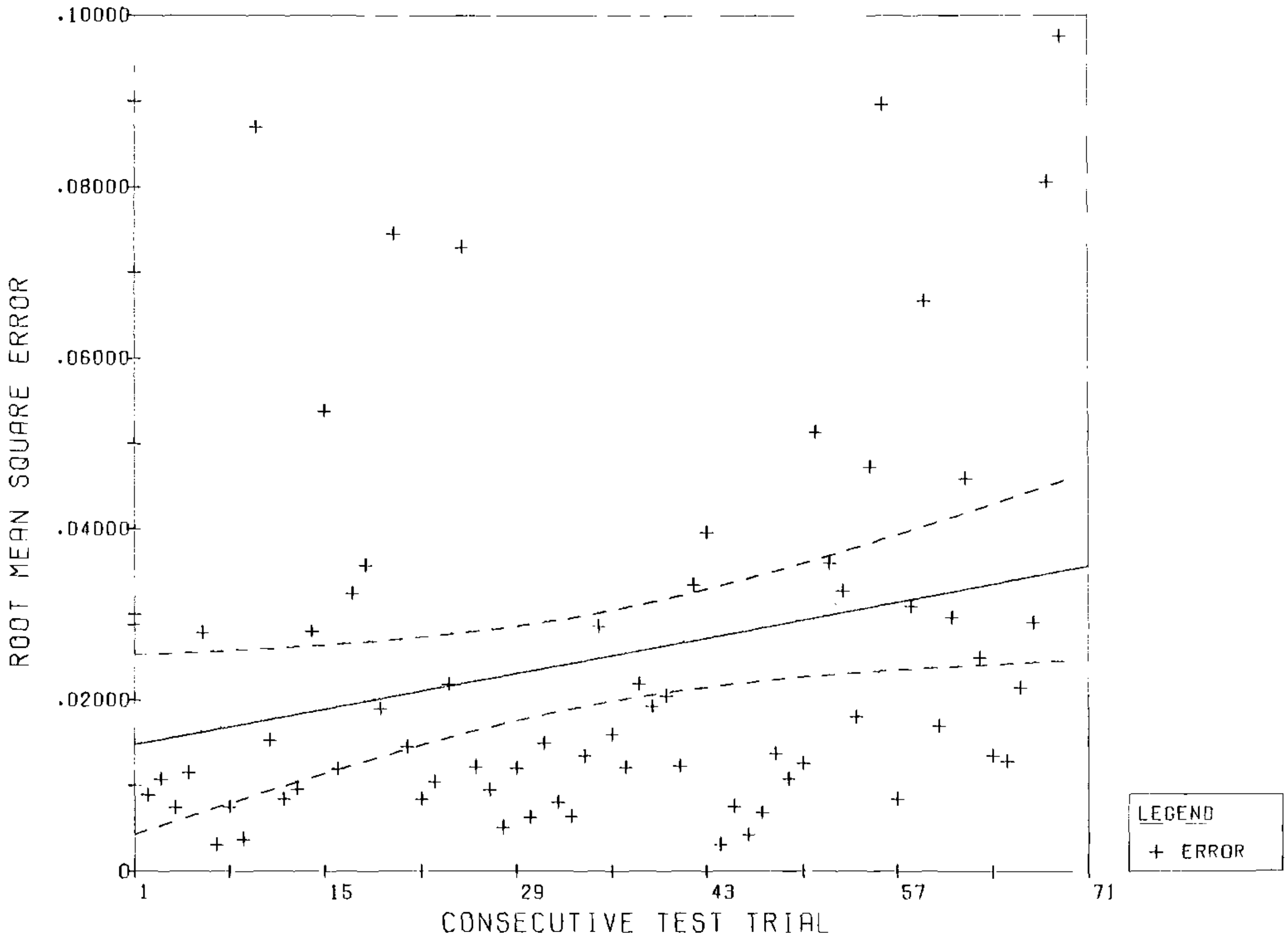


Figure A-2. Gun Crew (A-20) Data with Regression Line and Confidence Interval

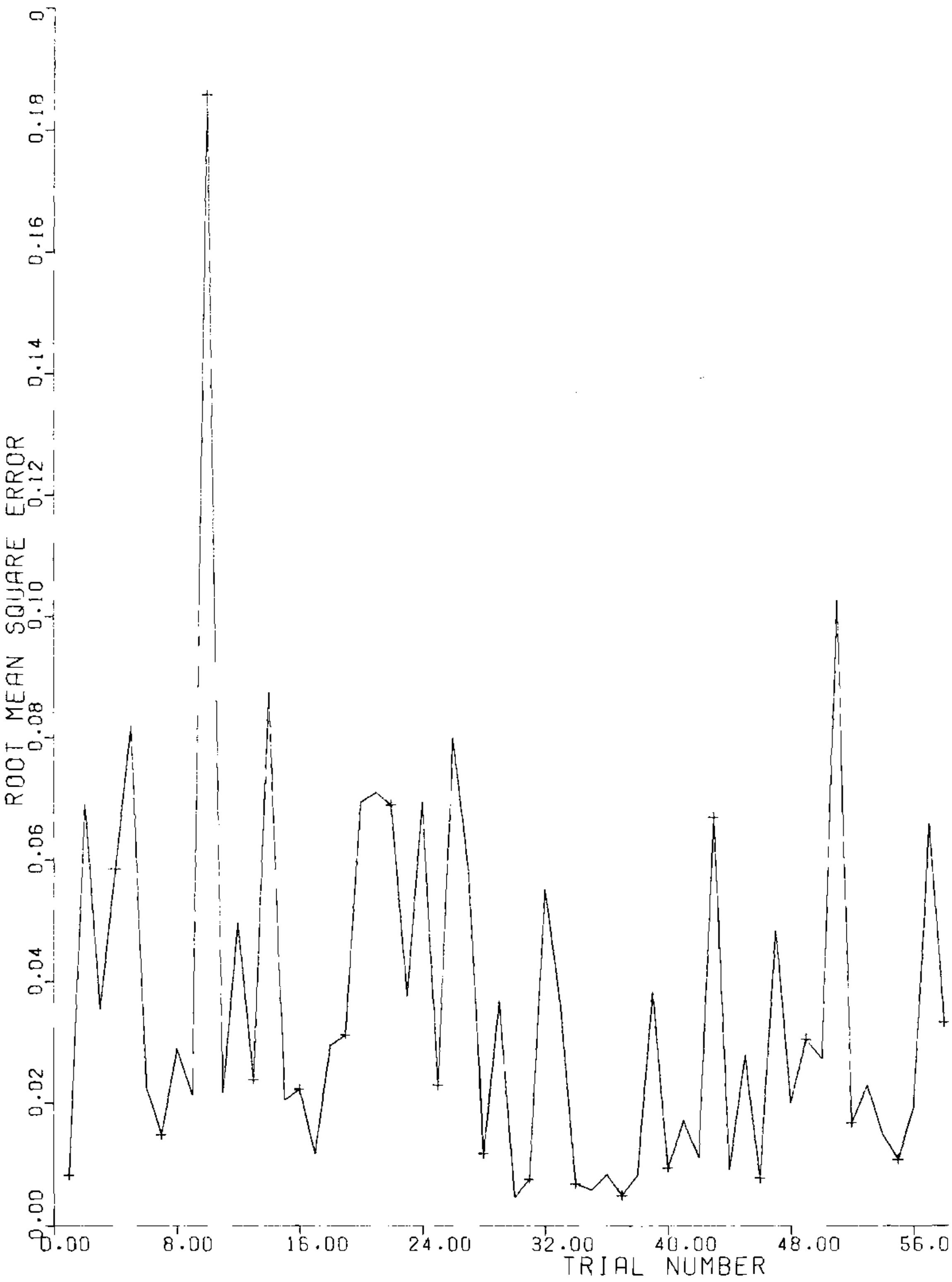


Figure A-3. Plot of Data for ITV Gun Crew (3-19)

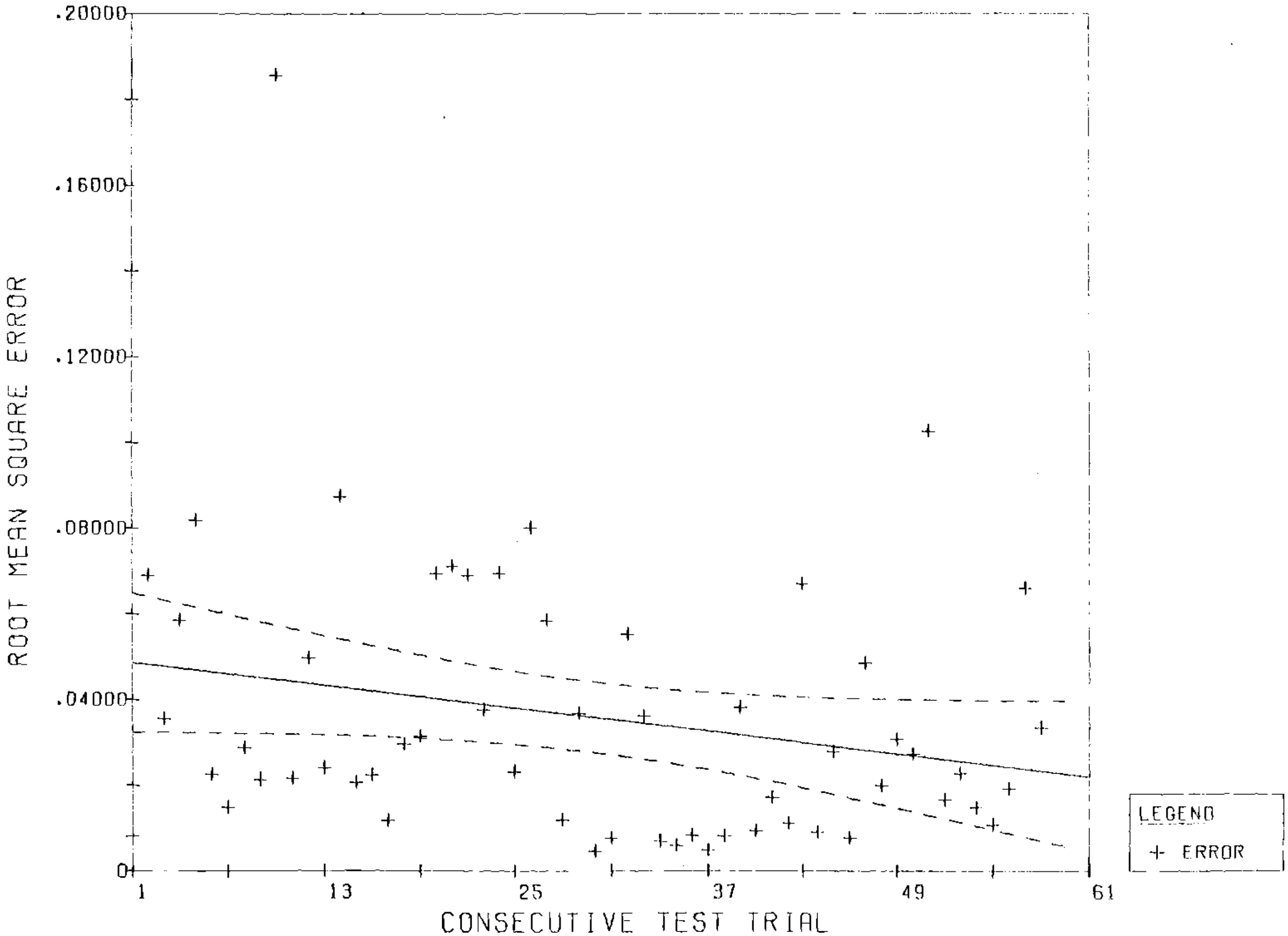


Figure A-4. Gun Crew (3-19) Data with Regression Line and Confidence Interval

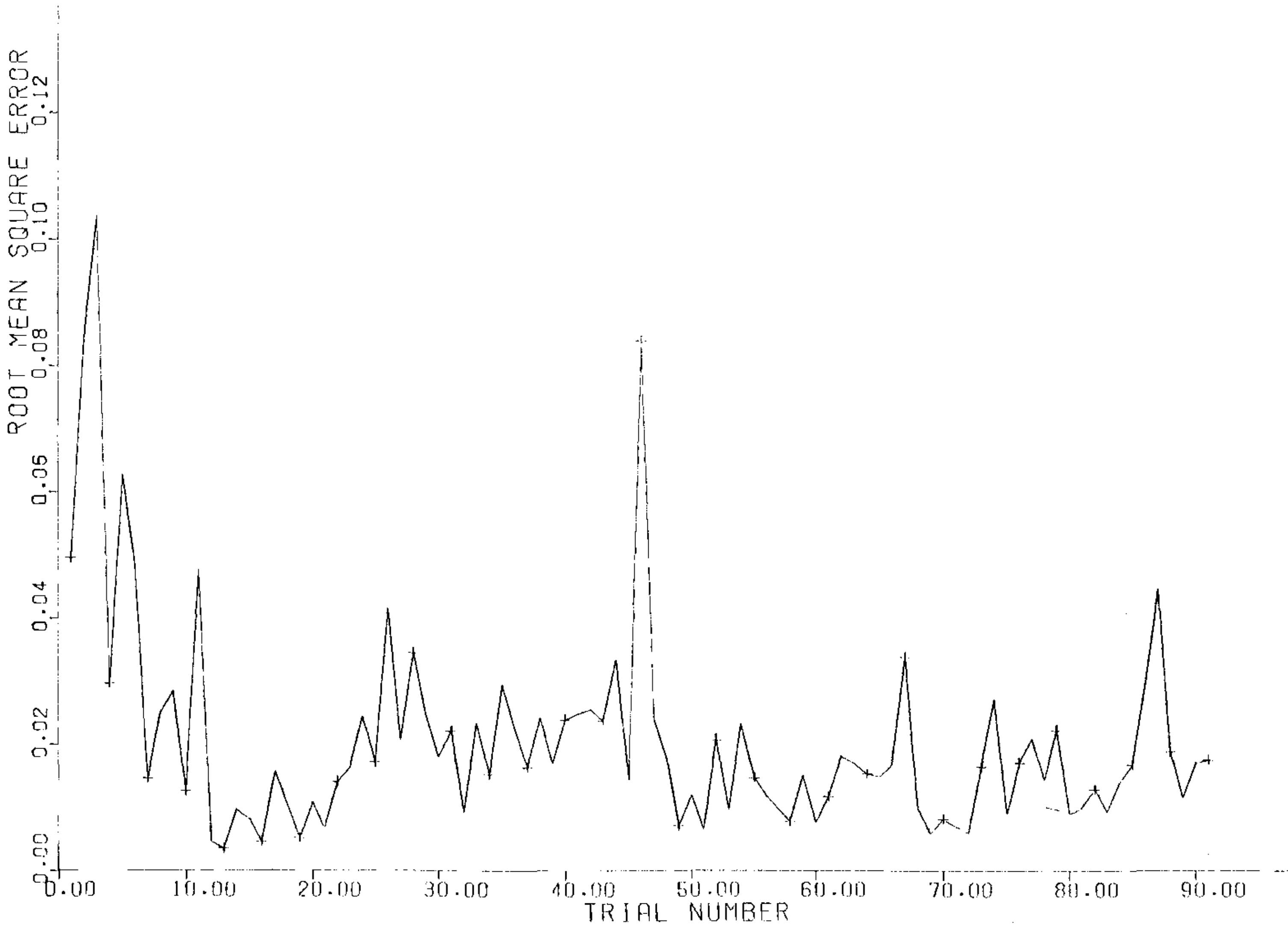


Figure A-5. Plot of Data for ITV Gun Crew (D-17)

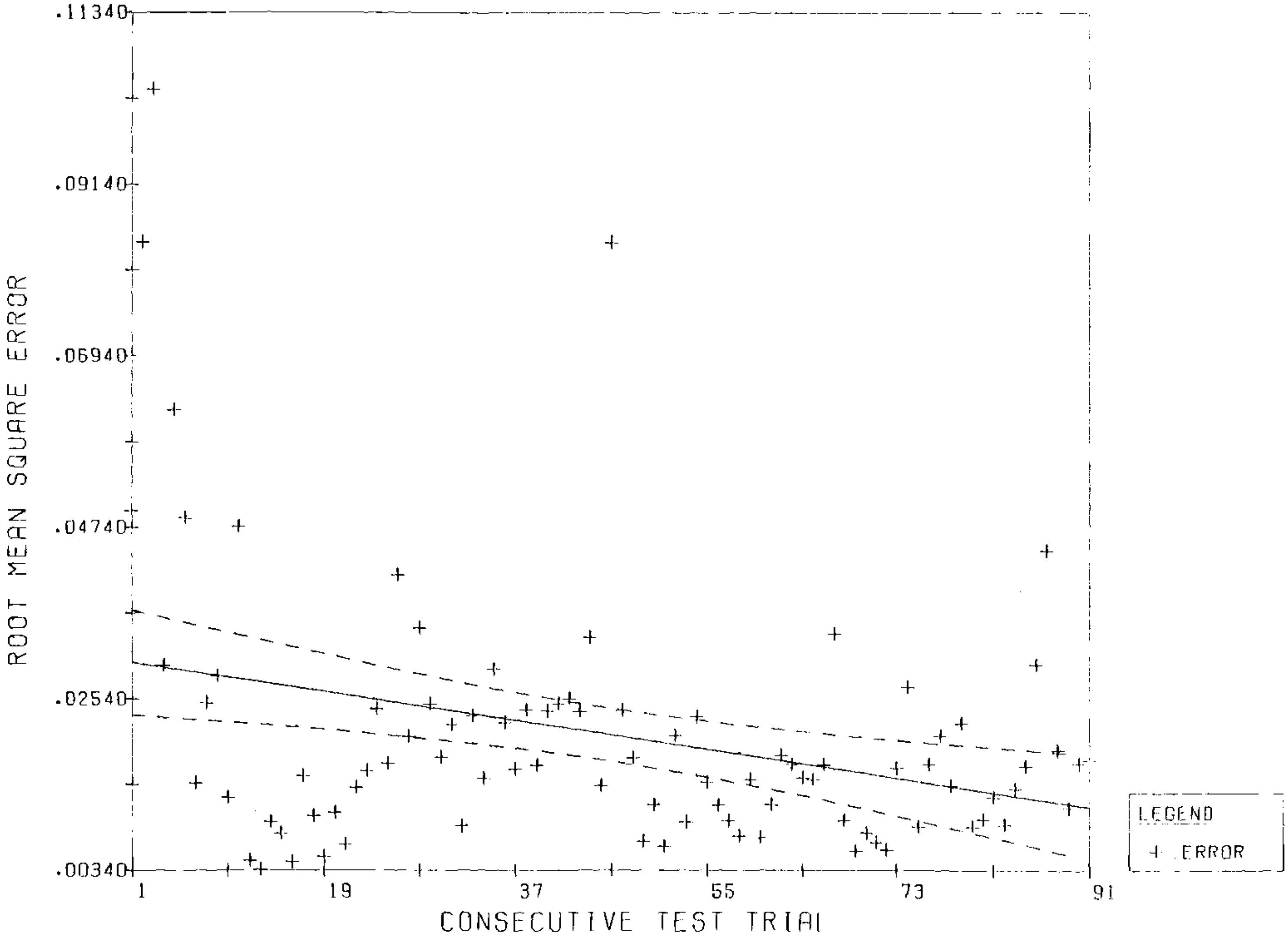


Figure A-6. Gun Crew (D=17) Data With Regression Line and Confidence Interval

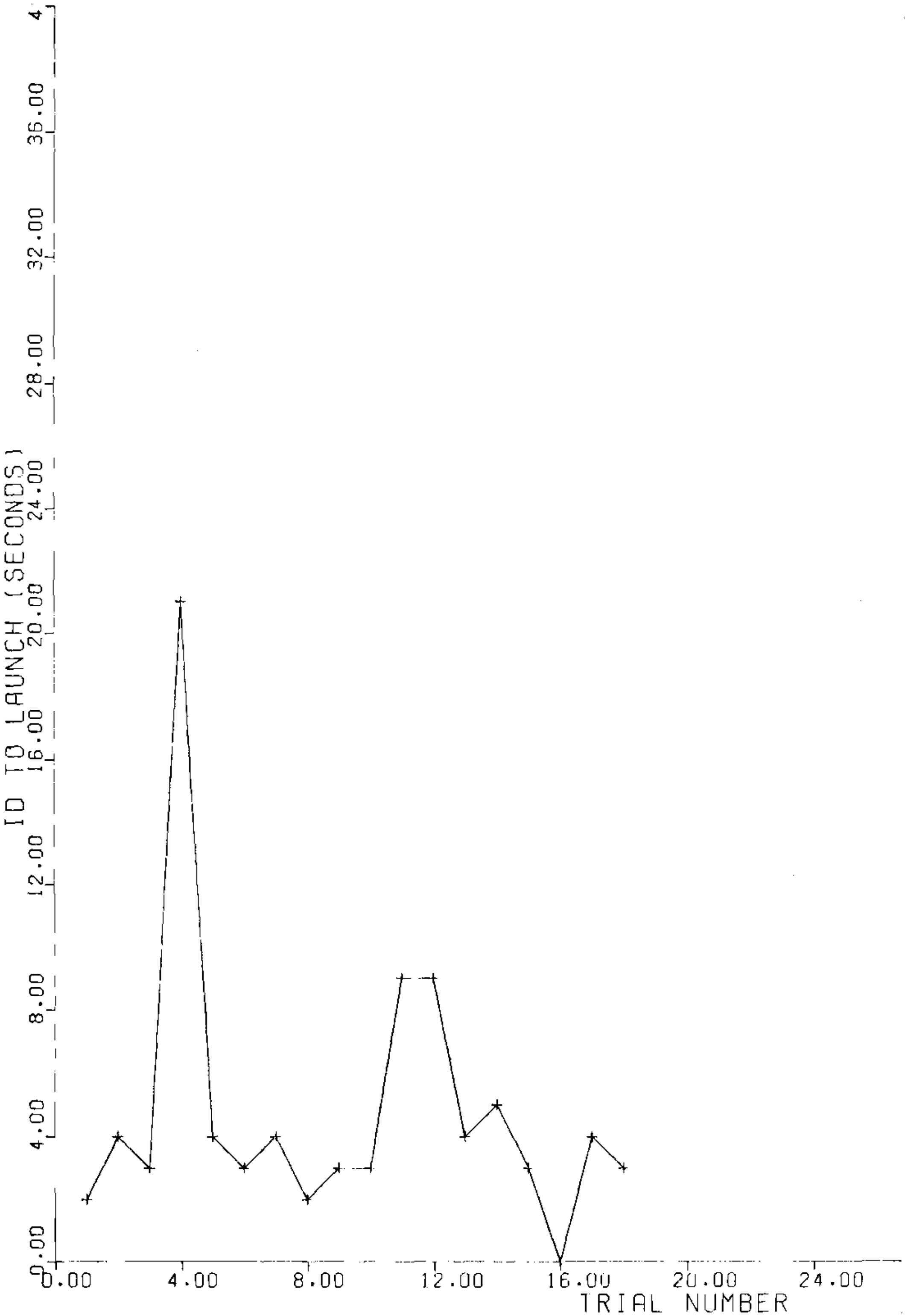


Figure A-7. Plot of Data for Dragon Gun Crew 0308(T_2)

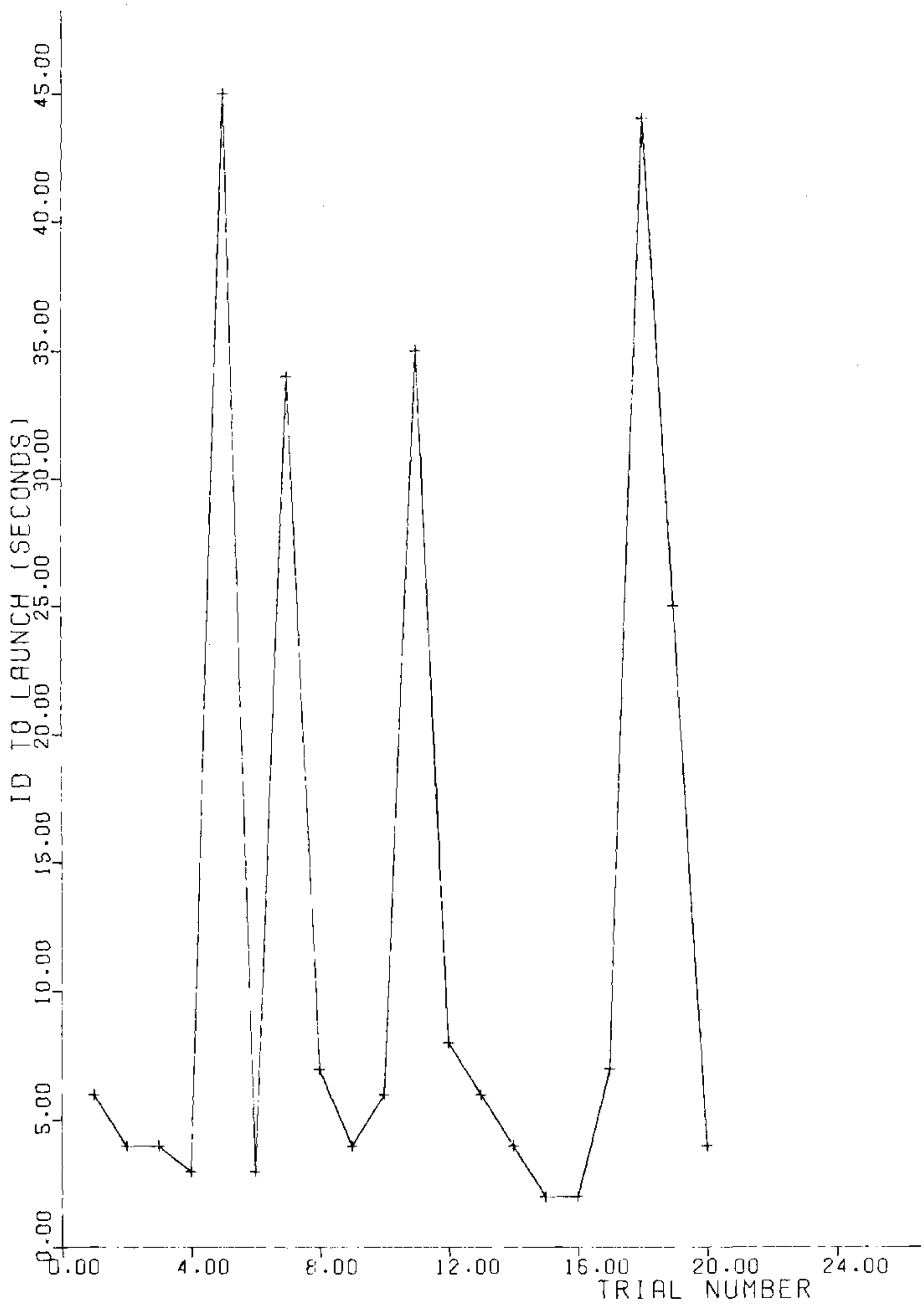


Figure A-3. Plot of Data for Dragon Gun Crew 2319(T2)

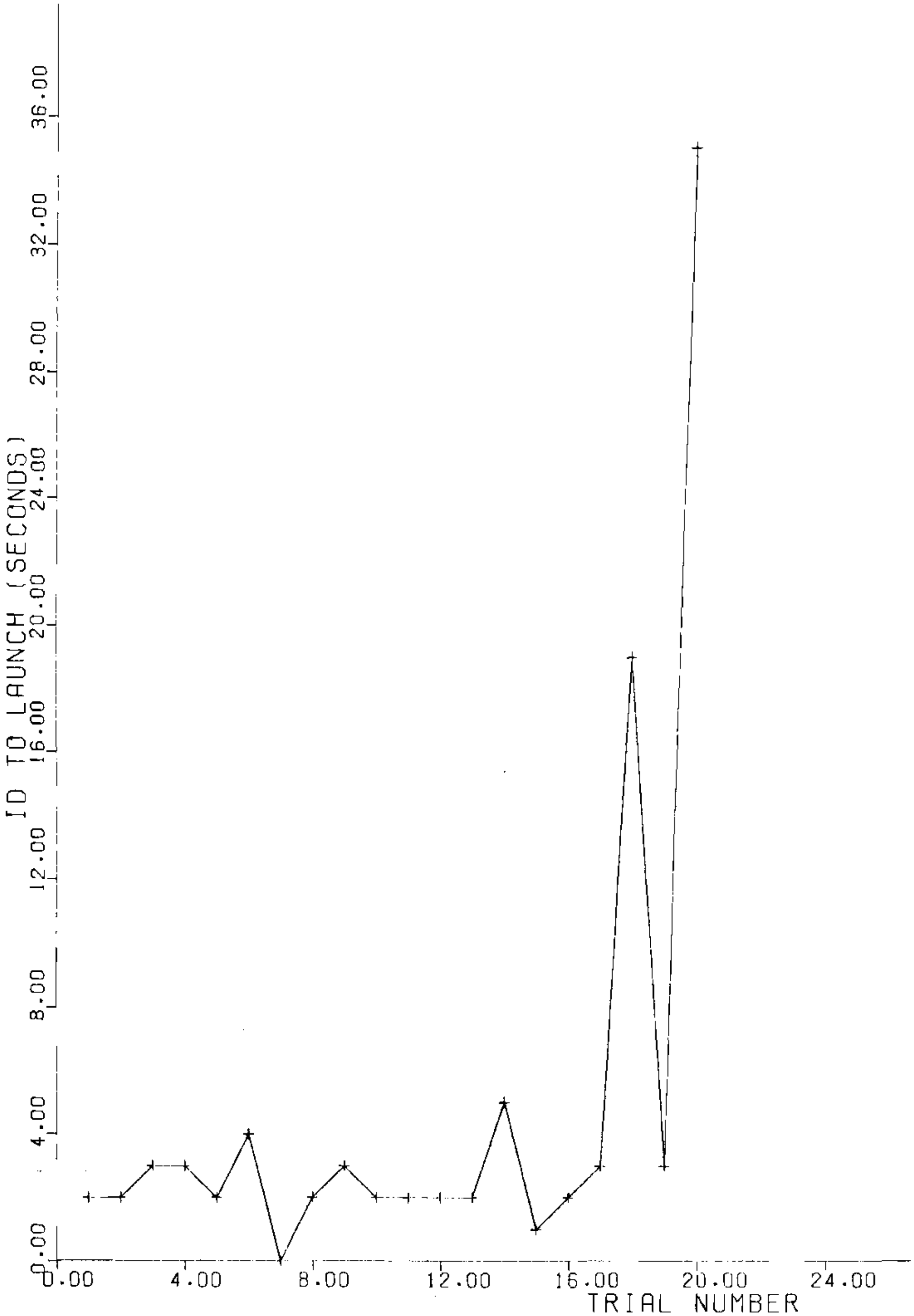


Figure A-9. Plot of Data for Dragon Gun Crew 2319(T4)

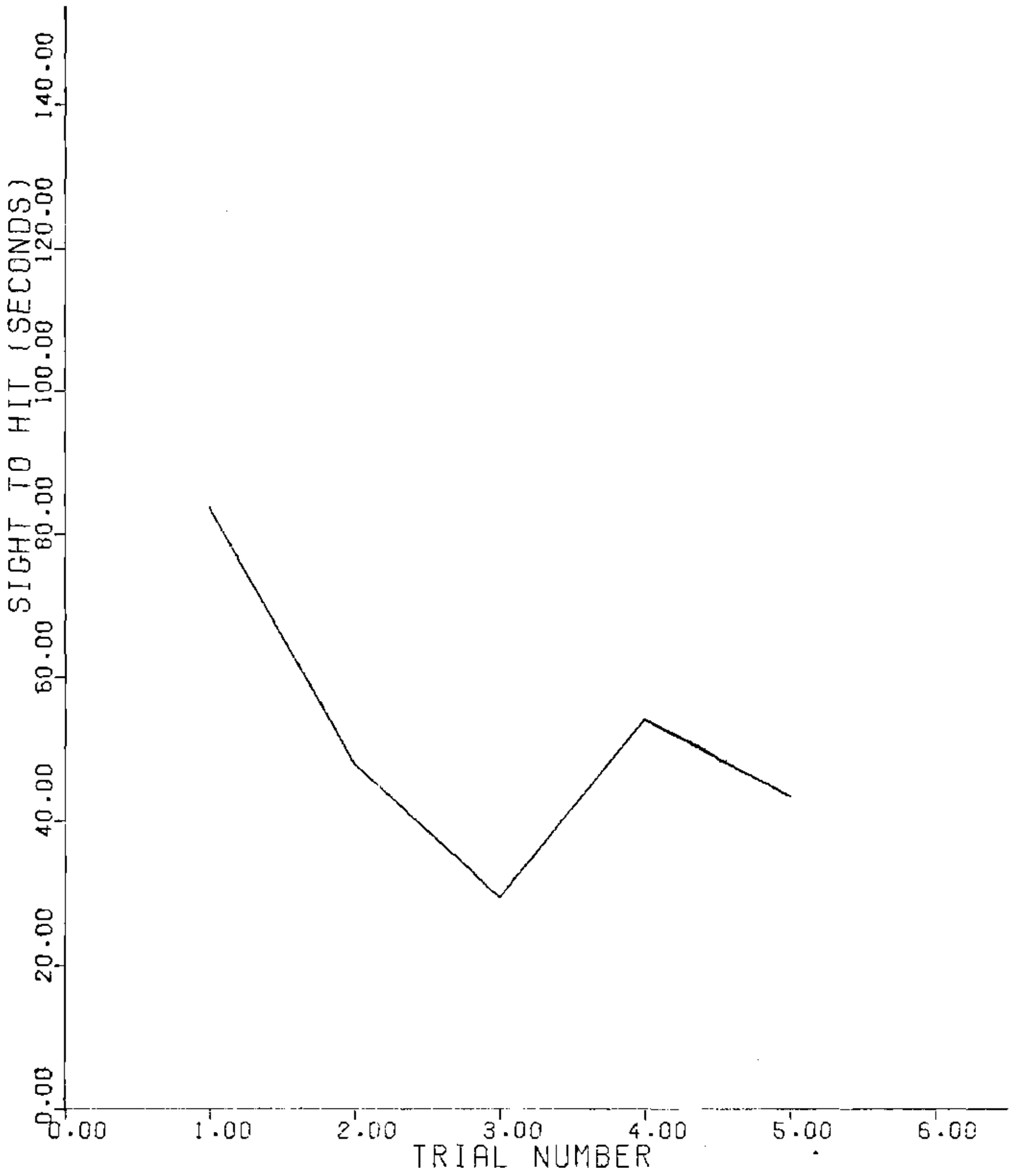


Figure A-10. Plot of Data for Project Stalk Crew 1 (TTC)

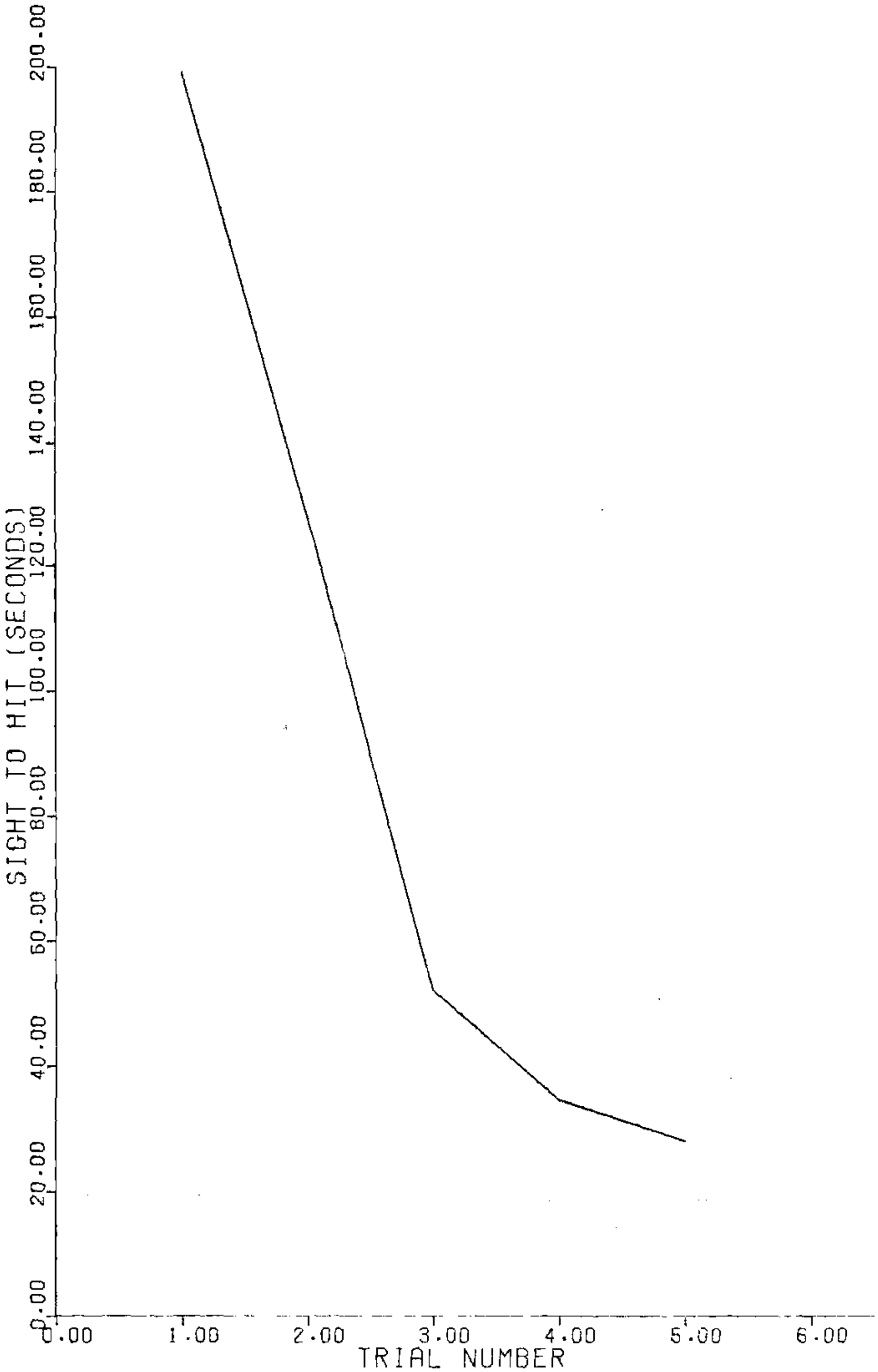


Figure A-11. Plot of Data for Project Stalk Crew 2 (TTC)

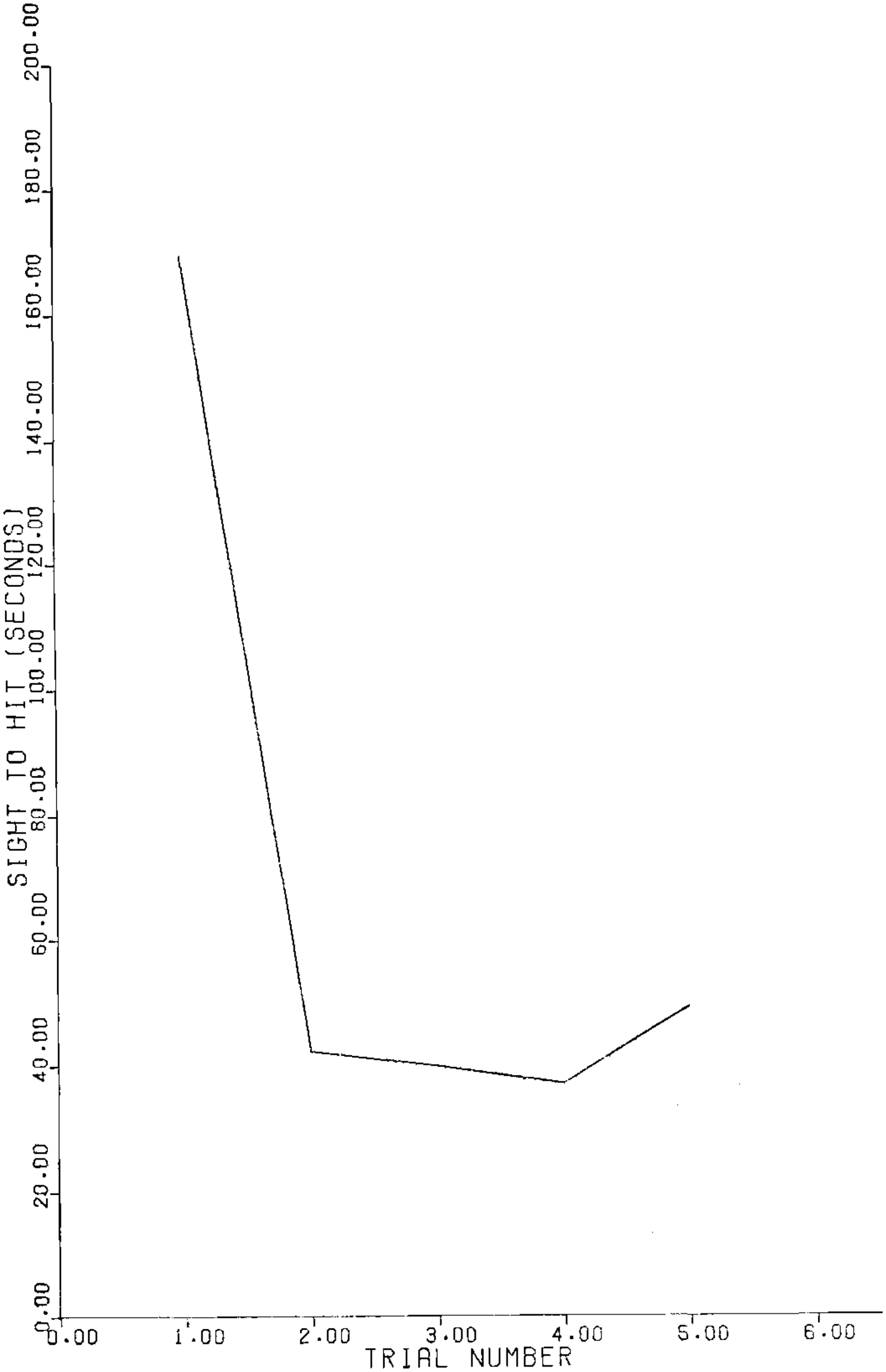


Figure A-12. Plot of Data for Project Stalk Crew 3 (TTC)

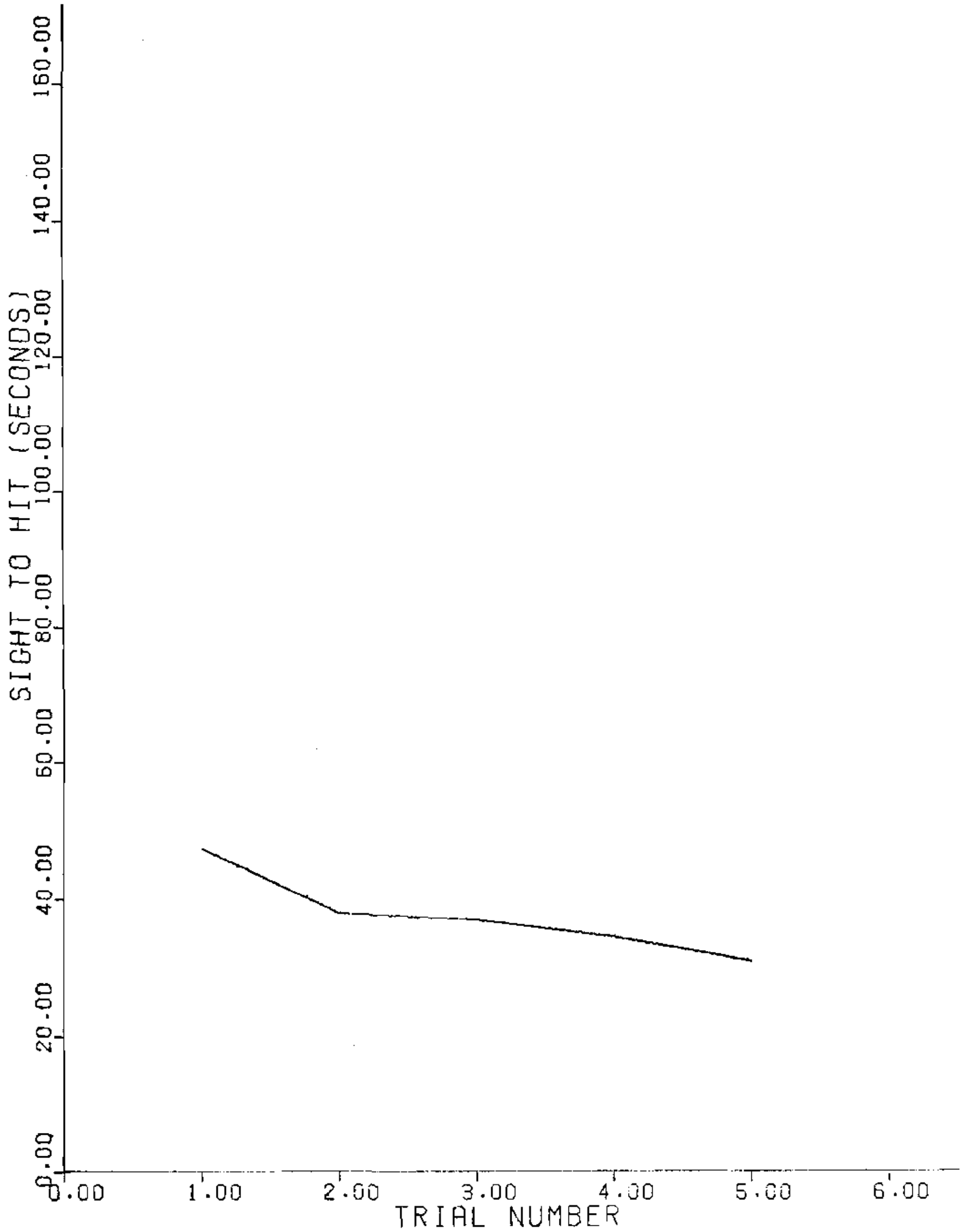


Figure A-13. Plot of Aggregate Data for all Crews for Project Stalk (TTC)

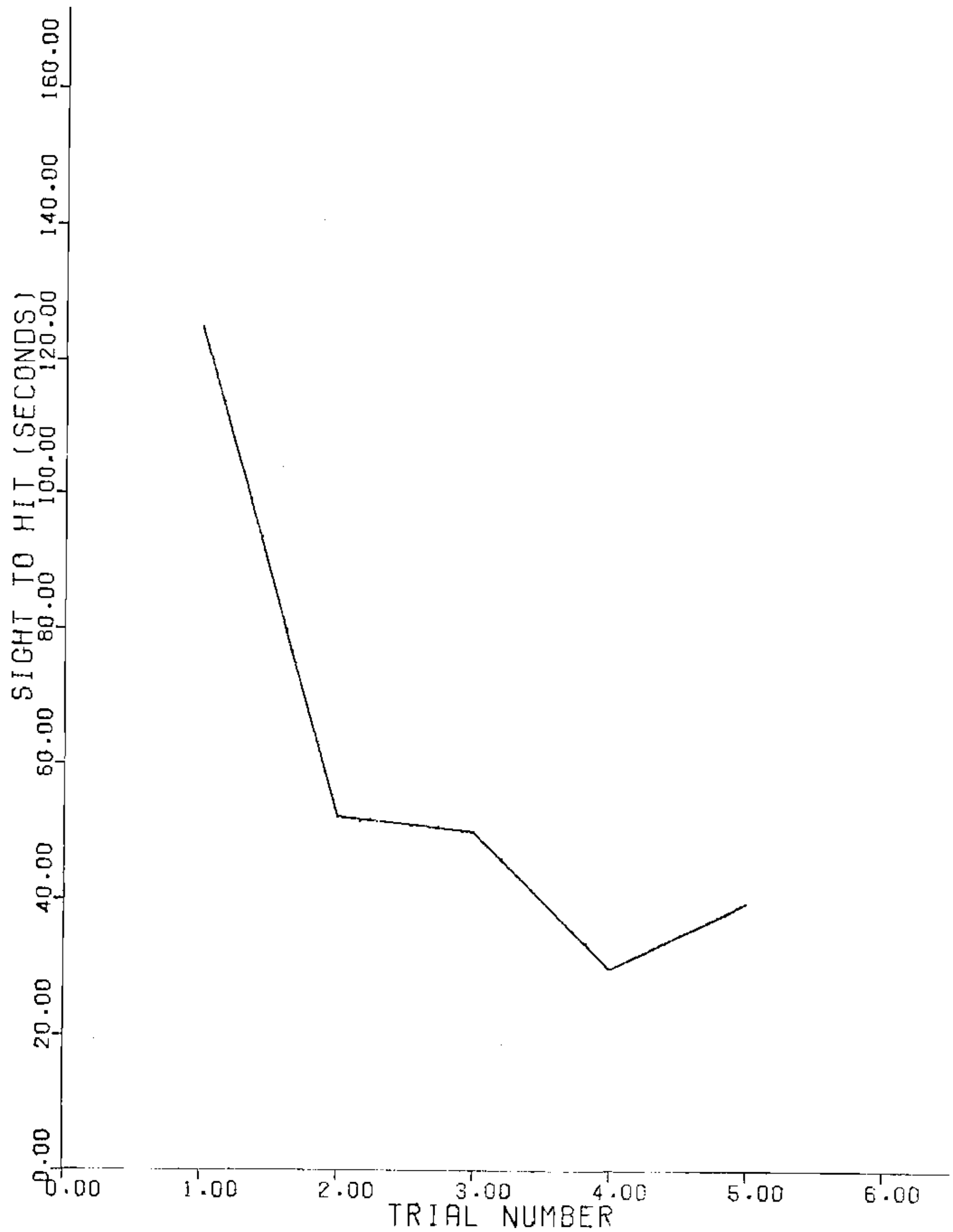


Figure A-14. Plot of Data for Project Stalk Crew 1 (TC)

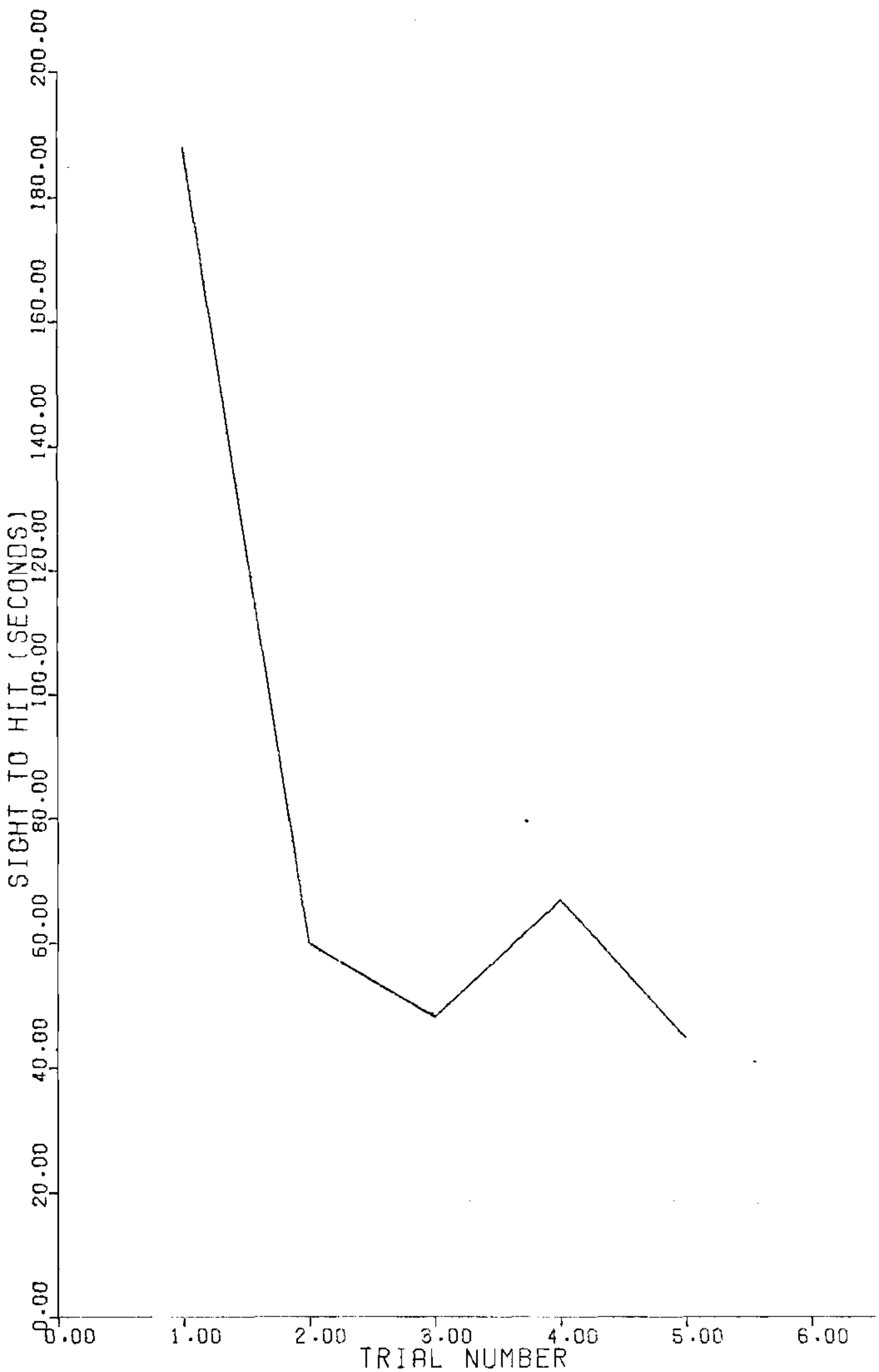


Figure A-15. Plot of Data for Project Stalk Crew 3 (TC)

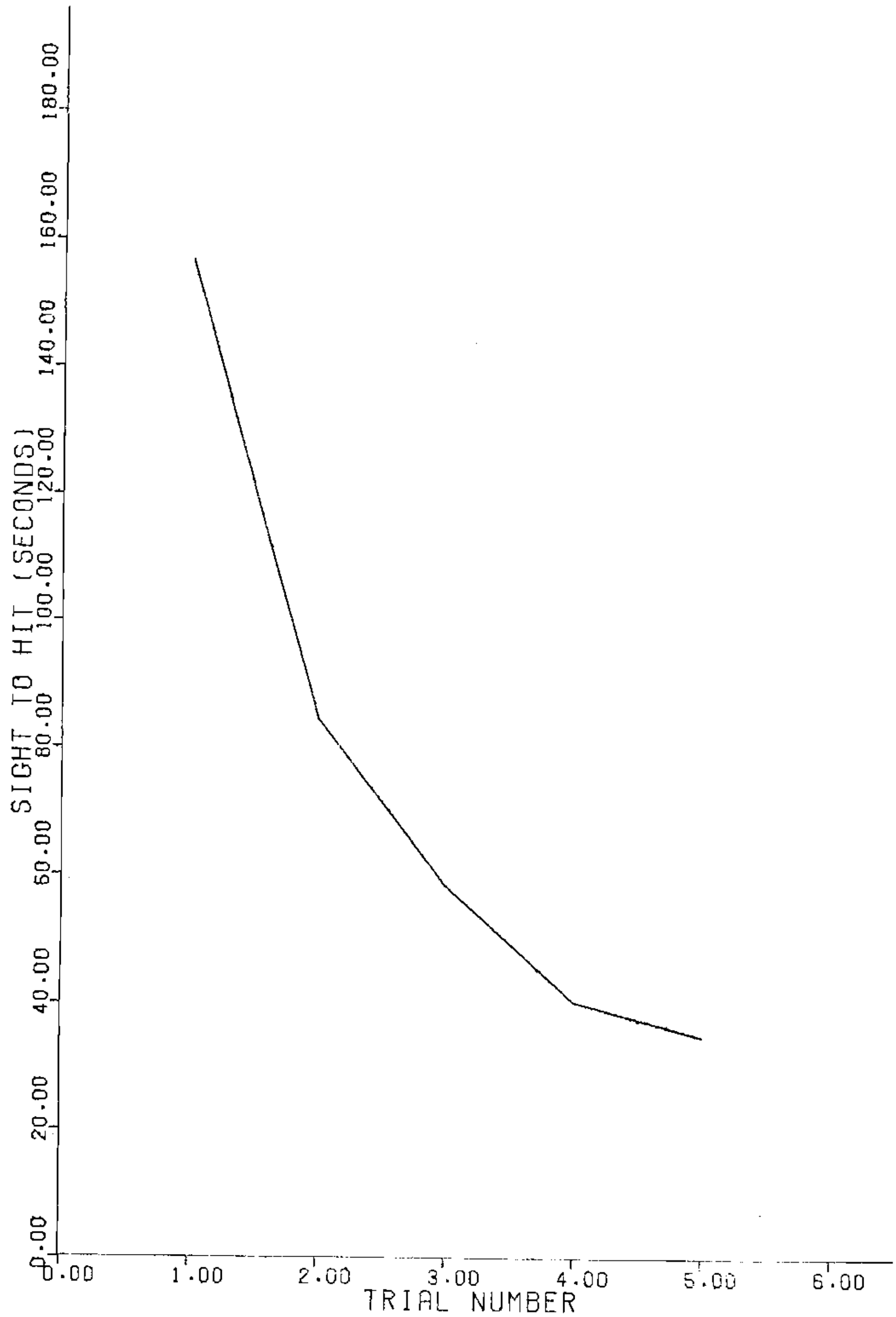


Figure A-16. Plot of Data for Project Stalk Crew 11 (TC)

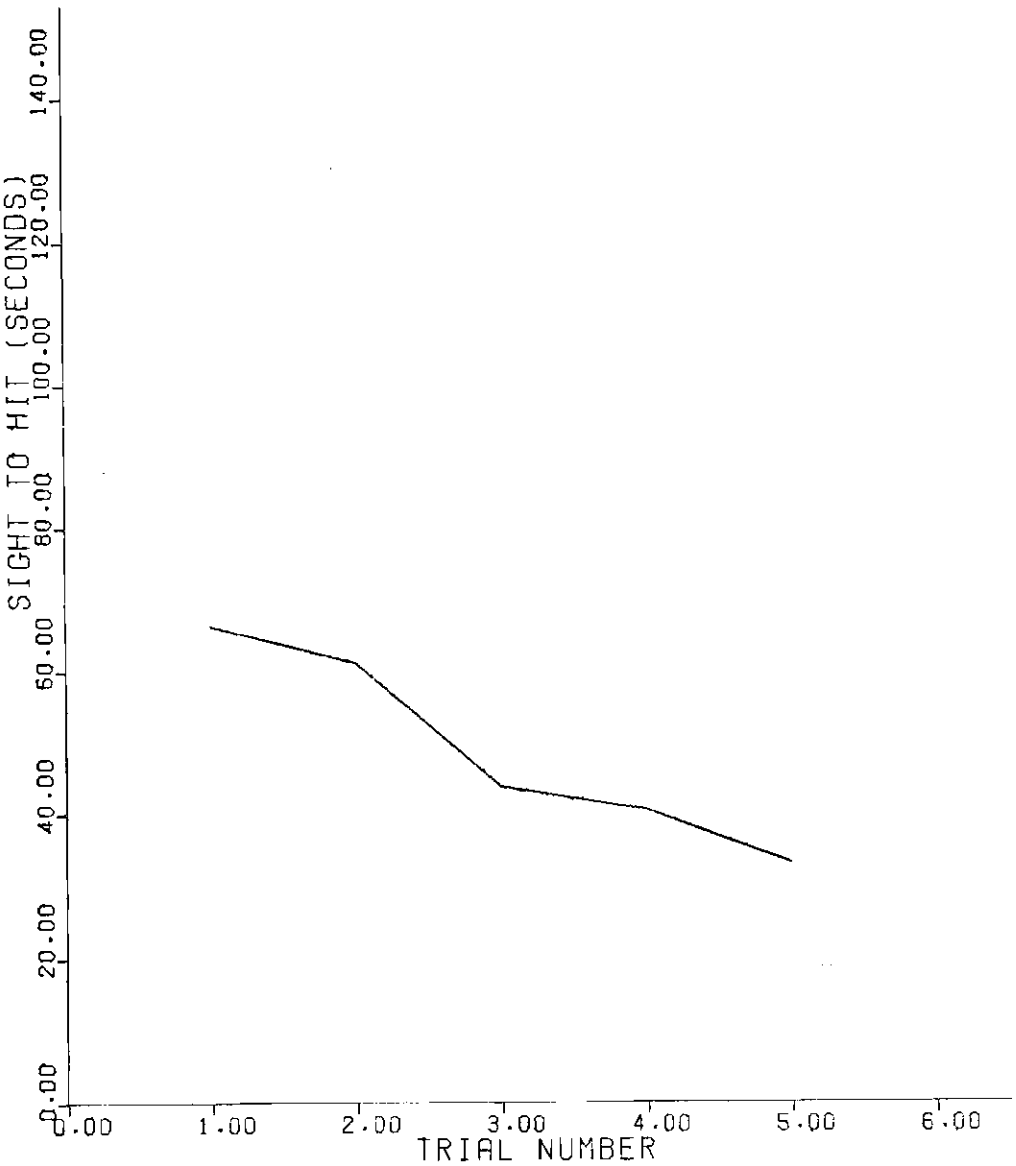


Figure A-17. Plot of Aggregate Data for All Crews for Project Stalk (TC)

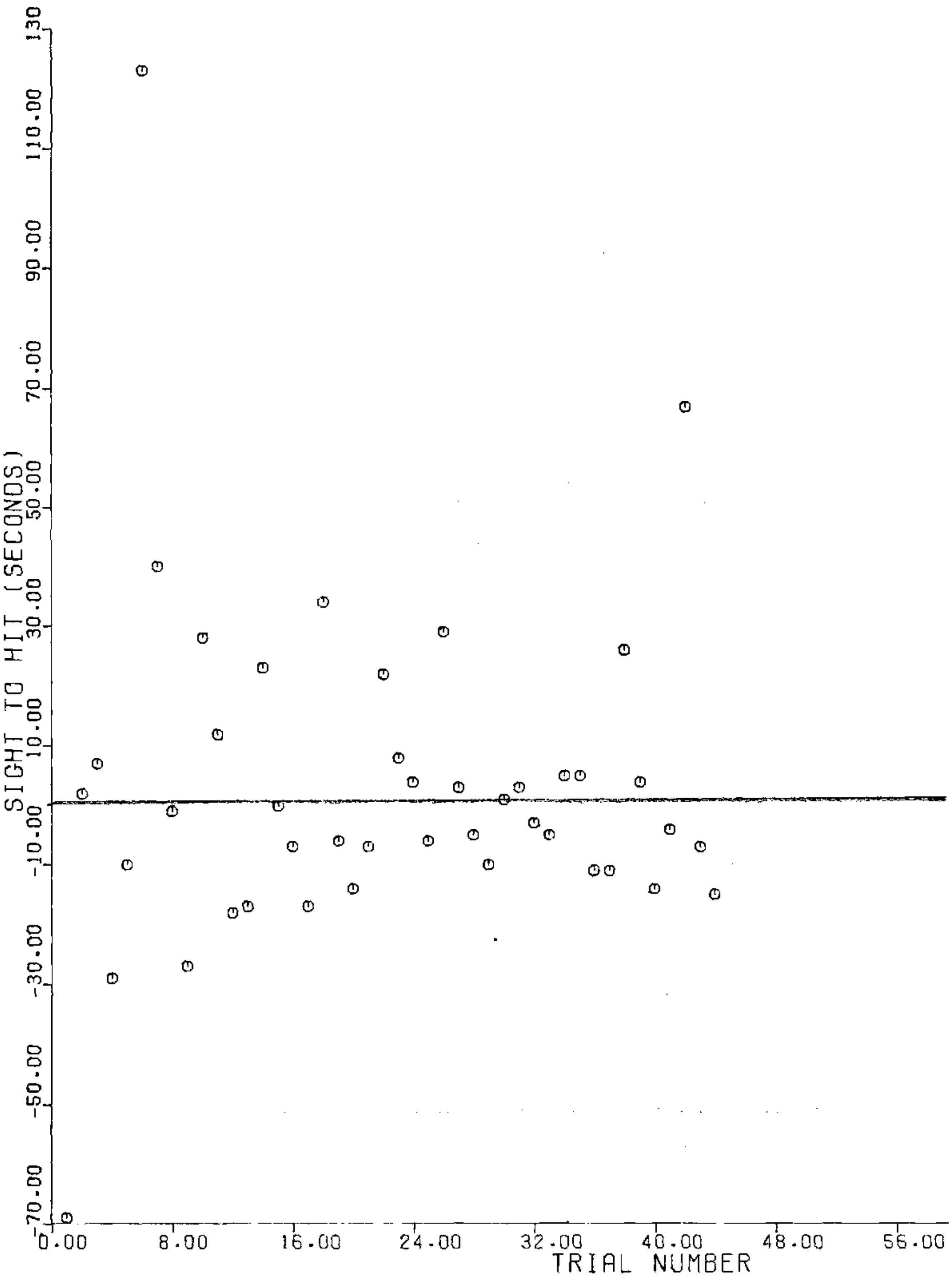
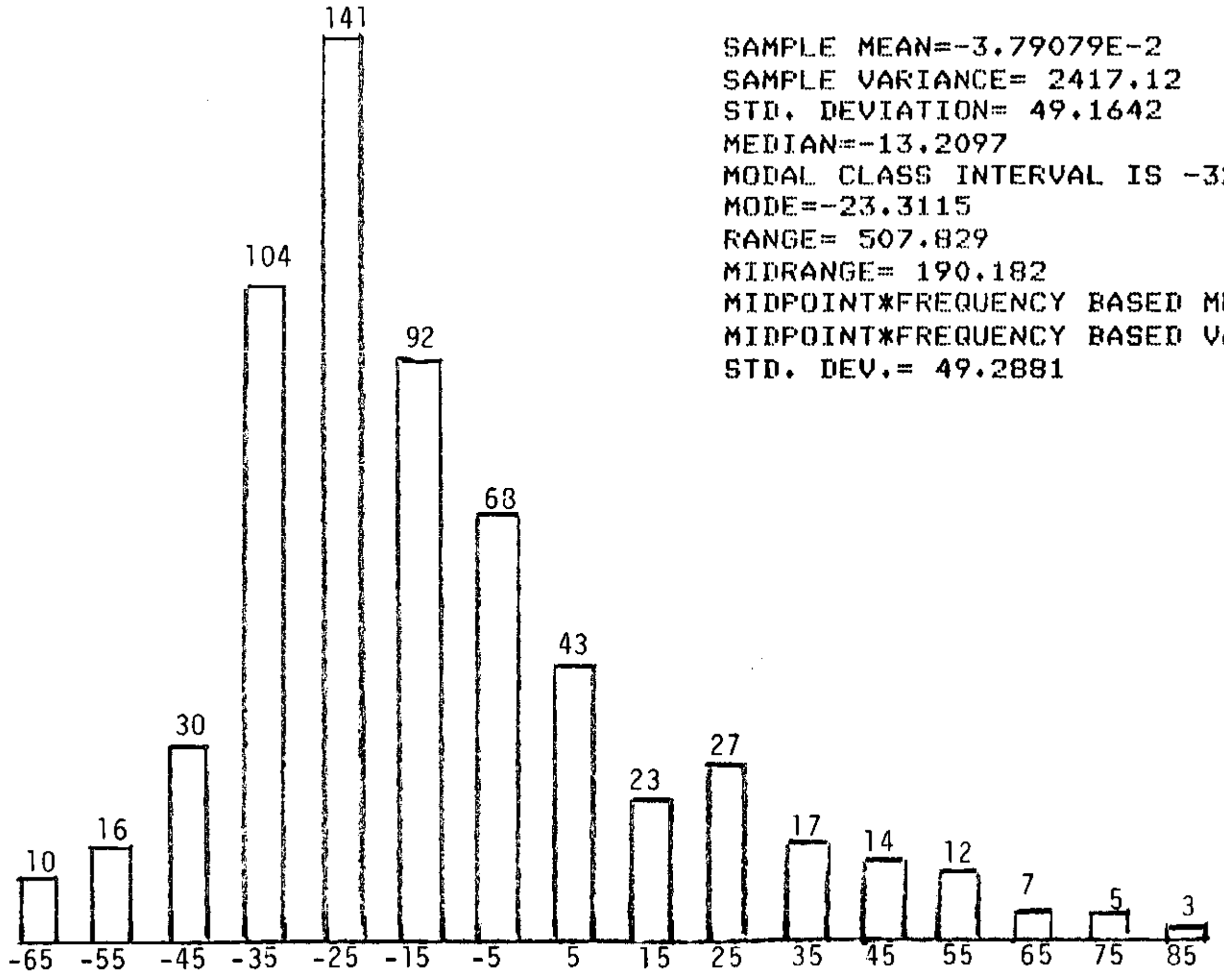


Figure A-18. Overall Plot of Average Residuals



SAMPLE MEAN=-3.79079E-2
SAMPLE VARIANCE= 2417.12
STD. DEVIATION= 49.1642
MEDIAN=-13.2097
MODAL CLASS INTERVAL IS -32 TO -22
MODE=-23.3115
RANGE= 507.829
MIDRANGE= 190.182
MIDPOINT*FREQUENCY BASED MEAN=-9.63665E-2
MIDPOINT*FREQUENCY BASED VARIANCE= 2429.32
STD. DEV.= 49.2881

Figure A-19. Histogram of Residuals for Adjusted Sample Data (TTC)

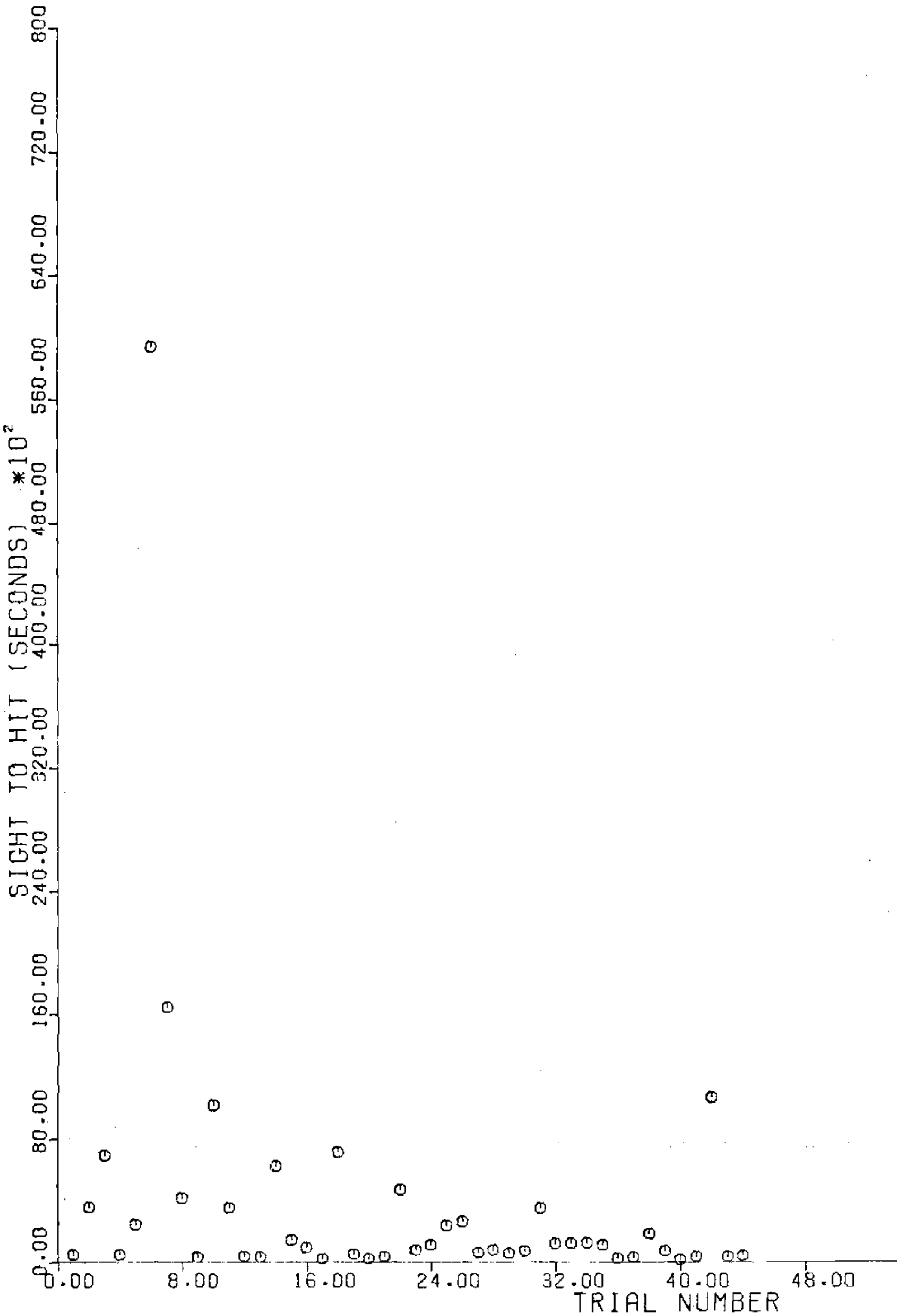


Figure A-20. Plot of MS_E at Each Trial

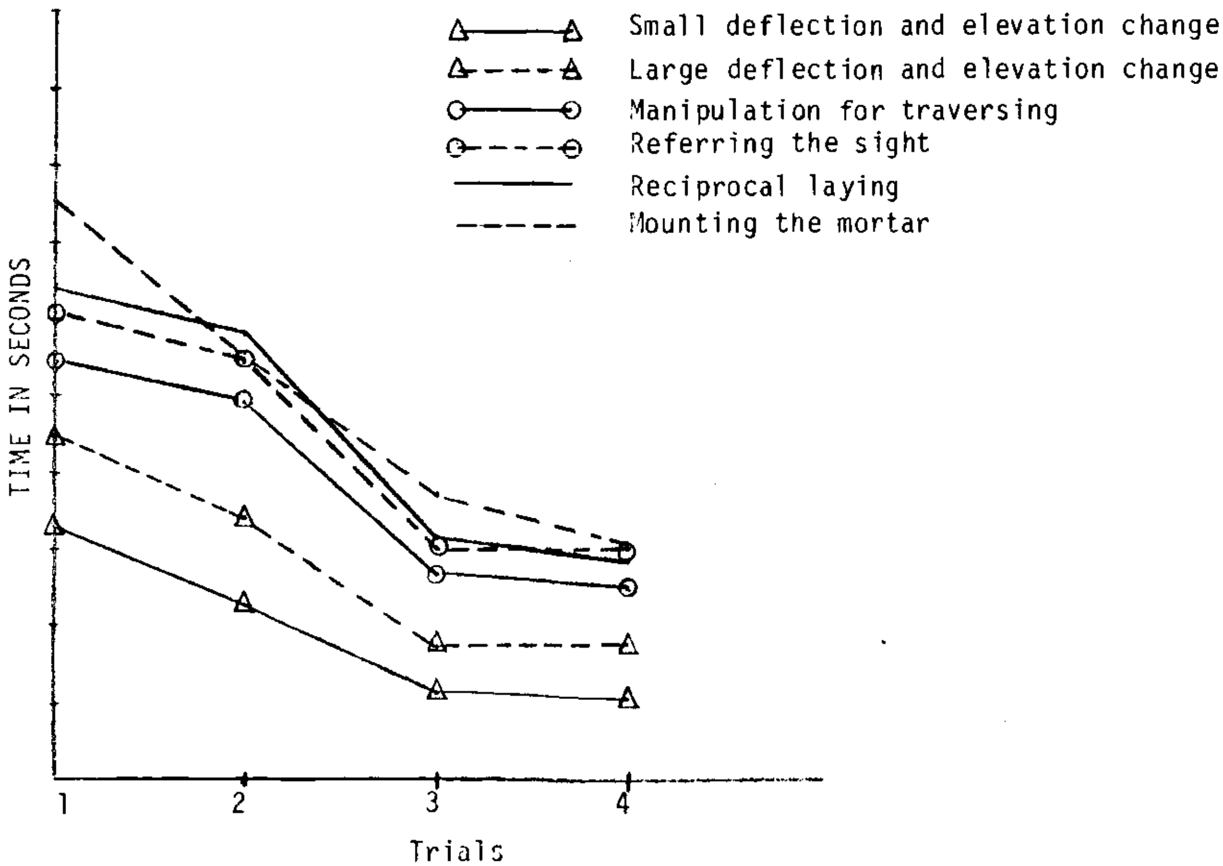


Figure A-21. Gunner's Examination Times (seconds) (44)

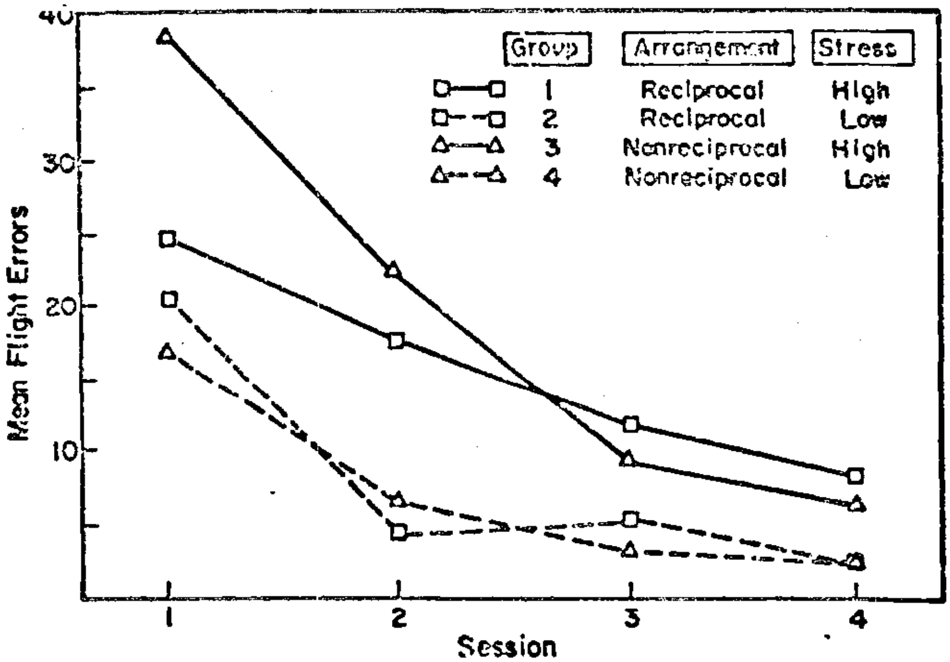


Figure A-22. Flight Errors by All Groups of Experiment VIII (45)

APPENDIX B

This appendix contains a FORTRAN IV listing of the program used to provide parameter estimates used in SPSS subprogram Nonlinear. To execute program, the user must provide the number of observations, starting values for parameters, actual observations, and trial numbers for each observation.

```
      PROGRAM PARAMS(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      DIMENSION OBS(700),TIME(700)
      READ*,N,A,B,(OBS(I),I=1,N),(TIME(I),I=1,N)
C
C     THIS PROGRAM SOLVES FOR PARAMETERS "A", "B" BY
C     MINIMIZING THE SUM OF SQUARED ERRORS USING A
C     GRADIENT TYPE SEARCH PROCEDURE.
C
      DO 11 K=1,100
      Q1=Q2=0.0
      DO 12 I=1,N
      F1=(0.0+(1.0/TIME(I)**B))
      F2=(0.0+(A/(TIME(I)**B))*ALOG(1.0/TIME(I)))
      F11=F11+F1*F1
      F12=F12+F1*F2
      F21=F21+F2*F1
      F22=F22+F2*F2
      Q1=Q1+(OBS(I)-(A/(TIME(I)**B)))*F1
      Q2=Q2+(OBS(I)-(A/(TIME(I)**B)))*F2
   12 CONTINUE
C
C     SOLVE FOR ELEMENTS OFDIRECTION VECTOR THAT WILL
C     IMPROVE OUR ESTIMATES OF PARAMETERS "A" AND "B".
C     FIND "D1" AND "D2" BY SOLVING A 2X3 MATRIX.
C
      F111=1.0
      F121=F12/F11
      Q11=Q1/F11
      F211=1.0
      F221=F22/F21
      Q21=Q2/F21
C
C     CONDUCT MATRIX ADDITION TO OBTAIN ZERO COEFFICIENT
C     FOR D1 IN SECOND EQUATION.
C
      F112=F111
      F122=F121
      Q12=Q11
      F212=F211-F111
      F222=F221-F121
      Q22=Q21-Q11
C
C     GET COEFFICIENTS OF D2 IN BOTH EQUATIONS AT
C     SAME VALUE
C
      F113=F112*(F222/F122)
      F123=F122*(F222/F122)
      Q13=Q12*(F222/F122)
      F213=F212
      F223=F222
      Q23=Q22
```

```
C
C     CONDUCT MATRIX ADDITION TO OBTAIN ZERO COEFFICIENT
C     FOR D2 IN FIRST EQUATION
C
      F114=F113
      F124=F123-F223
      Q14=Q13-Q23
      F214=F213
      F224=F223
      Q24=Q23
C
C     PUT IN STANDARD FORM WHERE COEFFICIENT OF D1 IN
C     EQUATION 1 EQUALS 1 AND COEFFICIENT OF D2 IN
C     EQUATION 2 EQUALS 2 AND FIND THE VALUES FOR D1
C     AND D2 RESPECTIVELY
C
      F115=F114*(F122/F222)
      F125=F124
      Q15=Q14*(F122/F222)
      F215=F214
      F225=1.0
      Q25=Q24/F224
      H=0.0-1.0
      IF (F115 .GT. H) GO TO 13
      F115=F115*H
      Q15=Q15*H
   13 IF (F225 .GT. H) GO TO 14
      F225=F225*H
      Q25=Q25*H
   14 D1=Q15
      D2=Q25
C
C     FIND MAXIMUM DISTANCE, VMIN, TO PROCEED IN NEW
C     DIRECTION FROM CURRENT PARAMETER VECTOR TO GAIN
C     AN IMPROVEMENT IN MINIMIZING  SUM OF SQUARED ERRORS
C
      W=1.0
   21 A1=A
      B1=B
      A2=A+(W*.5)*D1
      B2=B+(W*.5)*D2
      A3=A+W*D1
      B3=B+W*D2
      QA1=QA2=QA3=0.0
      DO 15 I=1,N
      QA1=QA1+(OBS(I)-(A1/(TIME(I)**B1)))**2
      QA2=QA2+(OBS(I)-(A2/(TIME(I)**B2)))**2
      QA3=QA3+(OBS(I)-(A3/(TIME(I)**B3)))**2
   15 CONTINUE
      VAL1=QA1+QA3
      VAL2=2.0*QA2
```

```
      IF (VAL1 .EQ. VAL2) GO TO 18
      VMIN=0.5+.25*(QA1-QA3)/(QA3-2.0*QA2+QA1)
   18 AV=A+VMIN*D1
      BV=B+VMIN*D2
      QV=0.0
      DO 19 I=1,N
      QV=QV+(OBS(I)-(AV/(TIME(I)**BV)))**2
   19 CONTINUE
      VAL=QV-QA1
      IF (VAL .LT. .00001) GO TO 20
      W=W*.5
      WRITE(6,97)QV
   97 FORMAT(" "/"QV= ",F15.8)
      GO TO 21
   20 D11=D1
      D22=D2
      IF(D11 .GT. 0.0) GO TO 31
      D11=(0.0-1.0)*D11
   31 IF (D11 .GT. .000001) GO TO 32
      IF (D22 .GT. 0.0) GO TO 33
      D22=(0.0-1.0)*D22
   33 IF (D22 .LT. .000001) GO TO 16
   32 A=A+VMIN*D1
      B=B+VMIN*D2
   11 CONTINUE
   16 SE=(QA1/(N-2))**0.5

      WRITE (6,17) A,B,SE
   17 FORMAT(" "/"PARMA= ",F15.8,5X,"PARMB= ",
     CF11.8,5X,"STD DEV= ",F11.8)
      STOP
      END
```

APPENDIX C

This appendix contains an execution run for the Lightweight Company Mortar System sample data using the SPSS Nonlinear subprogram.

78/05/10.

```
VOGELBACK COMPUTING CENTER
NORTHWESTERN UNIVERSITY

S P S S - - STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

VERSION 7.0 -- JUNE 27 1977


            RUN NAME         BLAST
            FILE NAME        BLAST -- NONLINEAR REGRESSION PROGRAM
            VARIABLE LIST    X,Y
            INPUT FORMAT     FREEFIELD
            INPUT MEDIUM     CARDS
            N OF CASES       24
            NONLINEAR        VARIABLES=Y WITH X,NB=2
            MODEL            YHAT=B(1)*(1.0/X**B(2))
            PARAMETERS       B(1)=115.1387  $ B(2)=.58374261
                             BL(1)=0.0 $ BL(2)=0.0
                             BU(1)=200.0 $ BU(2)=5.0
                             FIX(1)=1.0 $ FIX(2)=1.0
            OPTION           4
            STATISTICS       5 $ 7 $ 9 $ 4 $ 8
            READ INPUT DATA

SPSS RELOADED

    00056200 CM NEEDED FOR NONLINEAR


    OPTION - 1
    IGNORE MISSING VALUE INDICATORS

    OPTION - 4
    ONLINE PRINT FORMAT
```

BLAST 78/05/10.

FILE ELAST (CREATION DATE = 78/05/10.) -- NONLINEAR REGRESSION PROGRAM

I T E R A T I O N   S U M M A R Y

NUMBER	SUM OF SQUARES	TIME
0	1.3549849E+04	0
1	1.3549849E+04	.076
2	1.3549849E+04	.057

T H E   L A S T   I T E R A T I O N

ITERATION NO. 2			BASE POINT	TEST POINT
SUM OF SQUARES			1.3549849E+04	1.3549849E+04
L			0	0
LAMBDA			0	0
GAMMA			1.0000000E+00	1.0000000E+00
ANGLE IN DEGREES			88.9266	82.7105
MAX. PIVOT REDUCTION			5.5830430E-01	5.5830429E-01
PAR.	1	B1	1.1513873E+02	S 1.1513873E+02
	2	B2	S 5.8374260E-01	5.8374260E-01

CUMULATIVE NO. OF FUNCTION CALLS = 3
ITERATION TIME = .057 SECONDS
CUMULATIVE TIME = .259 SECONDS

ITERATION TERMINATES
MAX. RELATIVE CHANGE IN A PARAMETER .LT. TOL(1) = 1.5000000E-08

FINAL FARAMETER VALUES

			FINAL VALUE
PAR.	1	B1	1.1513873E+02
	2	B2	5.8374260E-01

SUM OF SQUARES = 1.3549849E+04

N O N L I N E A R - P R O B L E M   S U M M A R Y

```
      24   CASES
       1   DEPENDENT VARIABLE(S)
       2   PARAMETERS
      50   ITERATION LIMIT
METHOD =   MARQUARDT
TOL1 =     1.5000000E-08   REL. CHANGE IN A PARAMETER
TOL2 =             0       REL. CHANGE IN SUM OF SQUARES
TOL3 =             0       RATIO TO INITIAL SUM OF SQUARES
TOL4 =     1.0000000E-06   PIVOT TOLERANCE
```

PARAMETERS

NO.	NAME	INITIAL VALUE	LOWER BOUND	UPPER BOUND
1	B1	1.1513870E+02	0	2.0000000E+02
2	B2	5.8374261E-01	0	5.0000000E+00

SETUP TIME = .126 SECONDS

FILE ELAST (CREATION DATE = 78/05/10.) -- NONLINEAR REGRESSION PROGRAM

V A R I A N C E C O V A R I A N C E M A T R I X

PARAMETER NO.			1	2
PAR.	1	B1	95.7256058	
	2	B2	.7942811	.0149210

C O N F I D E N C E L I M I T S O N L I N E A R H Y P O T H E S I S

			LOWER LIMIT	FINAL VALUE	UPPER LIMIT
PAR.	1	B1	9.5570839E+01	1.1513873E+02	1.3470662E+02
	2	B2	3.3943974E-01	5.8374260E-01	8.2804546E-01

EXPLORATION

			LOWER TEST/ UPPER TEST	SUM OF SQUARES	LINEAR EST. OF SUM OF SQUARES
PAR.	1	B1	1.1024676E+02	1.3825640E+04	1.3825640E+04
			1.2003071E+02	1.3825640E+04	1.3825640E+04
	2	B2	5.2266689E-01	1.3854565E+04	1.3825640E+04
			6.4481832E-01	1.3815990E+04	1.3825640E+04

FILE BLAST (CREATION DATE = 78/05/10.) -- NONLINEAR REGRESSION PROGRAM

F I N A L F U N C T I O N V A L U E S A N D R E S I D U A L S

ROOT MEAN SQUARE RESIDUAL = 2.4817377E+01 D.F. = 22
THIS IS THE SCALE UNIT IN THE GRAPH OF THE RESIDUALS.

CASE	VAR	PREDICTION		OBSERVATION	RESIDUAL
1	1	1.1513873E+02	1.5000000E+02	3.4861268E+01	3.4861268E+01
2	1	1.1513873E+02	6.7000000E+01	-4.8138732E+01	-4.8138732E+01
3	1	1.1513873E+02	1.2100000E+02	5.8612679E+00	5.8612679E+00
4	1	1.1513873E+02	9.0000000E+01	-2.5138732E+01	-2.5138732E+01
5	1	1.1513873E+02	1.2900000E+02	1.3861268E+01	1.3861268E+01
6	1	1.1513873E+02	1.0900000E+02	-6.1387321E+00	-6.1387321E+00
7	1	7.6824087E+01	1.1100000E+02	3.4175913E+01	3.4175913E+01
8	1	7.6824087E+01	4.7000000E+01	-2.9824087E+01	-2.9824087E+01
9	1	7.6824087E+01	1.1000000E+02	3.3175913E+01	3.3175913E+01
10	1	7.6824087E+01	6.9000000E+01	-7.8240873E+00	-7.8240873E+00
11	1	7.6824087E+01	1.1700000E+02	4.0175913E+01	4.0175913E+01
12	1	7.6824087E+01	9.9000000E+01	2.2175913E+01	2.2175913E+01
13	1	6.0632495E+01	7.6000000E+01	1.5367505E+01	1.5367505E+01
14	1	6.0632495E+01	2.3000000E+01	-3.7632495E+01	-3.7632495E+01
15	1	6.0632495E+01	6.1000000E+01	3.6750501E-01	3.6750501E-01
16	1	6.0632495E+01	3.5000000E+01	-2.5632495E+01	-2.5632495E+01
17	1	6.0632495E+01	6.2000000E+01	1.3675050E+00	1.3675050E+00
18	1	6.0632495E+01	5.3000000E+01	-7.6324950E+00	-7.6324950E+00
19	1	5.1259383E+01	6.2000000E+01	1.0740617E+01	1.0740617E+01
20	1	5.1259383E+01	2.1000000E+01	-3.0259383E+01	-3.0259383E+01
21	1	5.1259383E+01	6.2000000E+01	1.0740617E+01	1.0740617E+01
22	1	5.1259383E+01	3.6000000E+01	-1.5259383E+01	-1.5259383E+01
23	1	5.1259383E+01	5.8000000E+01	6.7406168E+00	6.7406168E+00
24	1	5.1259383E+01	5.0000000E+01	-1.2593832E+00	-1.2593832E+00

FILE   BLAST   (CREATION DATE = 78/05/10.)   -- NONLINEAR REGRESSION PROGRAM

GRAPH OF RESIDUALS

```
  CASE     VAR          -2        -1         0        +1        +2
     1      1                                ***************
     2      1            *********************
     3      1                                ***
     4      1                      ***********
     5      1                                *******
     6      1                              ***
     7      1                                ***************
     8      1                    *************
     9      1                                **************
    10      1                             ****
    11      1                                *****************
    12      1                                **********
    13      1                                *******
    14      1                 ****************
    15      1                                *
    16      1                      ***********
    17      1                                **
    18      1                             ****
    19      1                                *****
    20      1                    *************
    21      1                                *****
    22      1                          *******
    23      1                                ****
    24      1                               **
```

TIME SINCE END OF THE LAST ITERATION =   .364 SECONDS
TOTAL TIME =   .623 SECONDS

RUN COMPLETED

NUMBER OF CONTROL CARDS READ 15
NUMBER OF ERRORS DETECTED   0

APPENDIX D

This appendix contains plots of the final fitted models selected in Chapter IV.

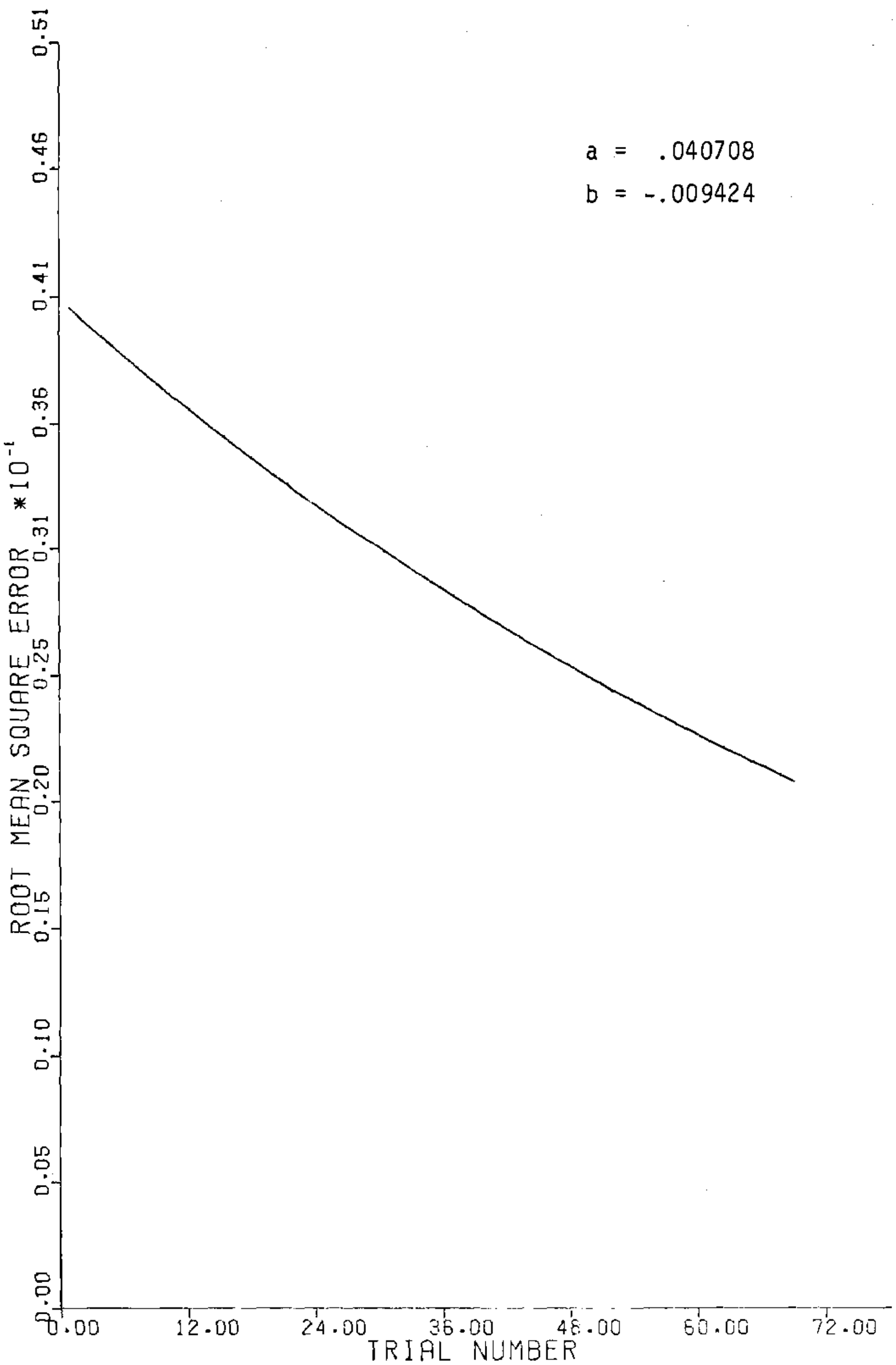


Figure D-1. $\hat{Y} = ae^{bt}$, Fitted Model for ITV(Sys 3) Data.

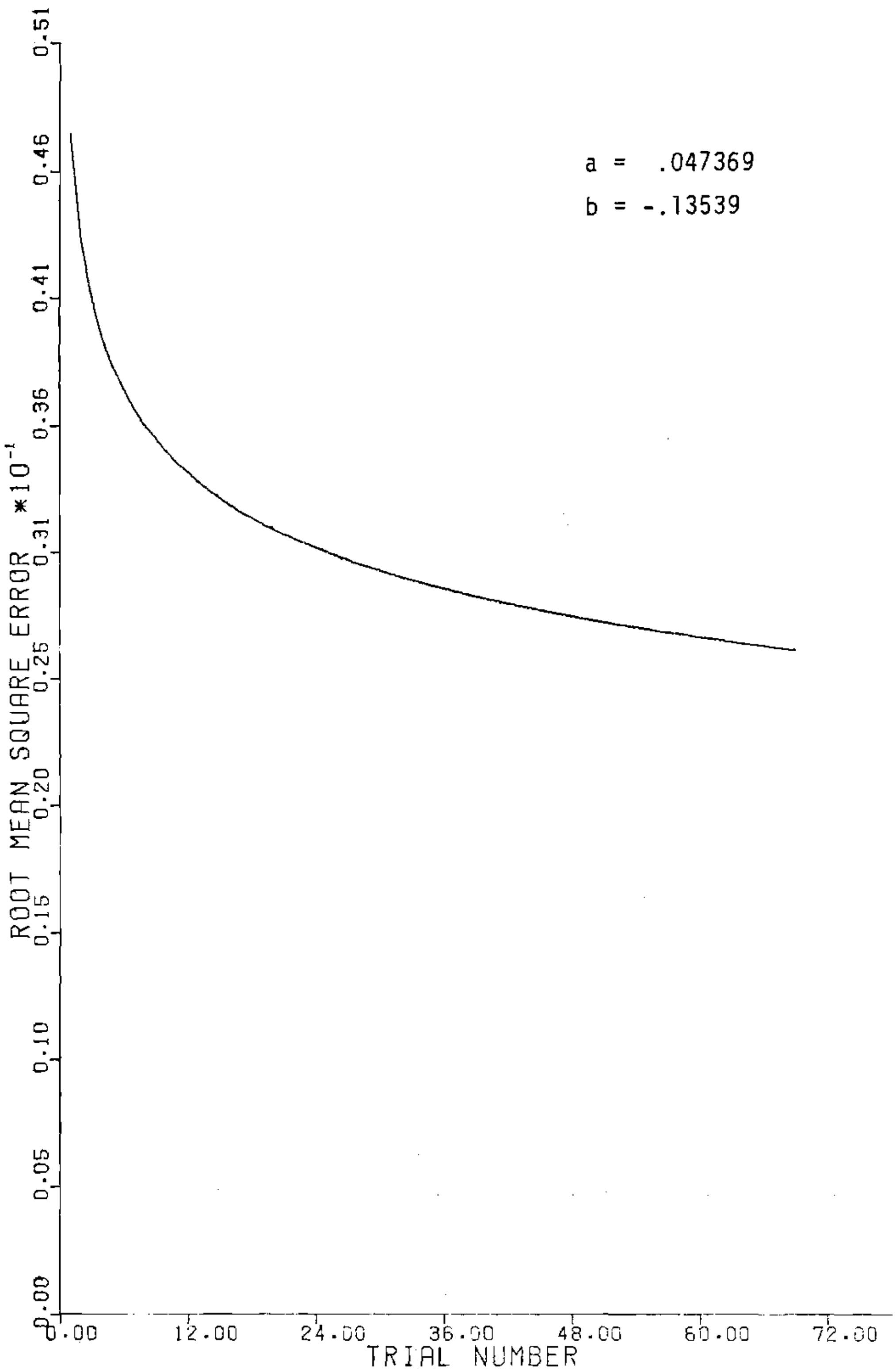


Figure D-2. $\hat{Y} = at^{-b}$, Fitted Model for ITV(Sys B) Data

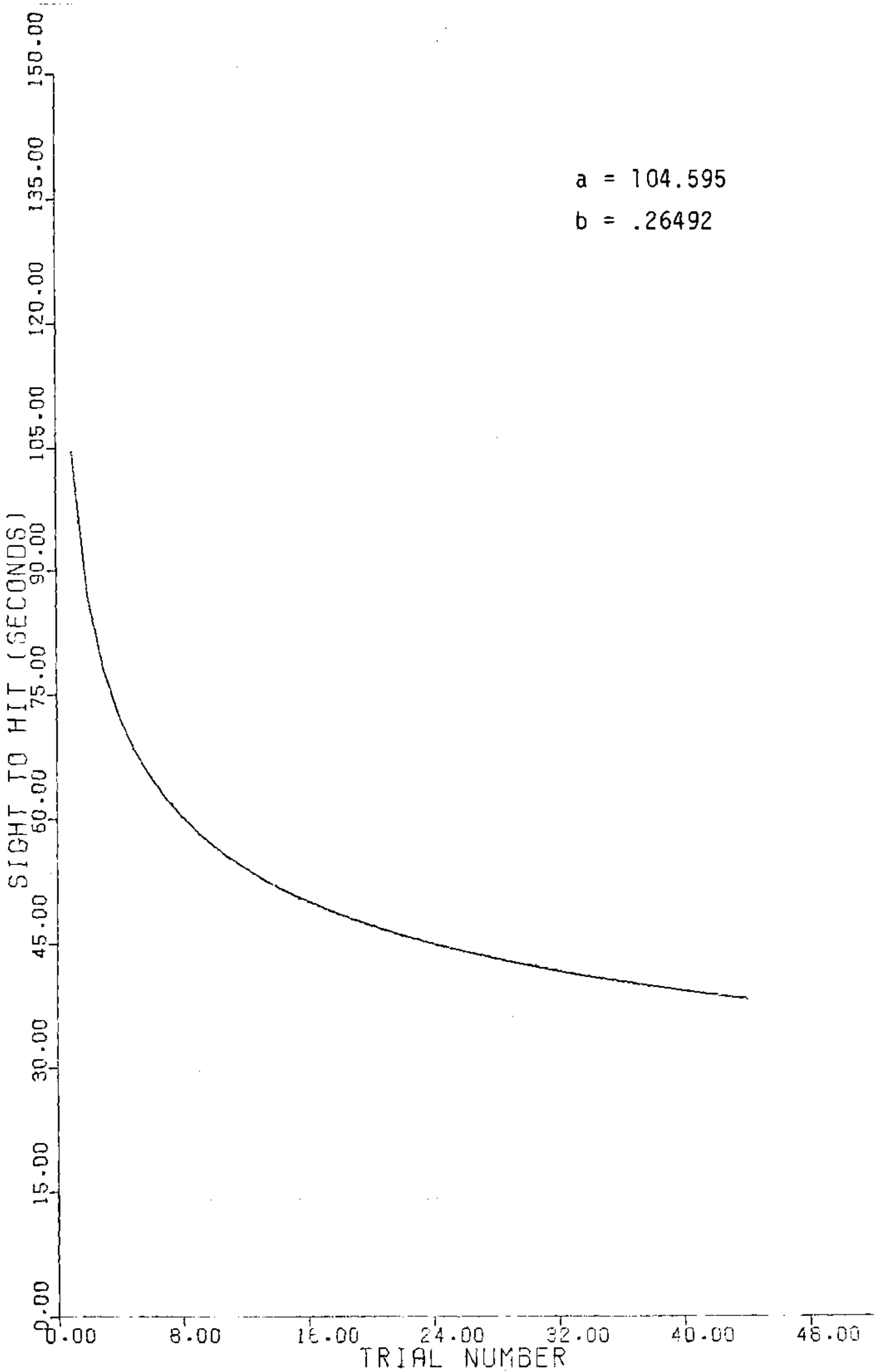


Figure D-3. $\hat{Y} = at^{-b}$, Fitted Model for TTC Data

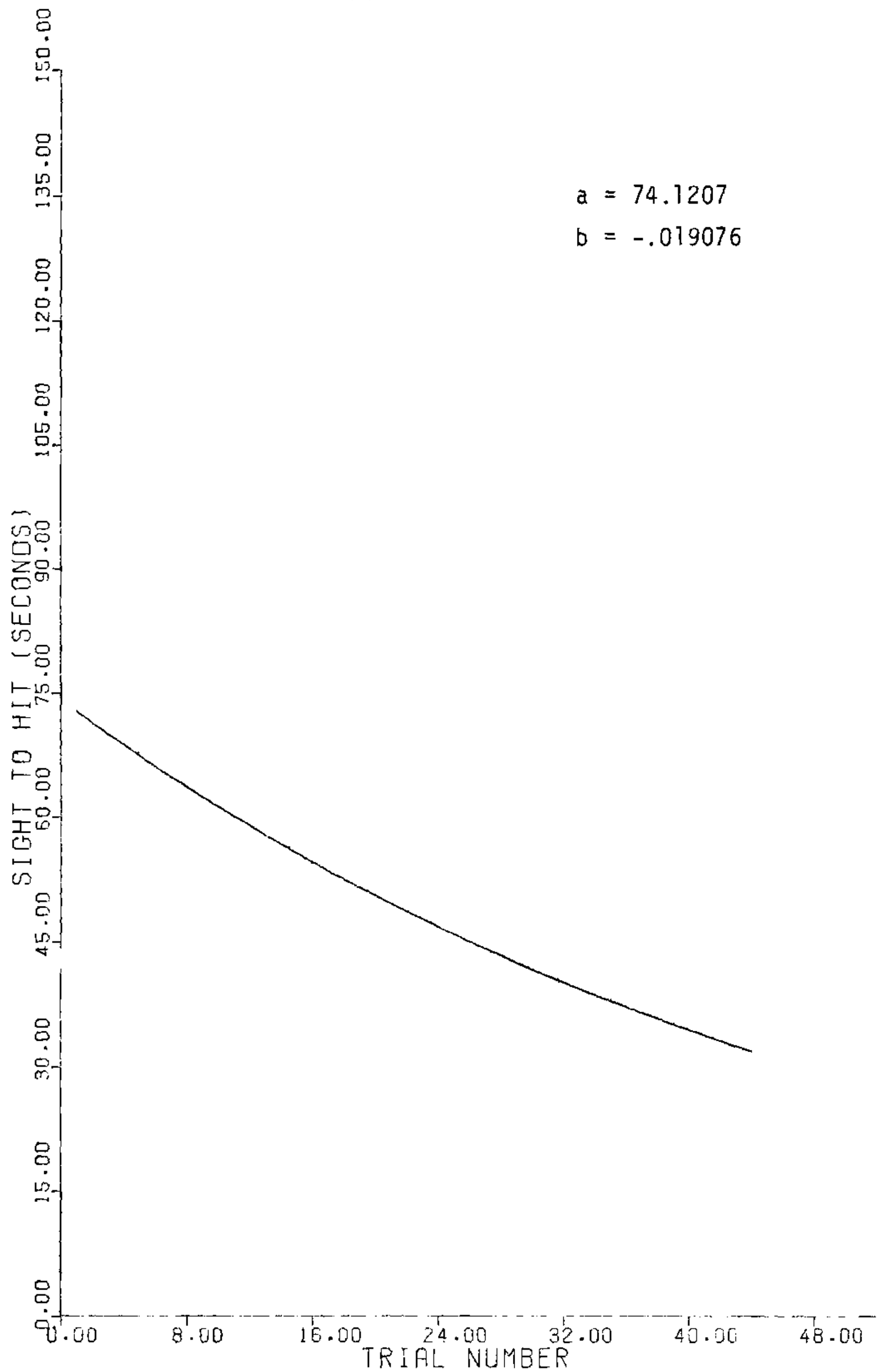


Figure D-4. $\hat{Y} = ae^{bt}$, Fitted Model for TTC Data

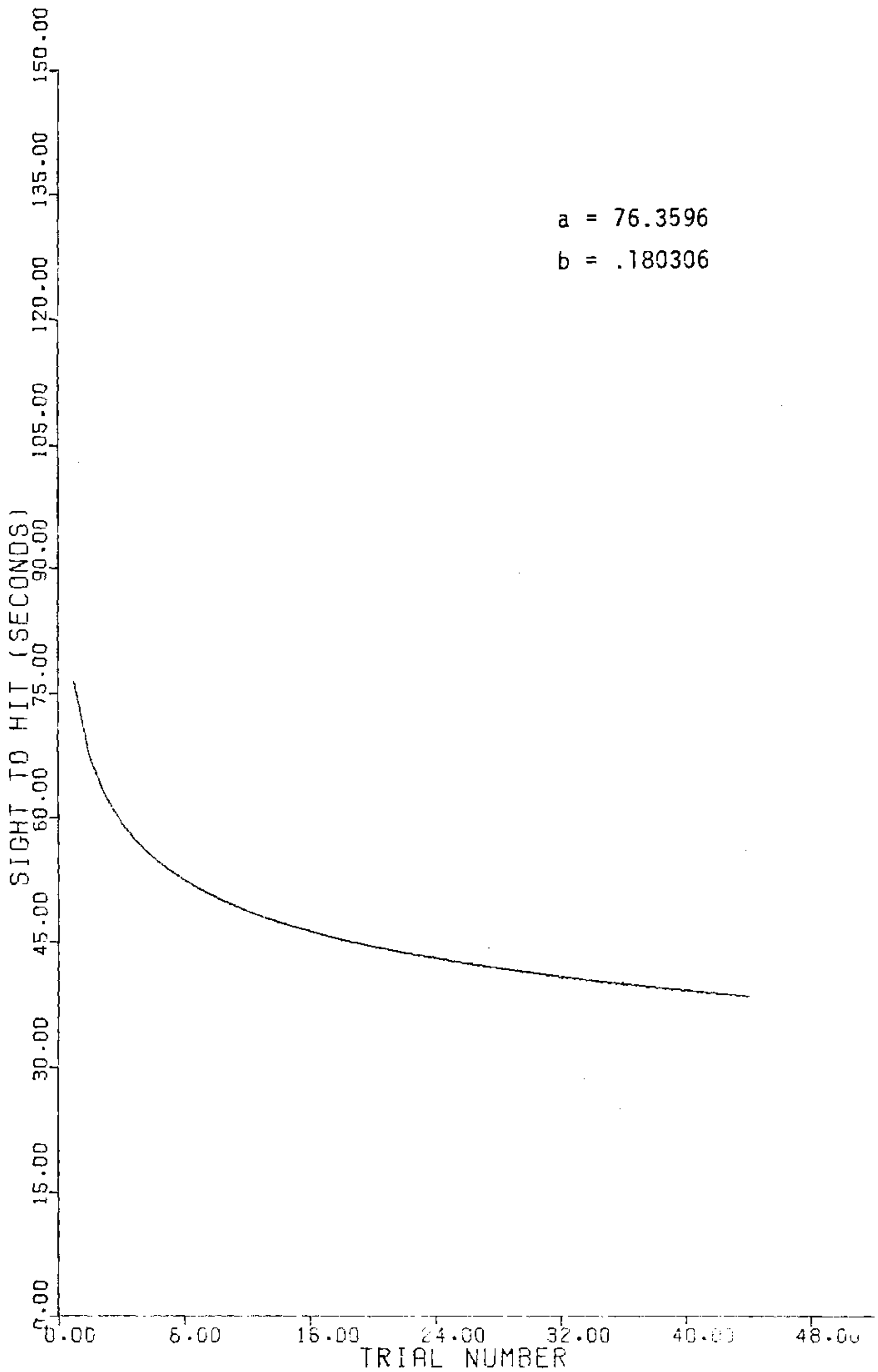


Figure D-5. $\hat{Y} = at^{-b}$, Fitted Model for TC Data

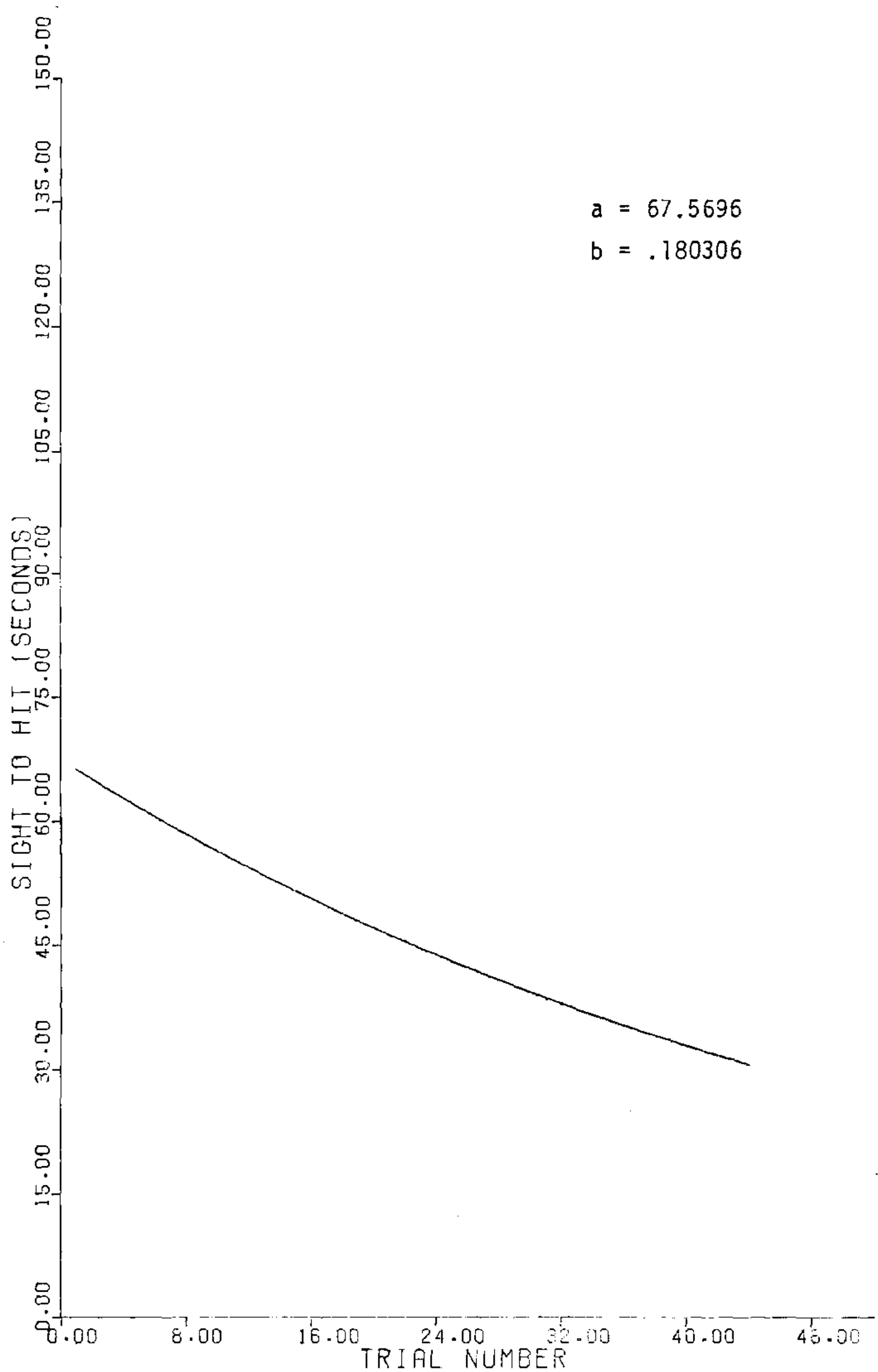


Figure D-6. $\hat{Y} = ae^{bt}$, Fitted Model for TC Data

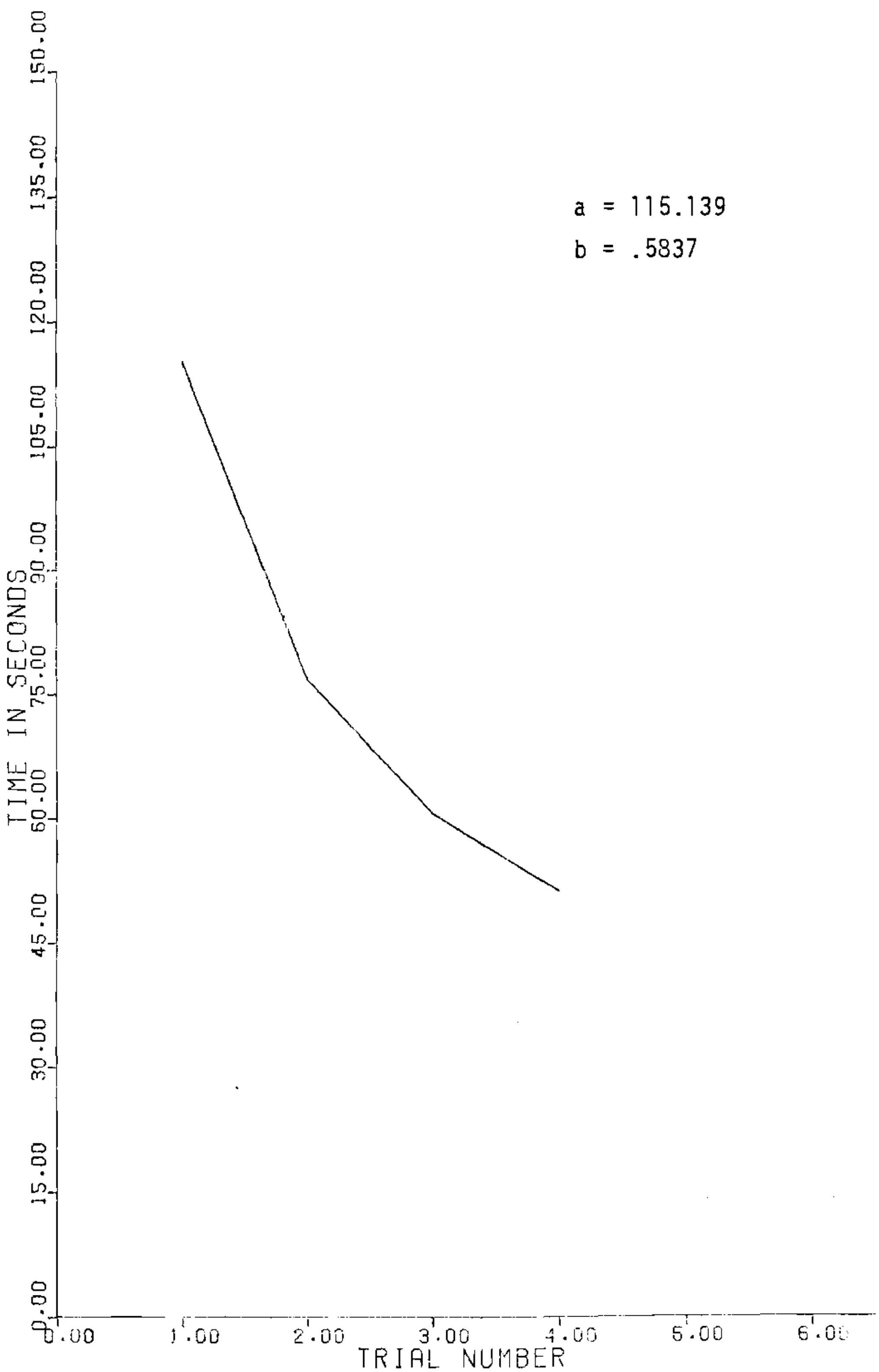


Figure D-7. $\hat{Y} = at^{-b}$, Fitted Model for LWCMS Data

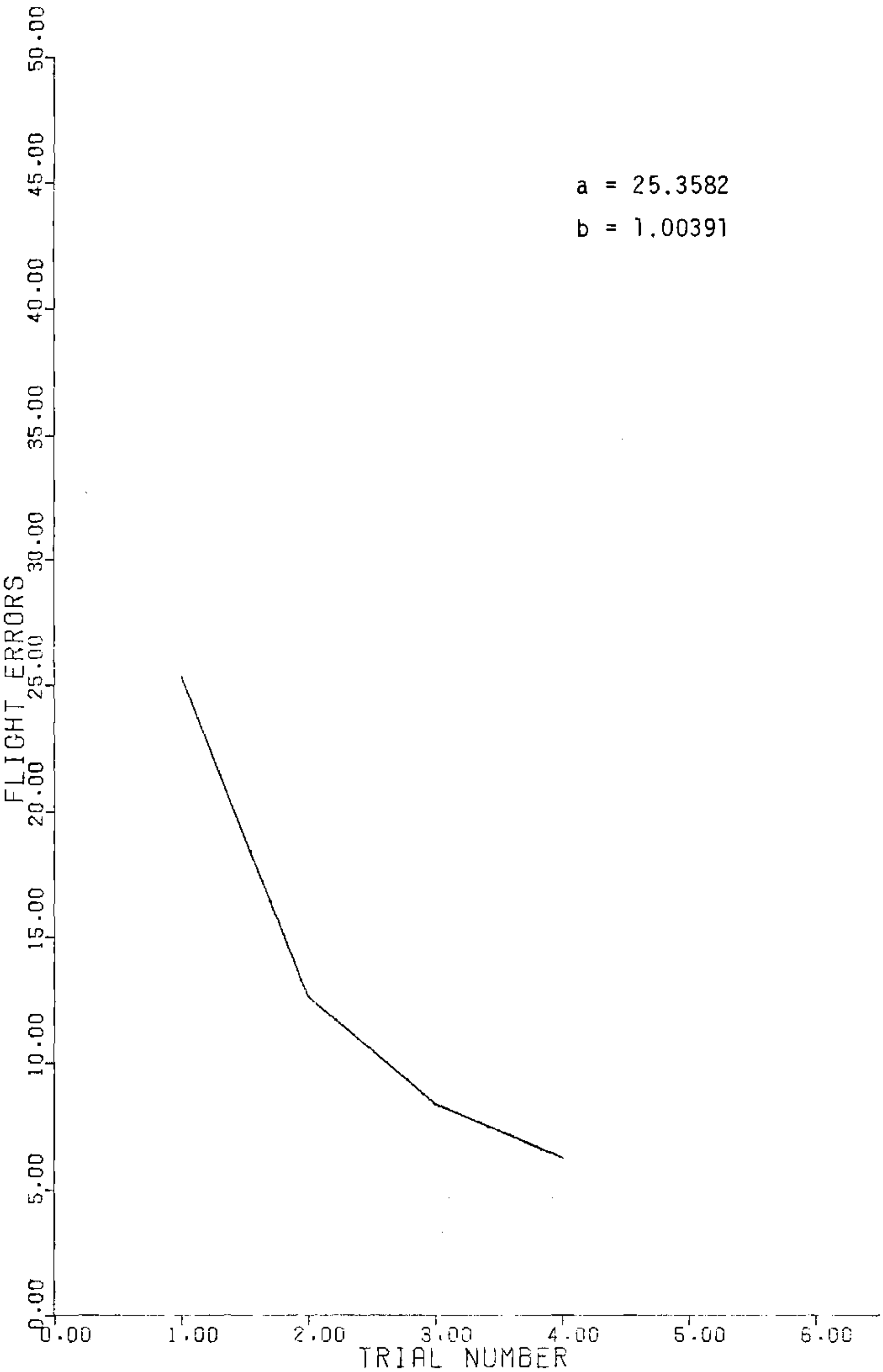


Figure D-3. $\hat{Y} = at^{-b}$, Fitted Model for Team Training (Ex VIII) Data

BIBLIOGRAPHY

1. AR 10-4, dated 30 December 1974, Organization and Functions, U.S. Army Operational Test and Evaluation Agency, Headquarters, Department of the Army, Washington, D.C.

2. AR 71-3, dated 30 November 1974, Research and Development and Force Development Testing, Headquarters, Department of the Army, Washington, D.C.

3. AR 1000-1, dated 5 November 1974, Basic Policies for Systems Acquisition by the Department of the Army, Headquarters, Department of the Army, Washington, D.C.

4. DOD Directive 5000.1, dated 1 September 1974, Acquisition of Major Defense Systems, Headquarters, Department of Defense, Washington, D.C.

5. DOD Directive 5000.2, dated 20 February 1974, The Decision Coordinating Paper (DCP) and the Defense Systems Acquisition Review Council (DSARC), Headquarters, Department of Defense, Washington, D.C.

6. Garry, R., and Kingsley, H.L. The Nature and Conditions of Learning, Prentice-Hall, Inc., New Jersey, 1970.

7. Burns, J.R., "A Comparison of Two Methods for Predicting Loss of Learning Due to a Break in Production," Report DARCOM-ITC-02-08-76-107, U.S. Army Materiel Development and Readiness Command, Alexandria, Virginia, 1976.

8. Mednick, Sarnoff A., Learning Englewood Cliffs, New Jersey: Prentice Hall, Inc. 1964.

9. Levine, G., and Burke, C.J., Mathematical Model Techniques for Learning Theories, Academic Press, New York, 1972.

10. Hull, C.L., Principles of Behavior, New York: Appleton, 1943.

11. Wright, T.P., "Factors Affecting the Cost of Airplanes," Journal of the Aeronautical Sciences, Vol. 3, February 1936.

12. Ilderton, R.B., "Methods of Fitting Learning Curves to Lot Data Based on Assumptions and Techniques of Regression Analysis"; Master's Thesis. George Washington University, August, 1970. AD-A011583.

13. Pichon, A.A., Jr., "The Development of a Predictive Model for First Unit Costs Following Breaks in Production," (USAF), Wright-Patterson Air Force Base, Ohio: Master's Thesis.

14. "Learning Curve Summary," United States Army Logistics Management Center, Fort Lee, Virginia: ALM-63-3126-H1(A).

15. Allison, J.D., "A Learning Model in Which All Members of a Work Group Do Not Benefit From Each Task Repetition," 685 Research Project; Texas A & M University, 1970.

16. Patton, J.D. Jr., Logistics Technology and Management, 1977.

17. Pegels, C.C., "On Startup or Learning Curves: An Expanded View," AIIE Transactions, Vol. I, No. 3.

18. Carr, G.W., "Peacetime Cost Estimating Requires New Learning Curves," Aviation, Vol. 45, April, 1946.

19. Guibert, P., Le Plan de Fabrication Aeronautique, Dunod, Paris, 1945 (English translation under the title, Mathematical Studies of Aircraft Construction, was made available by Central Air Documents Office, Wright-Patterson Air Force Base, Dayton, Ohio.

20. De Jong, J.R., "The Effects of Increasing Skill on Cycle Time and Its Consequences for Time Standards," Ergonomics, Vol. 1, No. 1, 1957.

21. Levy, F.K., "Adaption in the Production Process," Management Science, Vol. 11, No. 6, April, 1965.

22. Meister, D., Behavioral Foundations of System Development, John Wiley and Sons, Inc., New York, 1976.

23. Klaus, J. and Glaser, R., "Increasing Team Proficiency Through Training, 8. Final Summary Report, AIRE 1-6/68-FR, American Institutes for Research, May, 1968.

24. Wagner, H., et al, Team Training and Evaluation Strategies: State-of-the-Art, Technical Report HumRRO-TR-77-1, Human Resources Research Organization, Alexandria, Virginia, 1977.

25. Klaus, D.J. and Glaser, R., Increasing Team Proficiency Through Training,1. A Program of Research. Pittsburgh: American Institutes for Research, 1960.

26. Glaser, R., Klaus, D.J., and Egerman, K., Increasing Team Proficiency Through Training, 2. The Acquisition and Extinction of a Team Response. Pittsburgh: American Institutes of Research, 1962.

27. Egerman, K., Klaus, D.F., and Glaser, R., Increasing Team Proficiency

Through Training, 3. Decremental Effects of Reinforcement in Teams with Redundant Members. Pittsburgh: American Institutes for Research, 1962.

28. Egerman, K., Klaus, D.J., Increasing Team Proficiency Through Training, 4. A Learning-Theoretic Analysis of the Effects of Team Arrangement on Team Performance, Pittsburgh: American Institute for Research, 1963.

29. Klaus, D.J. and Glaser, R., Increasing Team Proficiency Through Training, 5. Team Learning as a Function of Member Learning Characteristics and Practice Conditions. Pittsburgh: American Institutes for Research, 1965.

30. Klaus, D.J., Grant, L.D. and Glaser, R., Increasing Team Proficiency Through Training, 6. Supervisory Furnished Reinforcement in Team Training. Pittsburgh: American Institutes for Research, 1965.

31. Short, J.G., Cotton, T.S., and Klaus, D.J., Increasing Team Proficiency Through Training, 7. The Simulation of Team Environments. Pittsburgh: American Institutes for Research, 1968.

32. Hoerl, A.E. Jr., "Fitting Curves To Data," Chemical Business Handbook, 1st Edition, 1954.

33. Draper, N., Smith, H., Applied Regression Analysis, John Wiley and Sons, Inc., New York, 1966.

34. Robinson, B., "SPSS Subprogram Nonlinear-Nonlinear Regression," Manual no. 433, Vogelback Computing Center, Northwestern University, 1977.

35. Hines, W.W. and Montgomery, D.C., Probability and Statistics in Engineering and Management Science, The Ronald Press Company, New York, 1972.

36. Beale, E.M.L., "Confidence Regions in Non-linear Estimation," Journal of the Royal Statistical Society, B-22, 41-76 (1960).

37. Beale, E.M.L., "The Computation of Measures of Non-linearity in Nonlinear Estimation," Report N1/M2-64, Admiralty Research Laboratory, Teddington, Egnland (1962).

38. U.S. Army Operational Test and Evaluation Agency, Improved Tow Vehicle (ITV), Operational Test, Washington, D.C.

39. U.S. Army Operational Test and Evaluation Agency, DRAGON, Operational Test, Washington, D.C.

40. Root, R.T., et al, "Initial Validation of REALTRAIN with Army Combat Units in Europe," Research Report 1191, U.S. Army Research Institute for the Behavioral and Social Sciences, 1976.

41. Banks, J.H., et al, "REALTRAIN Validation for Rifle Squads: Mission Accomplishment," Research Report 1192, U.S. Army Research Institute for the Behavioral and Social Sciences, 1977, AD A043515.

42. Hill, F.I., et al, "An Assembly of Project Stalk Data," Memorandum Report BRL-745, Vol. I, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland, 1954, AD 23768.

43. Hardison, D.C., et al, "A Partial Analysis of Project Stalk Data with Results of Single Tank Versus Single Tank Duels," Technical Note, BRL-980, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, 1955, AD 310-852.

44. U.S. Army Operational Test and Evaluation Agency, Lightweight Company Mortar System, XM224E1, Operational Test I, Washington, D.C. 1973.

45. Briggs, G.E. and Johnston, W.A., "Team Training," Technical Report: NAVTRADEVCEN 1327-4, Naval Training Device Center, Orlando, Florida, 1967, AD 660019.

46. Montgomery, D.C., Johnson, L.A., Forecasting and Times Series Analysis, McGraw-Hill Book Company, 1976.

47. Biel, W.C., Harman, H.H., and Sheldon, M.S., "Exercising Teams in Military Systems Through the Use of Simulation," Professional Paper SP-1739, System Development Corporation, October 1964.

48. Chesler, David J., "Computer-Assisted Performance Evaluation for Navy Anti-Air Warfare Training: Concepts, Methods, and Constraints," Research Report SRR 71-25, San Diego, California: Naval Personnel and Training Research Laboratory, May 1971.

49. Karcher, E.K., et al, "PRS Report 923 Unit Effectiveness Tests," Department of the Army, Personnel Research Section, January 1952.

50. Obermayer, Richard W., et al, "Combat-Ready Crew Performance Measurement System," AFHRL Final Report TR-74-108 (I), Air Force Human Resources Laboratory, December 1974.

51. Yealy, K.A., "Testing for Learning with Small Data Sets"; Master's Thesis, Georgia Institute of Technology, June, 1978.