R E P O R T   R E S U M E S

ED 020 116   56   SE 003 859

AN ADAPTIVE DECISION STRUCTURE FOR EDUCATIONAL SYSTEMS.

BY- ROE, ARNOLD
CALIFORNIA UNIV., LOS ANGELES

REPORT NUMBER 63-63   PUB DATE   DEC 63
EDRS PRICE MF-$0.75 HC-$7.76   192P.

DESCRIPTORS- *COLLEGE SCIENCE, *COMPUTER ASSISTED INSTRUCTION, *ENGINEERING EDUCATION, *SYSTEMS ANALYSIS, CURRICULUM PLANNING, EDUCATIONAL PROGRAMS, EDUCATIONAL PLANNING, TEACHING MACHINES,

THIS RESEARCH REPORT DEALS WITH THE THEORETICAL BASIS FOR DEVELOPMENT OF AN ADAPTIVE DECISION STRUCTURE FOR EDUCATIONAL SYSTEMS. VIEWING EDUCATION AS A WEALTH OR UTILITY-PRODUCING PROCESS, WHICH TRADITIONALLY HAS BEEN SHAPED BY PURELY INTUITIVE RATHER THAN ANALYTICAL DECISIONS, A PLAN FOR AN ANALYTICAL ADAPTIVE DECISION STRUCTURE BASED UPON FOUR ELEMENTS IS IDENTIFIED--(1) A DATA GATHERING AND UTILIZATION PLAN, (2) AN EXPLICIT CRITERION FUNCTION, (3) A SET OF DECISION RULES FOR ACHIEVING THE CRITERION, AND (4) A UTILITY EVALUATION FUNCTION WHICH RELATES SYSTEMS INPUTS AND OUTPUTS TO A VALUE SCALE OUTSIDE THE SYSTEM. THE UTILITY FUNCTION IS DEVELOPED IN TERMS OF THE OUTPUT OF AN EDUCATIONAL SYSTEM DEFINED AS THE LIFE-CYCLE PRODUCTIVE OUTPUT ATTRIBUTABLE TO THE EDUCATIONAL EXPERIENCE FOR ALL INDIVIDUALS WHO HAVE BEEN PART OF THE SYSTEM. THE CRITERION OF PERFORMANCE ESTABLISHED WAS THAT THE SYSTEM SHOULD OPERATE SO AS TO MAXIMIZE THE WORTH OF THE EXPECTED LIFE-CYCLE PRODUCTIVE OUTPUT OF ALL THOSE EDUCATED IN THE SYSTEM. DEVELOPED WAS A MULTI-STAGE OR CONTINUOUS SEQUENTIAL DECISION RULE FOR USE WITH NORMALLY DISTRIBUTED EXPECTED OUTPUTS. WITH THIS PROGRAM IT IS CONCEIVABLE TO ACHIEVE DECISIONS ON THE AMOUNT OF RESOURCES TO ALLOCATE TO THE DEVELOPMENT OF A PARTICULAR INSTRUCTIONAL PROGRAM. (DH)

DEPARTMENT OF ENGINEERING

DECEMBER 1963

an adaptive decision structure for educational systems

A. ROE

REPORT NO. 63-63

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Report No. 63-63
December, 1963

AN ADAPTIVE DECISION STRUCTURE FOR EDUCATIONAL SYSTEMS

by

Arnold Roe

TEACHING SYSTEMS RESEARCH PROJECT

H. W. Case A. Roe

Department of Engineering
University of California
Los Angeles 24, California

FOREWORD

The research described in this report, An Adaptive Decision Structure for Educational Systems, by Arnold Roe, was supported in part by a grant from the United States Office of Education, Department of Health, Education and Welfare under Title VII of the National Defense Education Act (NDEA Grant 7-04-138.01).

TABLE OF CONTENTS

LIST OF SYMBOLS

Symbols which are used only once in the text are not shown below. Some symbols are used in more than one context, in which case more than one definition appears next to that symbol.

A: Initial set-up costs

B: A forcing set

C: Sampling costs

D: Annual educational costs

D': Costs of an educational sub-unit

E: Expected value

F: A cumulative density function

K: Population with the highest mean

M: Survival factor

N: The size of the total available sample

P: Probability; also, productivity

R: Discount factor

S: The sum of sample values

T: Nominal learning time for an educational sub-unit

U: The number of repetitions of an experiment

V: Present worth of educational costs

W: Present worth of expected life-cycle productive output

X: A sample value

a: Age at which one starts an educational unit

b: Retirement age

c: A proportionality or weighting factor relating a sub-unit to the unit

d: A transform giving equivalent dollars for any given dates

f: A transform giving dollar values from productive output

g: Grades

h:	A transform relating current to previous grades
i:	(As a subscript) The identification tag for each of the samples from the j-th population; otherwise identifies an educational unit
j:	The identification tag for a population
k:	The total number of populations
ℓ:	Identifies an educational sub-unit
m:	(Subscripted) The a priori sample mean; otherwise signifies years of experience
n:	The trial number or sample number in a sequence of samples
p:	Relative sampling cost; also, a transform giving projected values from a history of previous values
q:	Order of a polynomial
r:	Discount rate
s:	Sample standard deviation
t:	Learning time
u:	Identifies the particular repetition of an experiment
w:	A transform relating grades to subsequent output
x:	A random variable
y:	Current date; also date student will complete educational unit
y':	Date of starting productive output
z:	A dummy variable
α:	Personality factor
β:	Factor describing history of past performance
μ:	Population mean
π:	A population
σ:	Population standard deviation
τ:	Years from the date of starting a given educational unit
Π:	Product sign

^: Designates group taking a specific educational unit

*: Designates a matched group not taking a specific educational unit

$: Reported median earnings

J: Idealized learning time for a given educational unit

~: Designates the number of remaining observations

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Sincere thanks go to Harry W. Case, William LeBold and William Diemer for making available their unpublished data on the graduates of the Schools of Engineering of the University of California; to D. L. Grant of the American Telephone and Telegraph Company for data on Bell System employees; to K. Srinivas for tracking down most of the sources of data on engineers' and craftsmen's earnings; to J. Osuga and G. Johnson for invaluable assistance in handling the mass of digital computations; to R. Wilson for an introduction to value iteration; and to the University of California for providing liberal access to computer facilities.

SECTION I

INTRODUCTION

In recent years engineers have become increasingly involved in the study of adaptive teaching systems. There are research groups involved in such studies at the University of Illinois' Coordinated Science Laboratory, at the Massachusetts Institute of Technology, in numerous other schools, and in engineering firms throughout the world.

A brief analysis will be made here of the earlier work, and some additional concepts on criteria functions, decision rules, and utility functions for adaptive educational systems will be introduced.

In an educational context, the word "system" is used to describe such diverse things as "The Blank County School System" and "Dash Publishing Company's Self-Instructional System for Slide-Rule Computations". For convenience in exposition "systems" will be roughly divided into four categories.

Micro-micro systems: concerned with a transformation of students' behavior by a single, relatively short sequence of learning items.

Micro systems: concerned with a transformation of students' behavior by a longer sequence of learning items, such as are encountered in a semester course.

Macro systems: a collection of micro systems characterized by a curriculum or curricula in a school, university or school district.

Macro-macro systems: related to the transformation of students' behavior by the total learning experience encountered during the students' lives.

Almost all of the previous studies on adaptive decision structures have been concerned with micro-micro systems. Criteria functions have not been explicitly stated, and therefore the decision rules have generally been non-optimal. Also, the transformations achieved by the system have been measured in abstract units (such as grades) which are left unrelated to value scales or "utility" outside of the system, thus precluding external evaluation of the system. The adaptive decision structure suggested in the following sections is completely general and is intended for systems of all sizes. However, since a utility function for the output of educational systems will be introduced, and since data for this utility function are most readily available for macro or macro-macro systems, the approach will be to start with the larger systems and work down to the micro-micro systems.

Adaptive decisions have always existed in educational systems. Course content and pedagogical techniques have changed in response to changes in the social, cultural, economic and technological aspects of the environment. However, the rules governing such change have seldom been explicitly recognized or stated, and the information needed for making decisions has often been incomplete. An adaptive decision structure is one which removes much of the decision-making function from the intuitive realm by providing a plan for accumulating relevant data and using these data according to a preconceived plan or decision rule to change or rearrange sub-elements of the system in order to achieve a predetermined criterion in some optimal fashion. Furthermore, this criterion should have some "utility" outside of the system. An adaptive decision structure therefore requires:

1. Data gathering and handling capability.
2. A criterion function.
3. Decision rules.
4. A utility function.

Only brief consideration will be given to the first of these requirements. The assumption will be made here that the amount of data that is ideally required for an adaptive decision structure is sufficiently voluminous to require the use of modern data-processing equipment. Whether or not this data-processing equipment is also used for presenting course content material directly to students (as in the computer-controlled "teaching machines") is a side issue to the main stream of thought. Questions of this type can be readily resolved by the techniques developed in the following sections. Another assumption that will pervade all of the subsequent discussion is that flexible scheduling (for individual students, at any time during the school year) is a desirable feature of an educational system using an adaptive decision structure. The merit of this assumption will become clearer when the utility function is described. The practical implementation of flexible scheduling will undoubtedly be enhanced by the use of data-processing equipment.

The main contributions of the following sections will be in the area of criteria functions, decision rules, and utility functions for adaptive decision structures, and can be summarized as:

a. The description of a criterion function for an adaptive decision structure in an educational system where two processes are being carried out simultaneously, namely, (a) students are learning subject matter, and (b) the system controllers are learning about the student's learning. Process (b) may include exploratory use of various alternative pedagogical procedures or subject matter, some of which may result in better student performance than others. In such a situation there is a trade-off between processes (a) and (b). The suggested criterion function is the sum of the net utility of all students' outputs, and obviously this function should be maximized.

b. The description of decision rules which tend to maximize the criterion function under different conditions of a priori information. In particular, some qualitative rules are obtained for the case of "total a priori ignorance", i.e., where there is no a priori information on the distribution of the net utility of students' outputs. Also, an extension is made to the procedure for two-stage sequential sampling from two normally distributed populations to include the case where the costs of taking or observing sample data is of some consequence. Of most interest is the development here of a computational backwards-induction solution for the multi-stage or continuous sampling procedure from k normal populations. This solution is applicable to problems outside the educational context and should be of interest in such fields as medical testing, agricultural experiments, production line evaluation and in many other fields where the criterion is to maximize the sum of net outputs. The procedure used can be generalized to binomial and other distributions.

c. The description of a utility function for converting such available measures as student grades, student learning time, teacher inputs, school capital and maintenance costs, etc. into a net value of the transformation effected by the system. Current measures of student output are used to derive a present worth of the student's expected life-cycle productive output (PWSELCPO) and these are compared with PWSELCPO for alternate system configurations.

In order to assign a value to the net output from an educational system one not only needs a utility function but also data to feed into such a function. Many of the necessary data are currently nonexistent or otherwise unavailable. Therefore, some rather strong restrictions must be imposed on the utility function so that it can operate with reduced precision with existing data. The important point to note at this stage is that in at least one case there is probably enough information available to start using an adaptive decision structure which includes a utility function. That case is in the field of engineering education, where most data exist relating students' performance in school to subsequent life-cycle productive outputs. A start must be made somewhere; otherwise there is little prospect that the additional data required for a more precise utility function will ever be accumulated. It may seem like a boot-strap operation to prescribe such a structure from incomplete information, but an adaptive decision structure is dedicated to making decisions in the face of uncertainty or incomplete information. Adaptive decision theorists suggest policy iteration [1] as a means of sequentially approaching the desired end.

SECTION II

BACKGROUND

A. The Three Approaches

In the last few years, many people have talked about the possibility of applying some of the tools of modern technology to the teaching-learning process. However, the "tools" that are proposed differ with the professional background of the proponents. For example, the psychologist is usually most interested in the learning theory approach, in which stimulus-response concepts are selectively applied to the micro-micro aspects of the educational process. Many experimenters and theorists have contributed to this approach (Thorndike, Hull, Skinner, Estes, etc.). Some of the most commonly quoted concepts in this approach are:

1. Principle of reinforcement: Certain environmental effects strengthen the behavior which has produced these effects (a correct response to a question, properly rewarded, will increase the probability that the correct response will be subsequently elicited on meeting the same or similar question).
2. Principle of gradual progression: Use a series of progressive approximations so as to lead, finally, to the required complex behavior. By giving reinforcement for each of the responses in the series making up the complex pattern, the desired behavior is gradually shaped.
3. Immediacy of reinforcement: Probability of future correct responses is inversely proportional to the time lapse between a response and its reinforcement. Furthermore opportunity for frequent responding and reinforcement helps maintain learner's interest and attention.

Some of these psychologically prescribed techniques may sound very similar to procedures which are currently used by many experienced instructors, and indeed they are. However, there may be a difference of degree. For example, Skinner breaks the learning sequence into extremely small steps -- generally short sentences -- and he has indicated that the only way economically to arrange the optimum conditions of reinforcement, immediacy, precision and frequency of response is in a teaching machine. There are problems with the learning theory approach:

a. The early theories are relatively simple and ignore many of the variables which affect human learning. Partly because of this, experimental attempts to confirm the theories with human subjects have not been spectacularly successful. More complex multivariable formulations have been slow in coming.

b. The reinforcement (the feedback of the "systems" approach) has been largely limited to the learner, and only haphazardly applied to the instructor, with the result that systematic improvements in instructional material or presentation methods are scarce.

Another approach, often proposed by engineers, is the systems approach in which people (as students and as teachers) are major components in the system. Generally, this approach emphasizes the "control" advantage of feedback to the student, to the instructor, and to the system evaluator (faculty or society). Feedback control system analogies are loosely used with emphasis on inputs, outputs, transform means, and system constraints. This approach has the following problems:

1. Educational "system" goals are difficult to express in operational terms.

2. Outputs are difficult to evaluate.
3. The function of time, a necessary element of most feedback control systems, has an ambiguous role in education.

Fundamental contributions from this approach will be limited until the above problems are resolved.

The third approach is the data-handling approach, which is relatively unconcerned with any specific learning theory or method for evaluating the educational system outputs. The proponents of this approach claim that with any given sub-set of teaching-learning procedures and with any given measure of output the use of modern data-handling and logic devices would permit much more extensive sampling of pertinent data and use of discriminative decision-making and that improvements in the teaching-learning process can be greatly accelerated, largely on an experiential basis.

None of the above approaches is completely independent of the others. Some balanced blend of the three will probably emerge. All approaches emphasize individual learning and the accumulation of knowledge about the teaching-learning process. All point toward increased mechanization of the bookkeeping chores (grading, record keeping, scheduling); and at least the latter two approaches point toward mechanization and possibly automation of the presentation of learning experiences to the student.

An early study at UCLA [2], and many subsequent experiments, indicated that certain kinds of mechanization are ill-advised, primarily because use of the mechanism does not yield "better" student learning than use of less expensive non-mechanized procedures, and some mechanized devices actually hamper student learning. Nevertheless, it is recognized that just to record and manipulate the multitude of contingent circumstances which affect the teaching-learning process, an efficient data-handling and logic device would be required. The modern digital computer and ancillary equipment have the desired capability, and it is therefore interesting to examine how some people have used the computer in teaching systems.

B. Computers in Teaching Systems

It will be noted that most of the computer-based teaching systems described below are micro-micro systems, concerned with small sequences of learning items. Historically, the impetus for the development of computer-based educational systems came from people primarily imbued with the learning theory approach, even though these people were often engineers or mathematicians working for companies dedicated to systems analyses or to computer design. Interest in the use of computers for larger (macro) systems has received later and less comprehensive consideration, and almost nothing has been done on input-output analyses for computer-based macro-macro systems. While the primary concern here is not with micro-micro systems (the so-called "teaching machines"), a review of the work on these micro-micro systems is revealing, because some fundamental problems arise in these smaller systems which are typical of all educational systems, regardless of size.

In 1958, Gustave Roth, Nancy Anderson and R. C. Brainerd of the IBM Research Center, following a suggestion from Dr. William J. McGill of Columbia University, used an IBM 650 computer to simulate a teaching machine. The group was primarily interested in the general characteristics of teaching machines and felt that it would be easier and perhaps less expensive to simulate different kinds of teaching machines with an available computer than to actually construct a number of different kinds of teaching machines. This was the first of a series of investigations in which it was suggested that the computer was valuable for educational research purposes but uneconomical as a regular training device.

Counting, addition, subtraction, multiplication and division in binary arithmetic were taught to individual students via a typewriter input-output station. The machine verifies the student's inputs digit by digit and signals him only when he makes a mistake. The computer program allows for individual differences in skill level and rate of learning. If a student is making no errors, he is given an option to skip 2, 1, or no problems.

When the student makes an error, the choice of the next problem depends on the number of errors the student has made on that section of the binary arithmetic course. If the student made fewer than 5 errors, the computer presents a problem at the same difficulty level as the last problem he completed correctly. If he makes more than 5 errors, he is presented a problem similar in difficulty to one of the first problems in that section of the course. Therefore, branching forward is at the student's option, and branching backward is based on some a priori decision written into the computer program.

Work by the group was discontinued in 1959, but in 1961 a new group under William Uttal resumed work on computer-based teaching. Encouraged by Professor Merrill Flood of the University of Michigan, the group believed that they could demonstrate the economic feasibility of computer-based teaching systems by providing multiplex student input-output stations per computer. Currently, the group is using a transistorized IBM 1410, a multiplexer, four input-output buffers, a card punch and reader, one psychomotor skill station (for teaching stenotyping), six typewriter stations (for teaching statistics and German), a real time clock, and an IBM 355 digital disc storage unit with an IBM 652 control unit which provides a random access audio memory (used for the stenotype and German language training).

Some spectacular results have been obtained by Uttal's group. For example, in the statistics course a group of six students completed half the semester's work in an average of 5.3 hours with an average mid-term examination grade of 94.3%, whereas a group of eight matched students, taking a lecture course at a university from the same instructor who wrote the program for the computer-administered course, required 24 hours of class lectures plus an average of 25 hours homework to get an average grade of 58.4%.

In correspondence and conversations with Dr. Uttal, he admits that the control programs are arrived at largely on an intuitive basis and require a good deal of cut and try modifications. No attempt is made at optimizing the structure of the programs by experimentation. The computer is not being used to calculate anything, but rather is being used as a data throughput and comparison system.

At Bolt, Beranek and Newman, Inc., J. C. R. Licklider, J. A. Sweats, and associates have been using a Digital Equipment Corporation PDP-1 computer which can use either a typewriter or a cathode tube and light pen as an input-output station. Some of the early work by this group in teaching sound discrimination to sonar operators was unsuccessful, possibly because an a priori decision was made to use branching techniques for student acquisition of relatively meaningless non-verbal sounds which actually had very little sequential relationship. Licklider and Sweats' application of human engineering techniques is perhaps more important than their applications of learning theory to computer aided teaching. By careful consideration of the man-machine interface, they were able to reduce learning time by at least 50%. Of further interest are their attempts to teach relations between the symbolic and the graphical representation of mathematical functions by having the student explore the effect of changing the coefficients of an equation and watching the resultant change of the graphical representation on the oscilloscope screen. By careful attention to the multiplexing problem and by use of a special purpose computer, the team at Bolt, Beranek and Newman, Inc. has succeeded in bringing the cost of computer, ancillary eqipment, and overhead down to $1.50 per student-hour, and anticipate that these costs could be further reduced to less than a dollar per student-hour, which is well within the range of current teaching costs.

At the Coordinated Science Laboratory of the University of Illinois, engineers D. L. Bitzer, P. G. Braunfeld, and W. W. Lichtenberger have used the old ILLIAC computer in conjunction with two alpha-numeric student input stations, course material stored on an electronically scanned set of slides, and two TV tube output stations. The computer program is cleverly conceived to allow for individual student differences. The program provides the student with an opportunity to determine the branching procedure by giving him the option to call for "help" sequences or to transfer out of a "help" sequence at any point in the sequence. The student can also use the computer for computational work to help speed solutions to problems in which computational skill is not the primary objective of the lesson.

At the System Development Corporation, John Coulson and Harry Silberman initially used a Bendix G-15 computer, random access slide projector and buffering system, a typewriter input station and opaque screen output station. This was a single station system, but more recently SDC has been using a Philco 2000 computer with a twenty-station multiplexed system. The student station contains multiple choice buttons for student inputs and a numbered read-out window which guides the student to numbered items in a programmed text.

The new SDC installation is also the first to try to go beyond the micro-micro approach, in that consideration is given to using the computer as a data-handling device which would provide diagnostic information on student performance to the teacher-counselor and to the instructor-program writer and would provide scheduling and "systems evaluation" to the school administrator.

All of the computer-based systems mentioned above place considerable emphasis on flexibility in selecting items of instruction to present to the student. Different items can be presented to each student depending on his history of responses to previous items. However, there is a major flaw in all of the above-mentioned procedures. A fixed set of rules as set down in the computer programs controls the teaching-learning process. These rules are usually intuitively determined and their effectiveness is seldom verified by systematic experimentation. Almost all of the people mentioned above relate how much time they spend changing elements of their computer control programs procedures for evaluating and modifying the a priori elements of these programs. This is somewhat surprising, since most of these experimenters agree that feedback on student progress could be used for on-line alteration of the curriculum sequence or pedagogical procedures, and would probably have more important long-range cultural significance than the simple feedback (knowledge of results fed back to the student) currently in use.

One of the earliest proponents of a variable, rather than a fixed, decision process in a teaching system was Gordon Pask of Systems Research, Ltd. His earlier work on "self-organizing" systems led to his propounding [3] the idea of a self-organizing teacher, (automaton) whose first problem is to find a language common to both itself and the student so that the two can "talk" to each other. To establish such conversational interaction, the automaton must be capable of theorizing and model building, and by trying different strategies (arising from different "theories") to eventually build a model which relates the automaton to the student in a satisfactory manner. Then it can effectively communicate new concepts to the student. Pask suggests that such an adaptive teaching machine can be designed in complete ignorance of how students learn. Essentially, the automaton pragmatically discovers how students learn by trying to get students to perform specified tasks.

Pask fails to mention two important criteria in his description of the self-organizing teacher. He does not hint at what would constitute an optimum procedure for trying different strategies, nor does he specify the criterion for determining what constitutes a satisfactory relationship between automaton and student.

The machines which Pask's associates have actually built are very cleverly designed training devices, but they do not incorporate the self-organizing concepts suggested above. Rather, they are adaptive at the same level as the computer-controlled devices mentioned on earlier pages; i.e., they adjust the difficulty level of the instructional material to the performance level of the individual student. One of the earliest adaptive devices developed by Pask's group was for radar operator training, [4] but the best known device is the Solartron Automatic Keyboard Instructor (SAKI) for training operators of keypunch machines. "SAKI" demonstrates that, at least for special purpose teaching situations, certain decision functions can be performed by compact electronic devices far less complex than the digital computers employed by other research groups.

A student using "SAKI" views an exercise line consisting of alpha-numeric characters which are illuminated one at a time, each for a different length of time. Simultaneously, the student attempts to replicate the characters by depressing the keys on a key-punch machine. A separately illuminated display of the keyboard layout indicates to the student the correct key to depress at the same time that a particular exercise character is being illuminated. This helpful information may be withheld, either completely or partially. If completely withheld, the keyboard layout display lamps are not illuminated; if partially withheld, these lamps are illuminated after a delay period, i.e., some milliseconds after the exercise character has been illuminated.

Unfortunately, the published article [5] which describes the mathematical model of "SAKI" has a number of errors and ambiguities which make a meaningful description of the internal mechanisms of the device impossible. These errors are discussed in Appendix A.

One encounters similar inconsistencies in later papers by Pask. However, of more serious consequence is Pask's use of a probability decision process in his adaptive systems. Every time (t) that a teaching routine must be selected from a set of available routines, a calculation is made for each teaching routine of the probability, $P_j(t)$, that the j-th routine will yield good results. The probability, $P_j(t)$, is based on the history of pay-offs, ρ_j, obtained from prior use of the j-th teaching routine. The probability of the selection of a particular routine is proportional to the $P_j(t)$. This is in essence a Monte Carlo sampling mechanism, and it can be demonstrated that for stationary rules for mapping the ρ_j into P_j, as $t \to \infty$, the average system pay-off will asymptotically approach

$$\frac{\sum \left(\bar{\rho}_j\right)^2}{\sum \left(\bar{\rho}_j\right)}$$

An obviously better procedure than that suggested by Pask would be one where the average system pay off asymptotically approached the supremum of the means of the ρ_j. Such a procedure will be discussed in Section III.

Most recently (July, 1962) the M.I.T. Press published a book, A Decision Structure for Teaching Machines, based on the Ph.D. dissertation of Richard D. Smallwood [6] (Electrical Engineering Department, M.I.T.). Before outlining his decision structure, Smallwood makes some rather strong assumptions.

1. It is possible to specify a matrix of blocks of instructional information, where rows represent the logical sequence of concepts and columns represent alternate forms of information within each row.
2. The probability that a student will respond correctly to a given block is equal to the fraction of students who have previously responded correctly on that block, regardless of the previous histories of learning experiences of the students.
3. Even though a "logical" ordering of blocks must exist, the probability of responding correctly on a block is considered to be independent of the sequence of blocks which were previously seen by the student and independent of his score in those blocks.

Smallwood makes other assumptions about the validity of certain theories of learning (reinforcement, self-pacing, small item size, etc.) which are not really essential for the development of his decision structure and only serve to limit the applicability of that decision structure.

The object of the decision process is to select which one of the instructional blocks from the matrix of possible blocks of information to present next to a given student.

The decision process has as its criterion: maximize the individual student's expected score until this score is above a (arbitrarily) specified minimum level; thereafter minimize the student's expected time to finish the course. The decision as to which block of information a student would be shown next was made as follows:

1. Toss a coin. If "heads", assign student to block for which the average score of previous students' responses was highest (or time was lowest, depending on which part of the compound criterion is governing the process at that instant).
2. If coin toss comes out "tails", assign student to block which has been given to previous students the least number of times.

Smallwood also suggests an alternate decision process, namely, that confidence intervals on the parameters determining the average score for each block be estimated, and when "too great a difference in the confidence intervals" for the different blocks exists, that the block with the largest confidence interval on the average score be selected.

Neither of the decision processes given above actually meets the stated criterion. The arbitrary choice of coin tossing to determine when to use the "maximizing" rule and when to use the "information gathering rule" is obviously non-optimal. Also, choosing the block with the largest confidence interval ignores the fact that this block may also have one of the smaller average scores. Thus, in both schemes, the process may choose blocks which result in sub-maximum scores with unnecessary frequency.

Furthermore, there is a contradiction between the criterion and the reasons given for using the particular decision processes. If the criterion is to maximize a specific individual student's expected score, then one should <u>always</u> assign this student to the block which has the highest average score of previous students' responses (this is similar to the decision process recommended by Bradt, Johnson & Karlin [7] for the two-armed bandit problem where there is but one play remaining). The implication one draws from the use of a "forced" choice (a non-maximizing choice) is that information gained from such "forced" choices will be of use in selecting the expected maximum block for later students. Therefore, the decision process does not adhere to the stated criterion of maximizing a particular student's expected score but rather implies that the criterion is to maximize the sum of all students' scores, i.e., maximize

$$S_n = {}_1X + {}_2X + \ldots + {}_nX \ldots + {}_NX$$

This point will be the key to the next section.

SECTION III

CRITERION AND DECISION RULES FOR AN ADAPTIVE SYSTEM

A. Criterion Function

Some confusion in discussions on adaptive systems could possibly be avoided if everyone took pains to describe the level or levels of adaptive behavior involved in each system. All of the devices described in the preceding section are called "adaptive devices" by their creators, but the level of adaptivity is not the same in all cases. For educational systems (regardless of size) the following levels of adaptive behavior are defined:

Zero Level Adaptive Behavior: A fixed, preconceived strategy (or pedagogy) is used for presenting to all students a fixed, preconceived set of courses or list of subject matter.

First Level Adaptive Behavior: A fixed strategy which uses an individual student's past history of performance to determine which particular course or list of subject matter from a preconceived set of such courses or subject matter is shown to that individual student.

Second Level Adaptive Behavior: The particular courses or list of subject matter which is shown to a particular student is determined by a fixed strategy which uses an individual student's past history of performance and the history of performance of all students who have previously gone through the system.

Third Level Adaptive Behavior: A set of strategies for presenting students with courses or lists of subject matter is available. The choice of a particular strategy for a particular student depends on the history of performance for each of the strategies.

(Separation of strategies and courses or lists of subject matter is a verbal convenience. Lists of subject matter could just as readily be considered sub-sets of strategies, in which case the ideogram is simplified. This also eliminates the distinction between second and third level adaptive behavior. Hereafter, use of the word "strategy" will imply both the pedagogical technique and the subject matter employed by the pedagogical technique.)

The zero and first level adaptive systems do not include provisions for data gathering or experimentation within the system. These systems are non-optimizing and their success is largely dependent on the subjective choice of the strategy.

With the exception of Smallwood's system, all of the computer-controlled micro-micro systems described in Section II fall into the zero or first level of adaptive behavior, even though it can be shown that elaborate data processing equipment need be used for such systems [8], [9].

Systems with higher levels of adaptive behavior must include provisions for storing information on students' performance and for experimenting, i.e., trying different strategies. In such systems students are simultaneously learners and "experimental subjects", and the traditional experimental approach of ignoring the effects on students who have been exposed to sub-optimal regimens should not be tolerated. It is this consideration which leads to the choice of the criterion: Maximize S_n, the sum of all students' net output. This criterion becomes increasingly important where changes in strategy (pedagogical techniques and/or subject matter) occur relatively frequently, so that the total number of students who could possibly be exposed to a given set of strategies is relatively small. Conversely, this criterion is needed for systems in which frequent change (hopefully towards the "better") is a desirable feature.

For second or higher level adaptive systems, the criterion stated above is equally applicable to micro-micro, micro, macro, and macro-macro systems. Therefore, in exploring the possible decision rules or procedures which could meet the stated criterion, the problem will be treated in a general way and no mention will be made of the size of the system. Later, when considering the problem of collecting data for systems which use the stated criterion, the size of the system will again be of some consequence, and systems of different size will have to be treated separately.

B. Sequential Decision Rules

For the general situation (independent of system's size) let ${}_nX_{ij}$ be a collection of random variables defined on a probability space $\mathfrak{J}$. ${}_nX_{ij}$ may be thought of as the random quantity that represents the n-th drawing in a sequence of drawings from a set of populations, $\pi_1, \pi_2, \ldots, \pi_j, \ldots, \pi_k$ where the subscript "i" indicates the i-th drawing from the j-th population. The π_j populations are specified by their cumulative distribution functions, $F_j(x)$. It is assumed that these functions of the random variables have expectations or means,

$$\mu_j = \int_{-\infty}^{\infty} x \, d\, F_j(x).$$

In the application to an educational system, the set of populations could represent different pedagogical procedures or different sequences of learning items, such as the blocks used by Smallwood. The random variable is considered, in some mysterious way, to represent the net return attributable to bringing together the n-th student and the j-th experience. Later, in Sections IV and V, an attempt will be made to unravel the mystery of how one finds X from such measurable descriptions as student learning time, teacher's time, student test scores, system capital costs, etc. It should suffice here to hint that X is likely to be a complicated functional of functions of random variables and that due consideration will have to be given to the stability of any decision process proposed for use in a real educational system; i.e., the decision process should preferably be one which guarantees that the error in the answer is no worse than the errors in the initial data, and conversely, one should not expect the solutions to have an error magnitude less than the errors in the initial data.

One more clarification is necessary at this point. The n students represent a set from a population of students. It is assumed that there is an isomorphic mapping from the set of the n available students to each of the population distributions and that each transformation is independent (though not necessarily dissimilar) from the others. Note that the "mapping" is from students to measures on the students, and the measures include all information on prior states of the students. That is, the X represents net returns or, if you will, a utility of the increase in performance ability as a result of being exposed to a particular educational experience (the transform).

Before considering adaptive decision procedures for maximizing S_n, some boundaries must be placed on the problem. Adaptive decision procedures will only be considered for the case where one desires to maximize S_n for an a priori set of possible strategies. In this scheme, non-contender strategies (i.e., those strategies with little chance of being selected as the "best" strategy) can be eliminated prior to the termination of the process, but new strategies can only be introduced for consideration before the process begins. Whenever a new contender comes to light, the problem is terminated and a new problem initiated. The same adaptive decision procedure may be used for the first and second problems, though it is more likely that a different decision procedure would be used for each, since more a priori information would be available for the second than for the first problem.

Herbert Robbins [10] in 1952 first focused attention on the problem of how to draw a sample ${}_1X, {}_2X, \ldots {}_nX$ from two populations if the object is to achieve the greatest possible expected value of the sum $S_n = {}_1X + {}_2X + \cdots + {}_nX$. Robbins indicated that this problem fits into the general context of sequential design of experiments, in which the size and composition of samples are not fixed in advance but are themselves functions of the observations, and as such, was the outgrowth of earlier work by Dodge and Romig [11] in double sampling inspection methods, and Wald's [12] theory of sequential design.

The available a priori information plays a most important role in the selection of a decision procedure, and some a priori knowledge conditions will be outlined below.

First, there is the "maximum ignorance" case, where there is no a priori knowledge of the distributions of the X_j, the relative magnitude of the μ_j, nor of the total number of students (max n = N) available prior to the termination of the process. Sub-classifications of this case occur for $n \to \infty$, and when "nature" can call a halt to the process at any n. Variations of this case occur for the process terminating at: N, a known constant; at N, given a known probability distribution on N; at $\mathfrak{f}(n)$.

Secondly, a priori knowledge may exist on the distributions of the X_j. The distributions may be binomial, gaussian, etc. It is conceivable that for a given problem some of the X_j will have one distribution and others of the X_j will have another distribution. The same sub-classifications given for the "maximum ignorance" case hold here too, namely; $n \to \infty$, N = unknown constant, N = known constant, and stopping at $\ell(n)$. Further sub-classification can be made for existence or non-existence of a priori estimates of the population parameters. In all of the above cases, the sampling process could be continuous until the end of the process, or, where the cost of making observations on samples is of consequence, the sampling could terminate before the end of the process. Some of the possible cases, separated according to the classifications given above, are shown in Figure 1. Those cases which will be discussed in more detail below are also indicated in Figure 1.

Case I A i. For the case of maximum ignorance, where the only thing known is that the distributions in $\mathcal{F}$ have finite means, no decision rule can be specified which will ensure that the sum of the net values of the observations S_n will be a maximum. However, if it is known that for each distribution there exists a second (or higher order) moment which is uniformly bounded, then C. L. Mallows and Herbert Robbins [13] suggest a decision rule which maximizes S_n in the sense:

$$P\left\{\lim_{n\to\infty} \frac{S_n}{n} = \mu_K\right\} = 1$$

or

$$\lim_{n\to\infty} \frac{E(S_n)}{n} = \mu_K$$

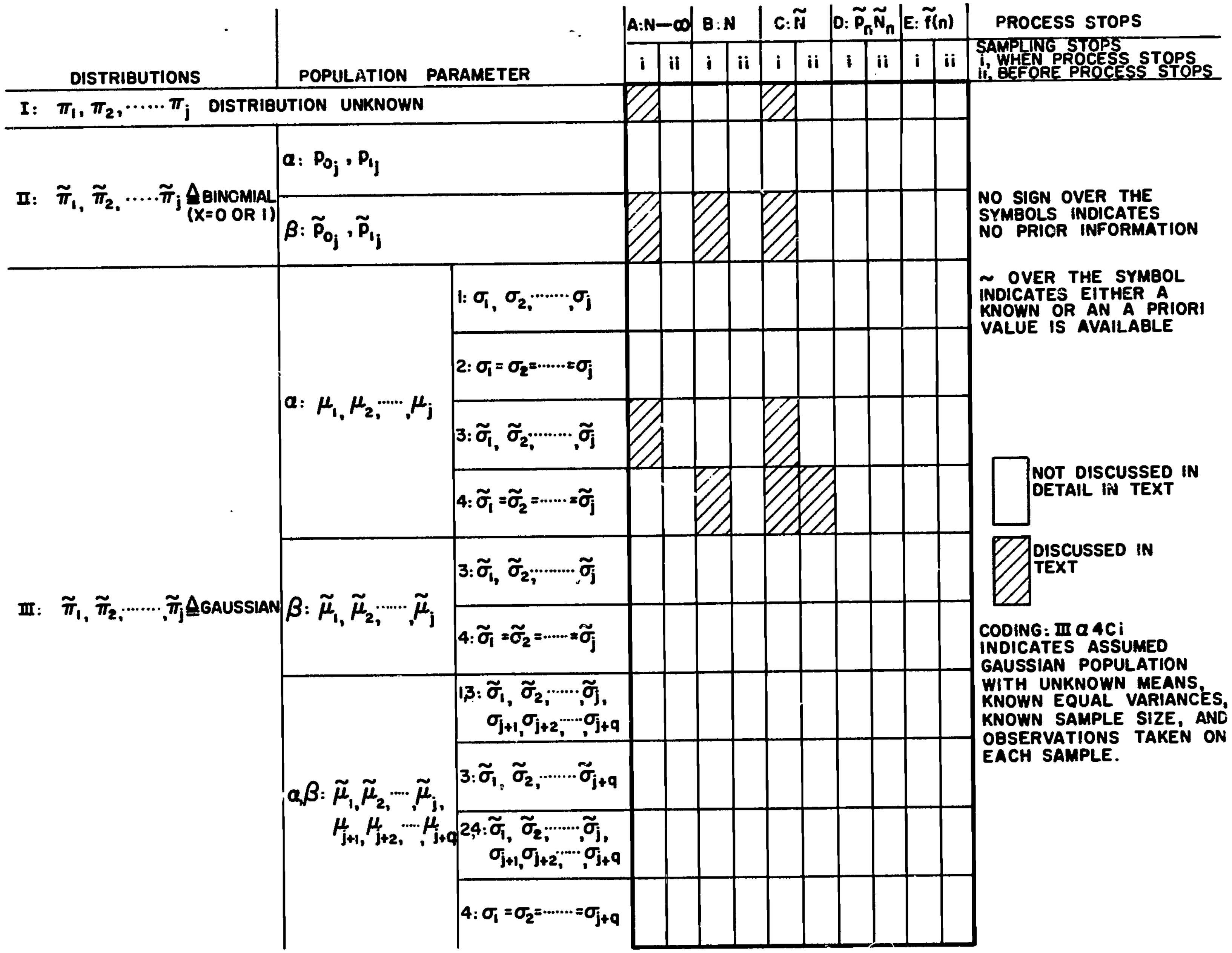
where μ_K is the mean of the population with the highest mean.

The recommended decision rule entails the following:

a. Specify a sequence $B_1, B_2, \ldots B_j \ldots$ of disjoint monotonic sequences of integers, with $B_j = B_{jh}$; $h = 1, 2, \ldots,$

$$b_{j1} < b_{j2} \ldots ; j = 1, 2, \ldots, k$$

$$b_{11} = 1$$



CASES OF A PRIORI INFORMATION
IN DECISION PROCESSES

FIGURE 1

$B_j \cap B_{j^*} = \phi$; $j \neq j^*$ (i.e., the intersection of two sequences is the empty set)

where

$$B_j \neq \phi$$

and

$$(b_{j2} - b_{j1}) < (b_{j3} - b_{j2}) < \ldots$$

to which we add that, for convenience,

$$b_{j1} = j.$$

b. If $n \in B$, use decision D_1: select the n-th observation from the j-th category.

c. If $n \notin B$, use decision D_0: select the n-th observation from the category which had the highest sample mean at the (n-1)th trial.

An observation selected according to D_0 is called free, and one selected according to D_1 is called forced.

Forced observations, made according to a predetermined sequence of inspection epochs, are required for the proof of

$$P\left\{\lim_{n\to\infty} \frac{S_n}{n} = \mu_K\right\} = 1$$

and also satisfy the intuitive notion that some such procedure should be used to reduce the small but finite probability that the selection process becomes "trapped" in a category which does not have the maximum mean. This possibility of being "trapped" is readily illustrated in the following simple example:

$$\mu_1 = 1; \qquad \mu_2 = 0$$

$${}_1X_{11} < 0; \qquad {}_2X_{12} \geqq 0$$

(selecting one observation from each category) that the n-th observation be made from the category which had the maximum sample mean at the (n-1)th trial.

The expected value of S_n before any observations are taken is:

$$E\left(S_N\right)_o = \sum_{j=1}^{k} \mu_j + \sum_{n=1}^{N-k} \sum_{j=1}^{k} (\mu_j)({}_nP_j)$$

where the first sum on the right-hand side of the equation is the expected value of the first k observations — one from each category -- and the second sum represents values from a branching tree, where at each junction point on the tree there exists a probability (${}_nP_j$) that one of the k categories will be selected. With no knowledge about the distributions, the ${}_nP_j$ cannot be estimated, and neither analytical nor computational solution exists for this case. However, a simulation study of some possible decision rules is revealing.

Although one of the conditions initially imposed on the decision rule for this case is that it should be usable for any set of populations for which the cumulative distribution functions have finite expectations, and there exist second (or higher) order absolute moments which are uniformly bounded, sets of k normally distributed populations having equal variances and equal contrasts between the means were selected for convenience in the simulation. (The computational work was done on the University of California IBM 7090, IBM 1401 and IBM 1620 computers.)

Since the current study is a part of an old and continuing search for an appropriate framework for adaptive educational systems, the decision rules suggested by earlier experimenters were included in the simulation study. Admittedly, some of these rules could, under certain conditions, be eliminated from consideration by analytical methods. However, these rules were examined for three

Then at n = 3, j = 2, and since

$$E(\overline{X}_2) = \mu_2 = 0 > {}_1X_{11}$$

it is possible that using D_o no further observations will be made from the j = 1 category which has the larger mean.

It was the concern over the possibility of "trapping", or as he put it: "the dangers ... that the decision process may eliminate some of the alternatives from consideration because of lack of data on the consequence of the alternatives", that led Smallwood to use the coin-tossing analog to select forced and free decision rules. The drawback of Smallwood's forcing rule is that no matter how much information is accumulated on the categories whose sample means $\overline{X}_j < \overline{X}_K$, the frequency of selecting from these j categories remains unchanged.

Robbins' B_j is completely arbitrary, within the limits defined above for B_j, and for $n \to \infty$ one set of B_j is just as good as another. However, the case of $n \to \infty$ is not of particular interest within the context of the type of evolving adaptive educational system that has been suggested earlier.

Case I B. No unique solution exists for this case.

Case I C i. For the case of a finite N, convergence with probability one cannot be demonstrated every time the problem is run as in Case I A i. The best that can be expected of a decision rule for finite N is that the expected value of S_n is maximum in some sense, i.e.,

$$\lim_{u \to \infty} \frac{E\left(S_{N,u}\right)}{Nu} \to \mu_K$$

where u is the number of times the problem is repeated, in which case there is an intuitive appeal to the decision rule which requires reasons: 1) the rules could possibly be used under conditions where they could not readily be eliminated by analytical methods; 2) even where analytical methods could theoretically be used, the analytical methods may be more cumbersome than the empirical methods; 3) some insight was desired on the magnitude of the difference resulting from using the different decision rules, including the admittedly inferior rules. The possibility existed that a "good" rule might be so much more difficult to implement in the real world as not to justify its use, particularly if the "inferior" rule yielded results not too much below those of the "good" rule.

The following sampling decision rules were tested:

RULE 1. For $n \leq k$ select one observation from each of the k categories. For $n > k$ select the n-th observation from that category which had the highest sample mean at $n - 1$.

RULE 2. For $n \leq k$ select one observation from each of the k categories. For $n > k$ flip an unbiased coin. If "heads", select the n-th observation from that category which had the highest sample mean at $n - 1$. If "tails", select the n-th observation from the category from which the least number of observations has been made.

RULE 3. For $n \leq k$ select one observation from each of the k categories. For $n > k$ flip an unbiased coin. If "heads", select the n-th observation from that category which had the highest sample mean at $n - 1$. If "tails", select the n-th observation from the category which had the highest product of the sample mean and the sample standard deviation at $n - 1$.

RULE 4. For $n \leq k$ select one observation from each of the k categories. For $n > k$ select the n-th observation from the category which had the highest product of the sample mean and the sample standard deviation (s) at n - 1.

RULE 5. For $n \leq k$ select one observation from each of the k categories. For $n > k$ select the n-th observation such that the probability that the n-th observation will come from the j-th category is:

$$P_j = \frac{\left({}_{n-1}\bar{X}_j\right)\left({}_{n-1}s\right)}{\sum_{j=1}^{k}\left({}_{n-1}\bar{X}_j\right)\left({}_{n-1}s\right)}$$

RULE 6. For $n \leq k$ select one observation from each of the k categories. For $n > k$ select the n-th observation such that the probability that the n-th observation will come from the j-th category is:

$$P_j = \frac{{}_{n-1}\bar{X}_j}{\sum_{j=1}^{k}\left({}_{n-1}\bar{X}_j\right)}$$

RULE 7. For $n \leq k$ select one observation from each of the k categories. For $n > k$ if $n \in B_j$, select the n-th observation from category j. If $n \notin B$, select the n-th observation from the category which had the highest sample mean at n - 1.

Rule 2 is Smallwood's decision rule. Rule 4 is derived from an untested suggestion by Smallwood. Rule 3 is a mixture of Rules 2 and 4. Rule 6 is Pask's decision Rule. Rule 5 is a mixture of Smallwood's Rule 4 and Pask's Rule 6. Rule 7 is Robbins' decision rule. Rule 1 is a simplification of Robbins' rule; i.e., it is the case where the set B is the empty set.

Another rule:

RULE 8. For $n \leq k$ select one observation from each of the k categories. For $n > k$ modify the forcing set B according to sub-rule Z; then if $n \in B_j$, select the n-th observation from category j. If $n \notin B$, select the n-th observation from the category which had the highest sample mean at n - 1.

Rule 8 was not tested because "sub-rule Z" could not be specified at this point in the investigations. It was hoped, however, that the simulation study would shed some light on possible sub-rules.

For Rule 7, the following arbitrary forcing set B was specified:

Category	Set B, k = 4		Set B, k = 6		Set B, k = 8	
A	9	36	9	36	9	36
B	11	44	10	40	10	40
C	14	53	11	44	11	44
D	18	63	14	48	12	48
E			17	53	14	50
F			18	63	15	53
G					17	58

At each of the u repetitions of the problem, the numbers in each column were randomly scrambled. For example, for the second iteration of the problem with k = 4, the forcing set was:

Category	Set B	
A	11	53
B	9	63
C	14	44
D	18	36

The integers in each column of Set B were selected so that no matter what combination of integers randomly appeared in the first and second columns, adherence would be made to the restriction that:

$$\left(b_{j2} - b_{j1}\right) < \left(b_{j3} - b_{j2}\right) ;$$

where

$$b_{j1} = j$$

A preliminary set of simulations was made to demonstrate how each rule behaved in individual iterations of the problem. An example is shown in Table 1, where Rule 1 was used with $k = 4$, $\mu_K = 85$, $\sigma = 10$, and contrast of 10. In the first run, "trapping" occurred in π_3. In the second run, all observations after the first k are taken from the category which has μ_K. In the third run, some switching between π_3 and π_4 occurs before all subsequent observations are taken from the category which has μ_K. The results of these preliminary simulations should be borne in mind during all the subsequent discussions, which will deal exclusively with averages or expectations over many repetitions of the same problem.

The "Expected Values" of $\frac{S_n}{n}$ were obtained from 500 iterations of the same problem. These $E\left(\frac{S_n}{n}\right)$ were obtained for values of N from 1 to 100 -- $\sigma = 10, 20, 30$; $k = 4, 6, 8$; contrasts of 5 and 10 -- and are summarized in Table 2. Also shown for each combination of N, k and contrast is the maximum expected $\frac{S_n}{n}$, i.e., the value that would be obtained if the first k selections yielded numbers equal to $\mu_1, \mu_2, \ldots, \mu_k$ and the subsequent $(N - k)$ selections all yielded numbers equal to μ_K.

Examination of Table 2 reveals that Rules 1 and 7 (derived from Robbins) yield consistently better results than Rule 2 (Smallwood) and Rule 6 (Pask), and also better results than the "mixed" Rules 3, 4, and 5. For reasons that are fairly obvious, the results of Rule 3 should approach the results of Rule 1, and the results of Rule 5 should approach the results of Rule 6. Rule 6 yields results which approach

TABLE I

SIMULATION OF RULE I

$k = 4$; $\mu_K = 85$; $\sigma = 10$; contrasts = 10

	n	π_j	$_nX_j$	$\frac{S_n}{n}$	$\mu_1=55$ $_n\bar{X}_1$	$\mu_2=65$ $_n\bar{X}_2$	$\mu_3=75$ $_n\bar{X}_3$	$\mu_4=85$ $_n\bar{X}_4$
Run 1:	1	1	48.5	48.5	48.5	----	----	----
	2	2	51.6	50.1		51.6	----	----
	3	3	84.8	61.7			84.8	----
	4	4	69.8	63.7			84.8	69.8
	5	3	68.1	64.6			76.5	
	6	3	78.3	66.9			77.1	
	7	3	64.0	66.5			73.8	
	8	3	66.5	66.5			72.4	
	9	3	74.1	67.3			72.6	
	10	3	66.9	67.3			71.8	
	11	3	80.7	68.5			72.9	
	12	3	66.0	68.3			71.9	
	13	3	69.8	68.4			73.3	
	14	3	86.6	69.7			73.0	
	15	3	[illegible]0.7	69.8	48.5	51.6	72.4	69.8
Run 2:	1	1	68.2	68.2	68.2	----	----	----
	2	2	66.7	67.4		66.7	----	----
	3	3	78.4	71.1			78.4	----
	4	4	97.9	77.8				97.9
	5	4	92.0	80.6				95.0
	6	4	85.5	81.4				91.8
	7	4	87.5	82.3				90.7
	8	4	99.3	84.4				92.4
	9	4	72.6	83.1				89.1
	10	4	90.3	83.8				89.3
	11	4	78.1	83.3				87.9
	12	4	77.6	82.8				86.7
	13	4	77.3	82.4				85.8
	14	4	80.7	82.3				85.3
	15	4	77.7	82.0	68.2	66.7	78.4	84.7
Run 3:	1	1	59.1	59.1	59.1	----	----	----
	2	2	51.8	55.5		51.8	----	----
	3	3	82.2	64.4			82.2	----
	4	4	79.0	68.0			82.2	79.0
	5	3	76.6	69.7			79.4	79.0
	6	3	70.2	69.8			76.3	79.0
	7	4	89.7	72.7				84.4
	8	4	77.8	73.3				82.2
	9	4	87.9	74.9				83.6
	10	4	64.4	73.9				79.8
	11	4	78.0	74.2				79.5
	12	4	98.1	76.2				82.1
	13	4	85.7	77.0				82.6
	14	4	73.0	76.7				81.5
	15	4	81.9	77.0	59.1	51.8	76.3	81.6

TABLE 2

EXPECTED VALUES

$$E\left(\frac{S_N}{N}\right) = \frac{1}{N}\left[\frac{1}{U}\sum_{u=1}^{U}(S_N)_u\right] \qquad u = 500 \quad \mu_K = 85$$

Contrast	k	N	Max $E\left(\frac{S_n}{n}\right)$	σ = 10 Rule 1	σ = 10 Rule 2	σ = 10 Rule 3	σ = 10 Rule 4	σ = 10 Rule 5	σ = 10 Rule 6	σ = 10 Rule 7	σ = 20 Rule 1	σ = 20 Rule 2	σ = 20 Rule 3	σ = 20 Rule 4	σ = 20 Rule 5	σ = 20 Rule 6	σ = 20 Rule 7	σ = 30 Rule 1	σ = 30 Rule 2	σ = 30 Rule 3	σ = 30 Rule 4	σ = 30 Rule 5	σ = 30 Rule 6	σ = 30 Rule 7
Contrast = 5	k = 4	20	83.5	81.8	79.1	79.8	78.1	77.7	77.8	81.1	79.8	78.4	79.1	78.1	77.8	77.9	79.5	79.4	78.0	79.0	77.7	78.2	78.6	78.3
		40	84.25	82.6	79.4	81.2	78.2	77.9	77.8	82.4	80.6	78.6	80.1	78.7	77.9	77.8	81.0	79.8	78.6	79.7	78.4	77.9	77.9	80.1
		60	84.5	82.9	79.5	81.9	78.3	78.0	77.8	82.8	81.1	78.8	80.6	78.9	77.9	77.8	81.2	80.1	78.5	80.0	78.5	77.8	78.0	80.2
		80	84.625	83.2	79.6	82.3	78.4	77.9	77.8	83.1	81.3	79.0	80.9	78.9	77.8	77.8	81.5	80.2	78.5	80.5	78.7	77.8	77.8	80.6
		100	84.7	83.3	79.6	82.6	78.4	78.0	77.9	83.3	81.4	79.2	81.3	79.0	77.9	77.9	81.8	80.3	78.6	80.8	78.7	77.9	77.8	80.9
	k = 6	20	81.25	79.6	75.3	75.0	73.5	73.5	73.6	76.4	77.5	74.4	74.8	73.6	73.6	73.8	75.1	75.9	74.3	74.1	74.1	73.8	74.6	74.1
		40	83.125	81.6	76.0	78.0	74.7	73.4	73.5	79.4	79.2	75.1	77.3	74.4	73.4	73.4	77.8	77.9	74.9	76.2	75.3	73.4	73.7	75.8
		60	83.75	82.3	76.3	79.4	75.4	73.4	73.5	80.5	80.2	75.3	78.6	74.9	73.4	73.6	79.2	78.7	75.2	77.6	75.8	73.5	73.8	77.5
		80	84.063	82.6	76.5	80.2	75.8	73.3	73.5	81.3	80.7	75.5	79.5	75.1	73.4	73.5	80.0	79.2	75.3	78.4	76.1	73.5	73.7	78.4
		100	84.25	82.9	76.6	80.8	76.1	73.2	73.4	81.9	81.0	75.8	80.0	75.3	73.4	73.4	80.7	79.4	75.5	79.0	76.3	73.3	73.5	79.0
	k = 8	20	78.0	76.6	71.4	69.5	68.7	68.4	68.9	70.6	73.9	70.6	69.2	68.3	68.1	69.2	69.2	73.2	69.7	68.9	68.3	69.1	69.9	68.6
		40	81.5	80.1	72.4	74.7	71.5	69.1	69.0	76.2	77.7	71.7	74.1	71.0	68.3	69.1	74.3	76.0	70.8	72.7	71.0	68.7	69.7	73.1
		60	82.67	81.3	72.9	77.2	72.9	69.2	69.1	77.3	79.0	72.1	76.3	72.3	68.6	69.3	75.6	77.3	71.4	74.7	72.1	68.9	69.8	74.4
		80	83.25	81.9	73.2	78.5	73.9	69.3	69.1	79.0	79.6	72.4	77.6	72.9	68.9	69.4	77.2	78.2	71.8	76.1	72.8	69.1	69.8	75.8
		100	83.6	82.3	73.5	79.5	74.1	69.4	69.1	80.1	80.2	72.6	76.5	73.3	69.0	69.4	78.5	78.5	72.0	77.0	73.5	69.1	69.7	77.2
Contrast = 10	k = 4	20	82.0	80.8	74.0	76.3	72.5	70.6	70.8	78.4	78.8	72.4	74.8	72.5	70.4	70.9	76.8	76.6	72.3	74.3	72.7	70.5	70.7	75.8
		40	83.5	82.4	74.2	79.8	73.9	71.4	71.5	80.9	80.5	73.0	77.9	74.0	71.2	71.7	79.3	78.4	72.7	76.9	74.0	71.7	71.4	77.6
		60	84.0	82.9	74.4	81.2	74.5	71.6	71.5	81.7	81.2	73.4	79.3	74.6	71.3	71.7	80.3	79.2	73.2	78.2	74.5	72.0	71.6	78.7
		80	84.25	83.3	74.6	81.9	74.9	71.6	71.5	82.3	81.6	73.7	80.2	74.8	71.3	71.5	81.0	79.5	73.4	79.1	74.8	71.9	71.3	79.4
		100	84.4	83.4	74.7	82.4	75.1	71.7	71.5	82.8	81.7	73.9	80.7	75.1	71.6	71.8	81.5	79.8	73.6	79.6	74.9	72.3	71.6	79.9
	k = 6	20	77.5	76.5	66.4	67.5	64.6	62.3	63.6	69.2	73.9	65.1	66.6	63.9	62.0	64.5	67.1	71.6	64.7	66.0	63.9	62.6	64.7	66.0
		40	81.25	80.4	68.0	74.6	69.4	63.9	64.2	75.7	77.9	66.8	73.3	68.2	63.4	64.9	74.2	75.5	66.0	72.1	68.0	63.6	64.6	72.4
		60	82.5	81.7	68.6	77.5	71.3	64.3	64.4	77.5	79.5	67.5	76.0	69.9	64.0	65.0	76.1	77.0	66.7	74.9	69.7	64.1	64.9	74.8
		80	83.125	82.4	68.9	79.1	72.3	64.7	64.4	79.1	80.3	68.1	77.6	70.9	64.6	65.0	77.8	77.9	67.0	76.6	70.7	64.5	64.9	76.2
		100	83.5	82.8	69.1	80.1	72.9	64.8	64.4	80.2	80.8	68.4	78.7	71.5	64.9	65.0	78.9	78.5	67.4	77.5	71.2	64.7	64.9	77.4
	k = 8	20	71.0	70.2	58.7	55.7	54.0	52.3	57.1	56.6	67.8	57.3	55.1	53.8	53.1	56.6	55.8	65.2	56.2	54.5	53.9	52.7	57.8	55.5
		40	78.0	77.1	61.5	68.6	63.4	57.1	58.8	68.8	74.5	59.8	67.4	62.8	57.2	58.4	67.3	72.2	58.8	66.0	62.7	57.3	59.6	66.0
		60	80.33	79.4	62.6	73.5	67.0	58.8	59.7	71.3	76.9	61.0	72.4	66.3	58.5	59.2	69.7	74.7	59.8	70.8	66.4	59.1	60.9	68.3
		80	81.5	80.6	63.1	76.1	68.9	59.6	60.0	74.3	78.2	61.7	74.9	68.0	59.4	59.7	72.9	76.0	60.6	73.2	68.3	59.9	61.3	71.6
		100	82.2	81.3	63.4	77.7	70.1	60.1	60.1	76.4	79.0	62.1	76.5	69.1	60.0	60.0	75.0	77.0	61.3	74.8	69.5	60.6	61.7	73.7

the theoretical value $\frac{\sum(\mu_j)^2}{\sum(\mu_j)}$. It is plainly useless to continue considering the rules suggested by Smallwood and Pask for their adaptive systems.

Focusing attention on Rules 1 and 7, it is observed that in the case of small contrasts and large σ, there is a relatively high probability that Rule 1 selects from sub-maximum categories in the early trials; therefore, Rule 7 shows up better than Rule 1. Where contrasts are large and σ is small, Rule 1 picks fewer sub-maximum categories than the number "forced" by Rule 7. This suggests that the smaller the contrasts, and the larger the σ, the more dense set B should be.

Also, as the number of categories, k, increases, Rule 7 yields lower results, since the number of "forced" selections increases directly as k increases, while the likelihood of finding the category with the maximum mean by the use of "forced" selection decreases with an increase in k.

Case I C ii. If the cost of taking observations, C, and the initial set-up cost for a category, A, are considered, the expected value of S_n before any obs[illegible]ations are taken is

$$E\left(S_N\right)_o = \sum_{j=1}^{k} \mu_j + \sum_{n=1}^{N-k} \sum_{j=1}^{k} \left(\mu_j\right)\left({}_nP_j\right) - \sum_{j=1}^{k} n_j C_j - \sum_{j=1}^{k} A_j$$

where

$$\sum_{j=1}^{k} n_j = N$$

If optimal stopping is permitted, say, at the n^* trial, where $k < n^* < N$, and the remaining $N - k - n^*$ observations are taken from the category with the highest sample mean

$$E\left(S_N\right) = \sum_{j=1}^{k} \mu_j + \sum_{n=1}^{n^*} \sum_{j=1}^{k} \left(\mu_j\right)\left({}_{n}P_j\right) + (N-k-n^*) \sum_{j=1}^{k} \left(\mu_j\right)\left({}_{n^*}P_j\right)$$
$$- \sum_{j=1}^{k} n_j^* C_j - \sum_{j=1}^{k} A_j$$

where

$$\sum_{j=1}^{k} n_j^* = n^*$$

which is unsolvable for the same reasons as given in Case I C i. Again, these questions were explored by a computer simulation of the use of the various rules on specified normal "test" populations. Three sampling costs were considered: no cost (where obviously one should never stop taking observations); a cost of one percent of the μ_K for all π_j, and a cost of ten percent of μ_K. Four values of N were selected: $N = n^*$ (the sequential selection process stops and no students remain to assign to the category with the largest sample mean); N = 100; N = 1,000; and N = 10,000. Instead of having to compare each line of $E\left(\frac{S_n}{n}\right)$ values with its maximum $E\left(\frac{S_n}{n}\right)$ as was the case in Table 2, the results of the first k observations were excluded from the summations (though not from the decision-making procedure) shown in Tables 3 and 4, with the result that the single standard of comparison is $\mu_K = 85$. The expectation is now $E\left(\frac{S_{n-k}}{N-k}\right)$, taken over five hundred iterations.

A preliminary examination of Rules 2, 3, 4, 5, and 6 under the above conditions again showed that these rules yield lower results than do Rules 1 and 7. Rule 1, of course, is the same as Rule 7 with the set B as the empty set. The set B used in Rule 7 for computing the expected values of Table 2 can be considered a moderately dense set and was used again for the computations of Tables 3 and 4.

TABLE 3

EXPECTED VALUES WITH OPTIMAL STOPPING

E: B = Empty Set
M: B = Moderate Set
F: B = Full Set

N = 100

			σ = 10						σ = 30					
			Cost = .01 μ_K			Cost = .1 μ_K			Cost = .01 μ_K			Cost = .1 μ_K		
		n^*	E	M	F	E	M	F	E	M	F	E	M	F
Contrast = 5	k = 4	5	81.7	81.6	81.4	81.6	81.6	81.3	79.6	78.8	78.9	79.5	78.7	78.8
		10	82.7	82.4	82.1	82.3	81.9	81.6	80.2	79.9	79.3	79.7	79.4	78.8
		15	83.0	82.9	82.4	82.1	82.0	81.5	80.6	80.2	79.6	79.7	79.4	78.7
		25	83.1	83.3	82.3	81.4	81.6	80.6	80.9	80.7	79.7	79.2	79.0	78.1
		50	83.0	83.2	80.8	79.3	79.5	77.2	80.7	80.7	79.3	77.1	77.0	75.7
		75	82.8	82.9	78.8	77.1	77.2	73.2	80.5	80.5	78.2	74.8	74.9	72.6
		100	82.5	82.6	76.8	74.8	75.0	69.1	80.1	80.2	76.7	72.5	72.6	69.1
	k = 8	10	81.5	80.5	80.5	81.3	80.3	80.3	75.0	74.8	73.8	74.8	74.6	73.6
		15	82.4	81.1	80.8	81.8	80.6	[illegible]	76.3	74.9	75.1	75.7	74.4	74.5
		25	82.9	82.1	79.9	81.5	80.7	78.5	78.7	77.1	75.9	77.3	75.7	74.4
		50	82.9	81.3	76.1	79.5	77.8	72.6	79.2	77.6	73.7	75.8	74.1	70.2
		75	82.7	80.6	71.4	77.1	75.0	65.8	79.2	77.4	70.4	73.6	71.8	64.9
		100	82.4	80.3	66.8	74.8	72.6	59.1	79.0	76.9	66.6	71.3	69.3	58.9
Contrast = 10	k = 4	5	82.5	81.9	82.0	82.4	81.9	81.9	75.6	75.9	75.9	75.5	75.8	75.8
		10	83.6	83.4	82.7	83.2	83.0	82.2	77.6	78.3	77.4	77.2	77.8	76.9
		15	83.7	83.7	82.4	82.8	82.8	81.6	78.8	79.6	77.9	78.0	78.7	77.0
		25	83.8	83.6	81.3	82.1	81.9	79.6	79.4	80.2	78.0	77.0	78.5	76.3
		50	83.7	83.2	77.5	80.0	79.6	73.8	79.5	80.4	75.7	75.8	76.8	72.1
		75	83.4	82.7	73.3	77.7	77.0	67.7	79.2	80.0	72.6	73.5	74.4	66.9
		100	83.2	82.5	69.3	75.5	74.8	61.7	78.9	79.7	69.1	71.2	72.1	61.5
	k = 8	10	82.8	81.0	81.4	82.6	80.9	81.3	72.9	71.5	71.7	72.8	71.3	71.6
		15	83.6	80.6	80.4	83.0	80.0	79.8	77.4	73.4	73.4	76.8	72.8	72.8
		25	83.8	81.1	77.3	82.4	79.7	75.9	79.1	76.4	71.9	77.7	75.0	70.5
		50	83.6	79.4	68.4	80.1	75.9	64.9	79.6	76.0	66.3	76.1	72.5	62.8
		75	83.3	77.9	58.9	77.8	72.4	53.3	79.4	75.2	58.1	73.9	69.6	52.6
		100	83.1	77.7	49.6	75.4	70.1	41.9	79.1	74.8	49.3	71.4	67.1	41.7

TABLE 4

EXPECTED VALUES WITH OPTIMAL STOPPING

N = 1,000

			σ = 10						σ = 30					
			Cost = .01 μ_K			Cost = .1 μ_K			Cost = .01 μ_K			Cost = .1 μ_K		
		n^*	E	M	F	E	M	F	E	M	F	E	M	F
Contrast = 5	k = 4	5	81.7	81.6	81.5	81.7	81.6	81.5	79.6	78.8	78.9	79.6	78.8	78.9
		10	82.8	82.5	82.4	82.7	82.5	82.3	80.3	80.0	79.5	80.2	79.9	79.5
		15	83.1	83.3	83.0	83.0	83.2	82.9	80.6	80.4	80.0	80.5	80.4	79.9
		25	83.3	83.9	83.6	83.2	83.7	83.5	80.9	81.0	80.4	80.8	80.8	80.3
		50	83.5	84.1	84.2	83.1	83.8	83.8	81.2	81.5	81.3	80.9	81.2	80.9
		75	83.5	84.3	84.2	83.0	83.8	83.7	81.2	81.8	82.0	80.7	81.3	81.5
		100	83.6	84.3	84.1	82.8	83.6	83.4	81.4	82.1	82.1	80.6	81.4	81.3
	k = 8	10	81.5	80.8	80.7	81.5	80.7	80.7	75.0	74.9	73.9	75.0	74.9	73.9
		15	82.6	82.1	81.8	82.5	82.0	81.8	76.5	75.3	75.7	76.5	75.3	75.7
		25	83.2	83.6	82.6	83.1	83.4	82.5	79.4	78.1	77.6	79.3	78.0	77.5
		50	83.4	83.9	83.3	83.1	83.6	83.0	80.2	80.0	79.0	79.9	79.7	78.7
		75	83.4	84.1	83.1	82.9	83.6	82.6	80.8	81.1	80.0	80.3	80.6	79.5
		100	83.4	84.1	82.8	82.7	83.4	82.1	80.8	81.3	80.1	80.1	80.6	79.4
Contrast = 10	k = 4	5	82.5	81.9	82.1	82.5	81.9	82.1	75.6	75.9	75.9	75.6	75.9	75.9
		10	83.7	83.7	83.4	83.7	83.7	83.4	77.7	78.7	77.9	77.6	78.6	77.8
		15	83.9	84.3	83.9	83.8	84.2	83.8	79.0	80.2	78.9	78.9	80.1	78.8
		25	84.1	84.5	84.3	83.9	84.3	84.1	79.6	81.0	80.2	79.4	80.8	80.0
		50	84.2	84.6	84.3	83.8	84.3	83.9	80.2	82.0	31.3	79.8	81.6	80.9
		75	84.2	84.6	83.9	83.6	84.1	83.4	80.3	82.3	81.9	79.7	81.8	81.4
		100	84.2	84.6	83.5	83.4	83.9	82.8	80.4	82.3	82.2	79.7	81.6	81.5
	k = 8	10	82.8	81.7	82.1	82.8	81.7	82.1	73.1	72.0	72.1	73.0	71.9	72.1
		15	83.7	82.6	82.7	83.6	82.6	82.7	77.7	74.8	75.1	77.6	74.8	75.1
		25	83.9	84.0	83.1	83.8	83.9	83.0	79.6	78.8	76.3	79.5	78.7	76.2
		50	84.0	84.3	83.1	83.7	83.9	82.8	80.4	80.9	78.8	80.1	80.5	78.8
		75	84.0	84.2	82.4	83.4	83.7	81.9	80.7	81.7	79.7	80.2	81.2	79.2
		100	84.0	84.2	81.1	83.2	83.5	81.0	80.7	81.8	79.4	80.0	81.1	78.7

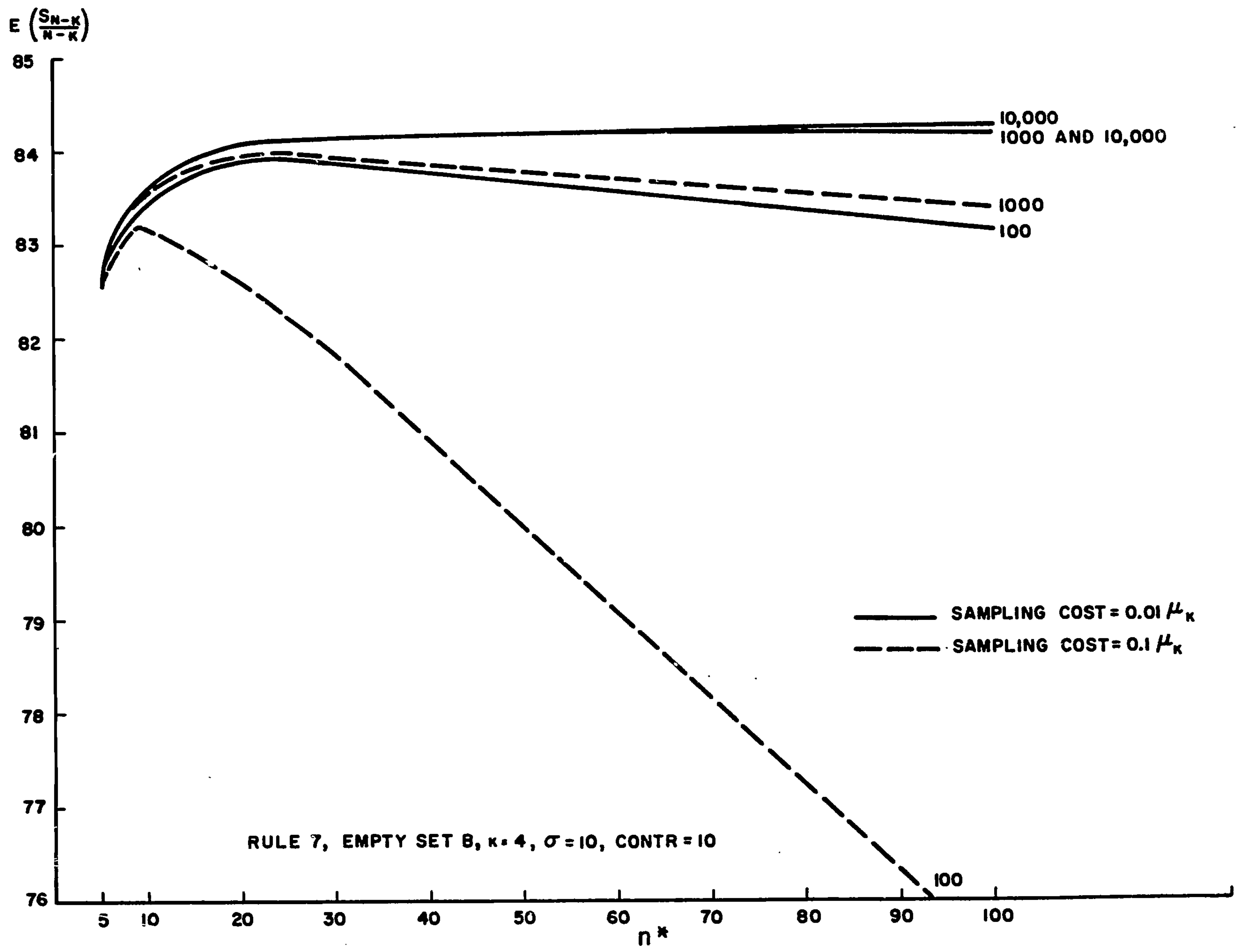
A "full set" B is one with each integer present, resulting in all observations being taken according to D_1, the "forcing" decision. Such a full set does not meet the restriction that

$$\left(b_{j2} - b_{j1}\right) < \left(b_{j3} - b_{j2}\right) < \cdots ,$$

but does present a convenient opposite to the other extreme case of an empty set. Also, use of such a full set is almost akin to those classical sequential sampling techniques which select one observation from each category prior to each decision step.

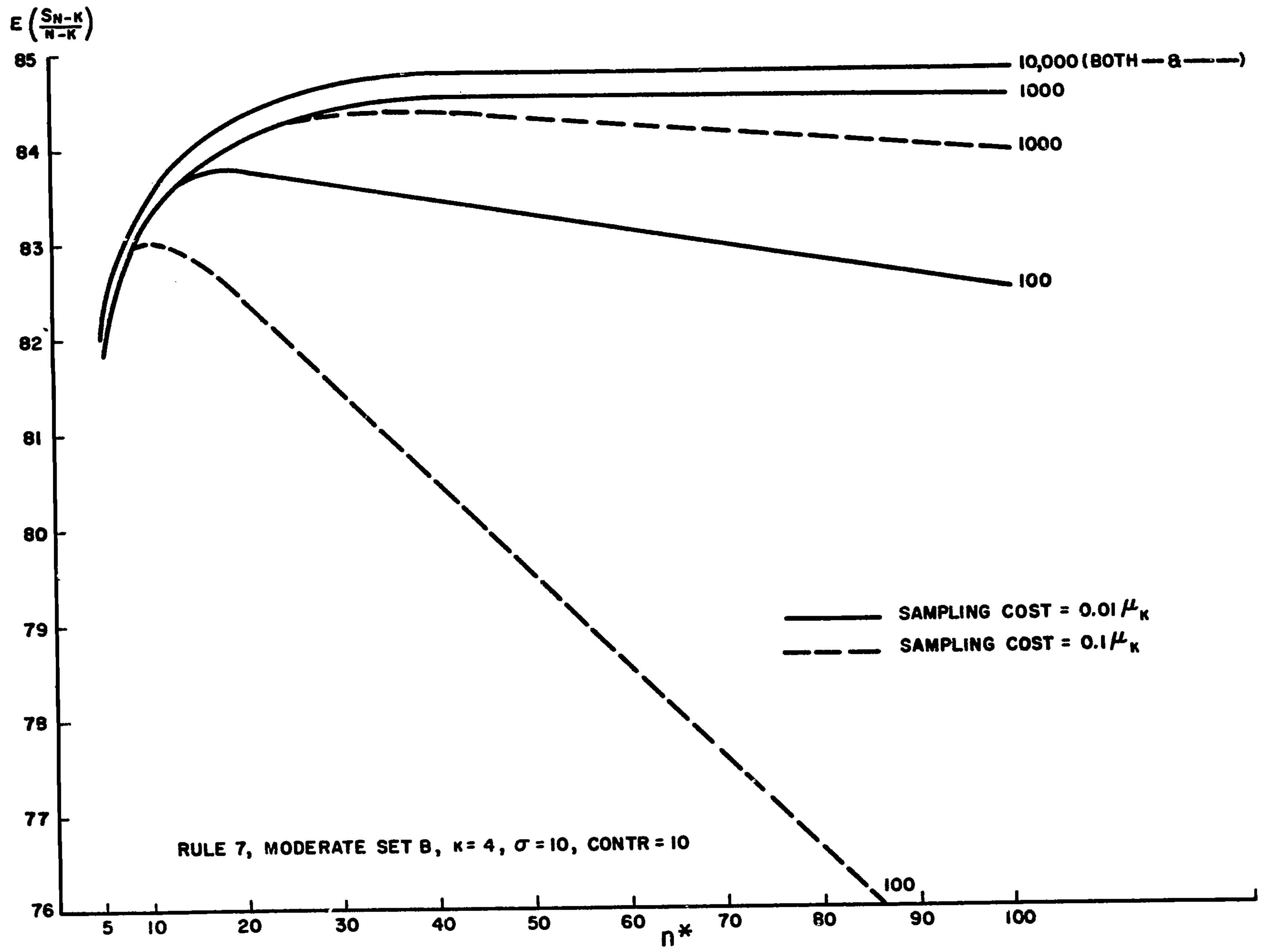
The general conclusion from this simulation is that both the density of set B and the optimum stopping point n^* are primarily dependent on the total number of students available, N, and less dependent on the size of k, σ, contrasts, and sampling costs (where these are moderate percentages of μ_K). This conclusion can be inferred more readily from some graphs than from Tables 3 and 4; and Figures 2, 3 and 4 illustrate results for the typical case of k = 4, $\sigma = 10$, and contrast = 10. From these figures it appears that the larger the N, the more dense should set B be, and the longer should one keep on sampling. If, however, N is determined by some decision process outside of the system -- i.e., the experiment may be terminated at any n = NH -- then Figure 5 shows that the empty set B is best.

The question now arises: for Case I, if one starts with Rule 7 and an a priori forcing set B_j, is it possible to modify this set as one gains information on the π_j? Since the forcing set is introduced to reduce the probability of being "trapped" in the wrong category, sample values are useless in determining what this set B_j should be, unless one wishes to make additional assumptions about the distributions of the π_j. A possible assumption is that all the π_j have the same distribution, only differing by the value of a parameter, say the



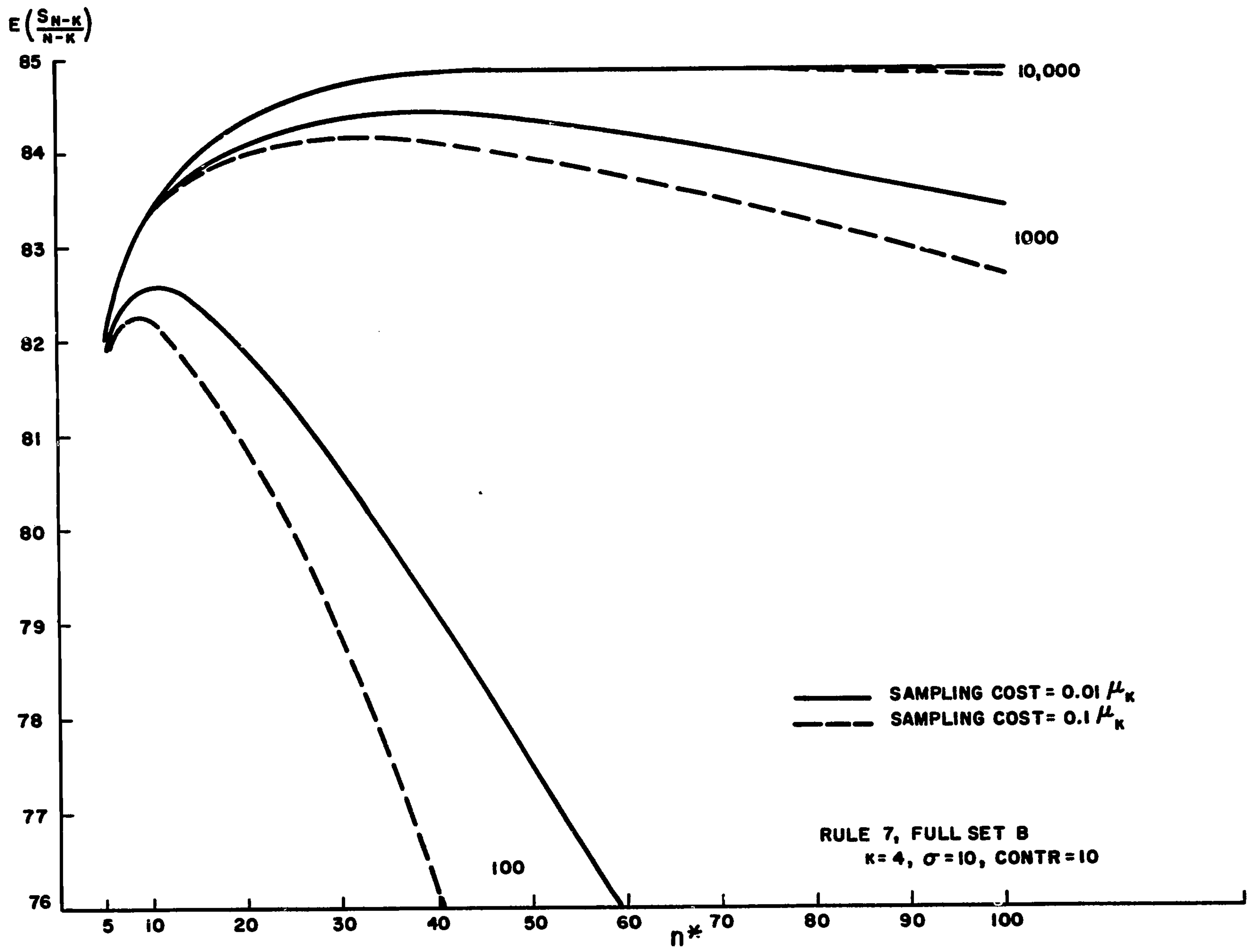
EXPECTED SUMS USING RULE 7,
EMPTY FORCING SET

FIGURE 2



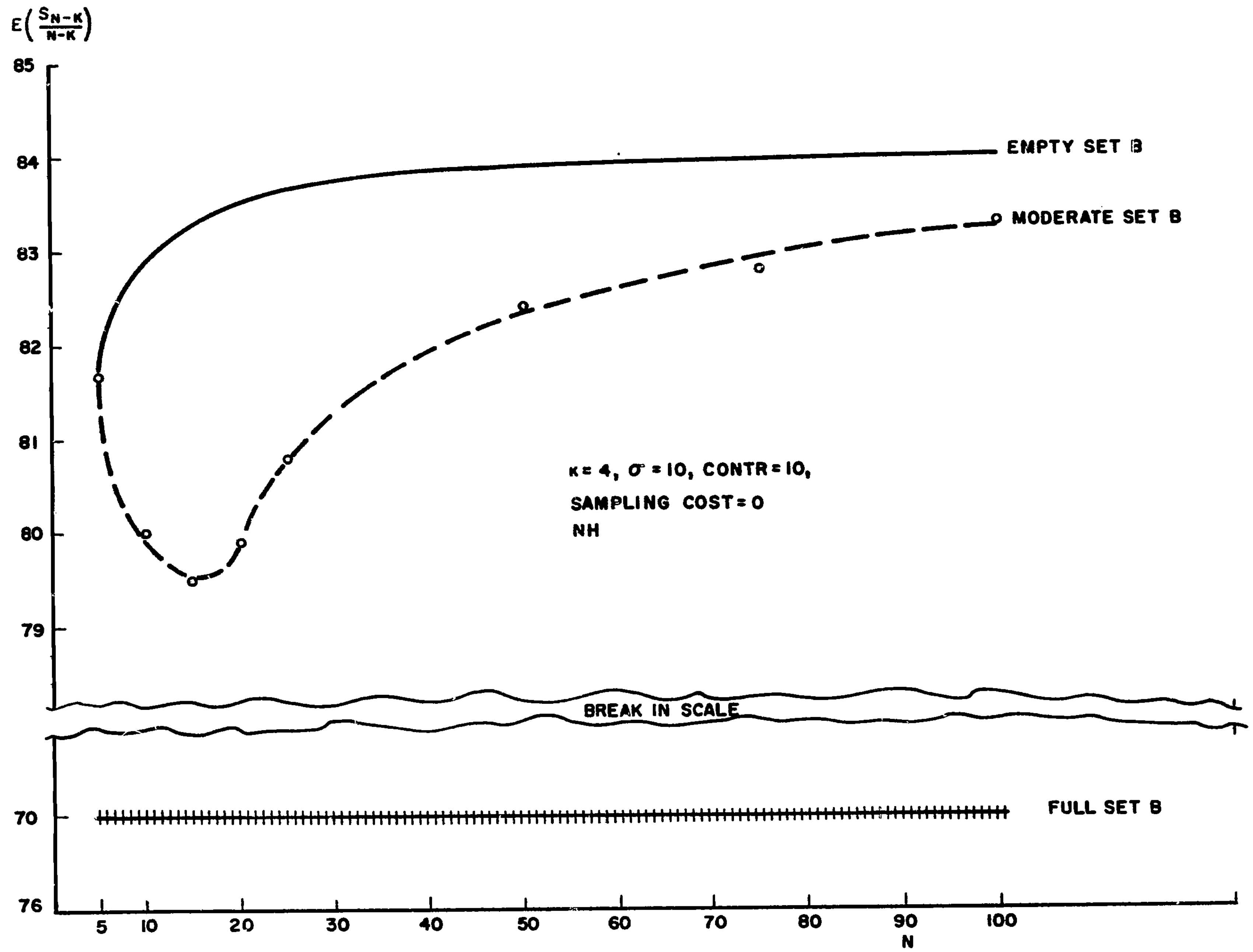
EXPECTED SUMS USING RULE 7,
MODERATE FORCING SET

FIGURE 3



EXPECTED SUMS USING RULE 7,
FULL FORCING SET

FIGURE 4



EXPECTED SUMS, NON-OPTIONAL STOPPING

FIGURE 5

means, μ_j. With this assumption, a point could be reached where sample values from the category with max n_j could be used to estimate the nature of the distributions. At this point, however, instead of changing set B_j it would probably be advisable to shift the problem to Case II, Case III, or other cases appropriate to the underlying distributions of the π_j.

Actually, it is hardly conceivable that Case I conditions could exist in any real educational system. Some Case II situations exist, but the preponderance of situations falls into Case III. Anticipating the results of later sections, it can be stated with a fair degree of certainty that the X's that will be used for the adaptive decision structure will be approximately normally distributed. Why then consider Case I? The reason is that the decision rules used for Case I require relatively little computational work (or hardware), whereas Case III decision rules may require a tour de force in computation and analytical techniques or hardware that does not currently exist. It is partly for this reason that the empirical studies in Case I were made in normal distributions, for if decision rules derived for the non-parametric case yield results not too inferior to those obtained from the more difficult Case III decision rules, then there could be some practical advantages to using the simpler rules.

Only brief mention will be made of the Case II problems, since in complexity they fall between Case I and Case III, and techniques developed for Case III can be used with some simplification for Case II.

Case II β. R. N. Bradt, S. M. Johnson, and S. Karlin [7] considered the special case of devising a sequential design which would maximize the expected value of the sum of n observations from two binomially distributed populations when the expected values of the distributions are known, though only an a priori probability is given to indicate which expected value is associated with which distribution. This special case was popularly called the "Two-Armed Bandit Problem" from its similarity to a familiar gambling situation.

R. Bellman [14] and M. Sakaguchi [15] couched the same special case in dynamic programming terminology.

Walter Vogel [16] considered the same special case and further examined this problem with the additional restriction that k observations are initially made on each of the two populations before the sequential sampling rule is employed [17].

Finally, Dorian Feldman [18] showed that for both a specified number, N, of observations and for an infinite number of observations, the optimum (in the expected value sense) decision rule is to always select the n-th observation from that category for which the Bayesian posterior probability at n - 1 is greatest.

Case III α C. Several approaches are available in the case where the π_j are assumed to be normally distributed and differing only in the (unknown) value of μ_j. One approach, often suggested, will be excluded from consideration at the outset. This is the two-action sequential approach of determining which of k categories has the highest mean and then assigning all remaining observations to that category. Bechhofer [19], Paulson [20], Fabian [21] and Dunnett [22] have made interesting contributions to this problem. In this approach, the problem of the trade-off between information gained from taking observations from categories with sample means less than $\max_j {}_n\bar{X}_j$ and the loss in expected return from taking such observations is handled by requiring the experimenter to state before the process begins values of δ^* and P^*, where δ^* is the smallest difference $\mu_K - \mu_{K-1}$ that is worth detecting, and P^* is the smallest acceptable value of the probability of selecting μ_K, when actually $\mu_K - \mu_{K-1} \geq \delta^*$. The difficulty with this approach is that when one wishes to maximize S_n, then δ^* and P^* are functions of the unknown μ_j and cannot be specified unless there are a priori measures on the μ_j. Furthermore, this approach requires taking observations from each category at each trial $n < n^*$, an obviously non-optimal procedure.

There is another two-action sequential approach which also requires taking observations from each category at each trial $n < n^*$ but which does not require a priori statements on δ^* or P^*.

Case III α 4 C ii. For X_{ij} normally distributed with equal known variances, let n^* be redefined as the number of trials, where each trial consists in taking one observation from each of the k populations; then ignoring set up and sampling costs the expected loss when making n^* trials is $n^* \sum_{j=1}^{k} \delta_j$, where $\delta_j = \mu_K - \mu_j$. If the sampling process stops at n^* and the expected loss from taking the remaining $(N-kn^*)$ observations from a $\pi_j \neq \pi_K$ is $(N-kn^*) \sum_{j=1}^{k} \delta_j ({}_{n^*}P_j)$, the total expected loss is:

$$E(L) = n^* \sum_{j=1}^{k} \delta_j + (N-kn^*) \sum_{j=1}^{k} \delta_j ({}_{n^*}P_j)$$

Maurice [23] considers the case where k = 2 and

$$E(L) = n^* \delta + (N - 2n^*) \delta \left({}_{n^*}P_j\right)_j \qquad j \neq K$$

and draws on Girschick's [24] earlier work indicating that sets of sample values $X_{11}, X_{12}, \ldots, X_{1n}$ and $X_{21}, X_{22}, \ldots, X_{2n}$ yielding sample means $\overline{X}_1$ and $\overline{X}_2$ can be identified with two population means μ_a and μ_b as:

$$\lambda = \frac{f(\bar{X}_1;\mu_a)\, f(\bar{X}_2;\mu_b)}{f(\bar{X}_1;\mu_b)\, f(\bar{X}_2;\mu_a)} \quad \text{where } f(\bar{X}_i;\mu_j) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left[-\frac{n}{2}\left(\frac{\bar{X}_i-\mu_j}{\sigma}\right)^2\right]$$

$$\therefore \lambda = \frac{\exp\left[-\frac{n}{2}\left(\frac{\bar{X}_1-\mu_a}{\sigma}\right)^2\right]\exp\left[-\frac{n}{2}\left(\frac{\bar{X}_2-\mu_b}{\sigma}\right)^2\right]}{\exp\left[-\frac{n}{2}\left(\frac{\bar{X}_1-\mu_b}{\sigma}\right)^2\right]\exp\left[-\frac{n}{2}\left(\frac{\bar{X}_2-\mu_a}{\sigma}\right)^2\right]} = \exp\left[\frac{n}{\sigma^2}\left(\bar{X}_1-\bar{X}_2\right)\left(\mu_a-\mu_b\right)\right]$$

$$= \exp\left[\frac{n\delta}{\sigma^2}\left(\bar{X}_1-\bar{X}_2\right)\right]$$

The sequential rule in this case is to continue sampling as long as $B < \lambda < A$. Since a loss results whether δ is positive or negative, $B = 1/A$ or sampling should continue as long as

$$\frac{1}{A} < \exp\left[\frac{n\delta}{\sigma^2}\left(\bar{X}_1 - \bar{X}_2\right)\right] < A$$

Taking the logarithm of this gives

$$-a < \frac{n\delta}{\sigma^2}\left(\bar{X}_1 - \bar{X}_2\right) < a$$

or

$$-\frac{a}{\delta}\sigma^2 < \sum_{i=1}^{n}\left(X_{i1} - X_{i2}\right) < \frac{a}{\delta}\sigma^2$$

The average expected sample number (ASN), designated here by n^{**}, for assumed large N is:

$$n^{**} = \frac{\sigma^2\, a\,(\exp[a]-1)}{\delta^2\,(\exp[a]+1)}$$

and

$$P = \frac{(1 - \exp[-a])}{(\exp[a]-\exp[-2]a)} = \frac{1}{\exp[a]+1}$$

Maurice then substitutes n^{**} for n^* and P for ${}_{n^*}P_j$ in the expression for E(L), maximizes with respect to δ, minimizes with respect to a/δ, and finds a solution of the form:

$$D_1: \text{ If } \sum_{i=1}^{n}\left(X_{i1} - X_{i2}\right) > 0.5842\sqrt{N}\,\sigma$$

make all subsequent observations from π_1.

$$D_2: \text{ If } \sum_{i=1}^{n}\left(X_{i1} - X_{i2}\right) < -0.5842\sqrt{N}\,\sigma$$

make all subsequent observations from π_2.

$$D_o: \text{ If } -0.5842\sqrt{N}\,\sigma \leqq \sum_{i=1}^{n}\left(X_{i1} - X_{i2}\right) \leqq 0.5842\sqrt{N}\,\sigma$$

take another set of k observations.

However, in the current application, and in other industrial applications, another cost should be included, and that is the cost of taking observations. This cost has not been included in the formulations of Maurice and others, and is derived here in Case IIIα 4 C ii.*

<u>Case III α 4 C ii.*</u> The conditions for this case are the same as those for case III α 4 C ii, except there is the additional expected loss attributable to the cost of sampling, or taking observations. For the case of k = 2

$$E(L) = n^*\,\delta + n^*\,C + (N - 2\,n^*)\,\delta\left({}_{n^*}P_j\right)_j \qquad j \neq K$$

where C is the cost of taking observations on each pair of X_{i1}, X_{i2}. If C can be stated as a percentage of the δ, i.e., $C = p\,\delta$

$$E(L) = n^*\,\delta\,(1 + p) + (N - 2\,n^*)\,\delta\,({}_{n^*}P_j)$$

and letting $c = 1 + p$

$$E(L) = n^*\,\delta\,c + (N - 2\,n^*)\,\delta\,({}_{n^*}P_j)$$

Following Maurice's procedure, substitute n^{**} for n^* and P for ${}_{n^*}P_j$ in the above expression, and let:

$$\ell = \frac{a}{\delta}$$

$$\therefore\ E(L) = \frac{\delta c\,\sigma^2 \ell(\exp[\ell\delta]-1}{\delta(\exp[\ell\delta]+1)} + \left[N - \frac{2\sigma^2\ell(\exp[\ell\delta]-1}{\delta(\exp[\ell\delta]+1)}\right]\left[\frac{\delta}{(\exp[\ell\delta]+1}\right]$$

$$E(L) = \frac{N\delta}{\exp[\ell\delta]+1} - \frac{2\sigma^2\ell\left(\exp[\ell\delta]-1\right)^2}{\left(\exp[\ell\delta]+1\right)^2} + \frac{c\,\sigma^2\ell(\exp[\ell\delta]-1)}{(\exp[\ell\delta]+1)}$$

To solve for ℓ:

$$\frac{\partial E(L)}{\partial\delta} = \frac{N(\exp[\ell\delta]+1-\ell\delta\,\exp[\ell\delta])}{(\exp[\ell\delta]+1)^2} - \frac{2\sigma^2\ell^2\exp[\ell\delta]}{(\exp[\ell\delta]+1)^2}$$

$$+ \frac{4\sigma^2\ell^2\,\exp[\ell\delta]\,(\exp[\ell\delta]-1)}{(\exp[\ell\delta]+1)^3} + \frac{2c\,\sigma^2\ell^2\exp[\ell\delta]}{(\exp[\ell\delta]+1)^2}$$

Setting this equal to zero and substituting $x = \exp[\ell\delta]$, $\ell n\ x = \ell\delta$

$$\frac{N}{\sigma^2} = \frac{2\ell^2 x(3-x-c-cx)}{(x+1)(x+1-x\,\ell n\,x)}$$

$$\frac{\partial E(L)}{\partial\ell} = -\frac{N\delta^2\exp[\ell\delta]}{(\exp[\ell\delta]+1)^2} - \frac{2\sigma^2(\exp[\ell\delta]-1)}{(\exp[\ell\delta]+1)^2} - \frac{2\sigma^2\ell\delta\,\exp[\ell\delta]}{(\exp[\ell\delta]+1)^2}$$

$$+ \frac{4\sigma^2\ell\delta\,\exp[\ell\delta]\,(\exp[\ell\delta]-1)}{(\exp[\ell\delta]+1)^3} + \frac{c\,\sigma^2(\exp[\ell\delta]-1)}{(\exp[\ell\delta]+1)}$$

$$+ \frac{2c\,\sigma^2\ell\delta\,\exp[\ell\delta]}{(\exp[\ell\delta]+1)^2} = 0$$

$$\therefore\ \frac{N}{\sigma^2} = \frac{-2(x-1)(x+1)-2x\ell nx(x+1)+4x\ell nx(x-1)+c(x-1)(x+1)^2+2cx\ell nx(x+1)}{\delta^2\,x\,(x+1)}$$

Equating the $\frac{N}{\sigma^2}$ and simplifying yields

$$x\,\ell n\,x(4x-8)-cx\,\ell n\,x(x+1)(x+3)-2(x-1)(x+1)+c(x-1)(x+1)^2 = 0$$

or, in terms of the percentage of δ

$$4\,x\,\ell n\,x(x-2)\cdot(1+p)x\,\ell n\,x(x+1)(x+3)-2(x-1)(x+1)+(1+p)(x-1)(x+1)^2 = 0$$

Solving this equation for different values of p, and substituting these back into the expression for $\frac{N}{\sigma^2}$, the required solution for ℓ is found from

$$\ell = \frac{\sqrt{N}}{\sigma}\sqrt{\frac{(x+1)(x+1-x\,\ell n\,x)}{2x(2-2x-p-px)}}$$

The decision rule now is:

D_1: If $\sum_{i=1}^{n}(X_{i1}-X_{i2}) > \ell\sigma\sqrt{N}$ make all subsequent observations from π_1

D_2: If $\sum_{i=1}^{n}(X_{i1}-X_{i2}) < -\ell\sigma\sqrt{N}$ make all subsequent observations from π_2

D_o: If $-\ell\sigma\sqrt{N} \leqq \sum_{i=1}^{n}(X_{i1}-X_{i2}) \leqq \ell\sigma\sqrt{N}$ take another pair of observations

Table 5 below gives the x and ℓ solutions to the above equations for different values of p.

TABLE 5

TWO-STAGE SEQUENTIAL STOPPING CONSTANTS

p	x	ℓ
0.0	9.061169	0.584160
0.2	8.517213	0.536543
0.4	8.148601	0.498402
0.6	7.883984	0.467080
0.8	7.685562	0.440822
1.0	7.531641	0.418433
1.2	7.408969	0.399067
1.4	7.309016	0.382113
1.6	7.226079	0.367117
1.8	7.156191	0.353734
2.0	7.096525	0.341697
2.2	7.045010	0.330798
2.4	7.000092	0.320870
2.6	6.960585	0.311777
2.8	6.925579	0.303410
3.0	6.894347	0.295677
3.2	6.866313	0.288504
3.4	6.841013	0.281825
3.6	6.818063	0.275587
3.8	6.797157	0.269744
4.0	6.778032	0.264256
4.2	6.760470	0.259088
4.4	6.744288	0.254211
4.6	6.729332	0.249599
4.8	6.715466	0.245228
5.0	6.702573	0.241078

<u>Case III α 3 C i.</u> A case which is of particular importance in educational (and other) systems arises when, by the nature of the process involved, observations will be made on each of the available N students (or experimental subjects, Ss.) and the k populations are known to be independently normally distributed. In this case, the analytical solution for the problem of maximizing S_N involves evaluating a (k - 1) - multinormal distribution, tabulated values of which are not available for k > 3. However, by stating the problem in recursive form a numerical solution is feasible. Such a solution, using a backwards-induction technique, is developed here.

In this case, the $\pi_1, \pi_2, \ldots, \pi_j, \ldots, \pi_k$ populations are all independently normally distributed with random variables x_j, known variances σ_j, and unknown means μ_j. Let n_j be the number of observations from π_j at the n-th stage. Therefore, $n = n_1 + n_2 + \ldots + n_j + \ldots + n_k$. Let $\tilde{n} = N - n$ be the number of observations remaining after the n-th observation has been made, and $S_{\tilde{n}}$ is the sum of the remaining observations. A k-dimensional decision tree can be imagined where each branch point in the tree is identified by the k-tuple, $(n_1, n_2, \ldots, n_j, \ldots, n_k)$, corresponding to the number of previous observations taken from each π_j. Also, after n observations have been made, there will exist a k-tuple of sample means $({}_n\overline{X}_1, {}_n\overline{X}_2, \ldots {}_n\overline{X}_j, \ldots, {}_n\overline{X}_k)$ corresponding to that $(n_1, n_2, \ldots, n_j, \ldots, n_k)$ branch point actually obtained at the n-th stage. The sample means can be just as readily identified by the number of stages remaining, $\tilde{n}$, or instead of ${}_n\overline{X}_j$, ${}_{\tilde{n}}\overline{X}_j$ can be used. Given $(n_1, n_2, \ldots, n_j, \ldots, n_k)$ and $({}_{\tilde{n}}\overline{X}_1, {}_{\tilde{n}}\overline{X}_2, \ldots, {}_{\tilde{n}}\overline{X}_j, \ldots, {}_{\tilde{n}}\overline{X}_k)$ at each stage, the principle of optimality in dynamic programming [26] would indicate for this case that an optimal decision rule is one which maximizes the sum of the remaining observations, regardless of what path or what decision rules one followed in arriving at the two state k-tuples. Therefore, the problem can be restated as one where $S_{\tilde{n}}$ must be maximized at each stage, where:

$$\left[S_{\tilde{n}} \mid (n_1, n_2, \ldots, n_j, \ldots, n_k), ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}}\bar{X}_2, \ldots, {}_{\tilde{n}}\bar{X}_j, \ldots, {}_{\tilde{n}}\bar{X}_k)\right] = {}_{\tilde{n}-1}x + {}_{\tilde{n}-2}x + \ldots + {}_{\tilde{o}}x$$

and the expected value of $S_{\tilde{n}}$ is defined as:

$$E[S_{\tilde{n}}] \equiv E\left[S_{\tilde{n}} \mid (n_1, n_2, \ldots, n_j, \ldots, n_k), ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}}\bar{X}_2, \ldots, {}_{\tilde{n}}\bar{X}_j \ldots {}_{\tilde{n}}\bar{X}_k), D\right]$$

$$= E\left[({}_{\tilde{n}-1}x + {}_{\tilde{n}-2}x + \ldots + {}_{\tilde{o}}x) \mid D \right]$$

$$= E\left[{}_{\tilde{n}-1}x \mid D \right] + E\left[S_{\tilde{n}-1} \mid D \right]$$

where the decision rule D is: Select the (n + 1)st = ($\tilde{n}$ - 1)st observation from the π_j which has the maximum expected value of the remaining observations. This can be expressed by the recurrence relationship:

$$E[S_{\tilde{n}}] = \max_j \begin{cases} E[{}_{\tilde{n}-1}x_1] + E_1\left[S_{\tilde{n}-1} \mid (n_1+1, n_2, \ldots, n_j, \ldots, n_k), ({}_{\tilde{n}-1}\bar{x}_1, {}_{\tilde{n}}\bar{X}_2, \ldots, {}_{\tilde{n}}\bar{X}_j, \ldots, {}_{\tilde{n}}\bar{X}_k), D\right] \\ E[{}_{\tilde{n}-1}x_2] + E_2\left[S_{\tilde{n}-1} \mid (n_1, n_2+1, \ldots, n_j, \ldots, n_k), ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}-1}\bar{x}_2, \ldots, {}_{\tilde{n}}\bar{X}_j, \ldots, {}_{\tilde{n}}\bar{X}_k), D\right] \\ \vdots \\ E[{}_{\tilde{n}-1}x_j] + E_j\left[S_{\tilde{n}-1} \mid (n_1, n_2, \ldots, n_j+1, \ldots, n_k), ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}}\bar{X}_2, \ldots, {}_{\tilde{n}-1}\bar{x}_j, \ldots, {}_{\tilde{n}}\bar{X}_k), D\right] \\ \vdots \\ E[{}_{\tilde{n}-1}x_k] + E_k\left[S_{\tilde{n}-1} \mid (n_1, n_2, \ldots, n_j, \ldots, n_k+1), ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}}\bar{X}_2, \ldots, {}_{\tilde{n}}\bar{X}_j \ldots {}_{\tilde{n}-1}\bar{x}_k), D\right] \end{cases}$$

Using the implicit assumption of Raiffa and Schlaifer [25] that unknown population means be treated as a Gaussian distributed random variable with mean of ${}_{n}\bar{X}_j$ and variance of σ_j^2/n_j, i.e., $G\left[\tilde{\mu}_j \middle| {}_{n}\bar{X}_j, \sigma_j^2/n_j\right]$, then:

$$E[S_{\tilde{n}}] \cong \max_j \left\{ {}_{\tilde{n}}\bar{X}_j + E_j[S_{\tilde{n}-1}] \middle| (n_1, n_2, \ldots, n_j, \ldots, n_k), \right.$$

$$\left. ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}}\bar{X}_2, \ldots, {}_{\tilde{n}-1}\bar{x}_j, \ldots {}_{\tilde{n}}\bar{X}_k), D \right\} \text{ for } j = 1, 2, \ldots, k$$

$$= \max_j \left\{ {}_{\tilde{n}}\bar{X}_j + E_j[S_{\tilde{n}-1}] \right\}$$

where the new mean is given by

$${}_{\tilde{n}-1}\bar{x}_j = \frac{n_j({}_{\tilde{n}}\bar{X}_j) + {}_{\tilde{n}-1}X_j}{n_j + 1}$$

and is Gaussian distributed with mean ${}_{\tilde{n}}\bar{X}_j$ and variance

$$\frac{\sigma_j^2}{n_j} - \frac{\sigma_j^2}{n_j+1} = \sigma_j^2/n_j(n_j+1)$$

To solve the problem, the following backwards-induction procedure is employed:

(a) Start at the end of the decision tree, where $n_1 + n_2 + \ldots + n_j + \ldots + n_k = n = N$, and therefore $\tilde{n} = 0$. At $\tilde{n} = 0$, $E[S_{\tilde{0}}] = 0$, and there is no decision to make since no observations remain to be made.

(b) Move back down the decision tree to the $(n_1, n_2 \ldots, n_j, \ldots, n_k)$ state points located at $\tilde{n} = 1$.

At each of these state points

$$E[S_{\tilde{1}}] = \max_j \left\{ {}_{\tilde{1}}\bar{X}_j + E_j[S_{\tilde{0}}] \right\} = \max_j \left\{ {}_{\tilde{1}}\bar{X}_j \right\}$$

If one were moving forward along a path in the decision tree, then at $\tilde{n}=1$ the $(n_1, n_2, \ldots, n_j, \ldots, n_k), ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}}\bar{X}_2, \ldots {}_{\tilde{n}}\bar{X}_j, \ldots {}_{\tilde{n}}\bar{X}_k)$

would be known and it would be a simple matter to select $\max_j\{{}_{\tilde{1}}\bar{X}_j\}$. However, in the backwards-induction the $({}_{\tilde{1}}\bar{X}_1, {}_{\tilde{1}}\bar{X}_2, \ldots, {}_{\tilde{1}}\bar{X}_j, \ldots, {}_{\tilde{1}}\bar{X}_k)$ are not known, and therefore the exhaustive procedure of considering all possible combinations of $({}_{\tilde{1}}\bar{X}_1, {}_{\tilde{1}}\bar{X}_2, \ldots, {}_{\tilde{1}}\bar{X}_j, \ldots, {}_{\tilde{1}}\bar{X}_k)$ will be used. If each ${}_{\tilde{1}}\bar{X}_j$ is examined at q discrete points in the range between $-\infty$ and $+\infty$, then at each $(n_1, n_2, \ldots n_j, \ldots, n_k)$ state point there are q^k cells, arranged in a k-dimensional array. Each cell corresponds to one of the possible discrete combinations of $({}_{\tilde{1}}\bar{X}_1, {}_{\tilde{1}}\bar{X}_2, \ldots, {}_{\tilde{1}}\bar{X}_j, \ldots, {}_{\tilde{1}}\bar{X}_k)$ and in each cell can insert the value of the maximum of the means corresponding to that cell. Also, one can record the identification of the population associated with the maximum mean for each cell. Therefore D is exhaustively determined for each possible combination of $({}_{\tilde{1}}\bar{X}_1, {}_{\tilde{1}}\bar{X}_2, \ldots, {}_{\tilde{1}}\bar{X}_j, \ldots {}_{\tilde{1}}\bar{X}_k)$ at each possible $(n_1, n_2, \ldots, n_j, \ldots, n_k)$.

(c) Move back down the decision tree to the state points at $\tilde{n}=2$. Here,

$$E[S_{\tilde{2}}] = \max_j\left\{{}_{\tilde{2}}\bar{X}_j + E_j[S_{\tilde{1}}]\right\}$$

Since values for $E[S_{\tilde{1}}]$ have been stored in the q^k cells at $\tilde{n}=1$, the $E_j[S_{\tilde{1}}]$ is computed from the sum of the products of the stored values with its probability of occurrence. The distribution of the ${}_{\tilde{n}-1}\bar{x}_j$ is $G\left[{}_{\tilde{n}-1}\bar{x}_j \,|\, {}_{\tilde{n}}\bar{X}_j, \sigma_j^2/n_j(n_j+1)\right]$ and in order to find probability weightings for each of the q discrete points that the distribution range has been divided into, a quadrature based on a Hermite polynomial approximation to the integrand can be used, such that

$$\int_{-\infty}^{\infty} \exp[-z^2]\left\{\exp[z^2]\mathcal{G}(z)\right\}dz \cong \sum_{i=1}^{q}(\alpha_i)\exp[z_i^2]\mathcal{G}(z_i) = \sum_{i=1}^{q} W_i\,\mathcal{G}(z_i)$$

where z_i, α_i and W_i are tabled [27] for quadratures up to $q = 20$.

(d) Step (c) is repeated for $\tilde{n} = 3, 4, \ldots, N-k$; where N-k corresponds to the starting state point $(1, 1, \ldots, 1, \ldots, 1)$.

The solution then consists of q^k cells at each state point in the decision tree, each cell corresponding to one of the possible combinations of $(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_j, \ldots \bar{X}_k)$, and in each cell is D, telling which population to take the next instruction from. While a solution as outlined above is feasible, the computational time and output storage requirements are excessive. For example, just to store the D in each of the cells requires

$$\frac{q^k}{k!} \prod_{a=1}^{k} (N-1+a)$$

storage locations. The number of required storage locations can be reduced by observing that:

i. Initially, one observation is taken from each π_j

ii. At points in the decision tree where an equal number of observations have been taken from each population, and at the $\tilde{n} = 1$ stage, the next observation will be taken from the population having the largest sample mean.

This reduces the number of storage locations to

$$q^k \left[b - \frac{N}{2} + \frac{1}{k!} \prod_{a=1}^{k} (N-k-2+a) \right]; \text{ where } b = \begin{cases} 1, & \text{for N even} \\ 1.5, & \text{for N odd} \end{cases}$$

For equal population variances the number of storage locations can be further reduced by a factor of k. Nevertheless, for example, for $k = 3$, $q = 16$, $N = 500$, and equal variances, a minimum of 4.19×10^{10} storage locations are required!

It is possible to make a significant reduction in output storage space requirements and in computation time by the reparameterization described below:

Define a set of superscripts $(a, b, \ldots, h, \ldots, i)$ such that:

$$\bar{X}^a > \bar{X}^b > \cdots > \bar{X}^h > \cdots > \bar{X}^i$$

At each $\tilde{n}$-th stage reassign the set of superscripts to the π_j, $\bar{X}_j$ and σ_j. Therefore, a given superscript need not be associated with the same subscript from stage to stage. Also define

$$U_{\tilde{n}} = \left[U_{\tilde{n}} \,\middle|\, (n_1, n_2, \ldots, n_j, \ldots, n_k), ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}}\bar{X}_2, \ldots, {}_{\tilde{n}}\bar{X}_j, \ldots, {}_{\tilde{n}}\bar{X}_k) \right] = \frac{{}_{\tilde{n}}\bar{X}^a - \frac{1}{\tilde{n}} E[S_{\tilde{n}}]}{\sigma^a}$$

$$\therefore\ U_{\tilde{n}} = \min_j \left\{ \frac{{}_{\tilde{n}}\bar{X}^a - \frac{1}{\tilde{n}} \left({}_{\tilde{n}}\bar{X}_j + E_j[S_{\tilde{n}-1}] \right)}{\sigma^a} \right\}$$

$$= \min \left\{ \frac{\Delta_j}{\tilde{n}} + \frac{\tilde{n}-1}{\tilde{n}} \left(\frac{{}_{\tilde{n}}\bar{X}^a - \frac{1}{\tilde{n}-1} E_j[S_{\tilde{n}-1}]}{\sigma^a} \right) \right\}$$

where

$$\Delta_j = \begin{cases} 0, & \text{for } \pi_j = \pi^a \\ \dfrac{\bar{X}^a - \bar{X}_j}{\sigma^a}, & \text{for } \pi_j \neq \pi^a \end{cases}$$

For the sake of simplicity, the case of $k = 2$ will be used in the following exposition. Therefore π^a, σ^a, and n^a will correspond to the larger sample mean $\bar{X}^a$, and π^b, σ^b and n^b will correspond to the smaller sample mean $\bar{X}^b$. In terms of the new variables:

It is possible to make a significant reduction in output storage space requirements and in computation time by the reparameterization described below:

Define a set of superscripts (a, b, . . . , h, . . . , i) such that:

$$\bar{X}^a > \bar{X}^b > \cdots > \bar{X}^h > \cdots > \bar{X}^i$$

At each $\tilde{n}$-th stage reassign the set of superscripts to the π_j, $\bar{X}_j$ and σ_j. Therefore, a given superscript need not be associated with the same subscript from stage to stage. Also define

$$U_{\tilde{n}} = \left[U_{\tilde{n}} \mid (n_1, n_2, \ldots, n_j, \ldots, n_k), ({}_{\tilde{n}}\bar{X}_1, {}_{\tilde{n}}\bar{X}_2, \ldots, {}_{\tilde{n}}\bar{X}_j, \ldots, {}_{\tilde{n}}\bar{X}_k)\right] =$$

$$\frac{{}_{\tilde{n}}\bar{X}^a - \frac{1}{\tilde{n}} E[S_{\tilde{n}}]}{\sigma^a}$$

$$\therefore \; U_{\tilde{n}} = \min_j \left\{ \frac{{}_{\tilde{n}}\bar{X}^a - \frac{1}{\tilde{n}} \left({}_{\tilde{n}}\bar{X}_j + E_j[S_{\tilde{n}-1}]\right)}{\sigma^a} \right\}$$

$$= \min \left\{ \frac{\Delta_j}{\tilde{n}} + \frac{\tilde{n}-1}{\tilde{n}} \left(\frac{{}_{\tilde{n}}\bar{X}^a - \frac{1}{\tilde{n}-1} E_j[S_{\tilde{n}-1}]}{\sigma^a} \right) \right\}$$

where

$$\Delta_j = \begin{cases} 0, \text{ for } \pi_j = \pi^a \\ \dfrac{\bar{X}^a - \bar{X}_j}{\sigma^a}, \text{ for } \pi_j \neq \pi^a \end{cases}$$

For the sake of simplicity, the case of k = 2 will be used in the following exposition. Therefore π^a, σ^a, and n^a will correspond to the larger sample mean $\bar{X}^a$, and π^b, σ^b and n^b will correspond to the smaller sample mean $\bar{X}^b$. In terms of the new variables:

$$U_{\tilde{n}} | (\Delta, n^a, n^b, a) = \min_h \left\{ \frac{\Delta^h}{\tilde{n}} + \frac{\tilde{n}-1}{\tilde{n}} \left(\frac{{}_{\tilde{n}}\bar{X}^a - \frac{1}{\tilde{n}-1} E^h[S_{\tilde{n}-1}]}{\sigma^a} \right) \right\}$$

where

$$\Delta^h = \begin{cases} 0, \text{ for } h = a \\ \Delta = \dfrac{\bar{X}^a - \bar{X}^b}{\sigma^a}, \text{ for } h = b \end{cases}$$

To get a recursion in terms of U, expand the expression inside the ( ) brackets:

$$U_{\tilde{n}} = \min_h \left\{ \frac{\Delta^h}{\tilde{n}} + \frac{\tilde{n}-1}{\tilde{n}} \left(\frac{{}_{\tilde{n}}\bar{X} - E\left[{}_{\tilde{n}-1}\bar{x}^h\right]}{\sigma^a} + \frac{E\left[{}_{\tilde{n}-1}\bar{x}^h\right] - \frac{1}{\tilde{n}-1} E^h[S_{\tilde{n}-1}]}{\sigma^a} \right) \right\}$$

For h = a, two cases can arise; either ${}_{\tilde{n}-1}\bar{x}$ results in a ${}_{\tilde{n}-1}\bar{X}^a$ which comes from the same population as ${}_{\tilde{n}}\bar{X}^a$, or ${}_{\tilde{n}-1}\bar{x}$ results in a ${}_{\tilde{n}-1}\bar{X}^a$ which comes from a different population than ${}_{\tilde{n}}\bar{X}^a$. For the former case, the first quantity inside the ( ) brackets is equal to:

$$\int_{{}_{\tilde{n}}\bar{X}^b}^{\infty} \left(\frac{{}_{\tilde{n}}\bar{X}^a - {}_{\tilde{n}-1}\bar{x}^a}{\sigma^a} \right) G\left[{}_{\tilde{n}-1}\bar{x}^a \middle| {}_{\tilde{n}}\bar{X}^a, \left(\sigma^a\right)^2 \Big/ n^a(n^a+1) \right] d\,\bar{x}^a$$

In the latter case this quantity is equal to:

$$\frac{{}_{\tilde{n}}\bar{X}^a - {}_{\tilde{n}}\bar{X}^b}{\sigma^a} \int_{-\infty}^{{}_{\tilde{n}}\bar{X}^b} G\left[{}_{\tilde{n}-1}\bar{x}^a / {}_{\tilde{n}}\bar{X}^a, \left(\sigma^a\right)^2 \Big/ n^a(n^a+1) \right] d\,\bar{x}^a$$

transforming both expressions by $y = \dfrac{{}_{\tilde{n}}\bar{x}^a - {}_{\tilde{n}}\bar{X}^a}{\sigma^a}$ and combining terms:

$$- \int_{\Delta}^{\infty} y\, G\left[y \mid 0,\ 1/n^a(n^a+1)\right] dy + \Delta \int_{-\infty}^{-\Delta} G\left[y \mid 0,\ 1/n^a(n^a+1)\right] dy$$

$$= - \int_{-\Delta}^{\infty} y\, G\left[y \mid 0,\ 1/n^a(n^a+1)\right] dy +$$

$$\Delta\left(\int_{-\infty}^{\infty} G\left[y \mid 0, 1/n^a(n^a+1)\right] dy - \int_{-\Delta}^{\infty} G\left[y \mid 0, 1/n^a(n^a+1)\right] dy\right)$$

$$= \Delta \int_{-\Delta}^{\infty} (y + \Delta)\, G\left[y \mid 0,\ 1/n^a(n^a+1)\right] dy$$

transforming by $\delta = y + \Delta$ gives:

$$\Delta - \int_{0}^{\infty} \delta\, G\left[\delta \mid \Delta,\ 1/n^a(n^a+1)\right] d\delta$$

In the two cases described above, the second quantity in the ( ) brackets is equal to:

$$\int_{0}^{\infty} \left[U_{\tilde{n}-1}\left(\delta \mid n^a+1, n^b\right)\right] G\left[\delta \mid \Delta,\ 1/n^a(n^a+1)\right] d\delta\ +$$

$$\frac{\sigma^b}{\sigma^a} \int_{0}^{\infty} \left[U_{\tilde{n}-1}\left(\delta \mid n^b, n^a+1\right)\right] G\left[\delta \mid -\Delta,\ 1/n^a(n^a+1)\right] d\delta$$

Similarly, for h = b, under the two cases, the first quantity in the ( ) brackets is equal to:

$$0 + \int_{{}_{\tilde{n}}\bar{\bar{X}}^a}^{\infty} \left(\frac{{}_{\tilde{n}}\bar{\bar{X}}^a - {}_{\tilde{n}-1}\bar{x}^b}{\sigma^a}\right) G\left[{}_{\tilde{n}-1}\bar{x}^b \,\middle|\, {}_{\tilde{n}}\bar{\bar{X}}^b, \left(\sigma^b\right)^2\ n^b(n^b+1)\right] d\,\bar{x}^b$$

transforming by $y = \dfrac{{}_{\tilde{n}-1}\bar{x}^b - {}_{\tilde{n}}\bar{\bar{X}}^b}{\sigma^a}$:

$$- \int_{\Delta}^{\infty} (y - \Delta)\, G\left[y/0,\ \left(\frac{\sigma^b}{\sigma^a}\right)^2 \Big/ n^b(n^b+1)\right] dy$$

and transforming now by $\delta = y - \Delta$ gives:

$$-\int_0^\infty \delta\, G\left[\delta \,|\, -\Delta,\ \left(\frac{b}{a}\right)^2 \Big/ n^b(n^b+1)\right] d\delta$$

The second quantity in the brackets is equal to:

$$\frac{\sigma^b}{\sigma^a}\int_0^\infty \left[U_{\tilde{n}-1}\left(\delta \,|\, n^b+1, n^a\right)\right] G\left[\delta \,|\, -\Delta,\ \left(\frac{\sigma^b}{\sigma^a}\right)^2 \Big/ n^b(n^b+1)\right] d\delta$$

$$+\int_0^\infty \left[U_{\tilde{n}-1}\left(\delta \,|\, n^a, n^b+1\right)\right] G\left[\delta \,|\, \Delta,\ \left(\frac{\sigma^b}{\sigma^a}\right)^2 \Big/ n^b(n^b+1)\right] d\delta$$

Collecting terms:

$$U_{\tilde{n}} = \min_h \begin{cases} h=a: \dfrac{\tilde{n}-1}{\tilde{n}}\left(\Delta + \displaystyle\int_0^\infty \left[U_{\tilde{n}-1}\left(\delta \,|\, n^a+1, n^b\right) - \delta\right] G\left[\delta \,|\, \Delta, 1/n^a(n^a+1)\right] d\delta \right. \\ \qquad \left. + \dfrac{\sigma^b}{\sigma^a}\displaystyle\int_0^\infty \left[U_{\tilde{n}-1}\left(\delta \,|\, n^b, n^a+1\right)\right] G\left[\delta \,|\, -\Delta, 1/n^a(n^a+1)\right] d\delta \right) \\ h=b: \dfrac{\Delta}{\tilde{n}} + \dfrac{\tilde{n}-1}{\tilde{n}}\left(\displaystyle\int_0^\infty \left[U_{\tilde{n}-1}\left(\delta \,|\, n^b+1, n^a\right) - \delta\right] G\left[\delta \,|\, -\Delta, \left(\frac{\sigma^b}{\sigma^a}\right)^2 \Big/ n^b(n^b+1)\right] d\delta \right. \\ \qquad \left. + \displaystyle\int_0^\infty \left[U_{\tilde{n}-1}\left(\delta \,|\, n^a, n^b+1\right)\right] G\left[\delta \,|\, \Delta, \left(\frac{\sigma^b}{\sigma^a}\right)^2 \Big/ n^b(n^b+1)\right] d\delta \right) \end{cases}$$

$$= \min_{a,b} \begin{Bmatrix} T^a U_{\tilde{n}} \\ T^b U_{\tilde{n}} \end{Bmatrix}$$

An attempt was made to numerically solve the above integrals by using a Gaussian quadrature of the following type:

$$\int_{-1}^{+1} f(z)\,dz \cong \sum_{i=1}^{q} \alpha_i\, f(z_i)$$

where z_i and α_i is tabled [28] for q = 1 to 48. Since the limits on the integrals in $U_{\tilde{n}}$ are from 0 to ∞ and not from -1 to +1, the following transformation was employed.

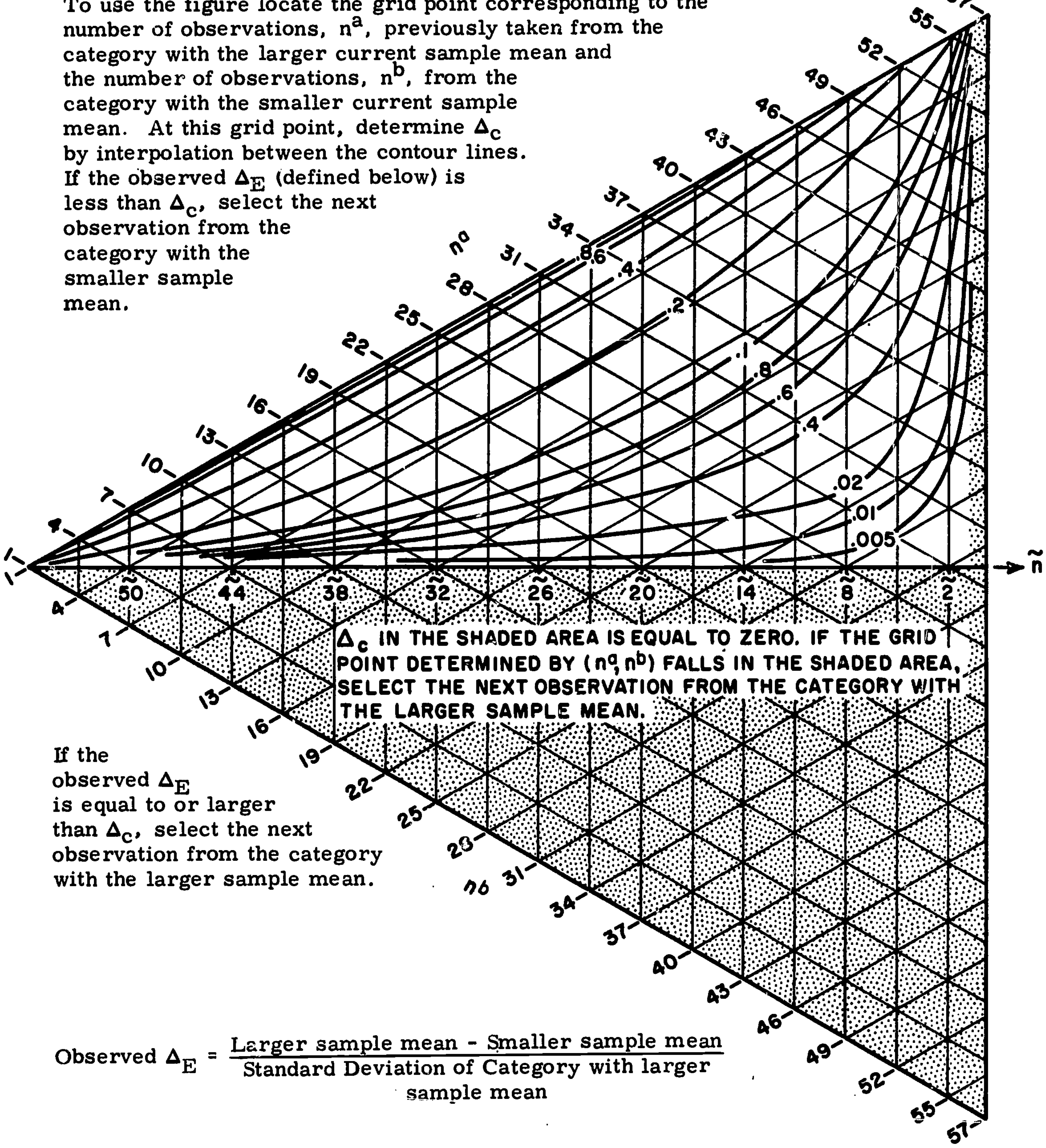
$$\delta = \frac{1+z}{1-z} \text{ with Jacobian } d\delta = \frac{2}{(1-z)^2}\, dz$$

This quadrature approximation of the integral is not accurate for value of Δ approaching ∞. At first it was felt that this would not be of serious consequence, since for very large Δ one would always select the population which contained $\bar{X}^a_{\tilde{n}}$. However, the error is multiplicative as one steps back through the N iterations of the recursion formula, and significant errors occur. This problem was overcome by using an approximately exponential grid spacing for the Δ, and at each Δ grid point approximating the integral by using tabled Gaussian probability values associated with seventeen equally spaced abscissa points in the range of $-3.2\ \sigma$ to $+3.2\ \sigma$.

The results confirm the intuitive notion that for $n^a > n^b$ one should always select the next observation from π^a. For $n^a \leq n^b$ there will exist a range of Δ, from $\Delta = 0$ to a critical Δ, "Δ_c", for which a choice from π^b has a smaller expected $U_{\tilde{n}}$ than does a choice from π^a. Therefore, on the decision tree one can associate with each (n^a, n^b) branch point a Δ_c. Having once determined the Δ_c for all branch points in the decision tree, the experimenter merely calculates the actual "Δ_E" obtained in his experiment at a particular (n^a, n^b) and compares the tabled Δ_c at that branch point with his Δ_E. If $\Delta_E < \Delta_c$ the next observation is taken from π^b. If $\Delta_E \geq \Delta_c$, the next observation is taken from π^a.

It is possible to show the results by a "topographic" map of the critical Δ_c surface, as illustrated in Figure 6. At present, one such map is required for each value of N. It remains to be seen whether some simple transformation of scales, in terms of N, can be used to obtain the Δ_c surface from one generalized map. The map, of course, can only be drawn for k = 2. For larger k, the values of Δ_c can be tabled or stored on magnetic tape.

The critical Δ_c surface is shown for $k = 2$, $\sigma^a = \sigma^b$, $N = 58$. To use the figure locate the grid point corresponding to the number of observations, n^a, previously taken from the category with the larger current sample mean and the number of observations, n^b, from the category with the smaller current sample mean. At this grid point, determine Δ_c by interpolation between the contour lines. If the observed Δ_E (defined below) is less than Δ_c, select the next observation from the category with the smaller sample mean.



If the observed Δ_E is equal to or larger than Δ_c, select the next observation from the category with the larger sample mean.

$$\text{Observed } \Delta_E = \frac{\text{Larger sample mean - Smaller sample mean}}{\text{Standard Deviation of Category with larger sample mean}}$$

DECISION SURFACE

FIGURE 6

Appendix B contains the flow diagram for the computer program used in the above solution. In its current form, the program requires an average of one-quarter of a second (on the IBM 7090) for the computation of each Δ_c. The program allows one to find a solution for any initial starting point on the decision tree. For example, an experimenter may have prior information on n_1 observations from π_1 and n_2 observations from π_2 before he decides to use the backwards-induction solution to determine an optimal path through the remainder of the decision tree. Also, the program accommodates problems in which the population variances are known but not equal. An interesting extension of the backwards-induction technique described above would be for the case where the variances are unknown.

In the application of the backwards-induction solution discussed above, the only thing of interest was the Δ_c. However, in other applications, the value of $E[S_{\tilde{n}}]$ is required, and therefore these values are also made available by the computer program.

It is conceivable that a library of solutions for different N and k can be obtained with this computer program. However, before any large scale project of this nature is undertaken, consideration should be given to the use of a hybrid analog-digital computer for the calculation of the numerous integrals encountered in this problem.

The final comment on this section is that even though some interesting decision rules have been developed for maximizing the sum of the net values associated with observations from k categories, considerable further work can be done in extending and generalizing both the two-stage and multi-stage sequential sampling plans. For specified N and k and Δ (or δ) it can be demonstrated that one or the other of the various decision rules discussed in this section yields the highest $E[S_{\tilde{n}}]$. However, the differences are not always large, and the significance of the difference between $E[S_{\tilde{n}}]$ obtainable with different decision rules cannot be evaluated without considering the precision of the basic data and the utility function employed in converting these data into "X" values. Therefore, it is now time to examine the hitherto mysterious "X" quantities used in this and the preceding sections.

SECTION IV

A UTILITY FUNCTION FOR THE OUTPUT OF EDUCATIONAL SYSTEMS

Up to this point, it has been suggested that in a situation where students are being "educated" and simultaneously being used as "experimental subjects", one should follow a decision rule which tends to maximize the net output of all students going through the system, that is, maximize

$$S_n = {}_1X + {}_2X + \cdots + {}_nX$$

Some decision rules which tend to give maximum S_n under different conditions of a priori knowledge were also suggested. However, the "net output", ${}_nX$, has remained ambiguous. This ${}_nX$ can be prescribed for different sets of conditions, some of which are given below.

A. Minimum Conditions

i. A nominally described teaching-learning program.

ii. A numerically scaled student performance measure, where equal distances on the scale have equal "value" and one end of the scale has "higher value" than the other end of the scale (a binary scale is permissible).

The number obtained for each student from the measure described in ii is the X for use in the decision rule.

The minimum conditions given above are typical of almost all currently reported educational experiments, where no attempt is made to specify the relationship between costs of education or the value of the subsequent life-productivity of the student and the school performance measure.

These minimum conditions may suffice for making decisions on micro-aspects of an educational system, but if one's actions are to make sense when judged from outside of the system, then the system inputs and outputs must be defined in value units which have currency outside the system. This is not a new problem, but one that has continuously plagued educators and has long been considered of fundamental importance. The views have often been despairing. For example, M. L. Jackson [29] noted the similarity between some engineering and some educational processes. He suggested that "the student is our 'product' in the manufacturing process of education. The raw material varies, sometimes in an uncontrollable manner. Classroom instruction is the process whereby the product is formed and this phase is of overall importance. The final product cannot be evaluated except after a number of years, and in most cases the feedback is obtained too late, or not at all". If what Jackson says is true, then very little meaningful analysis of such an educational process is feasible. If a current value for the output of the educational process stated in the same dimensions as the value of the inputs cannot be found, and if differences in the output cannot be related to specific differences in the transform, then the problem can only be resolved by insight and intuition.

The problem can be illustrated by a simple example; given the following data:

	Method A	Method B
Average Final Examination Score	80	90
Average Learning Time	9 months	6 months
Cost per Student	$1200	$2500

and the statement that differences between the average examination scores and the average learning times for the two methods are statistically significant, how does one determine which method to adopt?

How does one evaluate whether or not an increase in examination score of 10 points is worth an increase in cost per student of $1300? Or what "value" should be assigned to the three months' saving in learning time possible with Method B? Is one justified in using some combined measure of score and time, such as the commonly suggested final score divided by learning time? Why not use final score divided by the logarithm of learning time, or any other arbitrary weighting?

Partial answers to some of the aspects of this problem are found in the recent literature on the measurement of educational system outputs. Jones [30] used a rating of the individual graduate's subsequent "success" as evaluated by his peers and also the graduate's self-rating of satisfaction and achievement. Jones also attempted to obtain evaluations (from teachers of the graduates) on the contributions to society made by the individual graduate, and also on how these contributions compared with the teacher's subjective opinion of the potential capabilities of the graduate. However, there is some question as to the validity and reliability of the above measures.

Many investigators use life-cycle earnings of students as the measure which is (somehow or other) related to school performance, not because earnings are a more valid measure, but primarily because it is a more reliable and more readily available measure. Earnings are certainly not an ideal measure, since differences in income can be attributed not only to differences in the type, quality, and extent of education, but also to personality factors, regional factors, family contacts, etc. However, income has remained the most commonly used measure of the effect of education on student output.

Machlup [31] conceives the educational system as a knowledge-spreading industry and evaluates its economic efficiency. He calculates that this industry in 1958 produced goods and services worth $136.4 billion, and that all forms of education cost $60 billion, or almost 13% of the 1958 Gross National Product. He states that the total knowledge industry accounted for 29 percent of the Gross National Product and is now growing about two and one half times faster than the industries that produce all other kinds of goods and services.

Becker [32] studied rate of return from college education, allowing for the generally higher initial ability of the college student. He found that the rate of return on the investment in college education by urban white male students, including income foregone by the student while attending school was 12.5 percent in 1940 and 10 percent in 1950 before taxes. When the social cost of college education was added to the individual cost, the rate of return in both years was about 9 percent before taxes. Schultz [33] estimated that the rate of return on investment in college education in 1958 was 11 percent. He then calculated the total years of education in the labor force, gave appropriate weights to each level of education, and estimated that the return on the total investment in education was 17.3 percent. Schultz, like Becker, included income foregone in the total cost of education. Both Becker and Schultz calculated on the basis of total resource costs as well as on private resource costs.*

---

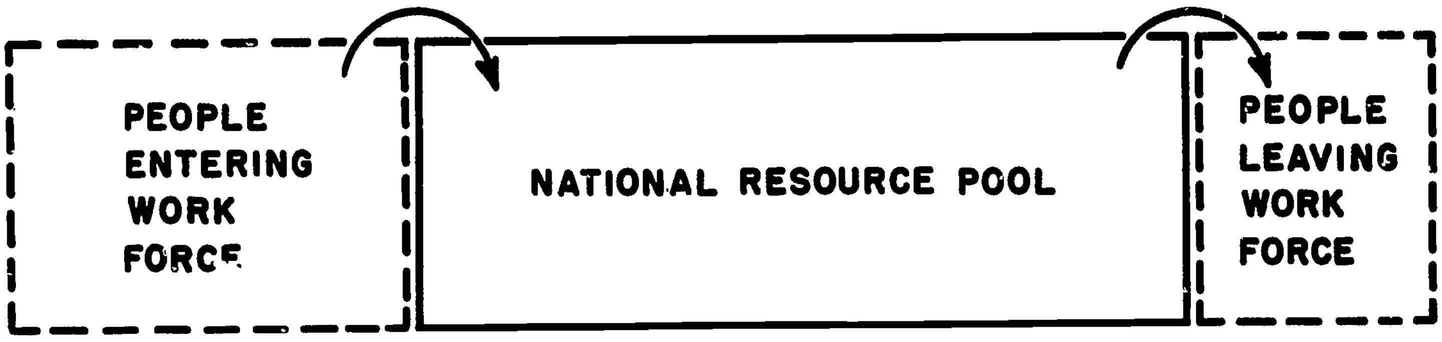
*Total resource costs include: (a) school costs incurred by society, i.e., teachers' salaries, supplies, "rental" of buildings and grounds, etc., (b) opportunity costs incurred by individuals, i.e., income foregone during school attendance and (c) incidental school-related expenditures paid by individuals, i.e., books, travel, etc.
Private resource costs include the same three components, except that in (a) above tuition and fees paid by individuals are substituted for society's costs, which are normally defrayed through taxation.

Hansen [34] has derived the internal rate of return for varicus levels of schooling from grade one to the completion of four years of college, and indicated that this measure provides a more useful method of ranking the economic returns to investment in schooling than do the more conventional lifetime or present value of lifetime income methods. Miller [35] computed the 1949 capital value of lifetime income according to years of schooling. Houthakker [36] estimated the present value of income streams associated with different levels of schooling on the basis of alternate discount values.

The view adopted here is that the investment which the individual and society make in education yields a return in the form of an increase (or decrease) in the contributions which the educated individual makes to his own well-being and to society throughout his later life and that current measures of student performance are indicators of the probable extent of these contributions. This view will be made more explicit, and methods for obtaining quantifiable input-output values will be suggested.

Imagine a "national resource pool" consisting of all the productive output,* instantaneous and accumulated (capital), of the population, as pictured in Figure 7. With a growing population, this pool can increase merely by the greater numbers of people entering the pool than leaving it, assuming the productive capacities of the entering and leaving persons are the same. In order for the people entering the work force to be able to perform most tasks, they require some training, at lease in the language and customs of the nation. Above this minimum -- let's say, unskilled laborer

*"Productive output" is here used in a very broad sense to cover any human activity which has social or private value. Later, a specific kind of productive output and a measure for such output will be described.



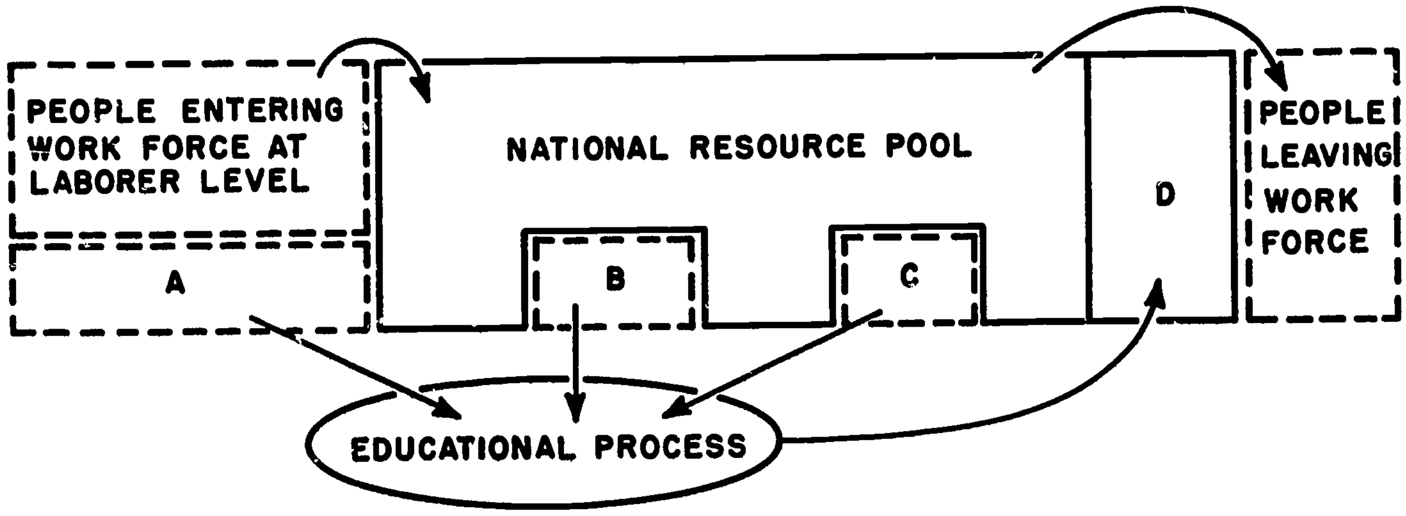
NATIONAL RESOURCE POOL

FIGURE 7

training -- the question arises as to how much of the national resource pool shall be withdrawn from active productive activity to increase the future productive output of the entering (or existing) work force. The question is similar to that propounded by Adam Smith in Wealth of Nations, (1776): How much benefit do I forego now in order to increase my benefits later? For example, in order to train prospective engineers, a certain number of "experienced" engineers must be withdrawn from active practice of their profession to "teach" the trainees. Simultaneously, a number of unskilled laborers must be withdrawn from the work force to become trainees, and also accumulated resources must be set aside for bricks and mortar to build schools, rather than, say, shoe factories. This can be illustrated as in Figure 8.

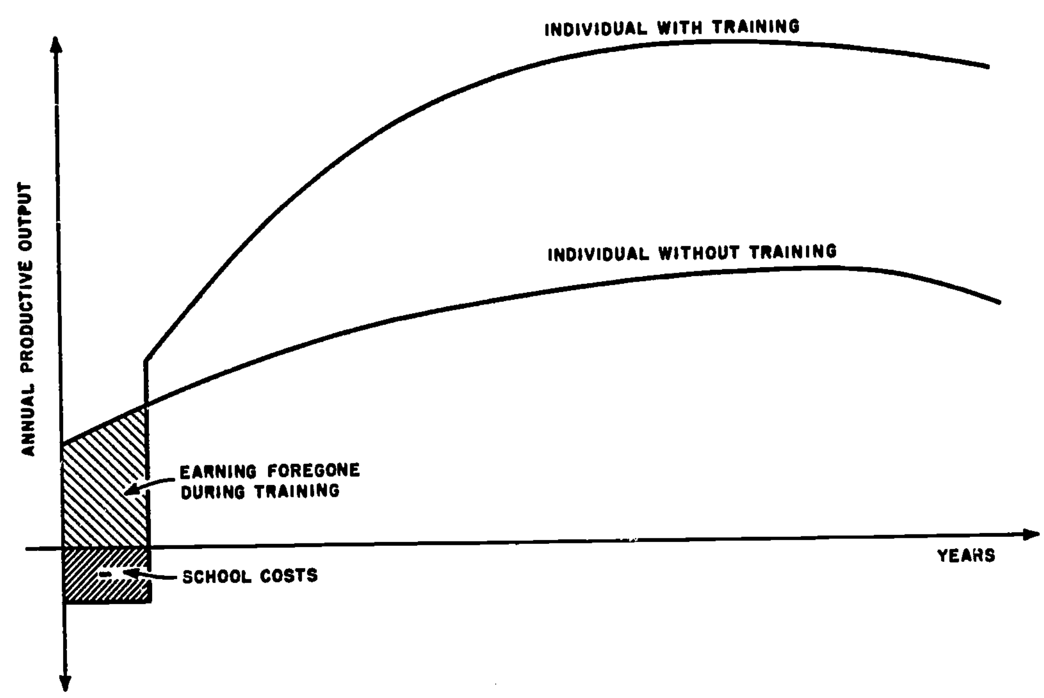
Presumably, after a time, the resource value of the trainees will be greater than the loss of withdrawing a, b, and c from the pool. A time-dependent relationship is needed to express this.

Figure 9 shows productive output vs. time for the "trainee" and for the same or equivalent person without training. The two curves form an interesting map, but the topography can be further



THE EDUCATIONAL PROCESS AS PART OF THE NATIONAL RESOURCE POOL

FIGURE 8



PRODUCTIVE OUTPUT VS. TIME

FIGURE 9

reduced to point values. Two possible point values are

$$\sum_m \hat{P}(m) \text{ and } \sum \overset{*}{P}(m)$$

where

$\hat{P}$: Annual productive output of "trainee".

$\overset{*}{P}$: Annual productive output of "non-trainee".

m: Years, from beginning of training or non-training bifurcation.

However, if a decision must be made at the bifurcation point whether to shunt an individual to the "trainee" or to the "non-trainee" path, the above simple point values may be inadequate since they ignore the fact that some of the annual productive output occurs closer to the bifurcation point than others. In short, the simple summation of annual outputs ignores the time value of productive output. It is suggested here that more reasonable point values of the productive output curves are given by:

$$\hat{W} = \sum_m \left[\hat{P}(m)\right] \left[R(r, m)\right]$$

$$\overset{*}{W} = \sum_m \left[\overset{*}{P}(m)\right] \left[R(r, m)\right]$$

where

$\hat{W}$: The present worth of the life-cycle† productive output of the "trainee".

$\overset{*}{W}$: The present worth of the life-cycle productivity of the "non-trainee".

R: The present worth discount factor.

r: The discount rate.

---

†"Life-cycle" productive output is another way of describing the productive output curve. It is an expression for P(1), P(2), ... P(m), ...

The X to use in the decision rule is:

$$X = \hat{W} - \overset{*}{W}$$

or, the present worth (at the age or date of bifurcation) of the difference in life-cycle productive output of the "trainee" and "non-trainee".

The recommendation to use a present worth discount factor on the life-cycle productive output is based on the following assumption:

ASSUMPTION 1. Productive output which becomes available n years from now has greater weight in influencing current decisions on the allocation of resources than does the same quantity of productive output which becomes available n + m years from now (where $n \geqq 0$, and $m > 0$.

Assumption 1 brings with it Condition 1.

CONDITION 1. In any specific situation where decisions are made using Assumption 1, an appropriate discount rate can be specified.

The choice of an appropriate discount rate requires human judgment, and in an educational system there is practically no way to prove an error in such judgment. Some comfort can be drawn from the hypothesis (which will be tested in the penultimate section) that many decisions are relatively unaffected by a change in the discount rate (within the range of usually selected values of 3-10%). Furthermore, there are commonly accepted guidelines for choosing a discount rate.† Nevertheless, the choice of using present worths of

†From the point of view of the "national resource pool", the minimum discount rate should be equivalent to the annual rate of growth of the national resource pool attributable to the growth of population. From an institutional point of view, the appropriate rate could be the prevailing rate on loans to the institution or the rate of return on other investments made by the institution.

life-cycle productive output in a decision rule, rather than, say, the abstract student performance measure of Minimum Condition A-ii is predicated on the belief that the effect of an error in judgment in the first case (using present worth) is less than the effect of an error in judgment in the second case (using Condition A-ii).

Returning now to the $\hat{P}(m)$ given above, it is seen that during the training or educational period productive output is consumed, i.e., withdrawn from the national resource pool. It will be convenient to treat this "negative" productive output as a separate quantity. Also, anticipating the form in which data on productive output is currently available, "m" is redefined to mean "years of experience". Therefore:

$$X = \hat{W} - \overset{*}{W} - V$$

$$= \sum_{m}\left[\hat{P}(m)\right]\left[R(r,m,T)\right] - \sum_{m}\left[\overset{*}{P}(m)\right]\left[R(r,m)\right] - \sum_{\tau}\left[D'(T)\right]\left[R(r,\tau)\right]$$

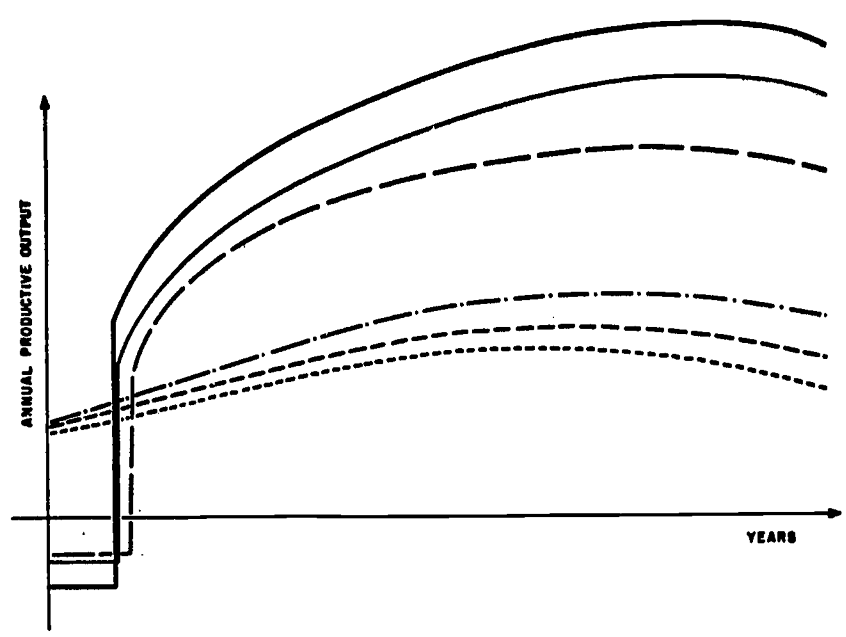
$$= \sum_{m=1}^{b-a-T}\left[\hat{P}(m)\right]\left[\left\{\frac{1}{1+r}\right\}^{T+m-\frac{1}{2}}\right] - \sum_{m=1}^{b-a}\left[\overset{*}{P}(m)\right]\left[\left\{\frac{1}{1+r}\right\}^{m-\frac{1}{2}}\right]$$

$$- \sum_{\tau=1}^{T}\left[D'(T)\right]\left[\left\{\frac{1}{1+r}\right\}^{\tau-\frac{1}{2}}\right]$$

where the redefined and new symbols are:

- $\hat{W}$: the present worth of the life-cycle productive output of the "trainee", excluding educational costs.
- V: The present worth of the educational costs.
- D': The annual educational costs.
- τ: Years from bifurcation date.
- T: Nominal time-span for education or training.
- a: Age at which individuals enter the system (age at bifurcation point).
- b: Retirement age.

In the foregoing, the effect of individual and educational differences on the $\hat{P}$, $\overset{*}{P}$, and D' has not been considered. If these differences are taken into consideration, then the productive output for a given individual will correspond to one of the family of curves shown in Figure 10. The question of individual and educational differences will now be examined in more detail, first under ideal and then under more realistic conditions. Furthermore, an attempt will be made to apply the concepts, expounded above for a macro-system, to sub-units of the macro-system.



EFFECT ON PRODUCTIVE OUTPUT FROM
INDIVIDUAL AND EDUCATIONAL DIFFERENCES

FIGURE 10

B. Ideal Conditions

If one could state the amount of productive output during each future year of a student's life attributable to specific personality factors and to specific performance scores on a specific version of a sub-unit of a total learning experience, given the history of performance score on all other sub-units, then a measure of the "gross value" of the student's performance in the sub-unit could be obtained from the present worth of the sum of these stated annual productive outputs. Furthermore, a "net value" could be obtained by subtracting from the "gross value" the present worth of the productive assets used in providing to the student the sub-unit of learning experience.

Explicitly, the conditions for the ideal case are:

i. A nominally described teaching-learning program, divided into various sequences of sub-units, with various versions of each sub-unit, each of which can be separately described and analyzed.

ii. A time span for completing i, and each sub-unit of i.

iii. A cost associated with providing each sub-unit of i.

iv. A student performance scoring procedure, in which the scores are related to those factors in the teaching-learning process which can be manipulated by the educator-experimenter and are independent of the student personality factors.

v. A personality rating procedure, in which the ratings are not affected by the teaching-learning program.

vi. The future increment in life-cycle productive output (of an individual with specified personality factors and history of performance) attributable to a specified performance in a specified version of a given sub-unit.

In this ideal case, the X used in the decision rule is given by:

$$_{n}X_{ij\ell} = \Delta W - V,$$

where

$$\Delta W = \sum_{m=1}^{b-a-\mathcal{J}} \Big[\Delta P(m, g, \alpha, \beta, j, \ell) \Big] \Big[R(r, m, \mathcal{J}) \Big]$$

and

$$R \Big[(r, m, \mathcal{J}) \Big] = \left\{ \frac{1}{1+r} \right\}^{+m-\frac{1}{2}},$$

where

$$V = \Big[D(t, j, \ell) \Big] \Big[R(r, t, \tau) \Big]$$

and

$$R\left[(r, t, \tau)\right] = \left\{\frac{1}{1+r}\right\}^{\tau - \frac{t}{2}},$$

where the new symbols used above are:

n: The number associated with each individual in the sequential sampling and decision rules.

ℓ: Sub-unit designation.

j: A version of the sub-unit.

i: The sequential number assigned to each "individual" in "j ℓ".

ΔP: The increment in life-cycle productive output attributable to going through the "j ℓ" sub-unit.

ΔW: The present worth of the increment in life-cycle productive output.

g: Student performance score.†

α: Student personality rating.

β: History of performance on other sub-units.

$\mathcal{T}$: Time span required by student to complete i.

t: Time span required by student to complete the "j ℓ" sub-unit.

D: The total costs associated with a sub-unit.

In this ideal case, the exact information on future productive output and on learning time for sub-units which come after the "j ℓ" sub-unit are presumed to be available at the instant when the student completes the "j ℓ" sub-unit. Since this is obviously impossible, estimates for ΔP and $\mathcal{T}$ must be found. Also, ΔP implies that in the ideal case the increment in productivity is directly measurable, something which is rarely possible. Most likely, ΔP will have to be

---

†"g" is independent of "t". If the performance specifications include a measure on speed, then this is reflected in the performance score.

derived from the difference of two P's. Consideration will first be given to the question of how to obtain estimates for P's and $\mathcal{J}$'s, and then the possible ways of obtaining ΔP will be considered.

If the life-cycle productive output of individuals who have previously gone through the "ℓ-th" sub-unit and who have the same g, α, β characteristics as the student who is currently completing the "ℓ-th" sub-unit are available, then it is suggested that an estimate of P(m) for the student can be obtained from projections of the P(m) of the "old grads". The data on past productivity from which the estimates of future productivity will be made is designated by $P(y', m, g, \alpha, \beta, j, \ell)$ where y' indicates the date on which "old grads" entered productive activity.

It is also possible to obtain an estimate of $\mathcal{J}$ for the current student by matching the student's history of "t" on all sub-units up to and including the "ℓ-th" sub-unit with the history of "t" of the "old grads" and then projecting from the $\mathcal{J}(y')$ of this matched group to an estimated $\mathcal{J}$ for the current student.

There are various methods for making forecasts, such as is suggested above for P and $\mathcal{J}$, from data on previous events to projected future events. All such forecasting methods presume a certain stability of the environment in which the events occur. Such stability does not necessarily mean that no change takes place, but rather if changes do occur, then the rate of change should be stable.

The practical application of much of what follows below depends upon the exactness of the forecasting and the ability to recognize when the assumptions of stability are being violated. Stated another way, the recommendation to make forecasts of future productive output from data of previous output is based on the following assumption:

ASSUMPTION 2. The factors which affect the relationship between an educational experience and subsequent productive output remain stable and discernible,

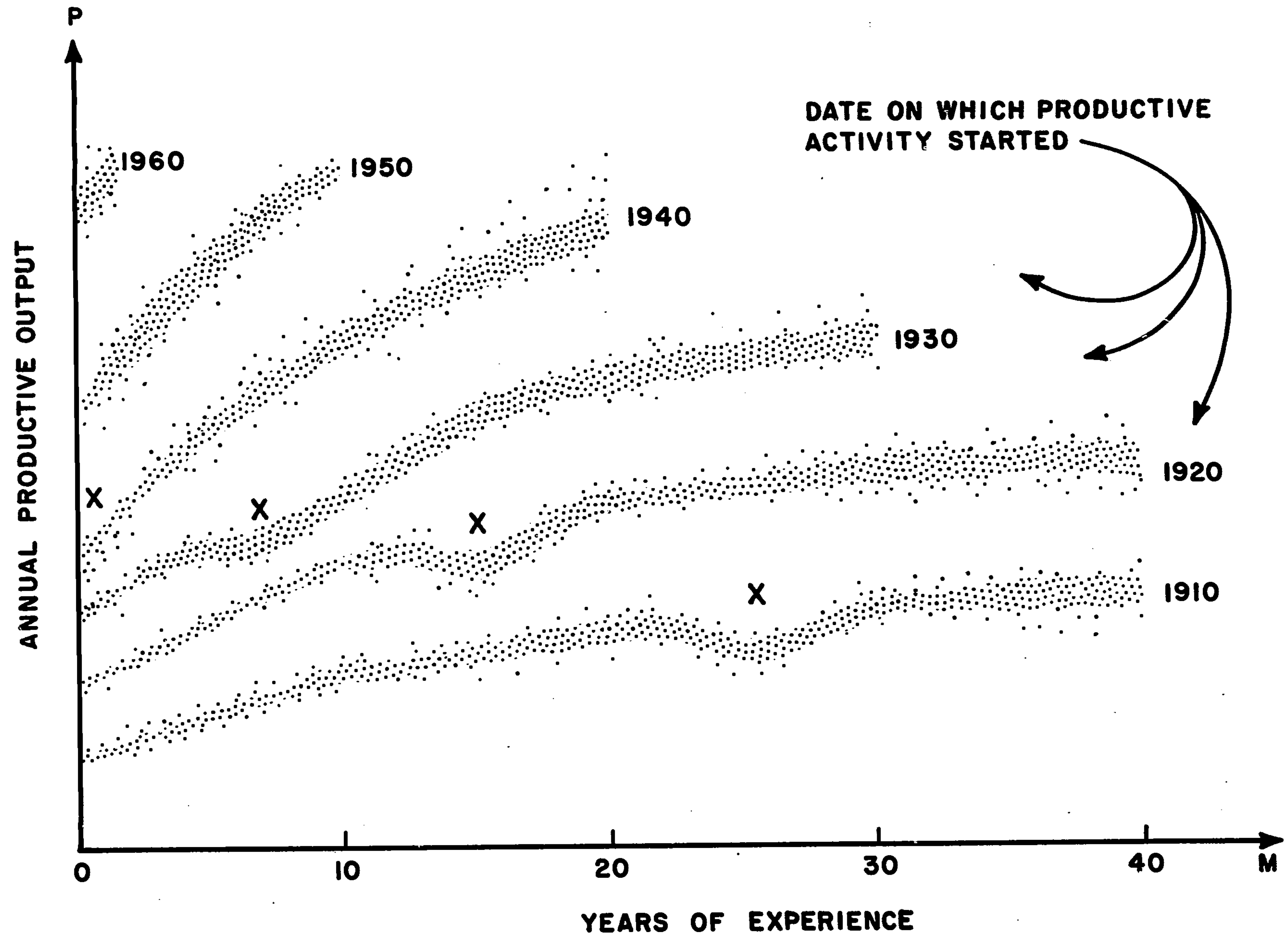
and

CONDITION 2. An appropriate transform can be specified for converting data on previous productive outputs to estimates of future productive output.

Since general concepts are being developed in this section, the possible transforms that could be used will not be discussed here. A detailed example of one such transform, for the life-cycle productive output of engineers, is given in Section V. Two comments, however, are pertinent at this point.

First, a word of caution about the indiscriminate use of mathematical curve-fitting techniques: a graphic display of the data may help in discerning anomolies or violations of Assumption 2. For example, Figure 11 shows a graphic plot of some hypothetical $P(y', m)$ for persons who started productive activity in the years (y') 1910, 1920, 1930, 1940, 1950, 1960. The dips in the curves at the cross-marks show the influence of the anomolous depression years.

Second, by the very nature of the data shown in Figure 11, as "m" increases, the number of data points available for forecasting $P(y, m)$ decreases. Therefore, the forecast of the productive output for the later years of experience are more subject to error. However, by using the sum of the present worths of the expected productive output for each year of experience in the calculation of ${}_nX_{ij\ell}$, the effect of the larger errors in forecasting the output of later years is partly offset by the relatively smaller weighting given to output of these later years.



SCATTER DIAGRAM OF HYPOTHETICAL PRODUCTIVE OUTPUTS

FIGURE 11

Some additional factors should now be considered, the first of which is a mortality factor. In making projections of expected life-cycle productive output from the record of individuals who have already had a productive output for "m" years, a mortality factor should be included to account for the probability that a student will be alive during each of the "m" years of his potential life-cycle of productive output. Also, a transform should be included to convert the various measures of productive input and output to one common measure, preferably to a monetary measure. It will be recalled that "productive output" is being used in this ideal case to cover any human activity which has social or private value and could include such diverse things as building bridges, writing scientific papers, receiving honors or prestige, enjoying leisure time, painting non-salable paintings, and so on. However, even in the ideal case, dimensional conformity is required of all the elements in an equation. A monetary measure, and more specifically, a dollar measure is recommended because much of the inputs and outputs that are likely to be considered are already measured in dollars. The use of a transform to convert all forms of productive output to dollar units brings with it the need for a transform that will convert dollar units of productive output reported in one year into dollar units of productive output reported in another year. In other words, adjustment must be made for the year to year fluctuation in the value of the dollar.

Consider now the problem of finding a ΔP, the increment in life-cycle productive output attributable to going through a particular sub-unit. In some rare cases it may be possible to match the productive output ($\hat{P}$) of individuals who have had the "ℓ-th" sub-unit with the productive output ($\overline{P}$) of individuals who have had all other sub-units except the "ℓ-th". Incorporating all of the above ideas gives the following modified ideal case.

C. Modifications to the Ideal Conditions

1. Conditions for the Modified Ideal Case Using $p(\hat{y}, \hat{P})$ and $p(\bar{y}, \bar{P})$:

i. A nominally described teaching-learning program, divided into various versions of each sub-unit, each of which can be separately described and analyzed.

ii. A time-span for completing i, and each sub-unit of i.

iii. A cost associated with providing each sub-unit of i.

iv. An objective, stationary, student performance scoring procedure (i.e., where scores obtained now have the same significance as scores obtained some years ago), where the scores are related to those factors in the teaching-learning process which can be manipulated by the educator-experimenter and are independent of the student personality factors.

v. A personality rating procedure, in which the ratings are not affected by the teaching-learning program.

vi. Data on the life-cycle productive output of individuals who have previously completed all sub-units of i. The data are sub-classified according to personality factors, history of performance on all sub-units of i, the date of entering productive activity, and for each year of experience.

vii. Data on the life-cycle productive output of individuals who have previously completed all except the "ℓ-th" sub-unit of i. The data are sub-classified as in vi.

viii. A transform which converts the data given by vi and vii into expected (or future) life-cycle productive output data.

ix. A transform which converts all measures of productive output to a common monetary measure.

x. A transform which converts monetary values reported for one year into the equivalent monetary value of any other year.

xi. Data on the probability of survival for individuals who do and for those who do not go through i. The probability of survival at age "a" at the bifurcation date is equal to one.

The new expression for X under these conditions is:

$$
\begin{aligned}
{}_{n}X_{ij\ell} &= \hat{W} - \bar{W} - V \\
&= \sum_{m}\Big[p\Big(\hat{y}, d\big(f(\hat{P}(y', m, g, \alpha, \beta, j, \ell))\big)\Big)\Big]\Big[R\Big(r, m, p\big(\hat{y}, \mathcal{J}(y', \alpha, \beta, j. \ell, t)\big)\Big)\Big] \cdot \\
&\qquad \cdot \Big[\hat{M}\Big(a, m, p\big(\hat{y}, \mathcal{J}(y', \alpha, \beta, j, \ell, t)\big)\Big)\Big] \\
&\quad - \sum_{m}\Big[p\Big(\bar{y}, d\big(f(\bar{P}(y', m, \alpha, \beta))\big)\Big)\Big]\Big[R\Big(r, m, p\big(\bar{y}, \mathcal{J}(y', \alpha, \beta, t)\big)\Big)\Big] \cdot \\
&\qquad \cdot \Big[\bar{M}\Big(a, m, p\big(\bar{y}, \mathcal{J}(y', \alpha, \beta, t)\big)\Big)\Big] \\
&\quad - \Big[D(t, j, \ell)\Big]\Big[R(r, t, \tau)\Big] \\
&= \sum_{m=1}^{b-a-p(\hat{y}, \hat{\mathcal{J}})}\Big[p\Big(\hat{y}, d\big(f(\hat{P}(y', m, g, \alpha, \beta, j, \ell))\big)\Big)\Big]\Big[\Big\{\frac{1}{1+r}\Big\}^{p(\hat{y}, \hat{\mathcal{J}})+m-\frac{1}{2}}\Big] \cdot \\
&\qquad \cdot \Big[\hat{M}\Big(a+m-\tfrac{1}{2}+p(\hat{y}, \hat{\mathcal{J}})\Big)\Big] \\
&\quad - \sum_{m=1}^{b-a-p(\bar{y}, \bar{\mathcal{J}})}\Big[p\Big(\bar{y}, d\big(f(\bar{P}(y', m, \alpha, \beta))\big)\Big)\Big]\Big[\Big\{\frac{1}{1+r}\Big\}^{p(\bar{y}, \bar{\mathcal{J}})+m-\frac{1}{2}}\Big] \cdot \\
&\qquad \cdot \Big[\bar{M}\Big(a+m-\tfrac{1}{2}+p(\bar{y}, \bar{\mathcal{J}})\Big)\Big] \\
&\quad - \Big[D(t, j, \ell)\Big]\Big[\Big\{\frac{1}{1+r}\Big\}^{\tau-\frac{t}{2}}\Big]
\end{aligned}
$$

where

$$\hat{\mathcal{J}} = \mathcal{J}(y', \alpha, \beta, j, \ell, t)$$
$$\bar{\mathcal{J}} = \mathcal{J}(y', \alpha, \beta, t)$$

and

$$\hat{y} = y + p(\hat{y}, \hat{\mathcal{J}}) - \tau$$
$$\bar{y} = \dot{y} + p(\bar{y}, \bar{\mathcal{J}}) - \tau$$

in which $\hat{y}$ and $\bar{y}$ can best be found by successive approximations.

The new and redefined symbols in the above expressions are:

W: Present worth of life-cycle productive output, exclusive of educational costs.

P: Annual productive output.

$\hat{}$: A sign to indicate that the symbol below the sign is associated with the individual who has had all sub-units in i.

$\bar{}$: A sign to indicate that the symbol below the sign is associated with the individual who has had all but the "1-th" sub-unit of i.

y: Current date.

y': Date on which individual who previously completed i started productive output.

m: Years of experience, since starting productive output.

p: A transform which operates on the history of past events to give an estimate of future events.

f: A transform which converts all forms of productive output to dollar values of the year that the output occurred in.

d: A transform which converts dollars of any given year into dollar values of any other specified year.

M: A mortality factor (or more correctly, a probability of survival factor).

In the event that the $p(y, \mathcal{J})$ are less than one-half year, the above formulation is considerably simplified, since we can ignore $\mathcal{J}$, τ, and t in the discount factor R. Thus:

$$ {}_{n}X_{ij\ell} = \sum_{m=1}^{b-a} \left[p\left(y, d\left(f(\hat{P}(y', m, g, \alpha, \beta, j, \ell))\right)\right) \right] \left[\left\{ \frac{1}{1+r} \right\}^{m-\frac{1}{2}} \right] \cdot \left[\hat{M}(a+m-\tfrac{1}{2}) \right] $$

$$ - \sum_{m=1}^{b-a} \left[p\left(y, d\left(f(\bar{P}(y', m, \alpha, \beta))\right)\right) \right] \left[\left\{ \frac{1}{1+r} \right\}^{m-\frac{1}{2}} \right] \left[\bar{M}(a+m-\tfrac{1}{2}) \right] $$

$$ - D(t, j, \ell) $$

Either of the above formulations may be adequate for the case where "ℓ-th" sub-unit under consideration is the last one in the sequence of sub-units of i, and also in the case where the student's performance in one sub-unit is independent of his performance in another sub-unit, an assumption which is often made for the sake of mathematical simplicity,† but one which seldom makes sense in most teaching-learning programs. The temptation to use simplifying assumptions is understandable, for in this case the logical move is to use the performance results on past and current sub-units to fill in the conditional probabilities of performance on future sub-units, a procedure which becomes exceedingly unwieldy and increasingly imprecise as the number of sub-units increases.

Since initially it may be difficult to accumulate enough $\bar{P}(y', m, \alpha, \beta)$ to use in obtaining satisfactory $p\left(y, P(y', m, \alpha, \beta)\right)$, two other possibilities should be examined.

†Smallwood and Pask both make this assumption in their adaptive system models.

One of the possibilities is that the proportional part that each sub-unit contributes to the overall subsequent productive output can be stated outright, in which case other conditions prevail:

2. Conditions for the Modified Ideal Case Using $p(\hat{y}, \hat{P})$, $p(\overset{*}{y}, \overset{*}{P})$ and Proportionality Factors.

i. Same as C-1-i.

ii. Same as C-1-ii.

iii. Same as C-1-iii.

iv. Same as C-1-iv.

v. Same as C-1-v.

vi. Data on the life-cycle productive output of individuals who have previously completed i. The data are sub-classified according to personality factors, history of performance on all sub-units of i, the date of entering productive activity, and for each of the years of experience.

vii. Data on the life-cycle productive output of individuals who did not go through i, but who had the same initial qualifications as those who went through i. The data are sub-classified according to personality factors, the date of entering productive activity, and for each of the years of experience.

viii. Same as C-1-viii.

ix. Same as C-1-ix.

x. Same as C-1-x.

xi. Same as C-1-xi.

xii. Proportionality factors which indicate the part that each sub-unit contributes to subsequent overall productive output.

For this case:

$$
\begin{aligned}
{}_{n}X_{ij\ell} &= c\,(\hat{W} - \overset{*}{W}) - V \\
&= c\Big(\ell, \sum_{m}\Big[p\big(\hat{y}, d(f(\hat{P}(y', m, g, \alpha, \beta)))\big)\Big]\Big[R\big(r, m, p(\hat{y}, \hat{J})\big)\Big] \cdot \\
&\qquad \cdot \Big[\hat{M}\big(a, m, p(\hat{y}, \hat{J})\big)\Big] \\
&\quad - c\Big(\ell, \sum_{m}\Big[p\big(\overset{*}{y}, d(f(\overset{*}{P}(y', m, \alpha)))\big)\Big]\Big[R(r, m)\Big]\Big[\overset{*}{M}(a, m)\Big] \\
&\qquad - \Big[D(t, j, \ell)\Big]\Big[R(r, t, \tau)\Big]
\end{aligned}
$$

where the new symbols are:

c: A proportionality or weighting factor, where

$$\sum_{\ell} c(\ell) = 1.0 \, .$$

$\overset{*}{y}$: Date on which individual who does not go through i starts productive output; also the bifurcation date.

Simplifications can be made in the above formulation if $p(\hat{y}, \hat{J})$ is less than one-half year.

It should be emphasized that "c" is a subjective measure. If objective measures are available, they would be used directly without introducing "c", as for example in the comparison of $\hat{P}$ and $\overline{P}$ given above. The difficulty with this subjective measure is that there is less concrete evidence and there are fewer guidelines available to help determine the magnitude of "c" than for any other element that enters into the determination of ${}_{n}X_{ij\ell}$. A common assumption, particularly where the sub-units are very small blocks of learning, is that each sub-unit has equal importance, and therefore all c-values are equal. Another common practice (for example, with the semester courses of a college or high school program) is to divide the sub-units into two major categories† and within each category, weight the sub-unit in direct proportion to the number of teaching hours allocated to that sub-unit. This practice assumes that within each category importance is related to teaching time and presumes that the amount of teaching time required for each sub-unit can be rationally resolved. D. Rosenthal, A. Rosenstein, and G. Wiseman [37] have suggested a novel way for a faculty committee to resolve the question of how to specify the relative (though still subjective) weighting of the sub-units. Nevertheless, the determination of "c" remains one of the more interesting areas for further research.

In a comprehensive application of an adaptive teaching system, one may have to settle for subjective approximative values for "c" when the system is inaugurated but include a feature for the accumulation of $\overline{P}$ data which, in time, can be used to supplant the use of "c". In many cases, the adaptive decisions will not be affected even by the choice of an inappropriate "c", particularly in those cases where:

$$\left.\begin{matrix} c(W - W) \\ c'(W - W) \end{matrix}\right\} >> V \quad \text{or} \quad \left.\begin{matrix} c(W - W) \\ c'(W - W) \end{matrix}\right\} << V$$

where "c" is the value actually used and "c'" is the unknown "correct" value. This contention will be examined further in Section VI.

Returning now to the problem of how to circumvent the dearth of data on $\overline{P}(y', m, \alpha, \beta)$, another possibility to consider is to forego the analysis on the sub-units of i and restrict oneself to making analyses for the entire i in which case neither $\overline{P}$ nor "c" is required.

† For example, one category could include all the laboratory and "non-academic" courses, while the other category could include all the lecture-recitation courses.

In this situation "j" could indicate a specific sequence of variations of the sub-units. If there are many such sub-units and variations of sub-units, then the number of "j" will be very large, and we are back to the old problem of fragmenting the P(y') into so many sub-divisions that very large numbers of P(y') will be needed to make reasonable forecasts of the future P(y). On the other hand, if there are few or no sub-units in i worthy of separate analysis (such as in short courses and in many industrial training situations), then this alternative is entirely reasonable. The conditions for this case are given below.

3. Conditions for the Modified Ideal Case Using $p(y, \hat{P})$ and $p(\overset{*}{y}, \overset{*}{P})$ for the Entire Learning Program.

i. A nominally described teaching-learning program.

ii. A time-span for completing i.

iii. A cost associated with completing i.

iv. Same as C-1-iv.

v. Same as C-1-v.

vi. Same as C-2-vi.

vii. Same as C-2-vii.

viii. Same as C-1-viii.

ix. Same as C-1-ix.

x. Same as C-1-x.

xi. Same as C-1-xi.

The formulation of X for this case is straightforward.

$${}_n X_{ij} = \hat{W} - \overset{*}{W} - V$$

$$= \sum_m \left[p\left(y, d\left(f\left(\hat{P}(y', m, g, \alpha)\right)\right)\right) \right] \left[R(r, m, \mathcal{J}) \right] \left[\hat{M}(a, m, \mathcal{J}) \right]$$

$$- \sum_m \left[p\left(\overset{*}{y}, d\left(f\left(\overset{*}{P}(y', m, \alpha)\right)\right)\right) \right] \left[R(r, m) \right] \left[\overset{*}{M}(a, m) \right]$$

$$- \sum_\tau \left[D'(\mathcal{J}) \right] \left[R(r, t, \tau) \right]$$

$$= \sum_{m=1}^{b-a-\mathcal{J}} \left[p\left(y, d\left(f\left(\hat{P}(y', m, g, \alpha)\right)\right)\right) \right] \left[\left\{ \frac{1}{1+r} \right\}^{\mathcal{J}+m-\frac{1}{2}} \right] \left[\hat{M}(a+m-\tfrac{1}{2}+\mathcal{J}) \right]$$

$$- \sum_{m=1}^{b-a} \left[p\left(\overset{*}{y}, d\left(f\left(\overset{*}{P}(y', m, \alpha)\right)\right)\right) \right] \left[\left\{ \frac{1}{1+r} \right\}^{m-\frac{1}{2}} \right] \left[\overset{*}{M}(a+m-\tfrac{1}{2}) \right]$$

$$- \sum_{\tau=1}^{\mathcal{J}} \left[D'(\mathcal{J}) \right] \left[\left\{ \frac{1}{1+r} \right\}^{\tau-\frac{1}{2}} \right]$$

If $\mathcal{J}$, τ and t are less than one-half year, the above formulation can be simplified to:

$${}_n X_{ij\ell} = \sum_{m=1}^{b-a} \left[p\left(y, d\left(f\left(\hat{P}(y', m, g, \alpha)\right)\right)\right) \right] \left[\left\{ \frac{1}{1+r} \right\}^{m-\frac{1}{2}} \right] \left[\hat{M}(a+m-\tfrac{1}{2}) \right]$$

$$- \sum_{m=1}^{b-a} \left[p\left(\overset{*}{y}, d\left(f\left(\overset{*}{P}(y', m, \alpha)\right)\right)\right) \right] \left[\left\{ \frac{1}{1+r} \right\}^{m-\frac{1}{2}} \right] \left[\overset{*}{M}(a+m-\tfrac{1}{2}) \right]$$

$$- D'(\mathcal{J})$$

The ideal case has been treated at some length because it represents an attainable set of conditions. Admittedly, currently available conditions are far removed from the ideal conditions, and it will be necessary to introduce additional assumptions to obtain a model that can be used today. The practical procedure would be to start using the strongest model that will work with the currently available data and simultaneously start gathering data in a form suitable for use in a model more closely approximating the ideal model.

The discrepancies between the ideal conditions and the conditions currently prevailing are given below:

a. Student performance scoring procedures generally are not objective, stationary, independent of student personality factors, nor related only to the factors in the teaching-learning process which can be manipulated by the educator-experimenter.

b. Personality rating procedures which are independent of the teaching-learning process and which are related to life-cycle productive output are not available.

c. Data on life-cycle productive output is not generally subclassified according to the (unavailable) personality factors, nor according to the (available) history of performance on all sub-units of i, nor are all the elements of an individual's output recorded.

D. Current Conditions

The question now arises: can a reasonable estimate of X be obtained from existing data? The answer depends in part on where the data are coming from. Some institutions have available fairly detailed information on individual graduates (see Section V on engineering graduates of the University of California); in other cases individual records are not available and only group mean or median figures are quoted. For example, original data on individuals in old Bureau of Census and Labor Department surveys have been lost or destroyed, and only group median figures are available. The answer to the question of whether reasonable estimates of X can be obtained from existing data depends also in part on the further assumptions one is willing to make in order to reconcile existing data with the modified ideal set of conditions given in C above.

For example, most of the old data on productive output are stated only in terms of dollar earnings, with no account being given to other possible signs of non-dollar productive output such as scientific publications, service to the community, etc. There are many ways of arguing this issue, from the one extreme which says that most apparently non-dollar productive output is eventually reflected in higher earnings, to the other extreme which says that our society accurately reflects the value it places on productive output by the dollar compensation it makes for such output. Both extreme views are certainly untenable for many individual cases but may be fairly accurate when median figures for large groups of individuals are considered. The assumptions that are suggested for the use of old data are:

ASSUMPTION 3. Annual earnings are an adequate measure of productive output.

ASSUMPTION 4. Where data on the annual earnings of individuals in a specified group are not available, the median annual earnings of the group can be used.

Using Assumptions 3 and 4,

$$f\Big(P(y', m, \ldots)\Big) = \$\,(y', m, \ldots)$$

where the \$ sign represents median annual earnings, in dollars.

Since dollars have different values in different years, in order to get some consistent value system, the following assumption is made:

ASSUMPTION 5. A stable reference for dollar values is the purchasing power (on a specified list of commodities and services) of the dollar.

Using this assumption, the following simple d-transform is suggested:

$$d(\ ) = \frac{CPI(y)}{CPI(y'+m)}\ (\ )$$

where CPI(z) is the Consumer Price Index for the z-th year. A word of caution about the use of CPI: from time to time the specified list of commodities and services used for evaluating the purchasing power of the dollar changes; also, the list is designed to reflect the normal purchases of the urban moderate income family, and the group whose $(y', m, ...) is being observed may not fall into this category.

Looking now at the ideal requirement that performance scores should be independent of personality factors, it becomes apparent that not only are the personality factors not specified in old data, but that these factors are inextricably mixed into the performance scores. This gives rise to the following further assumption for the use of old data:

ASSUMPTION 6. Personality factors need not be excluded from performance scoring procedures.

This gives rise to a new symbol, g', which represents performance scores that reflect both differences in the teaching-learning program and individual personality differences and eliminates α from the formulation of X.

It must furthermore be recognized that g' is usually not obtained from an objective scoring procedure, but rather from a relative ranking procedure and that the scoring procedure is not stationary. It is therefore necessary to make:

> <u>ASSUMPTION 7</u>. An adequate relationship can be found between previously recorded g' and currently observed g.

Stated another way, a transform "h" is needed, which serves to map elements in a set of g to elements in a set of g', or

$$g \leftrightarrow h(g')$$

This h-transform will, of necessity, be different for each specific application, and an example of one such transform is given in Section V.

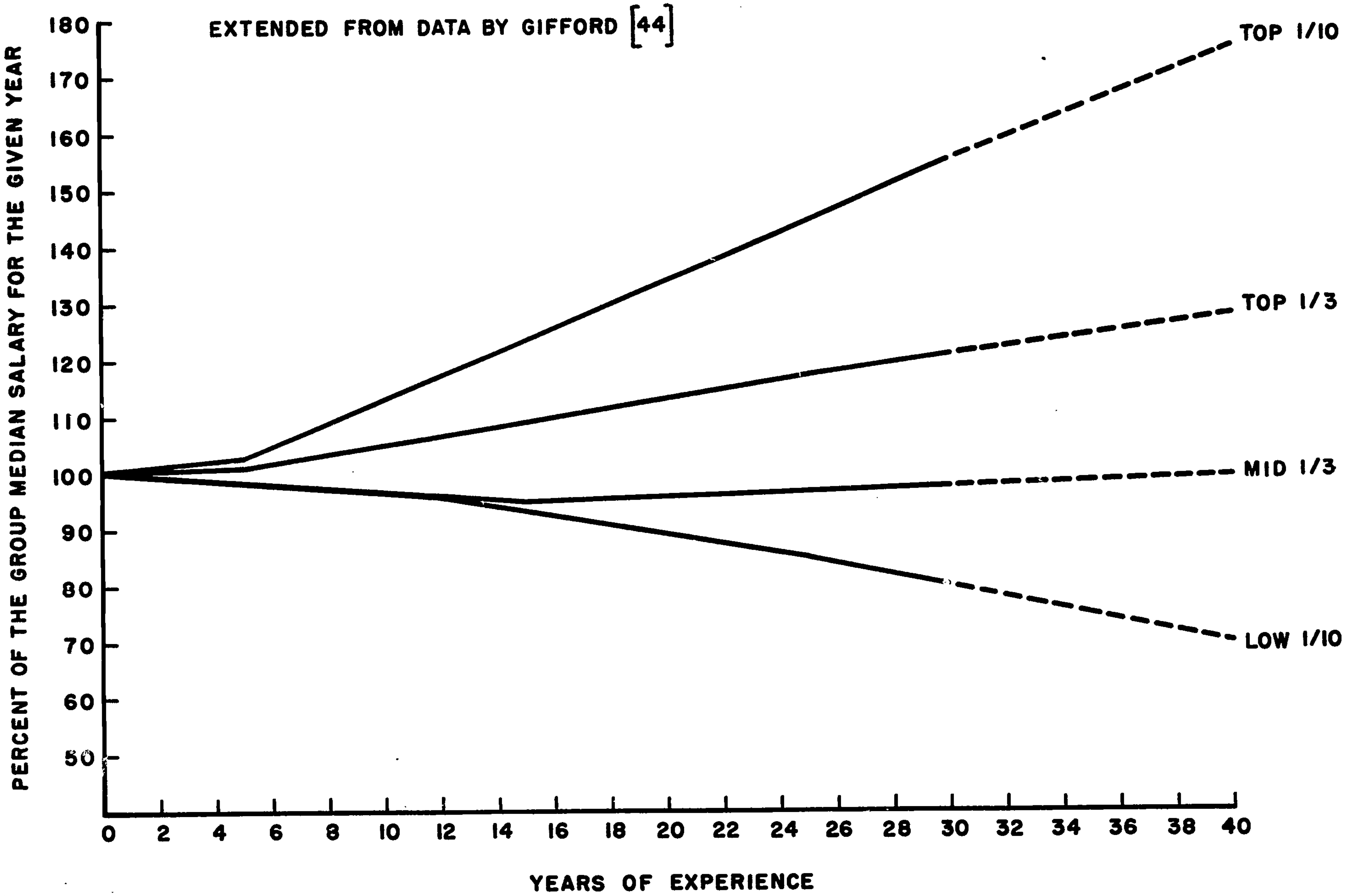
There is a further complication. Very often the median earnings data are not sub-classified according to g', but instead an overall median including all values of g' is quoted. However, it may still be possible to use these global median figures if, from independent sources, a relationship can be established between performance in school and subsequent life-cycle earnings. Then when finding $p\Big(y, P(y', m, g' \ldots)\Big)$, instead of using the $P(y', m, g' \ldots)$ which corresponds to the "g" of a current student, one would use the undifferentiated $P(y', m, \ldots)$ and a transform to obtain $p\Big(y, P(y', m \ldots)\Big)$. In order to do this another assumption must be made:

> <u>ASSUMPTION 8</u>. There is a discernible and independently verifiable relationship between performance in school and subsequent life-cycle productive output.

There have been many studies on the relationship between performance in school and subsequent productive output, the vast majority of which report no significant relationship. As a result, there exists a fairly prevalent feeling that such relationships do not exist or at best, can only be teased out by introducing such co-variables as family background, geographic area, personality factors, etc. However, a careful analysis of the studies on which the pessimistic feelings are based reveals that most of the studies were made on students who took their major in colleges of letters and science. This led to the speculation that training for professional practice (as in the case of engineering education) would be more highly correlated to later professional success than general education would be to later success in the variety of occupations in which a person could be engaged after such general education.

A re-sifting of the literature on such correlational studies was only partly encouraging. For example, Pierson [38] reported that for 320 engineering graduates examined, he found a correlation of 0.43 between their GPA on all college work and a rating of success in their professional life (rated by a faculty member who best knew the person in college). On the other hand, Haveman and West [39] indicate for the general college graduate, the low graders earn less than the high graders, but the highest graders are often in low-paid jobs such as teaching, etc. Some encouragement comes from Wallace [40] who, in 1954, studied alumni of the University of California Schools of Engineering and observed a slight tendency for higher salaries to go with higher grades.

Apropos to measures of productive output other than earnings, Taylor [41] investigated whether engineering undergraduate grades were predictive of later research activity. He used 239 cases and measured research performance by a three-category rating. The tri-serial correlation between these ratings and GPA was a dissappointing .06. But two apparently contradictory reports finally helped unravel the puzzle. LeBold [42] made a study of current monthly salaries of 3977 alumni of the Purdue University Engineering School and reported a positive relationship between income and scholarship for the group with 10 to 25 years of experience. On the other hand, Eurich [43] quoted two studies, one by the Hughes Aircraft Company and the other by the National Advisory Committee for Aeronautics (now absorbed by NASA), wherein for practicing engineers with 6 to 9 years of experience, no correlation was found between their achievement or salaries and their college records. The significant point was the different number of years of experience quoted in LeBold's and Eurich's reported studies. Could it be that the differentiation in earnings was related not only to school performance but also becomes more pronounced with increasing years of experience? Actually the answer had been given years before (1928) by Gifford and later quoted by Bridgman [44]. Gifford found that for the 3806 Bell Telephone System college graduate employees that were studied, higher salaries were associated with higher college standing, and lower salaries were associated with lower standing. Furthermore, the differences in salaries of the high college standing and the low college standing groups became increasingly apparent the longer they were employed. These findings are vividly demonstrated in Figure 12. However, a long time has passed since Gifford's study was made and that study had been based on data from the 1890's to the 1920's. More recently (1962) another study of 10,000 Bell Telephone System college graduates had been made by the American Telephone and Telegraph Company. The report on this study [45] indicated that the employees were divided into four groups: top tenth, top third, mid third, and lower third of their graduating class. When they were cross tabulated by salary thirds, a decided relationship between rank in graduating class and progress in the Bell System was evident. That is, 51 percent of those in the top graduating third were in the top salary third; 40 percent of those in the lowest graduating third were in the lowest salary third. After this encouraging report was



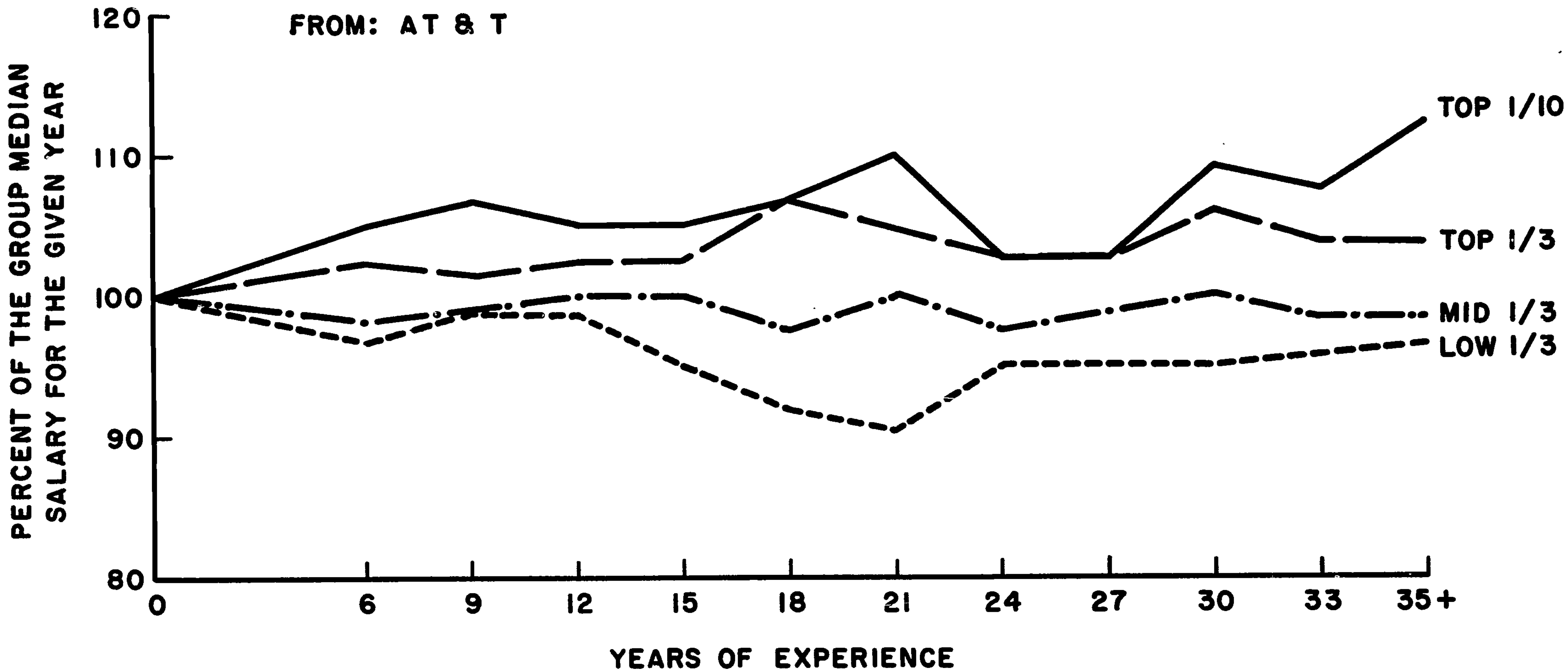
COLLEGE CLASS STANDING AND SUBSEQUENT
RELATIVE EARNINGS; FROM GIFFORD (1928)

FIGURE 12

received, the American Telephone and Telegraph Company investigators were prevailed upon to prepare a graph similar to Gifford's for use in this study. The new graph is shown in Figure 13. Note that the difference in median salaries of the top members of the class and the lower members is not so great in the recent study as it was in the earlier study.

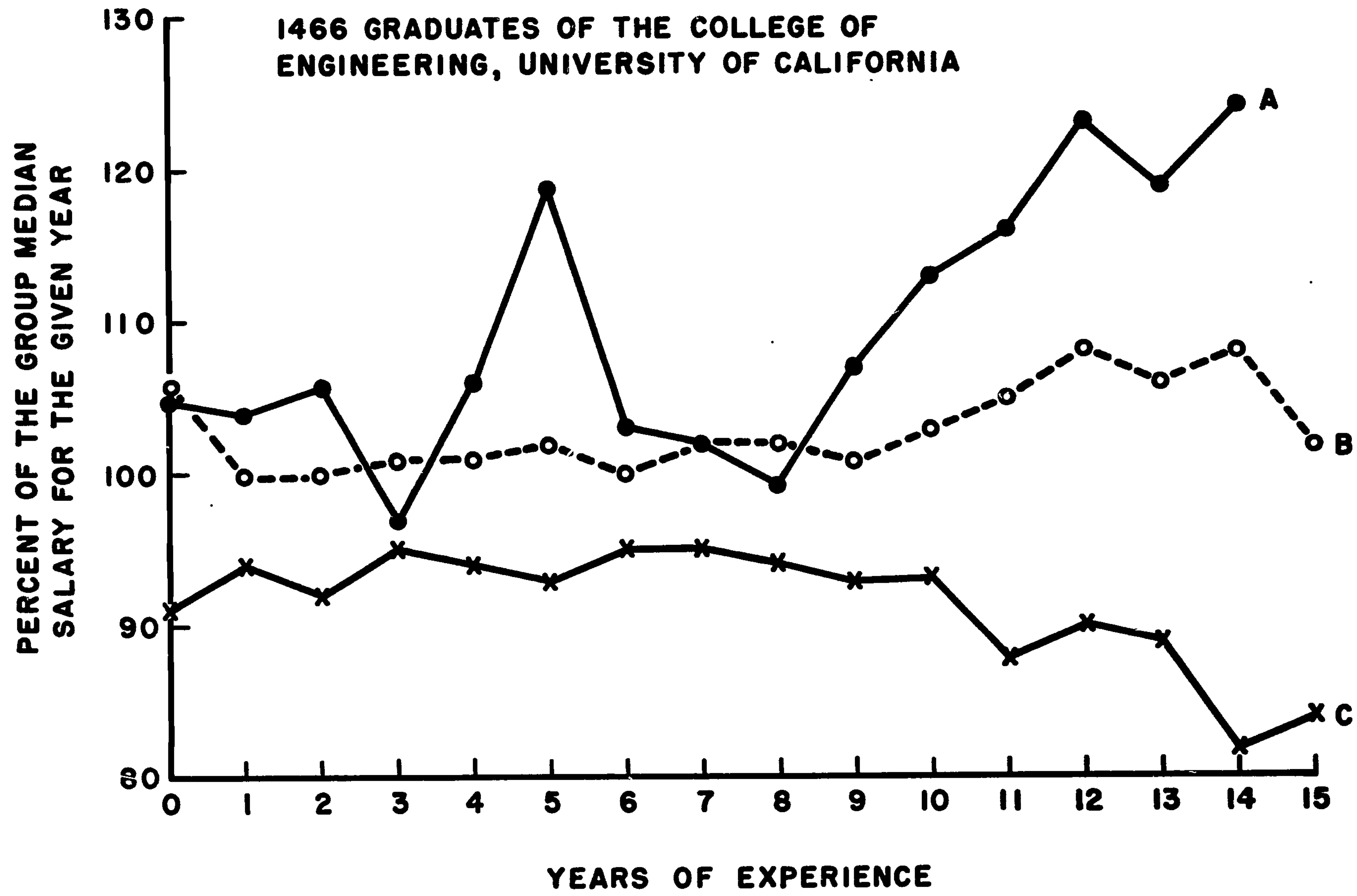
In the meantime, an analysis was made of the data that was very conveniently made available at this time from a comprehensive study of the engineering graduates of the University of California (Los Angeles and Berkeley) conducted by Harry Case, William LeBold, William Diemer and their associates. At the time this analysis was made (1963), data on 1466 graduates from the years 1947 through 1962 were available. All individuals reported their earnings for each year since graduation and also their average grades while in college.† A check on a sample of 170 graduates revealed that student-reported grades and grades actually recorded by the registrar correlated at 0.86, and hence the reported grades were thought to be sufficiently accurate for purposes of correlating school performance and the earnings received in later careers. The sample consisted of graduates of different years having different lengths of experience on the job. Because the purchasing power of money has itself changed during this period, all reported earnings were made comparable by converting them to equivalent dollars of 1962. Then the median earnings for each category of reported college grades were calculated as a percentage of the overall median for each year since graduation. This is shown in Figure 14. We note a similarity between the results of the

†Additional data is available on family background, high school experience, personal factors, etc., on these graduates, and members of Dr. Case's group are making their own analyses on how these other factors may co-vary with earnings and grades.



COLLEGE CLASS STANDING AND SUBSEQUENT
RELATIVE EARNINGS; FROM AT&T (1963)

FIGURE 13



COLLEGE GRADES AND SUBSEQUENT RELATIVE EARNINGS:
COLLEGES OF ENGINEERING, UNIVERSITY OF CALIFORNIA

FIGURE 14

University of California study and the later American Telephone and Telegraph Company study.

The two studies on the Bell System employees show a relationship between relative position in the class and subsequent earnings, and the University of California study indicates a relationship between grades and subsequent earnings. In these three studies, the measure of each student's school performance is relative to the performance of the student body as a whole. We should bear in mind that the factors which influence a student to perform at a level to place him in the top third of his class, or to get an A grade, may be the same factors which subsequently influence his earning power. Inborn intelligence, drive, competitiveness, ambition can be suggested as possible factors, and it is exactly these factors which are not directly manipulated in most educational experiments. Therefore, it is only with caution and with full cognizance of the implications of accepting Assumption 6 that one can recommend using a transform for modifying median expected life-cycle earnings to reflect different expected life-cycle earnings for students with different college performance scores. Such a transform, "w", depends on performance score and years of experience and operates on the undifferentiated or overall median expected life-cycle earnings:

$$w\Big(g, m, p\big(y, \hat{P}(y', m)\big)\Big)$$

or more likely:

$$w\Big(h(g'), p\big(y, \hat{P}(y, m)\big)\Big)$$

in the case where performance is not independent of personality factors, and where some change may also have occurred between the grading technique employed in determing "w" and that employed on the current students.

If, now, the further assumption --

ASSUMPTION 9. An individual's learning time and performance score for the sub-unit under examination is representative of the learning times and performance scores for that individual in all other sub-units --

is made, then β can be eliminated from the formulation, and the estimated time for completing i can be found by the following substitution:

$$p(y, \mathcal{J}) \approx \frac{tT}{T'}$$

where, it may be recalled,

t: time span actually required by a student to complete the sub-unit.

T': nominal time span for completion of the sub-unit.

T: nominal time span for completing i.

Lastly, and perhaps the most questionable, is:

ASSUMPTION 10. Each sub-unit contributes to future productive output in the same proportion that the nominal time span for completing each sub-unit bears to the nominal time span for completing the whole teaching-learning program.

This assumption gives:

$$c(\ell) = \frac{T'(\ell)}{\sum_{\ell} T'(\ell)} = \frac{T'(\ell)}{T}$$

For example, if it were ascertained that students spent approximately 7200 hours in and out of class in study and related activities during the normal four-year college period, and if the $(\hat{W} - \overset{*}{W})$ for a given student is $82,600, then the "output" for an average one-hour learning experience (including in and out of class time) would be $\frac{1}{7200}$ x \$82,400 = \$11.50.

To recapitulate, the conditions for using currently available old data are given below:

i. A nominally described teaching-learning program divided into various versions of each sub-unit, each of which can be separately described and analyzed.

ii. A nominal time span for completing i and each sub-unit of i.

iii. A cost associated with providing each sub-unit of i.

iv. A student performance scoring procedure.

v. A transform for relating current scoring procedures to previous scoring procedures.

vi. Data on the median life-cycle earnings for the group of individuals who have previously completed i. The data are sub-classified according to date of entering productive activity, and for each year of experience.

vii. Data on the median life-cycle earnings for the group of individuals with the same initial characteristics as the group in vi, but who have not gone through i. The data are sub-classified according to date of entering productive activity, and for each year of experience.

viii. A transform which converts the data given by vi and vii into expected (or future) life-cycle earning data.

ix. A transform which converts median expected life-cycle earnings into expected life-cycle earnings for individuals with different school performance records.

x. A transform which converts dollar earnings reported for one year into equivalent dollar values of any other year.

xi. Data on the probability of survival for individuals who do and for those who do not go through i.

xii. Proportionality factors which indicate the part that each subunit contributes to subsequent overall productive output.

xiii. An estimation of the total time required for the individual student to complete i.

Using the transforms suggested above for this case, the ${}_nX_{ij\ell}$ for a student with (g, t) is:

$$
\begin{aligned}
{}_nX_{ij\ell}(g,t) = \frac{T'(\ell)}{T} \sum_{m=1}^{b-a-\frac{tT}{T'}} & \left[w\left(h(g'),\ p\left(\hat{y},\ \frac{CPI(y)}{CPI(y'+m)}\ \hat{\$}(y',m)\right)\right)\right] \cdot \\
& \cdot \left[\left\{\frac{1}{1+r}\right\}^{\frac{tT}{T'}+m-\frac{1}{2}} \right] \left[\hat{M}\left(a+m-\tfrac{1}{2}+\frac{tT}{T'}\right) \right] \\
& - \frac{T'(\ell)}{T} \sum_{m=1}^{b-a} \left[p\left(\overset{*}{y},\ \frac{CPI(y)}{CPI(y'+m)}\ \overset{*}{\$}(y',m)\right)\right] \left[\left\{\frac{1}{1+r}\right\}^{m-\frac{1}{2}} \right] \cdot \\
& \cdot \left[\overset{*}{M}(a+m-\tfrac{1}{2}) \right] \\
& - \left[D(t,j,\ell) \right] \left[\left\{\frac{1}{1+r}\right\}^{\tau-\frac{t}{2}} \right]
\end{aligned}
$$

where

$$\hat{y} = y + \frac{tT}{T'} - \tau$$

It is appropriate, at this juncture, to examine how the data for the right-hand side of the above expression can be obtained.

SECTION V

DATA ON THE OUTPUT OF ENGINEERING EDUCATIONAL SYSTEMS

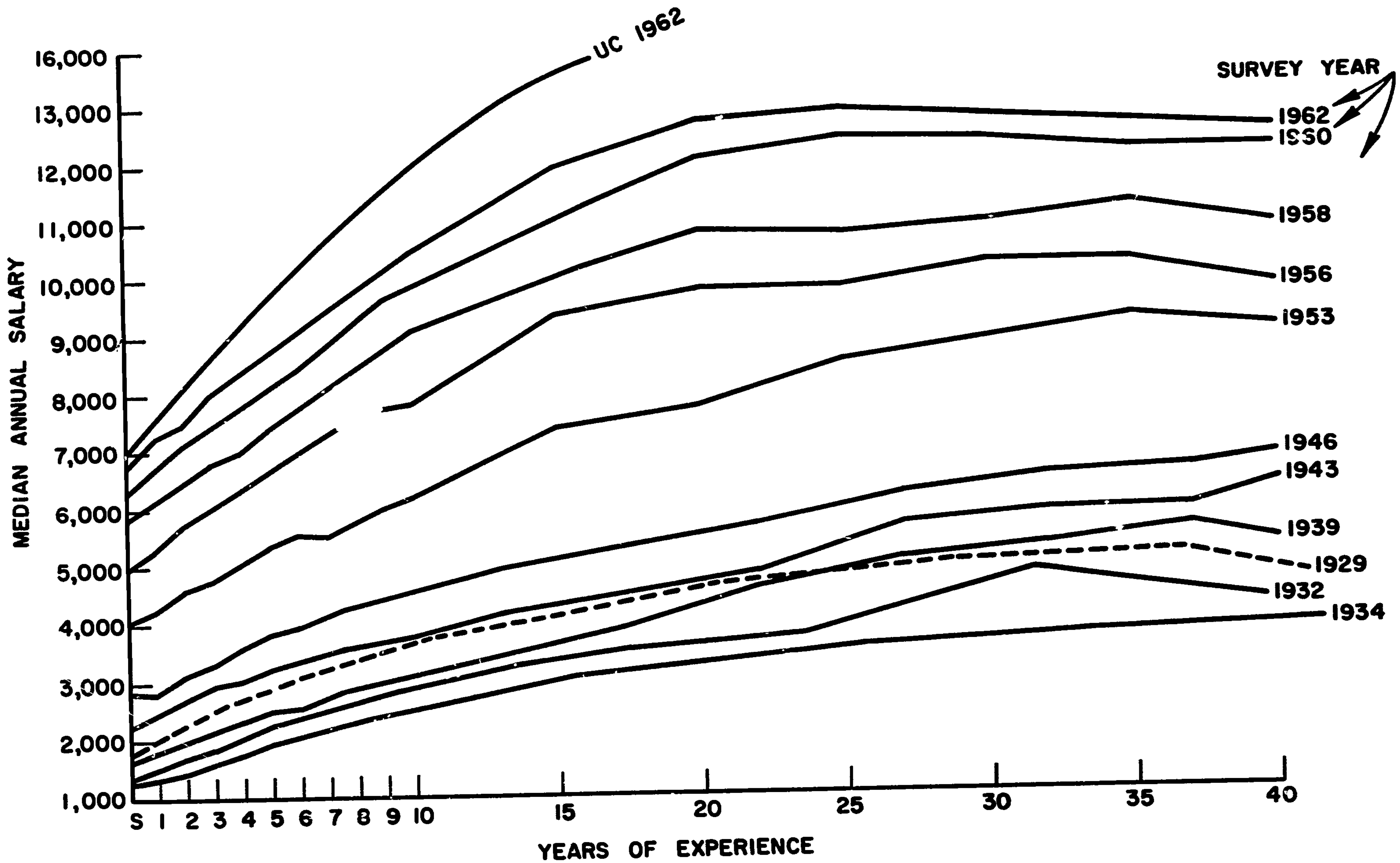
It has been pointed out in Section IV that the formula for obtaining ${}_nX_{ij\ell}(g, t)$ from existing old data would probably be most appropriate for educational or training situations which impart knowledge and skills that are direct use in later professional practice. Engineering education qualifies as such a teaching-learning situation. Furthermore, it turns out that the only group for whom relatively precise records of earnings have been kept over the past fifty-five years is the professional engineers. It is therefore within engineering education that the unique opportunity exists to immediately employ the valuation techniques described above.

A. National Data on Engineers

Engineers' salaries have been surveyed on a national basis since 1908. A composite picture of some of the survey results is shown in Figure 15. Table C-1 of Appendix C gives detailed information on the sources of earnings data and mentions the adjustments that have to be made in order to reconcile data from different sources. Also indicated in Figure 15 are the 1962 median salaries of engineering graduates from the University of California (Berkeley and Los Angeles Campuses).†

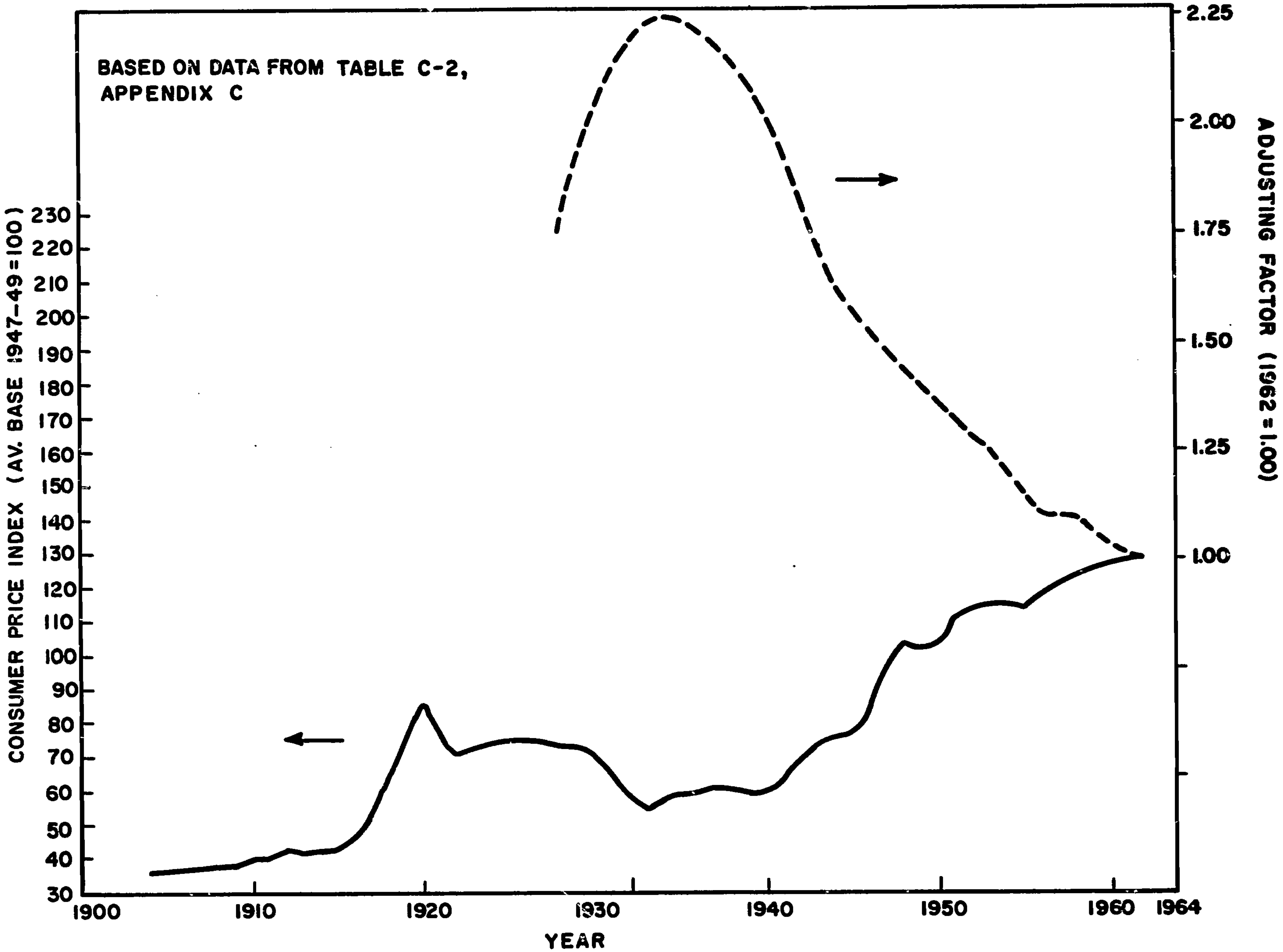
The salaries shown in Figure 15 are not directly comparable, since the purchasing power of the dollar changed during the reported period. Consumer Price Index figures and Adjusting Factors for different years are shown in Figure 16. In using the Consumer Price

†From unpublished data, University of California Engineering Graduate Study, courtesy of H. W. Case, William LeBold and William Diemer.



MEDIAN ANNUAL SALARIES FOR ENGINEERS

FIGURE 15



CONSUMER PRICE INDEX AND DOLLAR ADJUSTING FACTOR

FIGURE 16

Index, and the Adjusting Factor derived from it, one should be aware that the adjustment is approximate, since engineers' earnings tend to be higher than that of the urban moderate-income family whose living costs the C.P.I. is designed to measure.

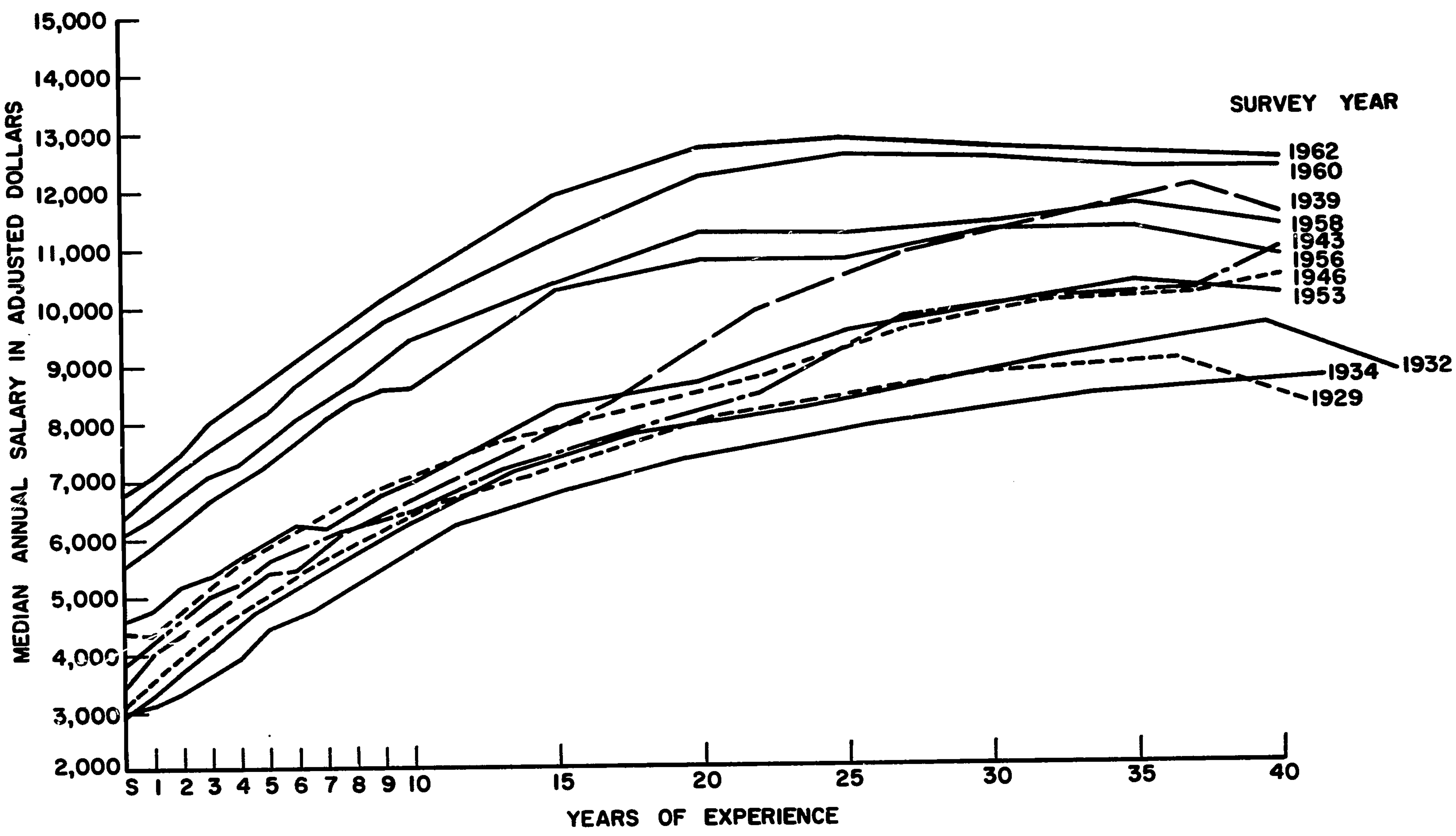
Figure 17 shows the reported salaries adjusted to 1962 equivalent dollars. Comparison of Figures 15 and 17 reveals that real purchasing power has increased less dramatically than dollar earnings.

Observe that Figures 15 and 17 show the median salaries versus years of experience for the different survey years. Not directly shown is the income of, say, the engineers who graduated in 1953. Their salaries are shown at zero years of experience on the 1953 curve and at five years of experience on the 1958 curve. By picking the data from the existing survey curves, life-cycle data for engineers who graduated in different years can be obtained. The unadjusted life-cycle earnings are shown in Figure 18. The adjusted life-cycle earnings are shown in Figure 19.

Earnings are seldom shown in this form, but this is the form needed for comparing life earnings of engineers who graduate at different times and is also necessary for projecting expected life earnings of graduates, of, say, the 1962 class. Shown in Figure 19 are the projected life-cycle earnings of the 1962 graduate. A middle, high, and a low estimate are indicated.

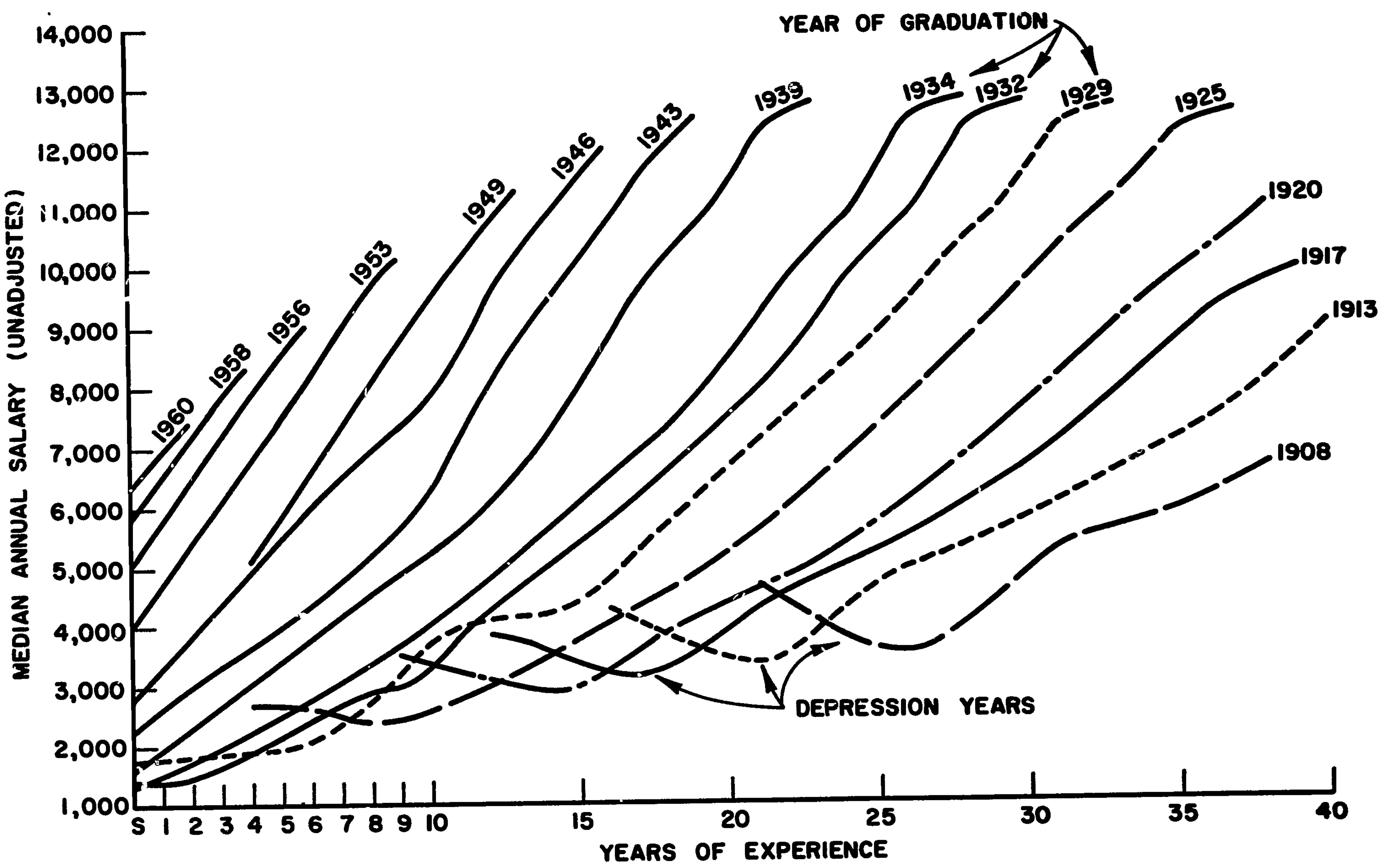
Based on the projections shown in Figure 19, the total expected life earnings for the "average" engineer graduating in 1962 is approximately $579,000. The present worth of the expected life earnings, adjusted for mortality† and discounted at different rates (3%, 4½%, and 6%), is shown in Figures 20 and 21. Figure 20 shows

† See next page.



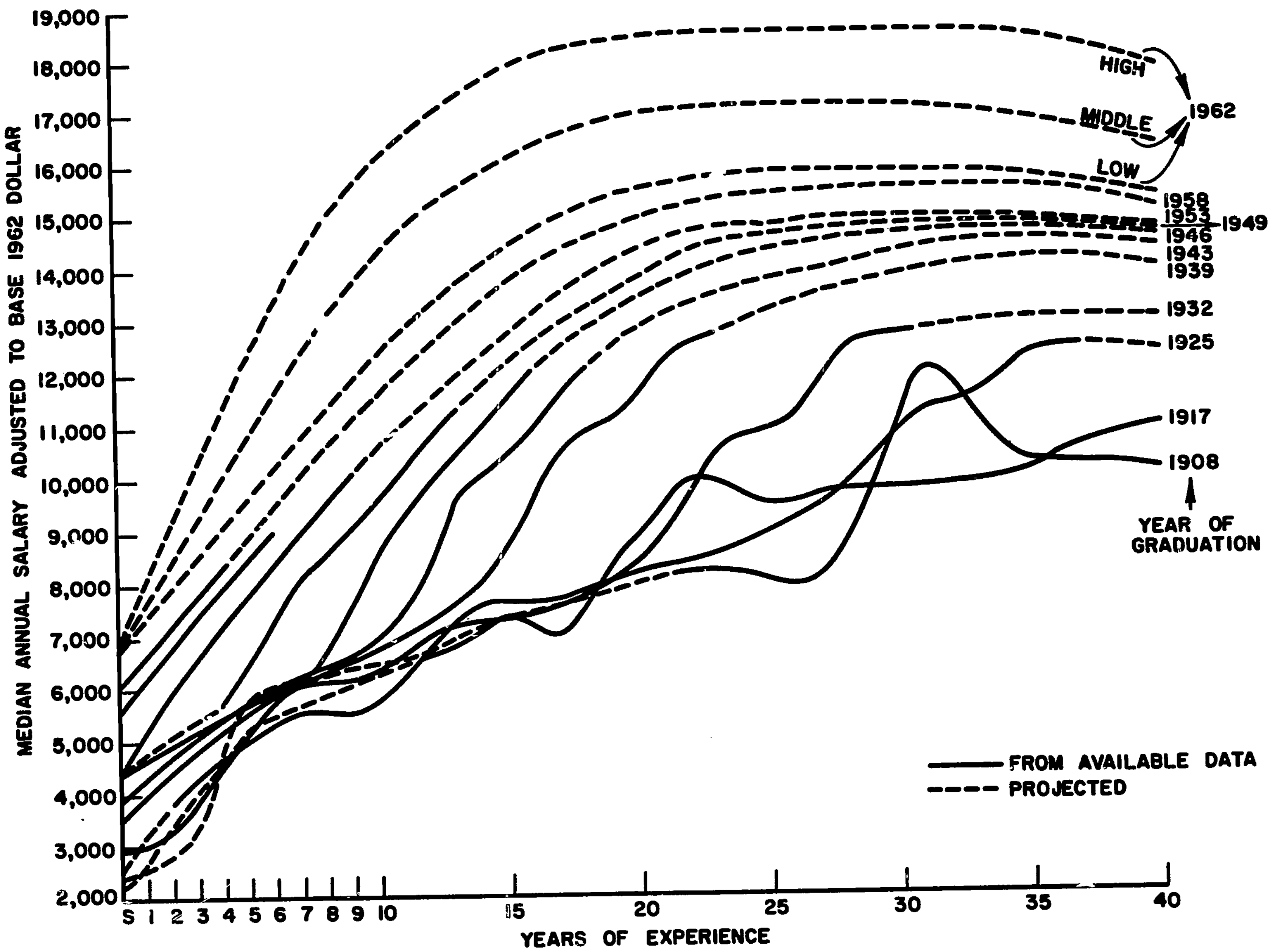
ADJUSTED MEDIAN ANNUAL SALARIES FOR ENGINEERS

FIGURE 17



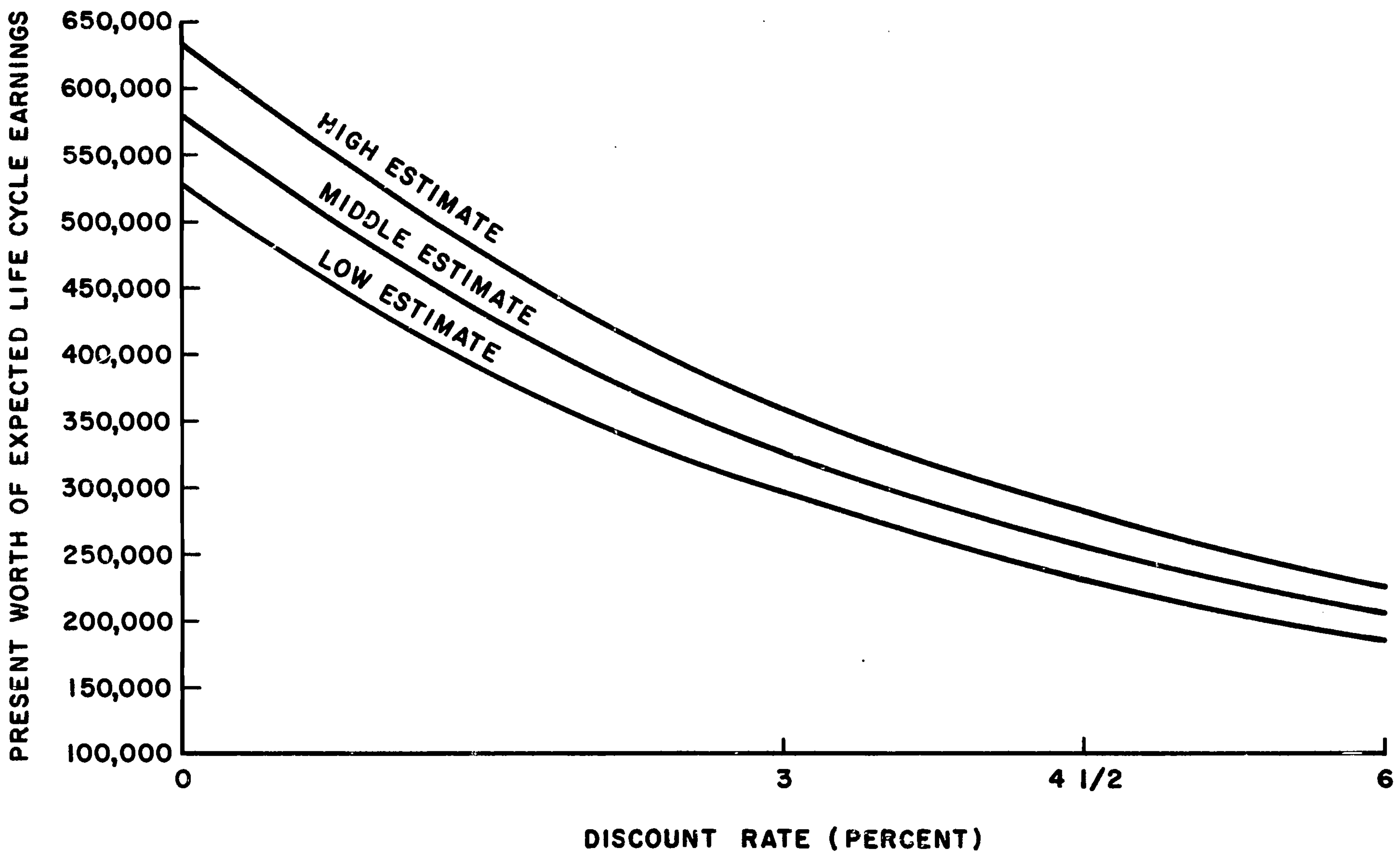
LIFE-CYCLE EARNINGS FOR ENGINEERS

FIGURE 18



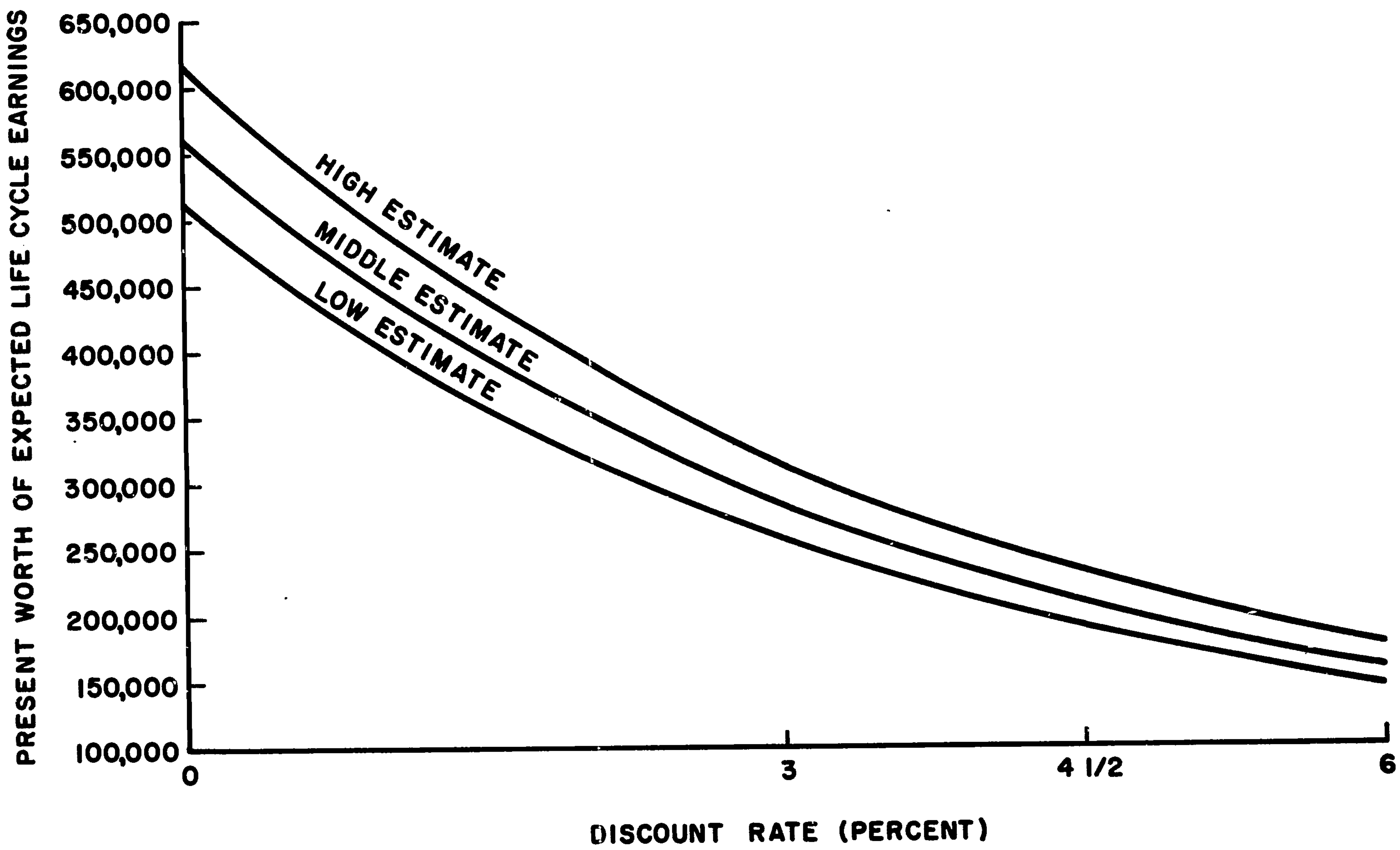
ADJUSTED LIFE-CYCLE EARNINGS FOR ENGINEERS

FIGURE 19



PRESENT WORTH AT AGE 22 OF EXPECTED LIFE-CYCLE EARNINGS FOR ENGINEERS

FIGURE 20



PRESENT WORTH AT AGE 18 OF EXPECTED
LIFE-CYCLE EARNINGS FOR ENGINEERS

FIGURE 21

the present worth at age 22, the supposed age of graduation from engineering school. Figure 21 shows the present worth at age 18, the supposed age at which a high school graduate would choose between going to engineering school or going to work.

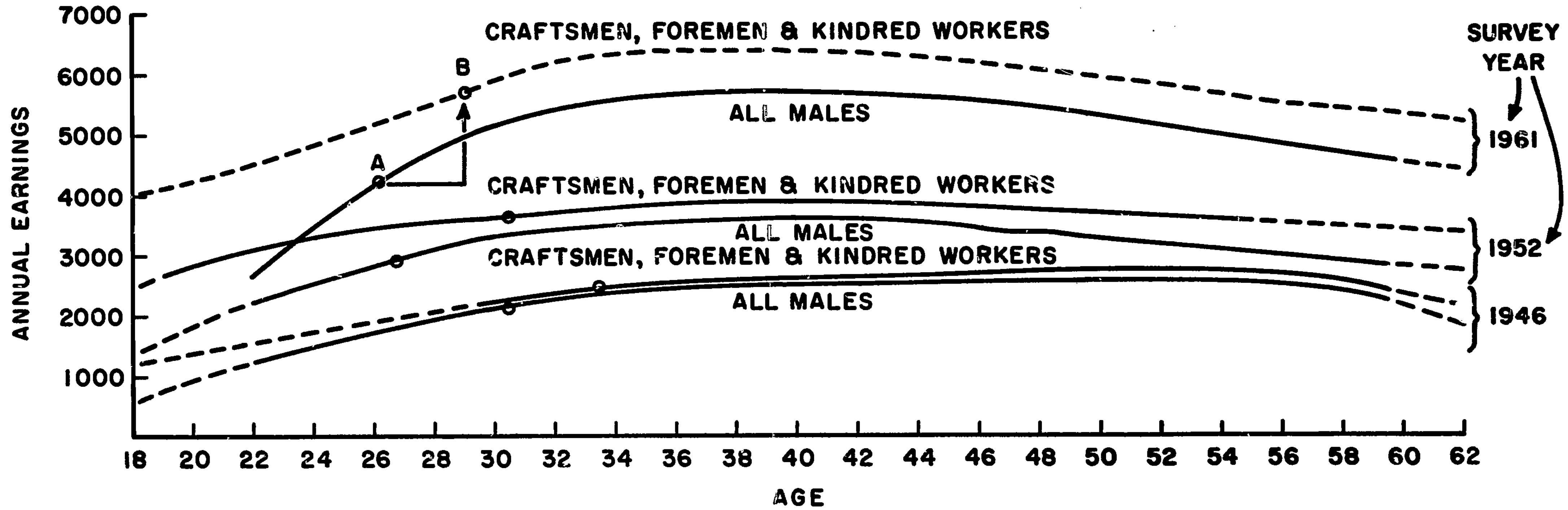
B. National Data on Comparison Group

On the assumption that an income somewhat more than the national median income would be earned by the high school graduate who had the ability to enter engineering school but instead chose to work, the median salary of craftsmen, foremen, and kindred workers was selected for comparison purposes. For this group, national salary surveys in relation to age are available for two years -- 1946 and 1951. For other years, only the overall median salary is reported. However, median income figures for all males by age are available, and these are used as shape curves to derive the salary curves for craftsmen, foremen, etc. The available data are given in Tables C-4 and C-5 of Appendix C.

Figure 22 shows two curves for each of the survey years 1946 and 1951. Notice that the salary curves for craftsmen, foremen, etc., closely parallel the income curve of all males except at the extremes where the latter curve drops off rapidly. Another observation is that the median earnings for craftsmen, etc., occurs at an age three years later than the median earnings for all males. Bearing these facts in mind, one can derive the salary curve for craftsmen, etc., for say, the year 1961 as follows: The earning curve for that year

---

† See Table C-3, Appendix C for sources of information and calculations of survival factor. Note that no adjustments were made for school attrition and rate of unemployment. The effect of unemployment is reflected in the basic data on median salaries, and it is assumed that the undetermined rate of unemployment remains constant.

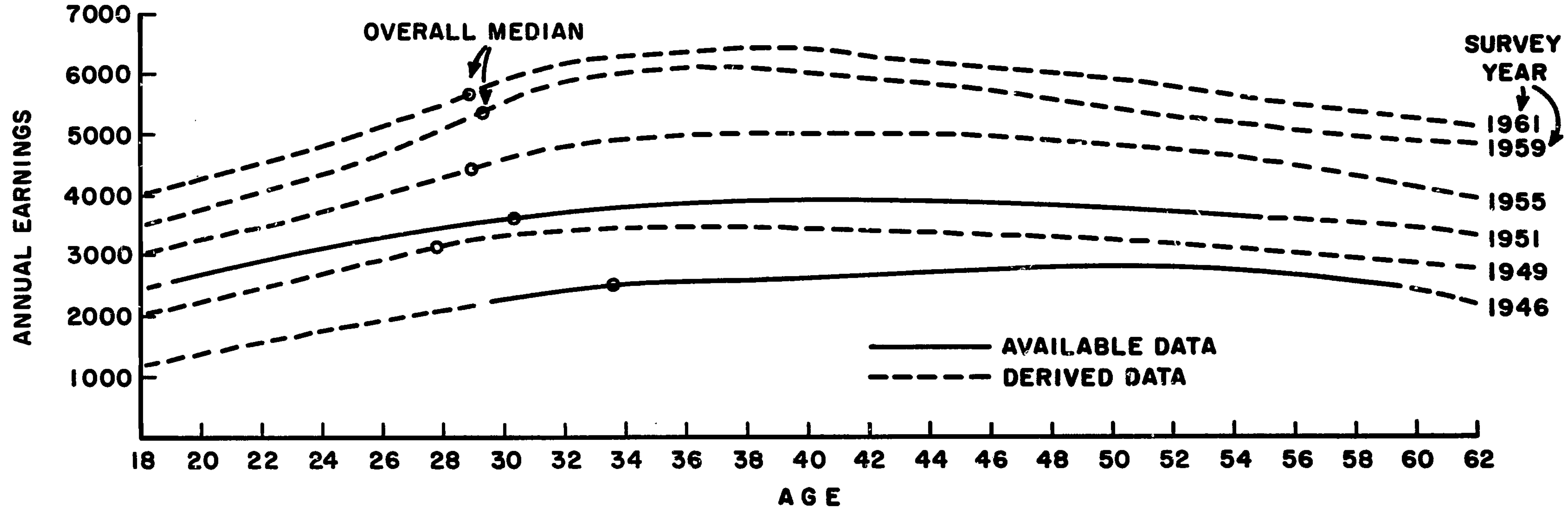


MEDIAN ANNUAL EARNINGS FOR COMPARISON GROUP

FIGURE 22

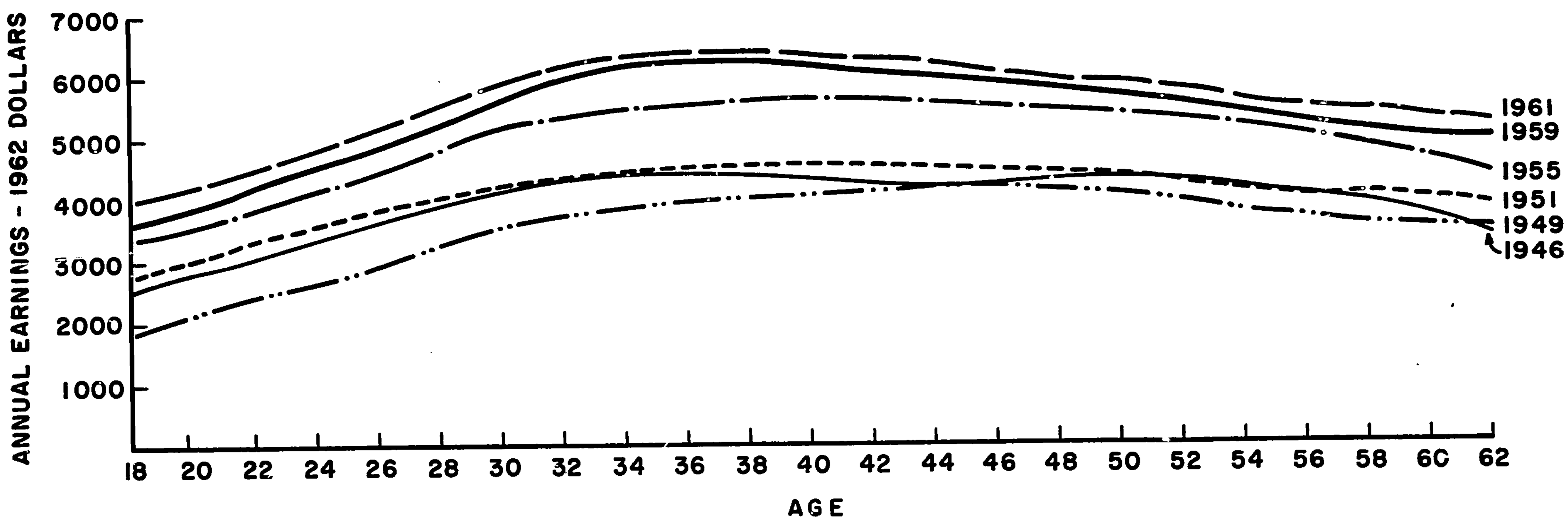
for all males is plotted; the overall median is located at Point A; the known overall median salary for craftsmen, etc., is located 3 years later at Point B; then a curve is drawn such that it passes this Point B and is parallel to the all-male's earning curve in the middle of the distribution and drops off only gradually at the extremes. This is the derived curve for craftsmen, etc., for the year 1961. Similar curves are drawn for years 1949, 1955, and 1959, and all these are shown in Figure 23. The curves shown in Figure 23 are not directly comparable since the dollar value is not the same over the years. Using the adjusting factors based on the Consumer Price Index, adjusted earnings were obtained and plotted in Figure 24. Figures 23 and 24 show median salaries versus years of experience for the survey years shown. As in the case of engineers, data from these curves were used to obtain life-cycle curves for craftsmen, etc., who finished high school in different years. Figure 25 shows the unadjusted life-cycle earnings, and Figure 26 shows the life-cycle earnings adjusted to 1962 equivalent dollars. Earnings of skilled workers are rarely shown in this form. Some previous efforts to derive craftsmen's life-cycle curves, such as done by DeHaven [46] and by Stewart [47] have been based on the assumption that beyond the apprenticeship period craftsmen income remains fairly constant. The life-cycle earnings curve of construction workers deemed by both Stewart and DeHaven to be representative of high school graduates who chose to work rather than go to engineering school is also shown in Figure 25. Figure 26 includes projections for the various years and three estimates (high, middle, and low) for the 1958 high school graduates.

The mid-estimate of the total expected life-earnings for an average skilled worker graduating from high school in 1958 (presumably the bifurcation date for the engineer who graduated from college



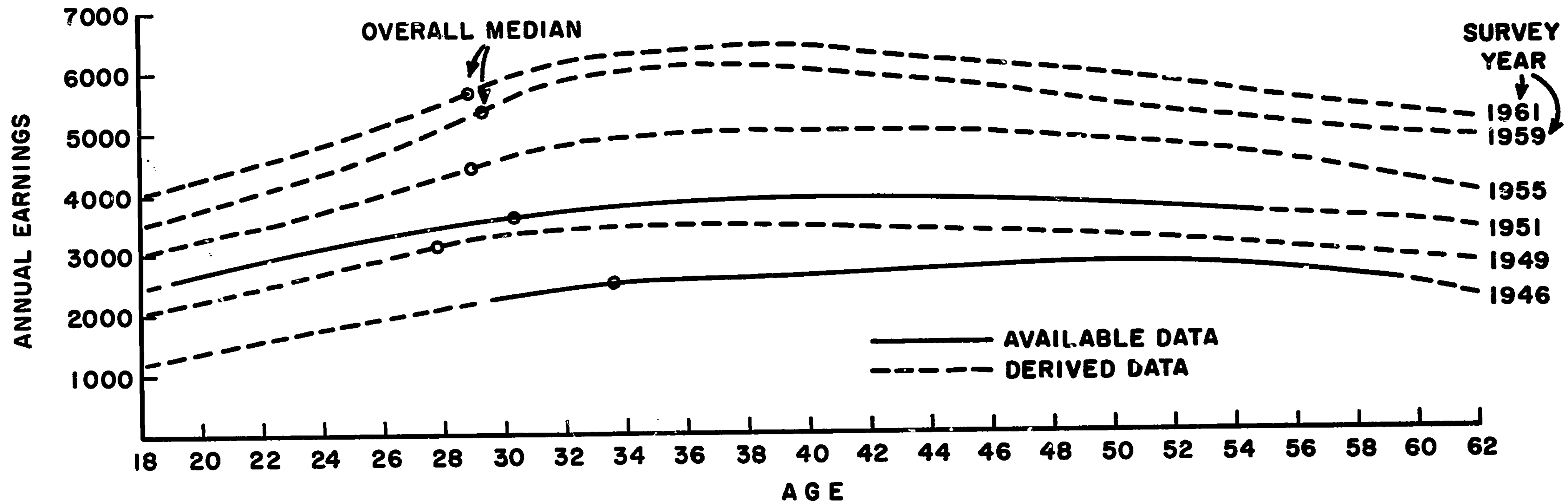
DERIVED MEDIAN ANNUAL EARNINGS
FOR COMPARISON GROUP

FIGURE 23



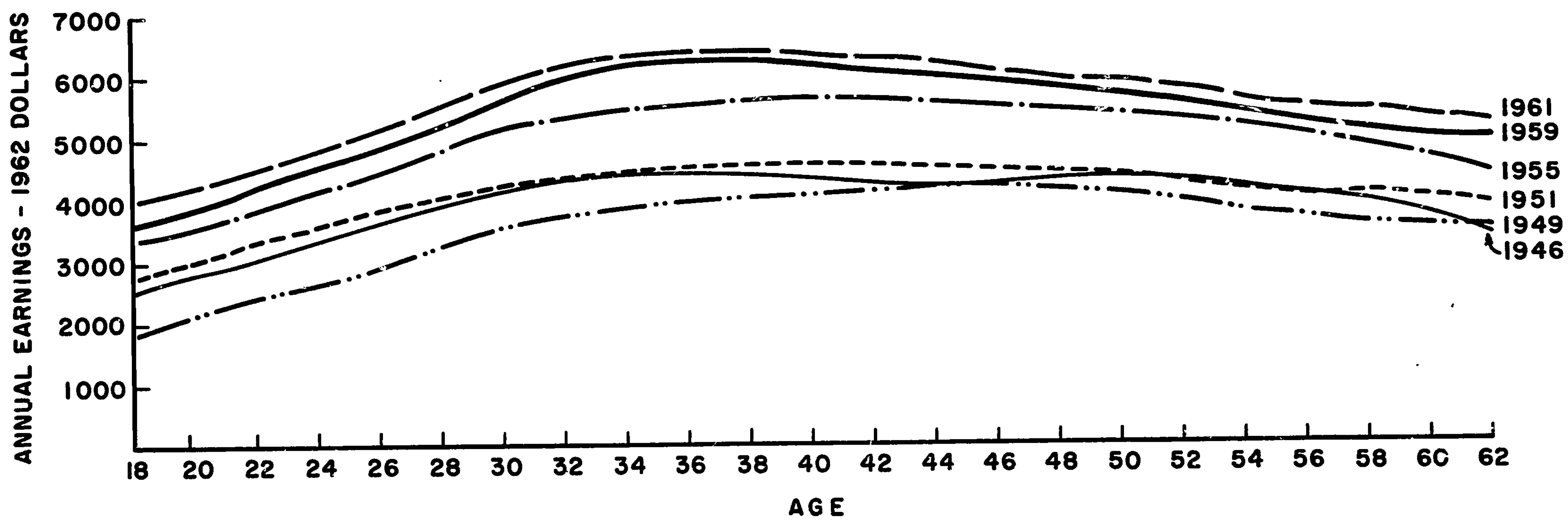
ADJUSTED MEDIAN ANNUAL EARNINGS
FOR COMPARISON GROUP

FIGURE 24



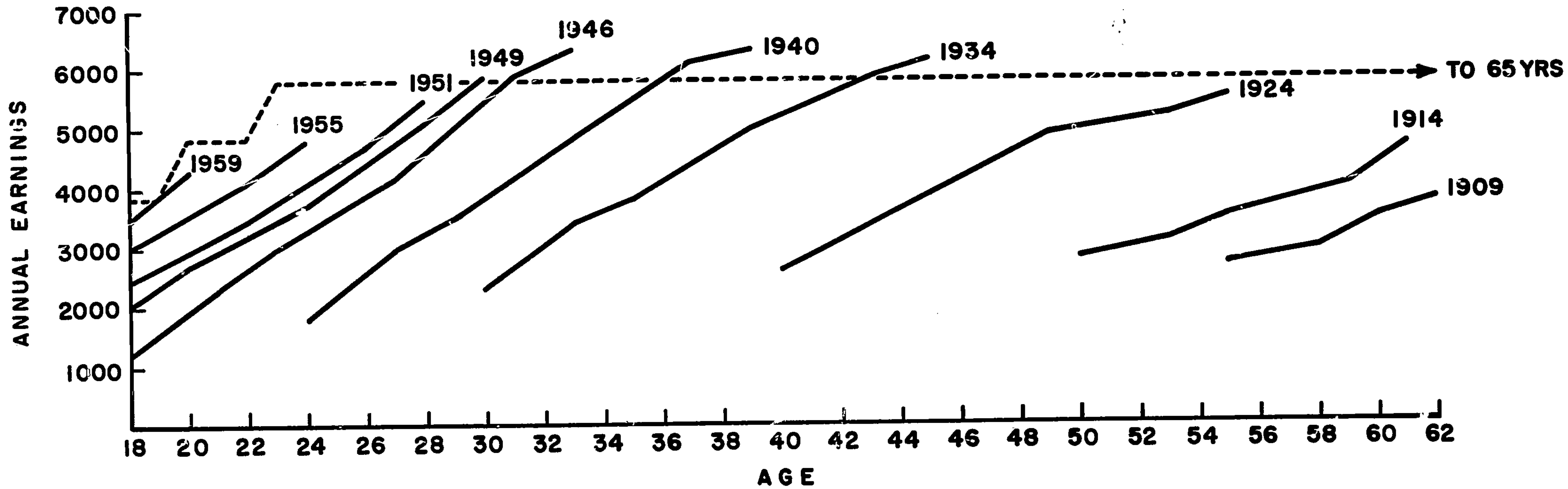
DERIVED MEDIAN ANNUAL EARNINGS
FOR COMPARISON GROUP

FIGURE 23



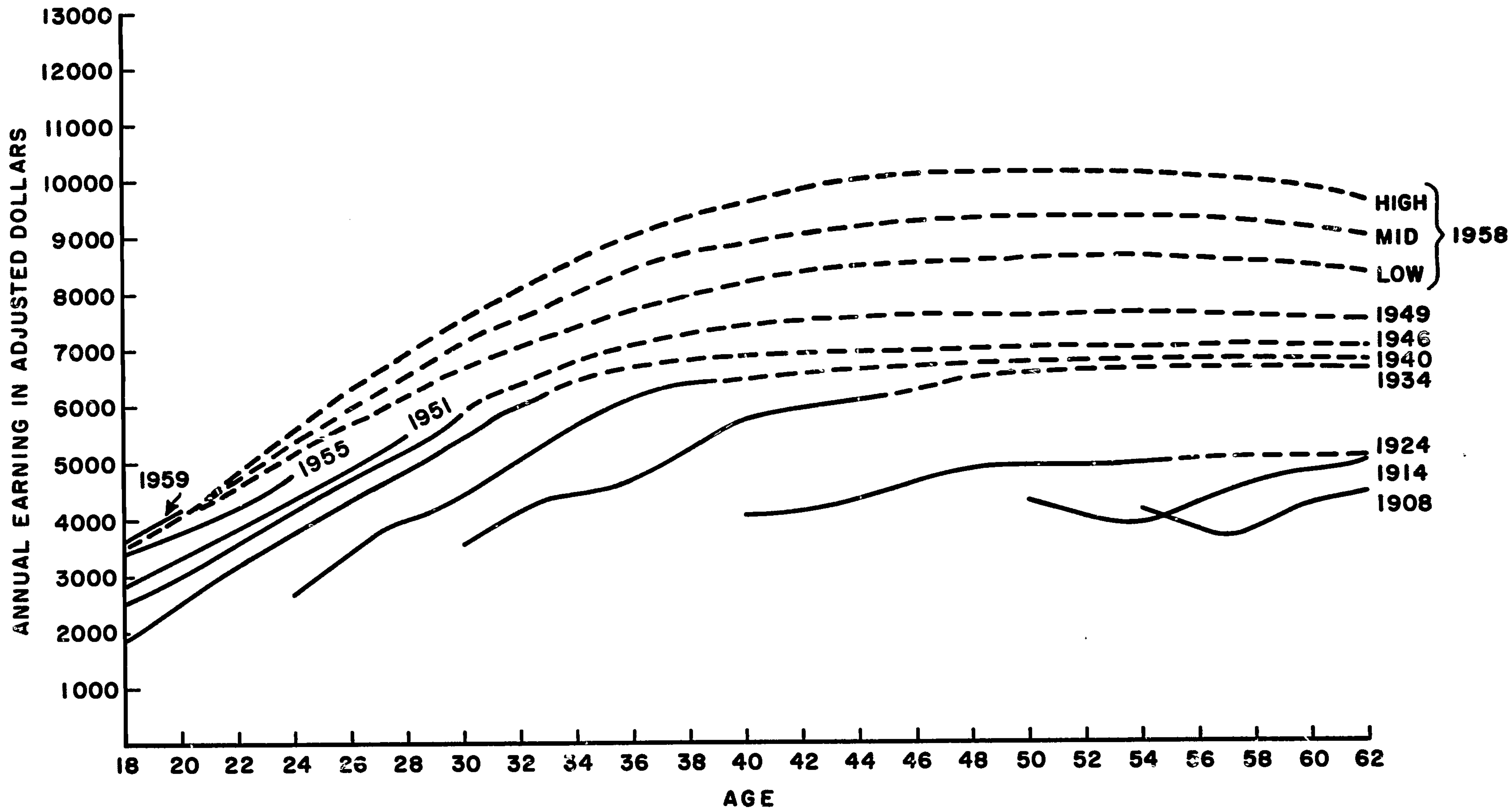
ADJUSTED MEDIAN ANNUAL EARNINGS
FOR COMPARISON GROUP

FIGURE 24



LIFE-CYCLE EARNINGS FOR COMPARISON GROUPS

FIGURE 25



ADJUSTED LIFE-CYCLE EARNINGS
FOR COMPARISON GROUP

FIGURE 26

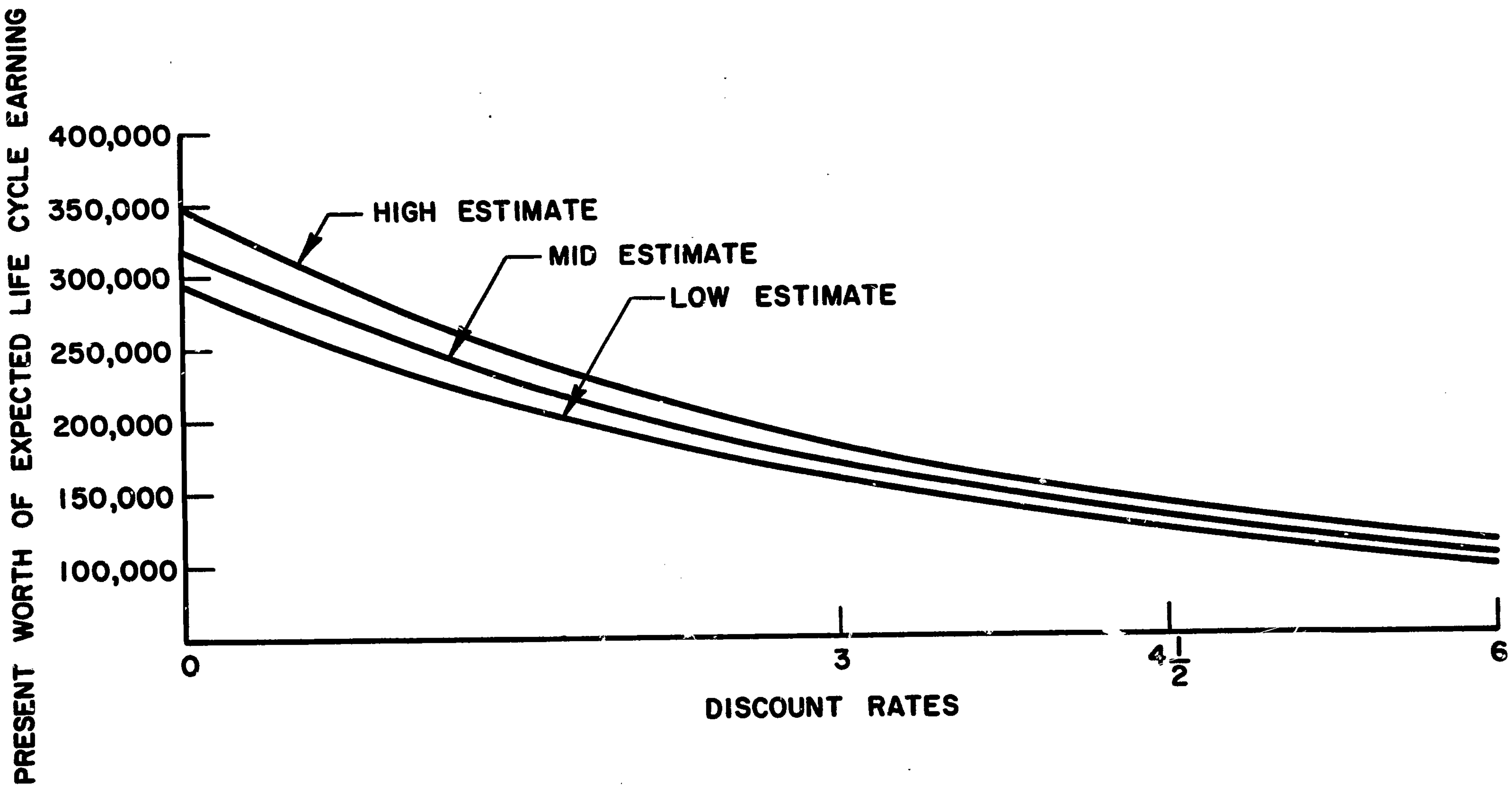
in 1962) is approximately $317,000. The present worth at age 18 of this expected life earnings, adjusted for mortality and discounted at different rates, is shown in Figure 27.

Room, board, transportation, and incidental expenses involved in the cost of an engineering education are excluded from the computation in this analysis. The assumption is made that the value of these items will be approximately the same for engineering students and the comparison group of working craftsmen.

Also, earnings foregone by students are not included, since the effect of the foregone earnings shows up in the calculation of the difference in the present worth of the expected life earnings for engineers and craftsmen. However, where students work part-time while going to school, part-time earnings should be included in the calculations.

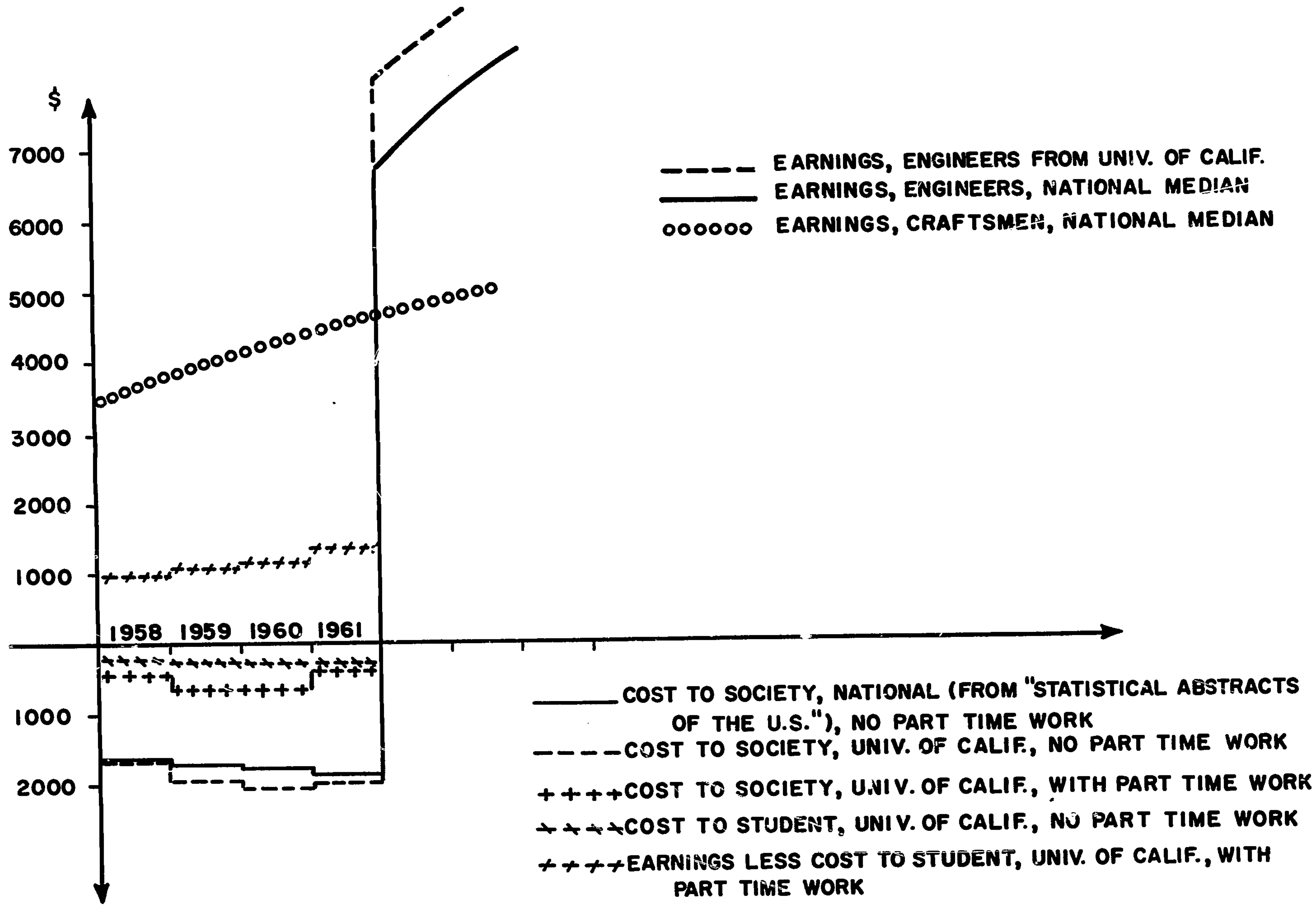
The primary concern here is with the total cost of education, including those costs borne directly by the student and those costs defrayed from public or private sources. Such costs are labelled "cost to society" to differentiate them from the personal cost to the student or his family. Typical costs and earnings are illustrated (to scale) in Figure 28. Figure 29 compares the present worth of expected life earnings and educational costs for engineers and craftsmen, at different discount rates. This figure shows a difference in the total expected life-cycle earnings of the (1962 graduate) engineer and the craftsmen amounting to approximately $236,000. At a discount rate of 4½% this difference shrinks to $73,000. The intersection of the two curves indicates that the internal rate of return on an engineering education would be approximately 17%.

At this point enough data have already been presented to perform some interesting macro-system studies. For example, if



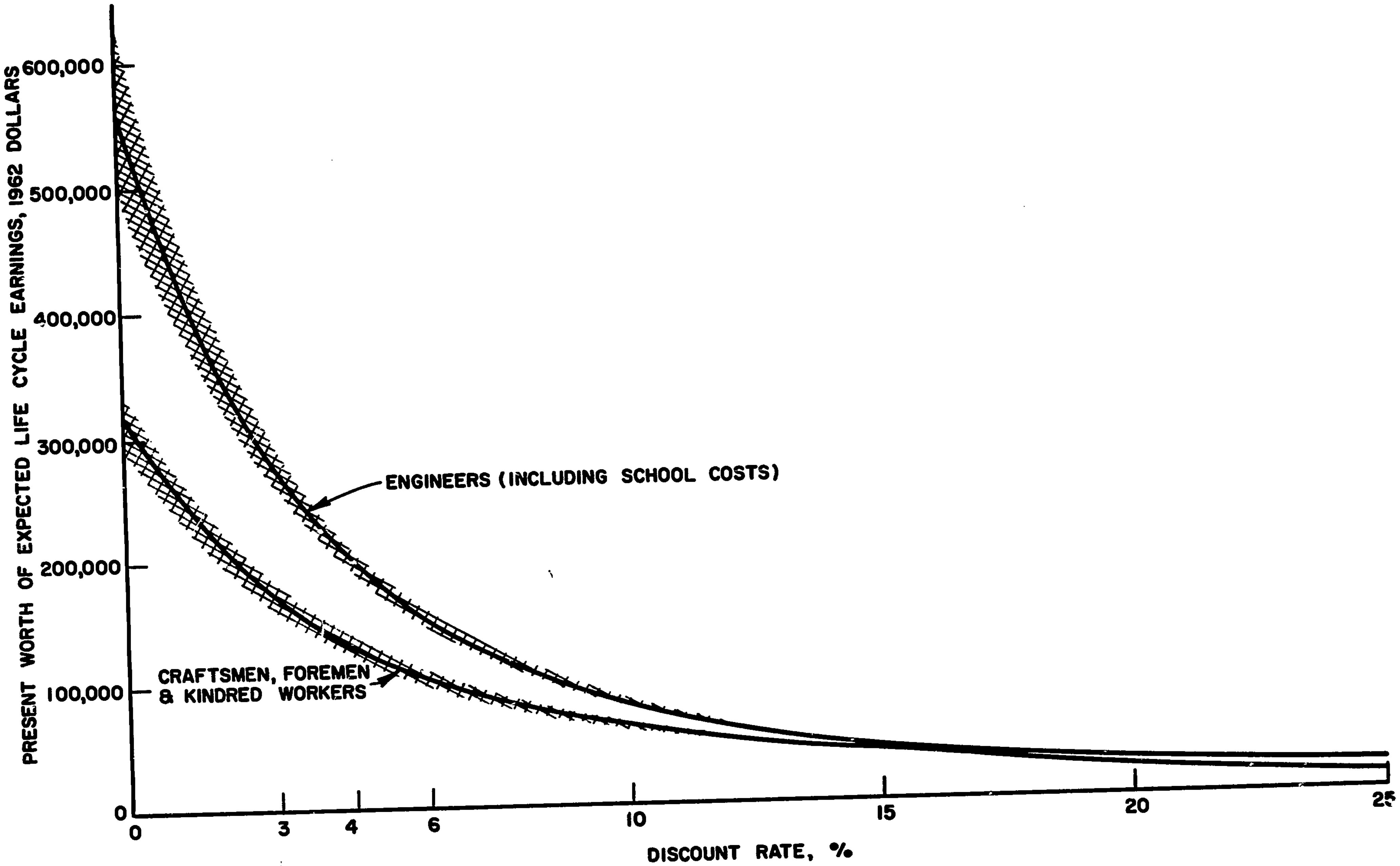
PRESENT WORTH AT AGE 18 OF EXPECTED LIFE-CYCLE EARNINGS FOR COMPARISON GROUP

FIGURE 27



EDUCATIONAL COSTS

FIGURE 28



PRESENT WORTHS OF EXPECTED LIFE-CYCLE EARNINGS FOR ENGINEERS AND COMPARISON GROUP

FIGURE 29

one wished to keep a specified difference between the expected present worth of an engineer's and a craftsman's life-cycle earnings and also maintain the same level of performance while the engineer is in college, but decrease the learning period from four years to three years, how much more could one afford to pay in educational costs? Or:

$$X(g, 4) \equiv X(g, 3)$$

$$\hat{W}(4) - \overset{*}{W} - V(4) = \hat{W}(3) - \overset{*}{W} - V(3)$$

$$\therefore\ V(3) - V(4) = \hat{W}(3) - \hat{W}(4)$$

$$\sum_{\tau=1}^{3}\left[D''(\tau)\right]\left[\left\{\frac{1}{1+r}\right\}^{\tau-\frac{1}{2}}\right] - \sum_{\tau=1}^{4}\left[D'(\tau)\right]\left[\left\{\frac{1}{1+r}\right\}^{\tau-\frac{1}{2}}\right]$$

$$= \sum_{m=1}^{b-a-3}\left[p\left(y-1, \frac{CPI(y)}{CPI(y'+m)}\ \$(y', m)\right)\right]\left[\left\{\frac{1}{1+r}\right\}^{3+m-\frac{1}{2}}\right] \cdot$$

$$\cdot\left[\hat{M}(a+m-\frac{1}{2}+3)\right]$$

$$- \sum_{m=1}^{b-a-4}\left[p\left(y, \frac{CPI(y)}{CPI(y'+m)}\ \$(y', m)\right)\right]\left[\left\{\frac{1}{1+r}\right\}^{4+m-\frac{1}{2}}\right] \cdot$$

$$\cdot\left[\hat{M}(a+m-\frac{1}{2}+4)\right]$$

In a numerical solution, using y = 1962, a = 18, b = 62, and r = 4½%, the right hand of the last equation gives approximately \$12,600, which is the expected worth of the average engineer's life-cycle income attributable to finishing school and starting to work one year earlier than is currently customary. Also, using the national average annual "cost to society", for an engineering education,

$$\sum_{\tau=1}^{4}\left[D'(\tau)\right]\left[\left\{\frac{1}{1+r}\right\}^{\tau-\frac{1}{2}}\right] \approx \$6,400 \text{ at } r = 4\tfrac{1}{2}\%$$

and

$$\sum_{\tau=1}^{3}\left[D''(\tau)\right]\left[\left\{\frac{1}{1+r}\right\}^{\tau-\frac{1}{2}}\right] \approx \$6,400 + \$12,600$$

If $D''(\tau)$ is a uniform annual figure, and $r = 4\frac{1}{2}\%$

$$D''\sum_{\tau=1}^{3}\left[\left\{\frac{1}{1+r}\right\}^{\tau-\frac{1}{2}}\right] = \$19,000$$

$$D'' = \frac{\$19,000}{2.801} = \$6,800.$$

Therefore, one could theoretically afford to spend up to \$6,800 for each of the three years in an accelerated program, as compared to approximately \$1,800 for each of the years in the normal four-year program.

Another variation of this problem is to calculate the additional amount of resources one would be willing to commit to education if these additional expenditures resulted in a student getting an M.S. instead of a B.S. degree in four years.

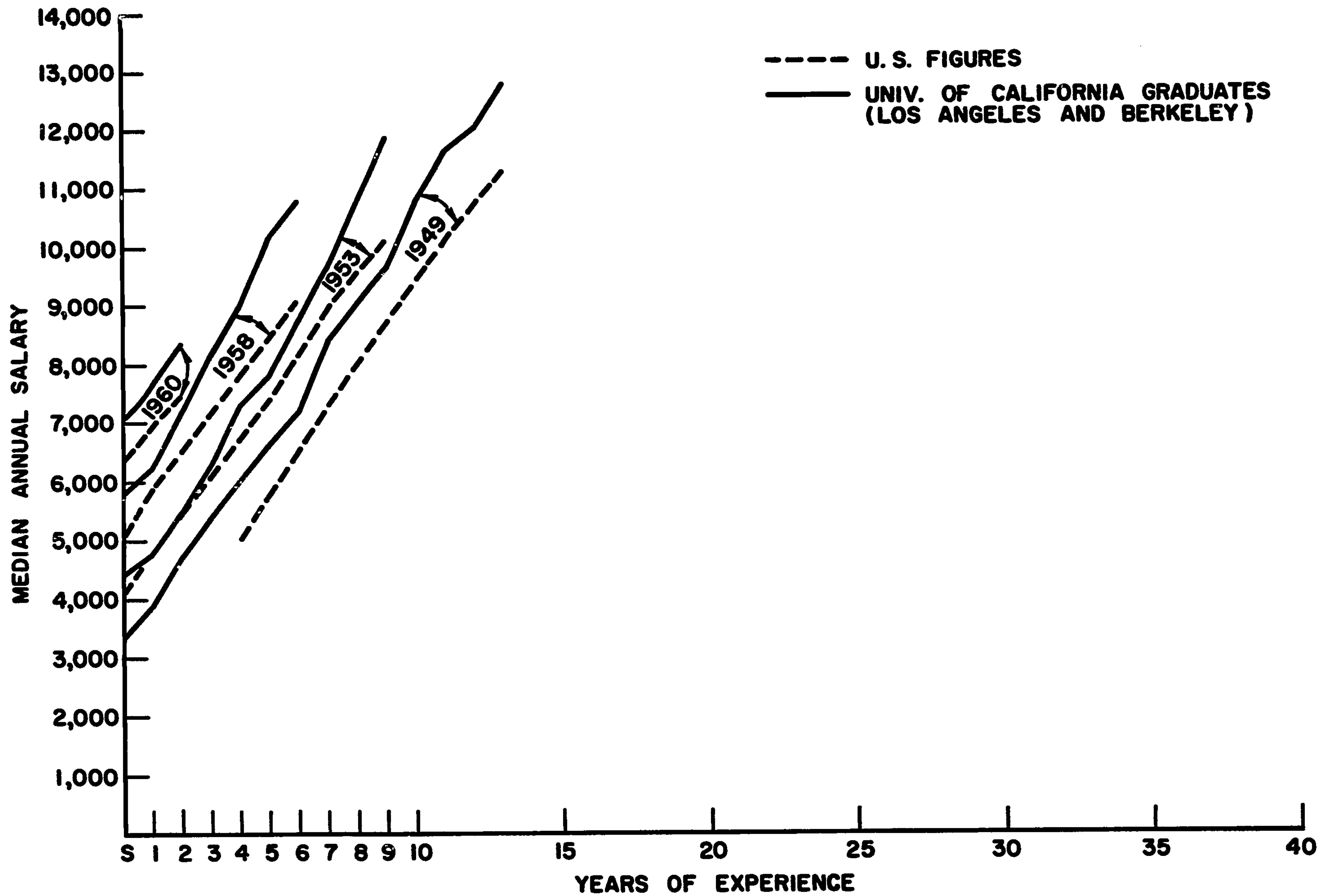
Somewhat more speculative, since it introduces the additional uncertainties of the relationship between school performance and subsequent professional performance, is the problem of calculating the additional amount of resources one would be willing to commit to education if these expenditures resulted in a student getting, say, an A average instead of a B average.

The above examples are sufficient to indicate the range of problems that could be investigated. Full treatment of such problems is left to a later work, since the primary concern here is how to use the input-output data in an adaptive decision situation.

C. University of California Data on Engineers

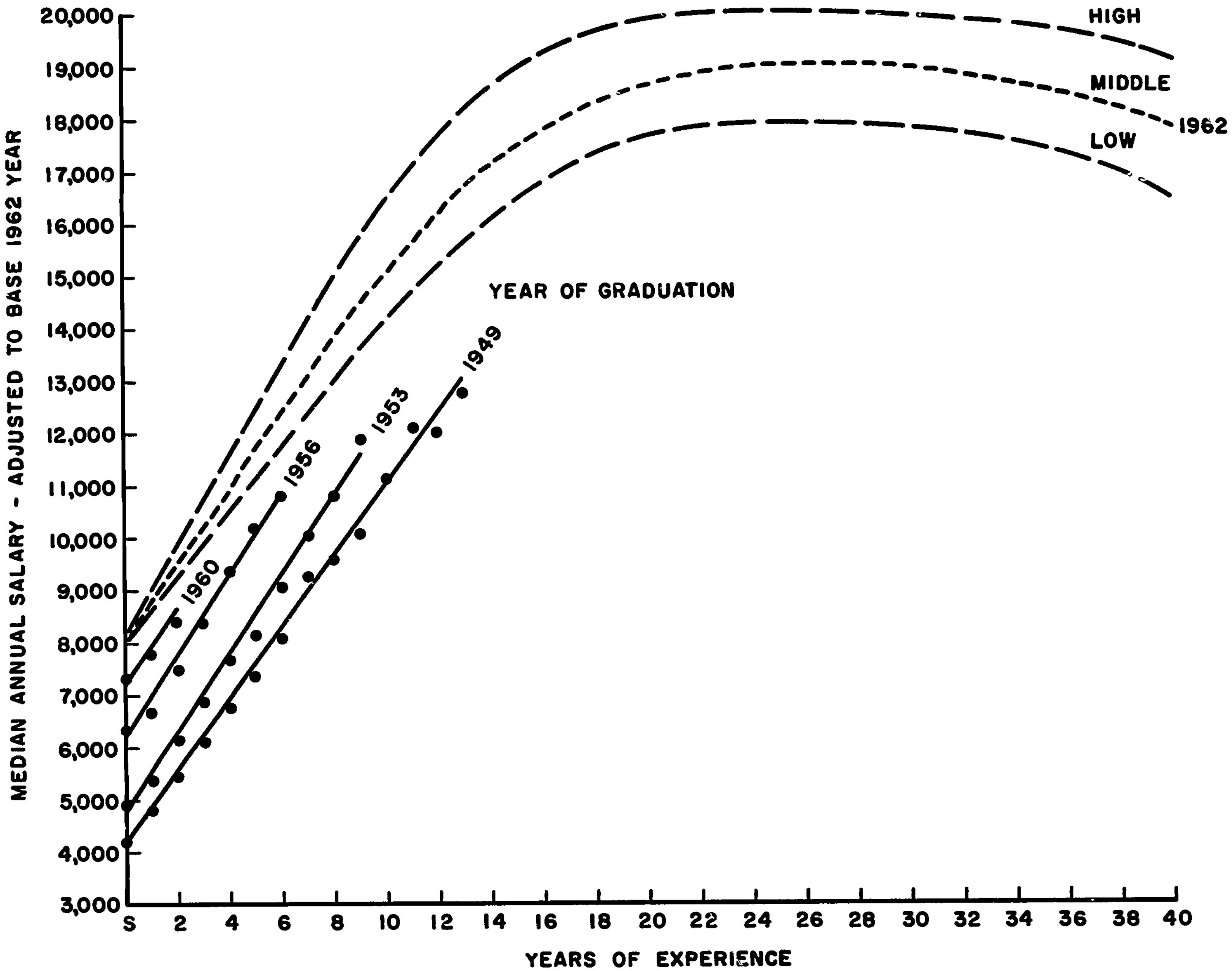
In an adaptive decision situation, one should, of course, use the data which are most relevant to the specific situation. For example, planners in an engineering school could, as a starter, use the national median earning figures for forecasting the expected life-cycle earnings of their graduates if no specific data on the earnings of graduates from that school are available. Where additional information is available it should be used. An illustration of the use of additional data is given below for the case of the graduates from the Berkeley and Los Angeles Colleges of Engineering of the University of California. A difference between the reported national median earnings of engineers and the median earnings of University of California engineering graduates for the survey year 1962 was already noted in Figure 15. A plot of the unadjusted median annual earnings by year of graduation shown in Figure 30 reveals that the University of California median figures are consistently higher than the national median. The University of California figures were adjusted for change in dollar values and re-plotted in Figure 31. Since the Los Angeles campus of the University of California had its first engineering graduates in 1949, the earning curves do not extend beyond thirteen years of experience. Therefore, the general shape of the national expected life-cycle earning curve (Figure 19) is used along with the available curves on University of California engineering graduates to project an expected life-cycle earning curve for the class of 1962.

An idealized set of performance correction factors (shown by solid lines in Figure 32) was derived from a combination of the American Telephone and Telegraph Company data (Figure 13) and the available data, covering a shorter span of years, from the University of California. Also shown in Figure 32 (by dotted lines) are the two extreme estimates for the performance correction factors, i.e., first, where it is assumed that no correlation between school performance and subsequent earnings exists, and therefore all



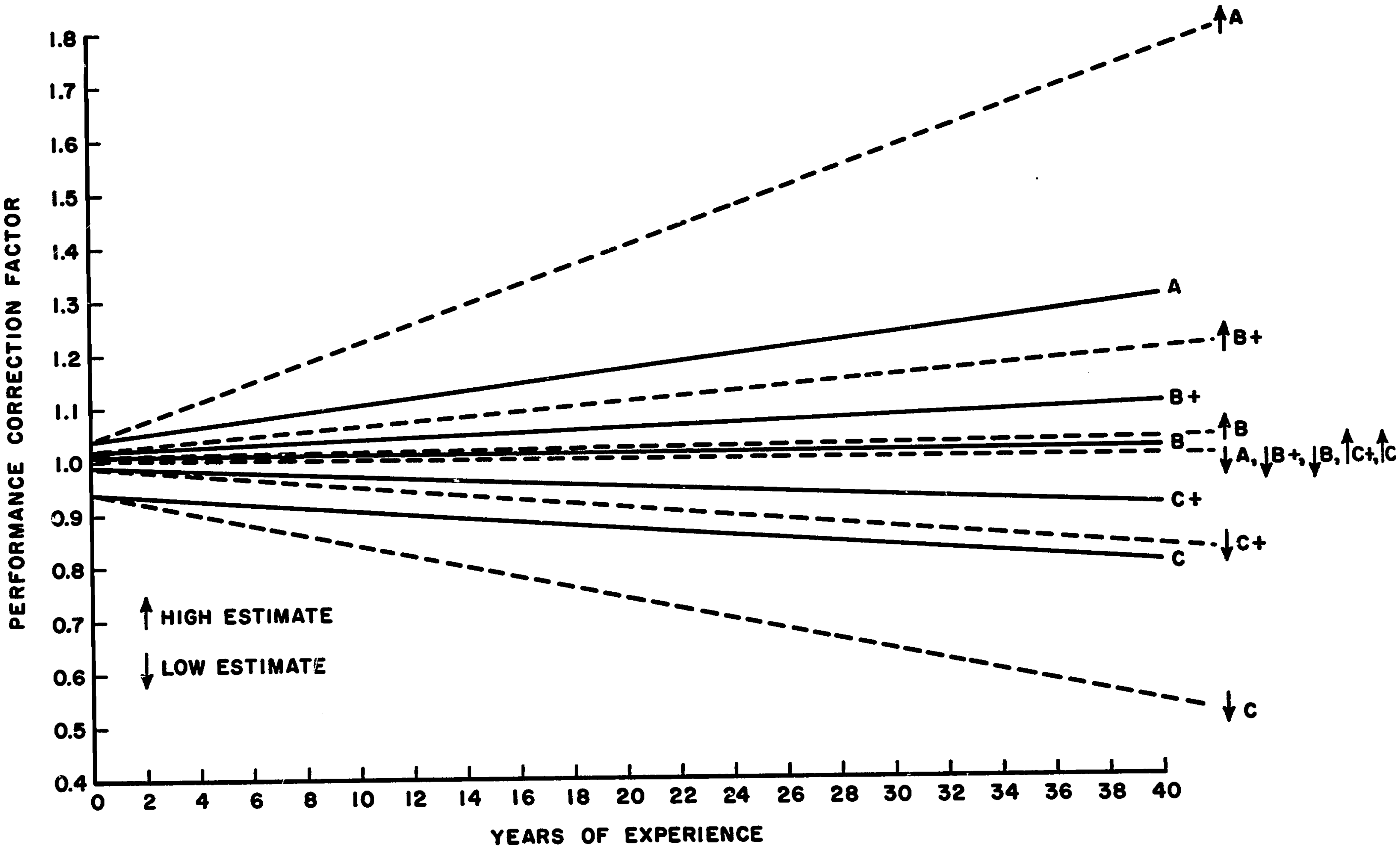
COMPARISON OF ANNUAL EARNINGS OF ENGINEERS

FIGURE 30



ADJUSTED MEDIAN ANNUAL EARNINGS FOR ENGINEERS
FROM THE UNIVERSITY OF CALIFORNIA

FIGURE 31



PERFORMANCE CORRECTION FACTORS

FIGURE 32

performance correction factors are equal to 1.0, and second, where Gifford's old results are used to estimate the correction factors.

The present worth of the expected life-cycle earnings is affected by differences in school performance scores, by the time it takes to complete the education, and by the discount rate. These three factors are used to modify the expected median life-cycle earnings for the University of California engineering graduate of the Class of 1962, and are displayed in Figures 33, 34, 35, 36. In each figure, the left-hand diagram is based on the assumption of no correlation between school performance scores and subsequent earnings, the right-hand diagram is based on Gifford's extreme performance correction factors, and the middle diagram is based on the idealized University of California performance correction factors. The shaded areas indicate the range of values between the high and low estimate of the median expected life-cycle earnings (see Figure 31).

Figures 33, 34, 35, 36 present (for the 1962 engineering graduate from the University of California) the solution for $(\hat{W} - \overset{*}{W})$ in the expression

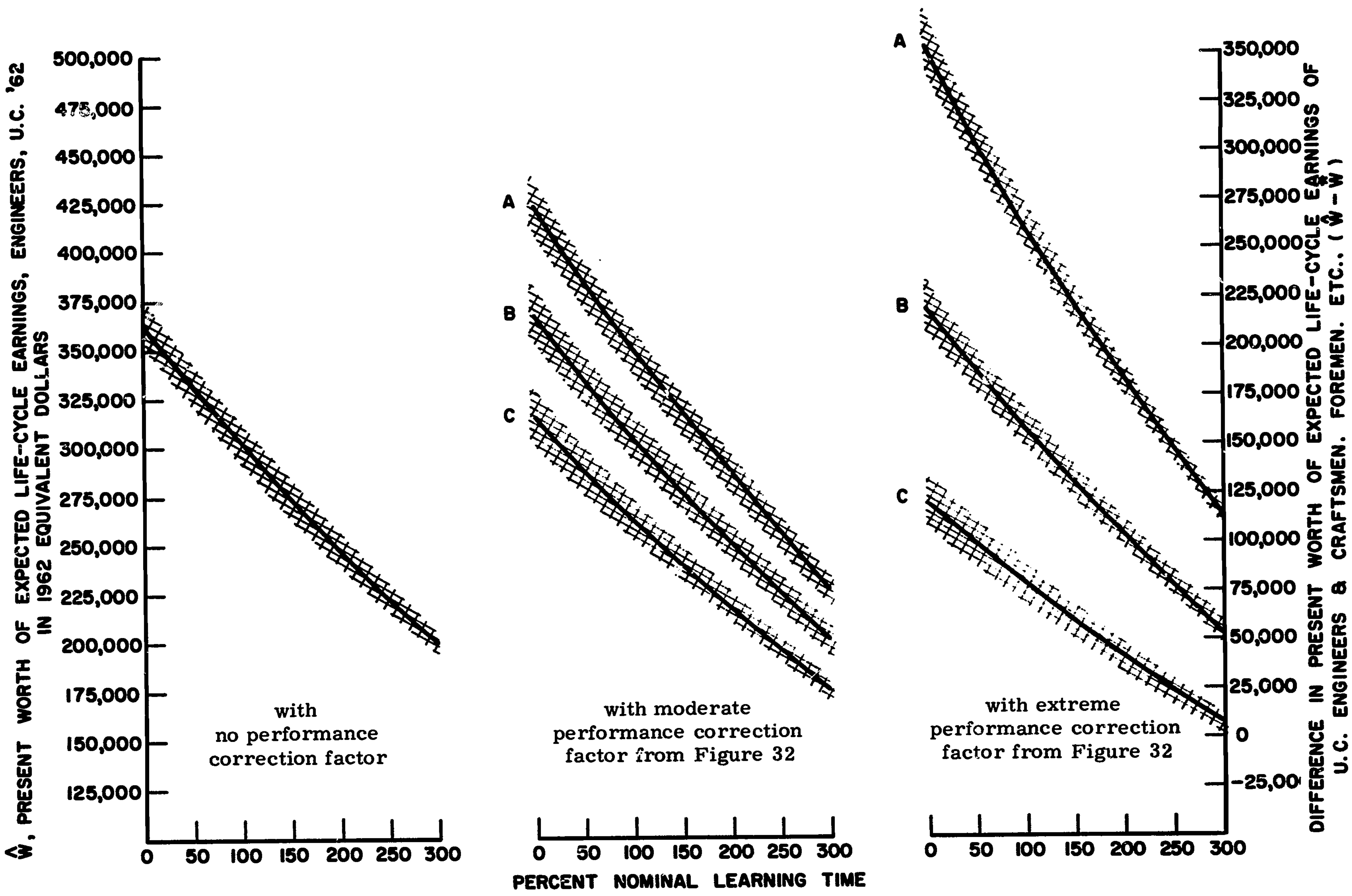
$$
{}_{n}X_{ij\ell}(g,t) = c(\hat{W} - \overset{*}{W}) - V
$$

$$
= \frac{T'(\ell)}{T} \sum_{m=1}^{b-a-\frac{tT}{T'}} \left[w\left(g, p\left(\hat{y}, \frac{CPI(y)}{CPI(y'+m)} \hat{\$}(y', m)\right)\right)\right] \left[\left\{\frac{1}{1+r}\right\}^{\frac{tT}{T'}+m-\frac{1}{2}} \right] \cdot
$$

$$
\cdot \left[\hat{M}\left(a+m-\tfrac{1}{2}+\frac{tT}{T'}\right)\right]
$$

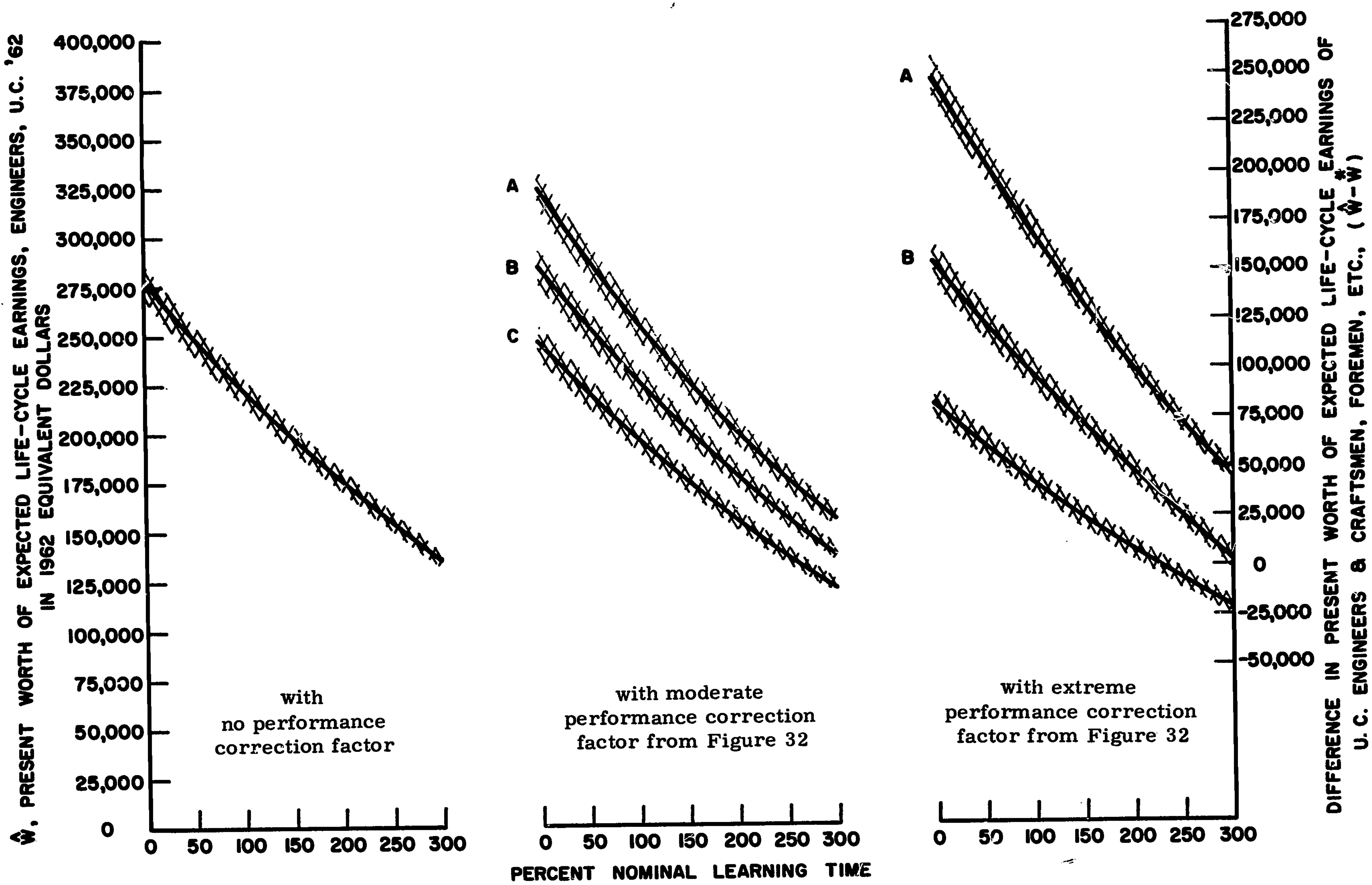
$$
- \frac{T'(\ell)}{T} \sum_{m=1}^{b-a} \left[p\left(\overset{*}{y}, \frac{CPI(y)}{CPI(y'+m)} \overset{*}{\$}(y', m)\right)\right] \left[\left\{\frac{1}{1+r}\right\}^{m-\frac{1}{2}} \right] \left[\overset{*}{M}(a+m-\tfrac{1}{2}) \right]
$$

$$
- \left[D(t, j, \ell) \right] \left[\left\{\frac{1}{1+r}\right\}^{\tau - \frac{t}{2}} \right]
$$



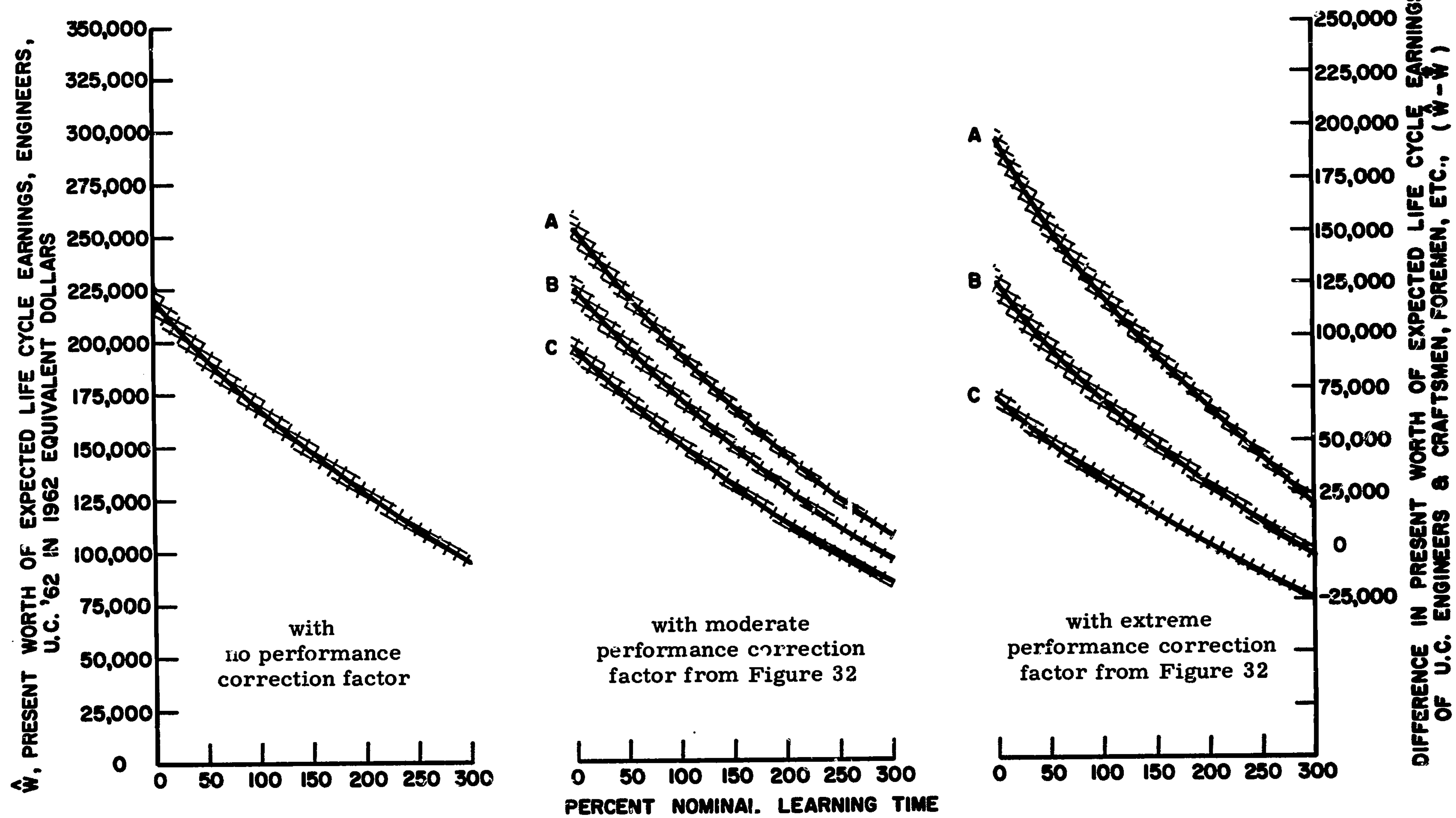
CORRECTED PRESENT WORTH OF EXPECTED
LIFE-CYCLE EARNINGS; 3% DISCOUNT RATE

FIGURE 33



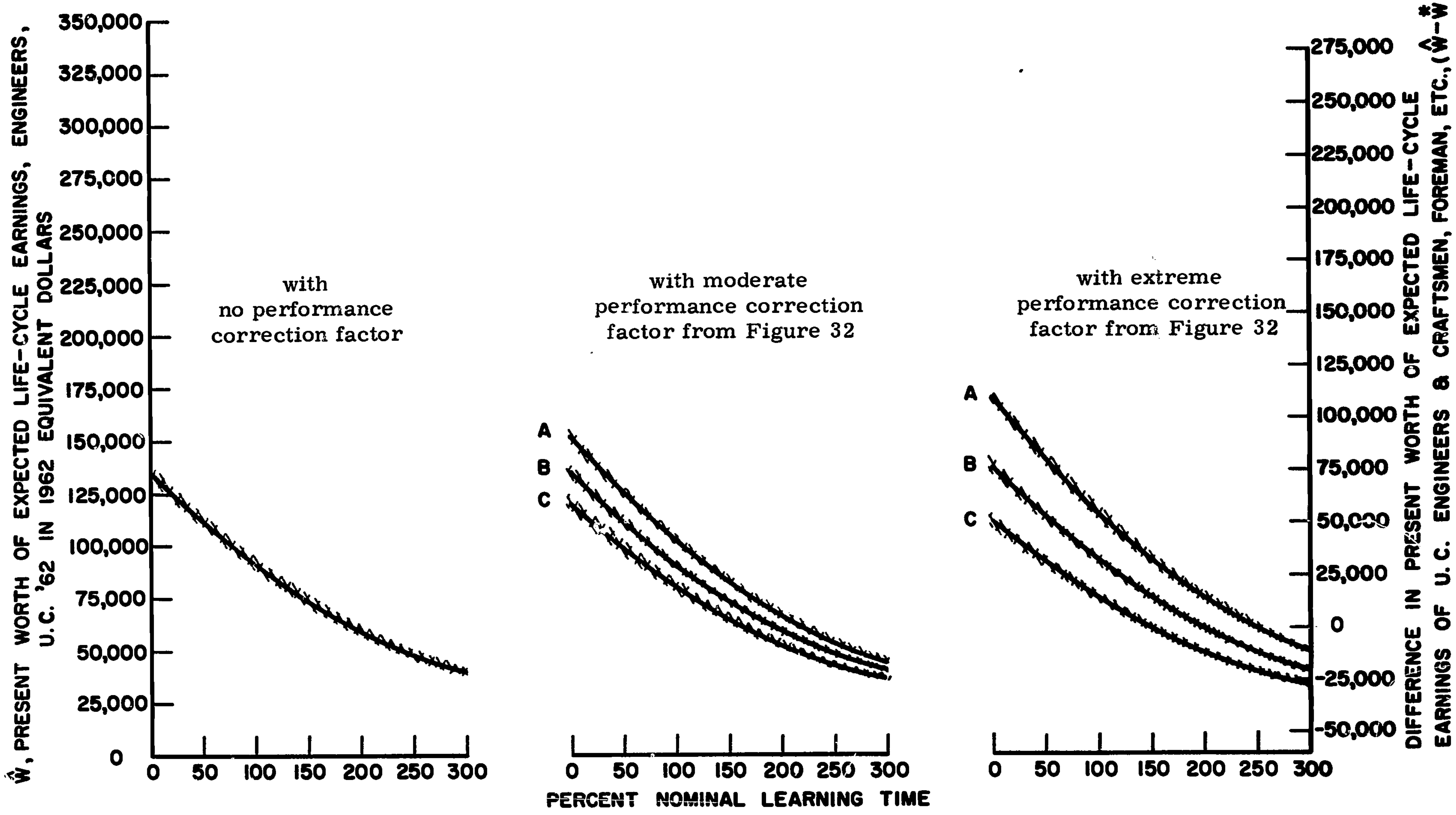
CORRECTED PRESENT WORTH OF EXPECTED LIFE-CYCLE EARNINGS; 4½% DISCOUNT RATE

FIGURE 34



CORRECTED PRESENT WORTH OF EXPECTED LIFE-CYCLE EARNINGS; 6% DISCOUNT RATE

FIGURE 35



CORRECTED PRESENT WORTH OF EXPECTED LIFE-CYCLE EARNINGS; 10% DISCOUNT RATE

FIGURE 36

In the figures, the abscissa is t/T' and the right-hand ordinate is $(\hat{W} - \overset{*}{W})$.

The results shown in these figures will be utilized in the following section where simulation will be made of an adaptive teaching situation using (g, t) data from an actual experiment with various decision rules, discount rates (r) and proportionality factors (c).

SECTION VI

SIMULATION OF AN ADAPTIVE DECISION STRUCTURE

There is an unfortunate aspect to the type of adaptive system that has been described in the preceding sections: its validity cannot be tested directly. The method given above for specifying the output of an educational system either has face validity, or none at all. Also, the appropriateness of a decision rule cannot be tested directly, since the identical naive (or unlearned) students are not available again for testing with alternate decision rules. Even if matched groups of students are available, in order to compare various decision rules one must either abandon the central concept that educational experiments should be conducted so as to maximize S_n, or else engage in the bootstrap operation of using a super-decision rule (up one rung in the ladder of levels of adaptivity) in order to find out which is the best decision rule (where the super-decision rule and the decision rule are likely to be one and the same).

There is a third, vicarious, alternative: use data from educational experiments which have been previously conducted without benefit of the criterion of maximizing S_n. The procedure for using existing data would be approximately as follows: take a random sample of size one from each category; convert the data into X scores; follow the specified decision rule in determining which category to take an observation from next; take a random sample of size one from this category, etc. This procedure could be repeated a number of times, and the distribution and expected value of S_n for a given decision rule could be determined and compared with the distribution and expected value of S_n for other decision rules. There is the further advantage that the results using the decision rule can be compared with the results obtained in the experiment for which the data were originally collected.

Since values of X have already been plotted for various combinations of g, t, and r for the engineering graduate of the University of California, it seemed most convenient to use data from an experiment conducted in an engineering school of the University of California. Furthermore, the experiment should have data on learning time and performance scores. Fortunately, the author had recently conducted an experiment which meets the above requirements [48]. The purpose of the experiment had been to determine the effectiveness of different branching procedures for self-instructional material. The precise nature of the branching procedure and the subject content for each category need not concern us for the simulation. However, it is of interest to note that a clear-cut decision could not be made in the original experiment as to which was the best category, since no category yielded the highest mean performance score and the lowest mean learning time.

The appropriate model for this experiment is:

$$ {}_{n}X_{ij\ell}(g,t) = c(\hat{W} - \overset{*}{W}) - V $$

$$ = \frac{T'(\ell)}{T} \sum_{m=1}^{b-a-\frac{tT}{T'}} \left[w\left(h(g'),\ p\left(\hat{y},\ \frac{CPI(y)}{CPI(y'+m)}\ \hat{\$}(y',m)\right)\right)\right] \left[\left\{\frac{1}{1+r}\right\}^{\frac{tT}{T'}+m-\frac{1}{2}} \right] \cdot $$

$$ \cdot \left[\hat{M}\left(a+m-\tfrac{1}{2}+\frac{tT}{T'}\right)\right] $$

$$ - \frac{T'(\ell)}{T} \sum_{m=1}^{b-a} \left[p\left(\overset{*}{y},\ \frac{CPI(y)}{CPI(y'+m)}\ \overset{*}{\$}(y',m)\right)\right] \left[\left\{\frac{1}{1+r}\right\}^{m-\frac{1}{2}} \right] \left[\overset{*}{M}(a+m-\tfrac{1}{2}) \right] $$

$$ - \left[D(t,j,\ell) \right] \left[\left\{\frac{1}{1+r}\right\}^{\tau-\frac{t}{2}} \right] $$

where

T = 4 years (7200 hours)

τ = 0 (the experiment was conducted during the first week of the freshman year)

a = 18 (assumed)

b = 62 (assumed)

$\hat{y}$ = 1962 (assumed)

$\overset{*}{y}$ = 1958

For $D(t, j, \ell)$, the following estimates were obtained:

$D(t, 1)$ = \$0.025 t + \$1.00

$D(t, 2)$ = \$0.025 t + \$1.05

Since the g in the experiment are given in percent and the g' shown in Figures 33, 34, 35, 36 are in letter grades, an h-transformation is required. This transformation was obtained by matching the relative frequency of reported grades for University of California engineering graduates with the relative frequency of the percentage scores obtained in the experiment. Then, combining the h-transformation with the w-transformation given by the heavy lines in Figure 32, it was found that

$$\left[w\left(h(g'), p\left(\hat{y}, \frac{CPI(y)}{CPI(y'+m)}\ \$(y'm)\right)\right)\right] = \left[1 + m\left(\frac{g^2}{700,000} - .007\right)\right] \cdot$$

$$\cdot \left[p\left(\hat{y}, \frac{CPI(y)}{CPI(y'+m)}\ \hat{\$}(y', m)\right)\right]$$

In the experiment, T' was estimated at 100 minutes. However, it is of interest to discover the effect of a choice of "c" on the results; therefore values of T' = 50, 100, 200 minutes will be used in the simulation. Also, r = .03, .045, .06 and .10 will be tried.

The original data and the calculated values of $X(g, t)$ for the different T' and r combinations are shown in Table 6. Also shown for each combination are the S_n.

In the simulation, the assumption is made that the cost of measuring and recording (g, t) for each student is small compared to X(g, t); therefore, instead of treating these costs as separate quantities, they are included in the $D(t, j, \ell)$. Four decision procedures are evaluated: Rule 1, Rule 7, the minimax rule, and the backwards-induction procedure. Furthermore, each of the procedures is used for two different total numbers of available students: one, corresponding to the number of students available from π_1, and π_2 in the original experiment, namely N = 58; two, corresponding to an assumed larger number of available students. In the simulation, the largest number of assumed available students that can be used is limited by the maximum N for which the backwards-induction procedure has been solved, namely N = 200. Since actual data are not available for N = 200 students, the assumption is made that the (g, t) measures on the students actually observed in the original experiment are representative of the distribution of such measures for each of the π_j and that random selections from the sample population will be approximately equivalent to random selections from the π_j.

Individual simulation runs were made with each of the decision procedures to check how the procedures behave in the particular rather than in the expected value sense. The results of these runs are shown in Appendix D.

Expected values were obtained for each decision procedure by taking the average of 500 iterations of each problem situation. These results are shown below in Table 7.

Before examining the results of the simulation, attention should be called to the small differences between the means of π_1 and π_2 shown in Table 6. These differences are approximately 0.2 standard deviations, and therefore different decision rules will not yield

TABLE 6

EXPERIMENT DATA AND CALCULATED UTILITIES

	π_1		π_2		r=.03						r=.045						r=.06						r=.10					
					T'=50		T'=100		T'=200		T'=50		T'=100		T'=200		T'=50		T'=100		T'=200		T'=50		T'=100		T'=200	
	Score	Time	Score	Time	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}	X_{11}	X_{12}
Mean:	70	146	82	138	6.11	6.54	16.89	17.58	38.44	39.66	0.97	1.38	6.60	7.25	17.86	19.00	-0.69	-0.30	3.30	3.91	11.26	12.31	-3.58	-3.24	-2.46	-1.95	-0.25	0.62
Std.Dev:	13	20	17	26	1.79	2.11	3.08	3.56	5.66	6.48	1.67	1.97	2.83	3.29	5.16	5.93	1.55	1.84	2.60	3.03	4.69	5.42	1.30	1.56	2.09	2.48	3.68	4.31
S_n/n					6.31		17.20		38.99		1.15		6.89		18.37		-0.52		3.57		11.73		-3.42		-2.23		0.14	
	64	157	72	120	5.55	7.74	16.02	19.53	36.96	43.11	0.43	2.50	5.78	9.04	16.47	22.12	-1.21	0.75	2.52	5.54	9.95	15.13	-4.05	-2.35	-3.15	-0.62	-1.36	2.82
	48	130	32	123	7.52	7.63	19.29	19.39	42.82	42.90	2.28	2.40	8.81	8.92	21.86	21.95	0.54	0.66	5.33	5.44	14.90	15.00	-2.55	-2.43	-0.83	-0.71	2.60	2.72
	92	140	76	128	6.00	7.55	16.49	19.34	37.47	42.92	0.87	2.30	6.23	8.84	16.95	21.93	-0.78	0.55	2.96	5.35	10.42	14.94	-3.62	-2.55	-2.71	-0.62	-0.92	2.62
	82	129	88	109	7.58	9.38	19.37	22.53	42.96	48.82	2.33	4.02	8.87	11.81	21.96	27.39	0.58	2.17	5.38	8.11	14.98	19.99	-2.52	-1.16	-0.79	1.48	2.65	6.73
	64	188	46	144	3.25	5.81	10.19	16.26	26.08	37.17	-2.59	0.69	0.54	6.03	6.77	16.71	-3.96	-0.94	-2.21	2.78	1.29	10.20	-6.25	-3.78	-6.78	-2.88	-7.85	-1.10
	84	144	64	146	5.89	5.77	16.37	16.24	37.34	37.18	0.76	0.65	6.12	6.00	16.84	16.70	-0.88	-0.99	2.86	2.74	10.31	10.18	-3.72	-3.82	-2.82	-2.92	-1.03	-1.14
	74	134	32	121	7.44	7.68	19.23	19.44	42.81	42.95	2.20	2.45	8.74	8.97	21.82	22.00	0.45	0.71	5.25	5.49	14.84	15.05	-2.65	-2.38	-0.92	-0.66	2.52	2.77
	68	157	82	143	5.55	5.85	16.02	16.31	36.97	37.25	0.43	0.73	5.78	6.07	16.48	16.77	-1.21	-0.91	2.52	2.81	9.96	10.25	-4.05	-3.75	-3.15	-2.85	-1.36	-1.06
	92	130	68	192	7.56	2.10	19.38	10.05	42.97	25.95	2.31	-2.74	8.86	0.39	21.96	6.63	0.56	-4.11	5.36	-2.36	14.97	1.15	-2.55	-6.40	-0.82	-6.93	2.63	-8.00
	84	187	38	148	3.54	5.70	12.74	16.15	31.15	37.04	-1.46	0.59	2.77	5.92	11.20	16.59	-2.97	-1.05	-0.26	2.67	5.17	10.09	-5.54	-3.88	-5.39	-2.98	-5.09	-1.20
	84	162	52	105	5.44	9.45	15.92	22.56	36.89	48.79	0.31	4.10	5.67	11.87	16.39	27.40	-1.33	2.25	2.41	8.18	9.86	20.03	-4.17	-1.07	-3.27	1.57	-1.48	6.81
	62	134	52	107	7.43	9.40	19.21	22.51	42.77	48.74	2.19	4.05	8.72	11.82	21.79	27.35	0.45	3.20	5.24	8.13	14.82	19.98	-2.65	-1.12	-0.93	1.52	2.51	6.76
	72	173	68	147	3.88	5.75	13.07	16.22	31.47	37.17	-1.12	0.63	3.11	5.96	11.53	16.68	-2.62	-1.01	0.09	2.72	5.51	10.16	-5.20	-3.85	-5.04	-2.95	-4.75	-1.16
	74	174	74	127	3.85	7.57	13.05	19.36	31.45	42.94	-1.14	2.32	3.08	8.86	11.51	21.95	-2.65	0.58	0.07	5.37	5.48	14.96	-5.22	-2.52	-5.06	-0.80	-4.77	2.65
	64	149	56	208	5.75	1.69	16.22	9.63	37.16	25.51	0.63	-3.14	5.96	-0.03	16.67	6.21	-1.01	-4.51	2.72	-2.77	10.15	0.73	-3.85	-6.80	-2.95	-7.33	-1.16	-6.40
	56	164	38	160	4.08	5.40	13.27	15.85	31.64	38.74	-0.90	0.29	3.31	5.62	11.72	16.29	-2.40	-1.35	0.31	2.37	5.71	3.79	-4.97	-4.18	-4.82	-3.28	-4.53	-1.50
	64	139	94	122	6.00	7.71	16.47	19.52	37.41	43.13	0.88	2.46	6.23	9.01	16.92	22.12	-0.76	0.71	2.97	5.51	10.40	15.12	-3.60	-2.40	-2.70	-0.67	-0.91	2.78
	72	131	82	130	7.52	7.50	19.31	19.30	42.88	42.89	2.27	2.25	8.81	8.80	21.89	21.89	0.52	0.50	5.32	5.30	14.91	14.90	-2.57	-2.60	-0.35	-0.87	2.60	2.58
	52	156	72	153	5.56	5.61	16.02	16.08	36.94	37.03	0.45	0.48	5.79	5.83	16.47	16.54	-1.19	-1.16	2.53	2.57	9.96	10.01	-4.02	-4.00	-3.12	-3.09	-1.34	-1.31
	68	131	40	77	7.51	11.51	19.30	26.00	42.87	54.96	2.27	6.08	8.81	15.13	21.88	33.23	0.52	4.15	5.32	11.28	14.90	25.52	-2.57	0.63	-0.85	4.26	2.59	11.48
	62	138	80	142	6.02	5.89	16.49	16.37	37.42	37.33	0.90	0.76	6.25	6.12	16.94	16.83	-0.74	-0.88	2.99	2.85	10.42	10.30	-3.57	-3.72	-2.67	-2.82	-0.89	-1.03
	52	157	80	146	5.54	5.79	16.00	16.27	36.92	37.23	0.42	0.66	5.76	6.02	16.45	16.73	-1.22	-0.98	2.51	2.75	9.93	10.20	-4.05	-3.82	-3.15	-2.92	-1.37	-1.13
	80	150	50	144	5.74	5.81	16.22	16.27	37.18	37.19	0.61	0.70	5.97	6.04	16.68	16.72	-1.03	-0.94	2.70	2.78	10.15	10.21	-3.87	-3.78	-2.97	-2.88	-1.18	-1.10
	70	110	60	143	9.39	5.85	22.52	16.31	48.78	37.24	4.03	0.73	11.81	6.07	27.37	16.76	2.19	-0.91	8.12	2.81	19.98	10.25	-1.14	-3.75	1.50	-2.85	6.74	-1.07
	60	173	70	167	3.87	3.98	13.08	13.17	31.43	31.56	-1.12	-1.02	3.09	3.20	11.51	11.63	-2.63	-2.52	0.09	0.19	5.49	5.60	-5.20	-5.10	-5.04	-4.94	-4.76	-4.65
	100	133	80	138	7.49	5.99	19.30	18.47	42.93	37.43	2.24	0.86	8.79	6.22	21.91	16.93	0.49	-0.78	5.29	2.95	14.90	10.40	-2.62	-3.62	-0.89	-2.72	2.56	-0.93
	50	134			7.42		19.19		42.73		2.16		8.71		21.77		0.44		5.23		14.80		-2.65		-0.93		2.51	
	50	133			7.45		19.22		42.76		2.21		8.74		21.79		0.47		5.25		14.83		-2.63		-0.90		2.53	
	70	168			4.00		13.20		31.59		-0.99		3.23		11.65		-2.50		0.22		5.83		-5.07		-4.92		-4.63	
	80	164			4.11		13.31		31.72		-0.89		3.34		11.77		-2.40		0.32		5.74		-4.97		-4.81		-4.52	
	76	112			9.34		22.48		48.75		3.99		11.77		27.33		2.14		8.07		19.94		-1.19		1.45		6.70	
	76	110			9.39		22.53		48.80		4.04		11.82		27.38		2.19		6.12		19.99		-1.14		1.50		6.75	

TABLE 7

$E(S_n/n)$ SIMULATION RESULTS*

		r=.03			r=.045			r=.06			r=.10		
		T'=50	T'=100	T'=200	T'=50	T'=100	T'=200	T'=50	T'=100	T'=200	T'=50	T'=100	T'=200
N=58	Rule 1	6.35	17.26	39.07	1.20	6.94	18.45	-0.48	3.57	11.78	-3.39	-2.16	0.18
	Rule 7	6.31	17.24	39.05	1.21	6.90	18.57	-0.48	3.62	11.87	-3.40	-2.18	0.21
	Minimax Rule	6.36	17.26	39.13	1.20	6.95	18.54	-0.44	3.67	11.85	-3.37	-2.17	0.27
	Backwards Induction	6.36	17.27	39.15	1.21	6.96	18.59	-0.42	3.69	11.92	-3.34	-2.15	0.29
	Original Method	6.31	17.20	38.99	1.15	6.89	18.37	-0.52	3.57	11.73	-3.42	-2.23	0.14
N-200	Rule 1	6.41	17.29	39.19	1.22	6.85	18.45	-0.47	3.71	11.92	-3.35	-2.12	0.25
	Rule 7	6.47	17.50	39.24	1.26	7.09	18.84	-0.41	3.78	11.92	-3.34	-2.12	0.39
	Minimax Rule	6.48	17.43	39.34	1.30	7.13	18.74	-0.39	3.80	12.08	-3.31	-2.02	0.41
	Backwards Induction	6.49	17.54	39.40	1.31	7.15	18.92	-0.38	3.82	12.12	-3.29	-2.00	0.42

*All results are in dollars.

Note: For Rule 7 the following forcing sets were used:

$$N = 58 \begin{cases} 3 & 10 & 28 \\ 5 & 14 & 36 \end{cases}$$

$$N = 200 \begin{cases} 3 & 9 & 22 & 48 & 100 \\ 4 & 11 & 24 & 50 & 102 \end{cases}$$

vastly different results. For example, for r=0.045 and T' =100, Table 6 reveals that $\overline{X}_1 = 6.60$ and $\overline{X}_2 = 7.25$, and the decision rule must yield an $E(S_n/n)$ somewhere between the two means. One measure of the effectiveness of a decision rule is given by

$$\text{Eff} = \frac{E(S_n/n) - \mu_1}{\mu_2 - \mu_1}$$

(assuming $\mu_2 > \mu_1$). Using the means given in Table 6 as approximations of the μ, Table 7 reveals that for r=0.045, T' =100, and N=200, the Eff of Rule 1 is 42%, the Eff of Rule 7 is 75%, the Eff of the Minimax Rule is 82%, and the Eff of the Backwards-Induction Rule is 85%. However, the absolute net expected value, in dollars per student, for the sub-unit of education investigated in the simulation, differs very little from one decision rule to another. For the r=0.045, T' =100 and N=200, these absolute values range from a minimum $6.85 for Rule 1 to a maximum of $7.15 for the Backwards-Induction Rule, a difference of less than 5%! Since the basic data used for obtaining the utilities are likely to have errors greater than ±5% it appears that when the differences between the true means of the π_j are small, the choice of the most effective decision rule will not give interestingly better results than the choice of a less effective decision rule. The above statement applies when the r, T', and N have been precisely determined.

Table 7 reveals that the choice of r and T' has a much greater effect on the absolute values of $E(S_n/n)$ than does the choice of a decision rule, and therefore any systems analysis which requires the use of the absolute value attributable to a given unit or sub-unit of education will be greatly affected by the choice of r and T'. For example, if the administrator of an educational system observed that the $E(S_n/n)$ was approximately $7.00 per student when comparing two different teaching methods, and having assumed an r=0.045 and a T'=100, he would probably be inclined to continue the sequential assignment of students to the two methods. However, if an r=0.10 and a T'=100 had been assumed, he may be confronted with an $E(S_n/n)$ of -$3.00, indicating the inputs outweigh the expected returns, in which case he would probably want to stop assigning students to the two methods and probably consider new alternatives.

To conclude, Table 7 indicates that for moderate to large N, the Backwards-Induction Rule yields better results than any of the other rules considered, regardless of the choice of r and T'. Therefore, this rule is recommended for use in adaptive educational systems, particularly since the value of r and T' would be fixed for all students involved in a given sequential assignment problem.

SECTION VII

CONCLUSION

Education is generally conceded to be a wealth or a utility producing process. It is also a process which traditionally has been shaped by intuitive rather than by analytical decisions. In the preceding sections, an attempt was made to show how an analytical adaptive decision structure can be built for educational systems. It was emphasized that such a structure rests on four cornerstones: a plan for gathering and using data; an explicit criterion function; a set of decision rules for achieving the criterion; and a utility function which relates system inputs and system outputs to a value scale outside of the system.

The utility function developed in the preceding sections defines the output of an educational system as the increment in life-cycle productive output attributable to the educational experience for all individuals who have been part of the system. An approximate measure of the average increment in productive output can be obtained by comparing the earnings of two matched groups of individuals, one of which has had the educational experience, the other of which has not. Such comparisons are relatively precise for large blocks of education, such as a college education versus no college education, and is less precise for smaller units of education, such as a semester course in a specific subject. The trend, over a number of past years, of the average increment in earnings of previous students is used to project the future expected increment in earnings of current students. For some educational experiences, such as the college training of professional engineers, a correlation can be found between performance in school and subsequent life-cycle earnings. In these special cases, the expected increment in life-cycle earnings of a current student can be adjusted by a school performance factor. The expected increment in earnings is distributed over the productive life-cycle, a span of perhaps forty to forty-five years. By discounting the future expected earnings, a single present worth of the entire expected increment in life-cycle earnings can be obtained. Similarly, a present worth of the total expenditures made in providing a student with an educational experience can be obtained. The difference between the two present worths represents the present worth of the net expected output per student of the system. By discounting the expected increment in earnings for each year of the productive life-cycle back to the date on which a student entered the educational system, an economic value can be associated with the amount of time it takes a student to complete the educational experience. All other things being equal, a student who completes a unit of education in three years would have each of the annual expected increments in earnings discounted one year less than if he completed the unit of education in four years. Using this time-value factor, and the school performance factor, it is possible for the first time to evaluate the possible trade-off between the student's learning time and performance level. Discounting expected earnings also reduces the effect of the uncertainties and errors that enter into the projection of future earnings.

The utility function is stated in sufficiently general terms so that the present worth of the expected increment in life-cycle productive output need not be measured only in terms of earnings. It is conceivable that adequate ec[illegible]ic measures can be found for such things as the expected increment in national and individual security attributable to an educational experience, or in the indirect contributions to the well-being of other individuals (say, from research discoveries), or for such indirect benefits as a pleasant work environment, longer vacations, a healthier life and other currently non-monetary benefits that may be attributable to an educational experience. Evaluation of the non-monetary measures becomes more important as the emphasis in a society shifts from monetary to non-monetary rewards for productive output (partly as a result of different tax rates for low and high earners).

Having established a plan for making an economic measure of the net output of an educational system, and having illustrated its use for University of California engineering students, the next important consideration is to establish an overall goal or criterion of performance for the system. The criterion that has been suggested here is that an educational system should operate so as to maximize the sum of the increment in the net present worth of the expected life-cycle productive output of all of the students who are being educated in the system. If one has prior knowledge of the costs and the expected gross outputs associated with different curricula or pedagogical techniques, then a straightforward input-output analysis can be made and that curricular configuration or those pedagogical techniques employed which will yield the maximum sum of the expected net outputs. One example where the costs and expected gross outputs could be readily anticipated is in a comparison of two-semester four-year college systems versus three-semester three-year college systems. However, in most situations of interest, accurate prior knowledge of the costs and expected gross outputs for different curricular or pedagogical techniques is not available. Therefore, some exploration or information gathering is necessary. If such exploration consists in trying different teaching methods or course content, then some students will be exposed to methods or content which may be inferior to other methods or curricular content, in that they yield lower present worths of the net increment in expected life-cycle productive output for those students. There is a trade-off between the probable loss attributable to assigning some students to inferior regimens during the information gathering phase, and the probable loss attributable to the failure to gather enough information as to which would be the best regimen for all future students. Therefore, decision rules are needed for assigning students to available curricular configurations or pedagogical methods in such a way as to meet the criterion of maximizing the sum of the net output of all students going through the system.

A number of possible decision rules have been examined in the preceding sections. For the case where no prior information exists as to the distribution of expected net outputs, some qualitative results have been obtained for specifying the set of "forced choices" first suggested by Robbins [10] in his statement of the sequential assignment problem. For the case where the distribution of expected net outputs is known to be normally distributed, a method has been developed here for including the cost of making observations on student performance during the information gathering period in a two-stage sequential decision procedure. Of most interest was the development in Section III of a multi-stage or continuous sequential decision rule for use with normally distributed expected net outputs. Since records are ordinarily kept on all students in an educational system, and not only on the first group of students who are assigned to specific curriculum, the multi-stage sequential assignment procedure is most appropriate. Where records on student performance are a necessary part of the system for reasons other than their use in a decision process, or where the cost of obtaining such records is very small compared to the net output, then the multi-stage sequential decision process gives better results than any other process. The solution to the multi-stage sequential assignment problem was accomplished by a backwards-induction, using numerical techniques to solve the multiple integrals that arise in the problem.

In the course of developing the framework for the adaptive decision structure for educational systems, a number of points arose which seem to warrant further investigation in order to improve the structure or extend its usefulness. First, in its current state of development, the multi-stage sequential assignment problem requires a separate set of calculations for each different estimated number, N, of students who will be going through a specified educational experience. For very large N, such computations can be excessively time consuming, even on the fastest available digital computer. Overall computation time could be reduced if the solution is carried out on a hybrid analog-digital computer.

Another fruitful avenue of investigation is to try to find a general solution in terms of N. Since the solution for different N results in surfaces which appear to have some regular features, such a general solution seems feasible.

The multi-stage sequential assignment problem has only been solved here for the case where the distributions are Gaussian and where the ratios of the variances are known. The solution can be further extended to include the case where the ratios of the variances are not known, and also to non-Gaussian distributions. However, it is felt that such extension will be of more interest in adaptive decision problems which arise outside of the context of the educational systems that were considered here.

Second, the utility function developed in Sections IV and V can be considerably enhanced by:

a. Careful studies to reveal those factors (in addition to school grades) which can be measured either before or while a student is engaged in an educational experience and which correlate with subsequent life-cycle productive output level.

b. More specific data on the life-cycle productive outputs of carefully matched "educated" and "non-educated" groups.

c. A means for including non-monetary indications of productive output.

d. The development of school performance measures which use an absolute scale, rather than such relative scales as obtained from the familiar bell-shaped curve. Some states have Regents' examinations and some professional schools have terminal examinations which are steps in the desired direction.

e. The accumulation of data on life-cycle productive outputs of students who have been exposed to different combinations of sub-units of a given educational program or have been exposed to different pedagogical procedures.

Even though the additional research outlined above would enhance the usefulness of the decision structure, it is possible to use the existing framework for some significant input-output analyses of educational systems, and it is also possible to inaugurate an adaptive decision procedure in some specific cases, such as in engineering education. Within the framework of the adaptive structure, it should be possible to make rational decisions on the amount of resources to allocate to the development of instructional material and on techniques that would permit a gradual shift from the lock-step grouping of students in semester length courses to a flexible scheduling scheme in which each student would progress through an eduational program as fast as possible, consistent with his own needs and the needs of the world in which he will some day become a productive member.

BIBLIOGRAPHY

1. Howard, Ronald A., Dynamic Programming and Markov Processes. Technology Press of M.I.T. and Wiley, New York, 1960.

2. Roe, A., M. Massey, G. Weltman and D. Leeds, "Automated Teaching Methods Using Linear Programs", University of California, Los Angeles, Department of Engineering, Report No. 60-105, December, 1960.

3. Pask, G., "The Self Organizing Teacher", Automated Teaching Bulletin, (Rheem-Califone Corporation, Los Angeles) Vol. 1, 1959, pp. 13-18.

4. Pask, G., "A Teaching Machine for Radar Training", Automation Progress, Vol. 2, 1957, pp. 214-217.

5. Pask, G., "The Teaching Machine", Overseas Engineer, Vol. 32, 1959, pp. 231-232.

6. Smallwood, Richard D., A Decision Structure for Teaching Machines. M.I.T. Press, Cambridge, Mass., 1962.

7. Bradt, R. N., S. M. Johnson and S. Karlin, "On Sequential Designs for Maximizing the Sum of n Observations", Ann. Math. Stat., Vol. 27, 1956, pp. 1060-1074.

8. Roe, A., "Format for Branching Programs in Automated Instruction", IRE Trans. on Educ. Vol. E-5, 1962, pp.131-135.

9. Roe, A., J. Lyman and H. Moon, "The Dynamics of an Automated Teaching System", in Applied Programmed Instruction, by Stuart Margulies and Lewis D. Eigen, editors, Wiley, New York, 1962, pp. 129-142.

10. Robbins, H., "Some Aspects of the Sequential Design of Experiments", Bull. Amer. Math. Soc., Vol. 58, 1952, pp. 527-535.

11. Dodge, H. F., and H. G. Romig, "A Method of Sampling Inspection", Bell System Technical Journal, Vol. 8, 1929, pp. 613-631.

12. Wald, Abraham, Sequential Analysis, Wiley, New York, 1947.

13. Mallows, C. L., and H. Robbins, Beating the Many-Armed Bandit. Columbia University, Department of Mathematical Statistics, New York, September, 1962.

14. Bellman, R., "A Problem in the Sequential Design of Experiments", Sankhya, Vol. 16, 1956, pp. 221-229.

15. Sakaguchi, M., "Dynamic Programming of Sequential Sampling Design", J. Math. Analysis and Applications, Vol. 2, 1961, pp. 446-466.

16. Vogel, W., "A Sequential Design for the Two-Armed Bandit", Ann. Math. Stat., Vol. 31, 1960, pp. 430-443.

17. Vogel, W., "An Asymptotic Minimax Theorem for the Two-Armed Bandit Problem", Ann. Math. Stat., Vol. 31, 1960, pp. 444-451.

18. Feldman, D., "Contributions to the 'Two-Armed Bandit Problem'", Ann. Math. Stat., Vol. 33, 1962, pp. 847-856.

19. Bechhofer, R., "A Sequential Multiple-Decision Procedure for Selecting the Best One of Several Normal Populations With a Common Unknown Variance, and Its Use With Various Experimental Designs", Biometrics, Vol, 14, 1958, pp. 408-429.

20. Paulson, E., "A Sequential Procedure for Comparing Several Experimental Categories With a Standard or Control", Ann. Math. Stat., Vol. 33, 1962, pp. 438-443.

21. Fabian, V., "On Multiple Decision Methods for Ranking Population Means", Ann. Math. Stat., Vol. 33, 1962, pp. 248-254.

22. Dunnett, C. W., "On Selecting the Largest of k Normal Population Means", J. Roy. Statis. Soc., Ser. B, Vol. 22, 1960, pp. 1-40.

23. Maurice, R., "A Different Loss Function for the Choice Between Two Populations", J. Roy. Statis. Soc., Ser. B., Vol. 21, 1959, pp. 205-213.

24. Girshick, M. A., "Contributions to the Theory of Sequential Analysis", Ann. Math. Stat., Vol. 17, 1946, pp. 123-143.

25. Raiffa, Howard and Robert Schlaifer, Applied Statistical Theory, Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1961.

26. Bellman, Richard, Dynamic Programming, Princeton University Press, Princeton, N. J., 1957.

27. Salzer, H. E., R. Zucker and R. Capuano, "Table of the Zeros and Weight Factors of the First Twenty Hermite Polynomials", J. Res. Natl. Bur. Stand., Vol. 48, 1952, pp. 111-115.

28. Davis, P., and P. Rabinowitz, "Abscissas and Weights for Gaussian Quadratures of High Order", J. Res. Nat. Bur. Stand., Vol. 56, 1956, pp. 35-37.

29. Jackson, M. L. "Evaluation of Teaching Effectiveness", J. Engrg. Educ., Vol. 50, 1960, pp. 866-868.

30. Jones, E. S., "College Graduates and Their Later Success", University of Buffalo Studies, Vol. 22, 1956, pp. 117-208.

31. Machlup, Fritz, The Production and Distribution of Knowledge, Princeton University Press, Princeton, N. J., 1962.

32. Becker, G. S., "Underinvestment in College Education", Amer. Econ. Rev., Vol. 50, 1960, pp. 346-354.

33. Schultz, T. W., "Education and Economic Growth", Chapter III in Social Forces Influencing American Education, Sixtieth Yearbook of the National Society for the Study of Education, University of Chicago Press, Chicago, 1961, pp. 46-88.

34. Hansen, W. L., "Total and Private Rates of Return to Investment in Schooling", J. Polit. Econ., Vol. 71, 1963, pp. 128-140.

35. Miller, H. P., "Annual and Lifetime Income in Relation to Education: 1939-1959", Amer. Econ. Rev., Vol. 50, 1960, pp. 962-986.

36. Houthakker, H. S., "Education and Income", Rev. of Econ. and Stat., Vol. 41, 1959, pp. 24-28.

37. Rosenthal, D., A. B. Rosenstein and G. Wiseman, "Information Theory and Curricular Synthesis", University of California, Los Angeles, Department of Engineering, Report EDP-3-63, 1963.

38. Pierson, G. A., "School Marks and Success in Engineering", Educ. and Psych. Meas., Vol. 7, 1947, pp. 612-614.

39. Haveman, Ernest and Patricia West, They Went to College, Harcourt, Brace, New York, 1952.

40. Wallace, Walter Paul, Engineering Studies at Freshman Level and After Graduation, Ph.D. Dissertation, College of Engineering, University of California, Los Angeles, 1954.

41. Taylor, C. W., "Some Variable Functioning in Productivity and Creativity". Paper presented at The Second University of Utah Research Conference on the Identification of Creative Scientific Talent, 1957.

42. LeBold, W. K., "A Study of the Purdue University Engineering Graduate", Purdue University, Lafayette, Indiana, Engineering Extension Department Extension Series No. 99, 1960.

43. Eurich, A. C., "Engineering, The Teaching of Engineering", J. Engrg. Educ., Vol. 53, 1963, pp. 273-278.

44. Bridgman, D. S., "Success in College and Business", Personnel Journal, Vol. 9, 1930, pp. 1-19.

45. College Achievement and Progess in Management, American Telephone and Telegraph Company, Personnel Research Section, New York, March, 1962.

46. DeHaven, J. C., "The Relation of Salary to the Supply of Scientists and Engineers", RAND Corporation Santa Monica, P-1372-RC, 1958.

47. Stewart, P. B., "Does Chemical Engineering Pay?" Chem. Engrg., Vol. 63, 1956, pp. 192-194.

48. Roe, A., "A Comparison of Branching Methods for Programmed Learning", J. Educ. Res., Vol. 55, 1962, pp. 407-416.

APPENDIX A

SAKI

A student using "SAKI" views an exercise line consisting of alpha-numeric characters which are illuminated one at a time, each for a different length of time. Simultaneously, the student attempts to replicate the characters by depressing the keys on a key-punch machine. A separately illuminated display of the keyboard layout indicates to the student the correct key to depress at the same time that a particular exercise character is being illuminated. This helpful information may be withheld, either completely or partially. If completely withheld, the keyboard layout display lamps are not illuminated; if partially withheld, these lamps are illuminated after a delay period, i.e., some milliseconds after the exercise character has been illuminated. If the subscript "j" identifies a particular exercise line (4 lines used in Saki), and the subscript "i" identifies the position of a character on a line (24 positions), then T_{ji} represents the interval of time allowed for illuminating the i-th character on the j-th exercise line, and E represents the delay time for illuminating the corresponding character in the helpful keyboard layout display. The symbols given here are those used by Pask.

According to Pask, a measure, $S_{ji}(t)$, (temporarily stored in the device as a potential) is obtained by:

a. Determining whether the response is correct or incorrect. Incorrect responses are arbitrarily assigned a $S_{ji}(t)$ value of minus one.

b. For correct responses, $S_{ji}(t)$ is the difference between the time allowed for illuminating the ji-th character on the exercise line and the actual response time.

c. $S_{ji}(t)$ will have an initial value of zero and a value of one at the end of the training process.

d. $1 \geq S_{ji}(t) > 0$. (A requirement which appears to contradict a and c.)

Furthermore, an average value of the quantities $S_{ji}(t)$, called θ, is obtained. θ would therefore have an initial value of zero and a final value of one.

A storage condenser is provided for each ji character and the potential at any instant on a condenser may be designated by an $a_{ji}(t)$ value. Initially, a charge of value "u" is placed on each condenser. If no move is made, or until a move is made for each ji-th character, the condenser is discharged exponentially through a high resistance. If a move is made, the condenser is charged through a resistance for a fixed time, t', by a potential, $S_{ji}(t)$. At the end of the training process the $a_{ji}(t)$ should all have a value of one.

To recapitulate,

$$^{\dagger}S_{ji}(t) = \begin{cases} T_{ji}(t-1) - \tau_{ji}(t) & \text{for correct response} \\ -1 & \text{for incorrect response} \end{cases}$$

where τ is response time

$^{\ddagger}1 \geq S_{ji}(t) > 0$ (probably true only for correct response)

$^{\dagger}\theta = \text{avg } S_{ji}(t)$ over all t

$^{\ddagger}T_{ji}(t) = (m + \theta)(a_{ji}) + u; \quad 0 < u \leq 1, \quad 0 < m \leq 1$

$^{\ddagger}E = \nu(a_{ji}); \quad 0 < \nu \leq 1$

---

† Inferred from verbal descriptions

‡ Explicitly defined by Pask

$$^{\dagger}a_{ji}(t) = a_{ji}(t-1)\exp\left[-\frac{\tau}{R_1C_1}\right] + \left(a_{ji}(t-1)\exp\left[-\frac{\tau}{R_1C_1}\right] + S_{ji}(t)\right)\exp\left[-\frac{t'}{R_2C_2}\right]$$

However, for fixed t'

$$a_{ji}(t) = \left(a_{ji}(t-1)\exp\left[-\frac{\tau}{R_1C_1}\right]\right)(1+k) + K\,S_{ji}(t)$$

where m, u, ν, R, C, K and (S = - 1) are all arbitrary constants.

The initial values are:

$$S_{ji}(0) = 0$$

$$\theta(0) = 0$$

$$a_{ji}(0) = u \leq 1$$

$$T_{ji}(0) = (m + \theta)(a_{ji}) + u = u(m + 1)$$

$$E(0) = \nu(a_{ji}) = \nu\, u$$

The final values are:

$$\theta(f) \rightarrow 1$$

$$a_{ji}(f) \rightarrow 1$$

As practice occurs:

i $T_{ji}(t)$ should diminish

ii E(t) should increase

Assume that the correct responses are made to the first $\hat{t}$ characters, with $\tau_{ji}(t) = T_{ji}(t-1)$ in each case. Then:

$$S_{ji}(\hat{t}) = 0$$

---

† Inferred from verbal descriptions

$$\theta(\hat{t}) = 0$$

$$a_{ji}(\hat{t}) < a_{ji}(o)$$

i' $\quad T_{ji}(\hat{t}) < T_{ji}(o)$

ii' and $\quad E(\hat{t}) < E(o)$

If now, at t + 1, $\tau_{ji}(t+1) < T_{ji}(t-1)$, i.e., a response is made in less than the allowed time, then:

$$S_{ji}(t+1) > 0 > S_{ji}(t)$$

$$(t+1) > 0 > (t)$$

$$a_{ji}(t+1) > a_{ji}(t)$$

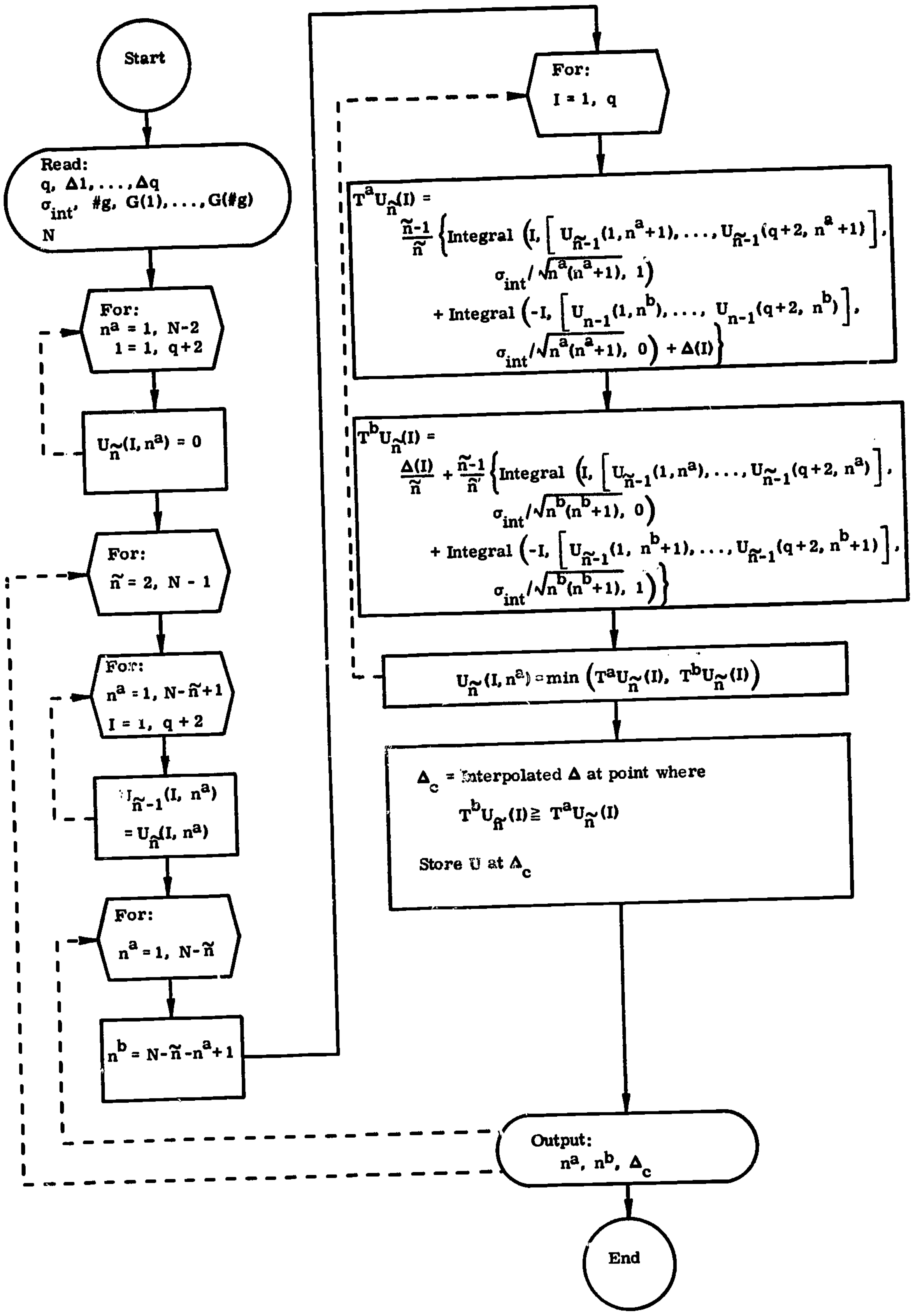
i'' $\therefore \quad T_{ji}(t+1) > T_{ji}(t)$

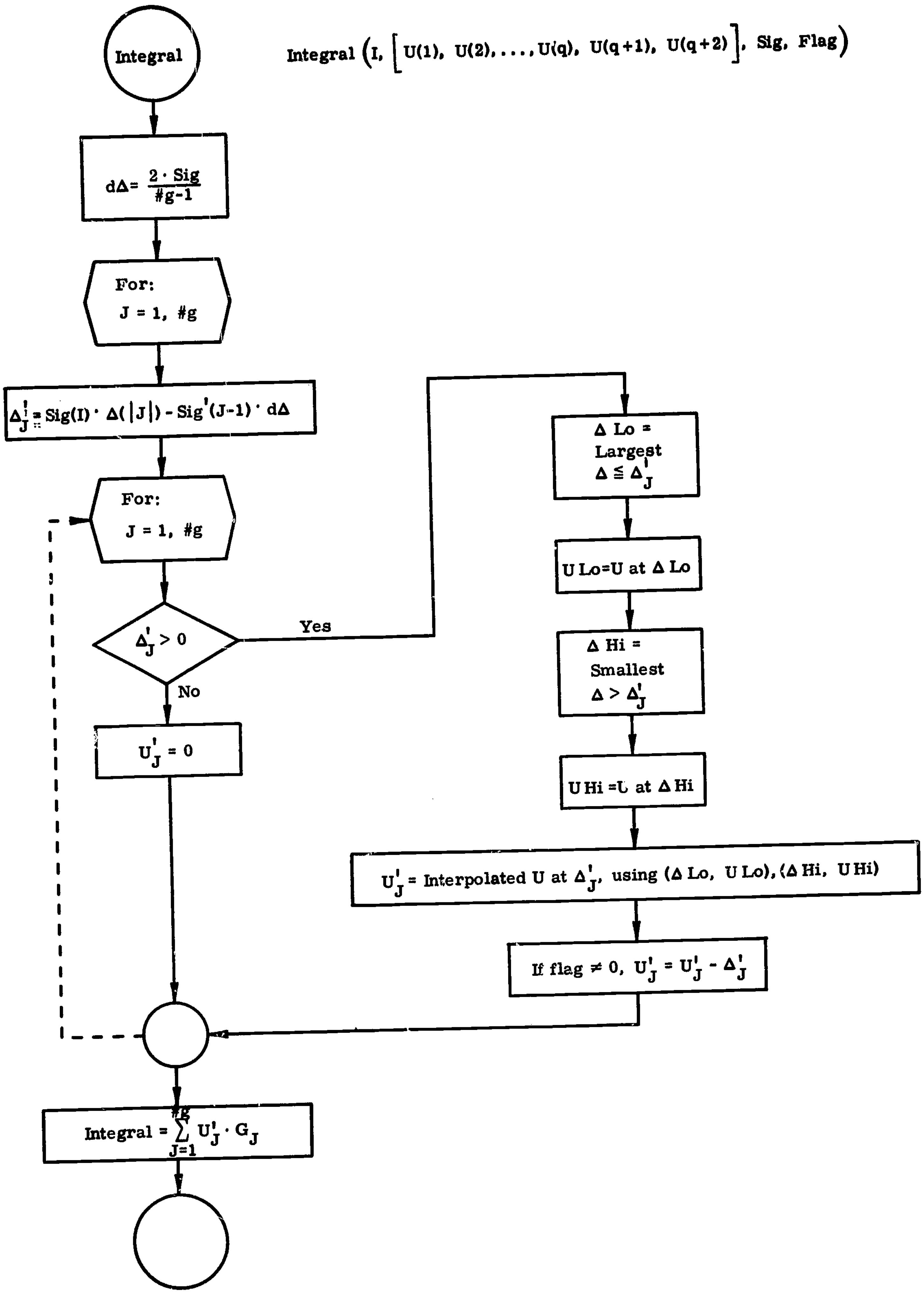
ii'' and $\quad E(t+1) > E(t)$

But note that ii' violates condition ii, and i'' violates condition i.

APPENDIX B

FLOW DIAGRAM FOR BACKWARDS INDUCTION




Integral
Integral (I, [U(1), U(2), . . . , U(q), U(q+1), U(q+2)], Sig, Flag)
dΔ = 2 · Sig / #g-1
For: J = 1, #g
Δ'_J = Sig(I) · Δ(|J|) - Sig'(J-1) · dΔ
For: J = 1, #g
Δ'_J > 0
Yes
No
U'_J = 0
Δ Lo = Largest Δ ≦ Δ'_J
U Lo=U at Δ Lo
Δ Hi = Smallest Δ > Δ'_J
U Hi =U at Δ Hi
U'_J = Interpolated U at Δ'_J, using (Δ Lo, U Lo), (Δ Hi, U Hi)
If flag ≠ 0, U'_J = U'_J - Δ'_J
Integral = Σ_{J=1}^{#g} U'_J · G_J

APPENDIX C

TABLE C-1

EARNINGS OF ENGINEERS

Years of Experience	YEAR OF SURVEY										
	1929[a]	1932[a]	1934[a]	1939[b]	1943[b]	1946[b]	1953[c]	1956[c]	1958[c]	1960[c]	1962[d]
< 1	1788	1332	1320	1610	2242	2842	4050	5000	5850	6300	6750
1	2172	1512	1392	1880	2464	2829	4250	5300	6125	6725	7025
2		1716	1488	2000	2702	3115	4600	5725	6475	7125	7425
3	2616	1872	1644	2182	2926	3324	4750	6050	6800	7475	8000
4		2172	1776	2318	3040	3600	5050	6350	7000	7800	8350
5	-		1984	2510	3250	3808	5325	6625	7400	8100	8725
6	3144	2460	2124	2522	3362	3986	5550	7000	7700	8450	9050
7				2818	3544	4246	5500	7300	8050	8900	9425
8	-	-	2388				5750	7600	8350	9250	9775
9	-	2808		-	-	-	6000	7750	8700	9625	10125
10	3720		-	3106	3754	4584	6200	7800	9100	9875	10425
11		-	2688	-	-	-					
12	-	-		-	-	-					
13	-		-	3404	4108	4944					
14	4068	3264	-	-	-	-					
15			3048	-	-	-	7400	9350	10000	11000	11900
16	-	-									
17	-	-	-	3866	4478	5274					
18		3540	-								
19		-	3276	-	-	-					
20	4620	-		-	-	-	7750	9800	10800	12075	12700
21			-								
22			-	4588	4874	5679					
23		3732									
24	-			-	-	-					
25	-		3504	-	-	-	8500	9800	10750	12400	12850
26											
27		-		5020	5612	6192					
28	4968	-									
29			-	-	-	-					
30			-	-	-	-	8850	10200	10900	12350	12700
31		4080									
32	-			5278	5816	6490					
33	-		3720								
34				-	-	-					
35		-		-	-	-	9200	10200	11200	12175	
36	5100	-									
37			-	5541	5888	6556					
38			-								
39		4224		-	-	-					
40	-			5298	6296	6754	9000	9750	10800	12175	12425
41	4656		3852								
42											
43		-									
44		3972									
45			-								
46			3408								

FOOTNOTES FOR TABLE C-1

(a) Source: Employment and Earnings in the Engineering Profession 1929-34. U. S. Dept. of Labor, Bureau of Labor Statistics, Bulletin No. 682, 1941. Table 64. Data in the above study were collected by mail questionnaires. 52,829 returns were used in the analysis. The sample was assumed to be representative of all engineers in the U. S. The original figures reflected "monthly earned median income from engineering work for time actually employed". These have been converted here to yearly figures. Some of the figures on Table B-1 are shown bounded by an upper and a lower dash. These upper and lower dashes indicate the range of the grouping of years of experience in the original report. For example, the fourth entry in the first column, 1929, is 3144, and it has an upper dash at year 5 and a lower dash at year 8. Thus the figure 3144 is the median of the group with 5 to 8 years of experience. The last figure in the columns is the salary for that year of experience and beyond.

(b) Source: Employment Outlook for Engineers. U. S. Dept. of Labor, Bureau of Labor Statistics, Bulletin No. 968, 1949, Table D-13. Data were collected by mail questionnaires. The sample was assumed to be representative of all U. S. engineers. The figures available in this report are for median base monthly salaries for the different engineering specialties, by years of experience. To obtain a composite figure for all engineers, the different specialties were summed across at each year level of experience and the average obtained. Since the proportion of the different specialties in the total sample was not the same, an attempt was made to weigh the different specialties proportionately in obtaining the composite figure. For the survey year, 1946, both weighted and unweighted composite figures were found and plotted on a graph. The curves were practically the same. Hence only the unweighted composite figures were calculated and these converted into yearly earnings.

(c) Source: Professional Income of Engineers, 1960. Engineers Joint Council, New York. Page 13. Data were collected by mail questionnaires. The sample was assumed to be representative of all U. S. engineers. Figures reflect "median annual base salary including cost of living allowance and bonus if considered part of salary". Figures in the original report were listed by years since B. S. degree. Here it is assumed that the year of completion of B. S. degree was the year of entry into work. Beyond the 10th year of experience salary figures are listed every five years. These are for terminal years and not for grouped years as in the case of (a) and (b) above.

(d) Source: Professional Income of Engineers, 1962. Engineers Joint Council, New York. Page 15. Same comments as for (c) above.

TABLE C-2

CONSUMER PRICE INDICES AND ADJUSTING FACTORS

Year	C.P.I.†	Adj. Factor‡	Year	C.P.I.†	Adj. Factors‡	Year	C.P.I.†	Adj. Factor	Year	C.P.I.†	Adj. Factor‡
1908	38.5‡		1922	71.6		1936	59.3		1950	102.8	1.25
1909	38.5‡		1923	72.9	1.76	1937	61.4		1937	110.0	1.16
1910	40.6‡		1924	73.1		1938	60.3	2.11	1952	113.5	1.13
1911	40.6‡		1925	75.0		1939	59.4	2.16	1953	114.4	1.12
1912	43.2‡		1926	75.6	1.69	1940	59.9		1954	114.8	1.11
1913	42.3	3.01	1927	74.2		1941	62.9		1955	114.5	1.12
1914	42.9		1928	73.3	1.75	1942	69.7		1956	116.2	1.10
1915	43.4		1929	73.3		1943	74.0	1.73	1957	120.2	1.06
1916	46.6		1930	71.4		1944	75.2	1.70	1958	123.5	1.04
1917	54.8		1931	65.0		1945	76.9	1.66	1959	124.6	1.03
1918	64.3	1.99	1932	58.4	2.19	1946	83.4	1.54	1960	126.5	1.04
1919	74.0		1933	55.3	2.32	1947	95.5		1961	127.8	1.00
1920	85.7		1934	57.2	2.24	1948	102.2		1962§	128.0	1.00
1921	76.4		1935	58.7		1949	101.8	1.26			

‡Adjustment factor for a given year is obtained by dividing the C.P.I. for 1962 by the C.P.I. for the given year.

Source: Historical Statistics of the U.S. -- from colonial times to 1957. Bureau of U.S. Census, Statistical abstracts of the U.S., 1962.

†For the years 1908-1912, the cost of living index (Federal Reserve Bank of N.Y.) with base year 1913=100 was adjusted to agree with the 1947-49=100 base.

§Early

TABLE C-3

SURVIVAL FACTORS

	Professional Workers		Skilled Workers	
(a) Age	(b) Deaths per 100,000	(c) Survival Factor	(b) Deaths per 100,000	(c) Survival Factor
18	95	.99853	138	.99862
19	95	.99694	138	.99724
20	95	.99599	139	.99585
21	96	.99503	139	.99446
22	96	.99407	140	.99306
23	95	.99312	140	.99166
24	95	.99217	141	.99025
25	94	.99123	142	.98883
26	93	.99030	143	.98740
27	93	.98937	144	.98596
28	95	.98842	148	.98448
29	104	.98738	157	.98291
30	114	.98624	167	.98124
31	125	.98499	176	.97948
32	135	.98364	184	.97764
33	145	.98219	200	.97564
34	155	.98064	225	.97339
35	175	.97889	250	.97089
36	200	.97689	275	.96814
37	220	.97469	300	.96514
38	250	.97219	325	.96189
39	275	.96944	375	.95814
40	300	.96644	415	.95399
41	350	.96294	450	.94949
42	400	.95894	500	.94449
43	450	.95444	550	.93899
44	525	.94919	625	.93274
45	575	.94344	675	.92599
46	650	.93694	750	.91849
47	725	.92969	825	.91024
48	800	.92169	925	.90099
49	980	.91189	1000	.89099
50	1000	.90189	1100	.87999
51	1100	.89089	1225	.86774
52	1225	.87864	1425	.85349
53	1350	.86514	1500	.83849
54	1475	.85039	1650	.82199
55	1600	.83439	1775	.80424
56	1750	.81689	1925	.78499
57	1922	.79767	2081	.76418
58	2075	.77692	2250	.74164
59	2225	.75467	2450	.71718
60	2425	.73042	2650	.69069
61	2650	.70392	2900	.66168
62	2886	.67506	3137	.63031

(b) Source: Inter- and extra-polated from Table 2 in Monyama, I. M. and Guralnick. Occupational and Social Clan Differences in Mortality. Trends and Differentials in Mortality. Milbank Memorial Fund, New York, 1956.

$$(c)_a = \frac{100,000 - \sum_{18}^{a} (b)}{100,000}$$

TABLE C-4

EARNINGS OF CRAFTSMEN, FOREMEN AND KINDRED WORKERS

Age	1946	1949	1951	1955	1959	1961
14-24	-		2684*			
25-34	2202[a]		3592[b]			
35-44	2629[a]		3913[b]			
45-54	2753*		3731[b]			
55-64	2456*		3544[bc]			
over-all median	2433[h]	3114[d]	3627[b]	4423[e]	5355[f]	5640[g]

* From unpublished data, Bureau of the Census, U.S. Dept. of Commerce.

a From Miller, H. P., The Income of the American People. Wiley, 1955 (Table 25, page 54).

b Bureau of the Census, Current Population Reports, Consumer Income, Series P-60, No. 11 (Table B).

c For age group 55 and beyond.

d Bureau of the Census, Current Population Reports, Consumer Income, Series P-60, No. 7 (Table 19).

e Bureau of the Census, Current Population Reports, Consumer Income, Series P-60, No. 23 (Table 5).

f Bureau of the Census, Current Population Reports, Consumer Income, Series P-60, No. 35 (Table 25).

g Bureau of the Census, Current Population Reports, Consumer Income, Series P-60, No. 39 (Table 29).

h Bureau of the Census, Current Population Reports, Consumer Income, Series P-60, No. 3 (Table 16).

TABLE C-5

EARNINGS OF ALL U. S. MALES

Age	1946[a]	1949[b]	1951[c]	1955[d]	1959[e]	1961[f]
14-19	406	410	434	416	411	399
20-24	1247	1726	2259	2223	2612	2654
25-34	2098	2754	3288	3886	4774	5045
35-44	2535	2951	3617	4255	5320	5726
45-54	2575	2751	3280	4138	4852	5321
55-64	2285	2366	2840	3440	4190	4597
65 +	1625	1016	1008	1337	1576	1758
over-all median	2134	2346	2952	3354	3996	4189

Source: Bureau of the Census, Current Population Reports, Consumer Income, Series P-60, Nos. (for a) 3, Table 10; (for b) 35, Table G; (for c) 11, Table 3; (for d) 23, Table 3; (for e) 35, Table 23; (for f) 39, Table 25.

APPENDIX D

TABLE D-1a

SIMULATION WITH RULE 1 - FIRST RUN

n	π_j	${}_nX_j$	S_n/n	$\mu_1=6.60$ ${}_n\bar{X}_1$	$\mu_2=7.25$ ${}_n\bar{X}_2$
1	1	5.98	5.98	5.98	0.00
2	2	9.04	7.51	5.98	9.04
3	2	8.84	7.96	5.98	8.94
4	2	-0.02	5.96	5.98	5.95
5	1	3.31	5.43	4.65	5.95
6	2	6.12	5.55	4.65	6.00
7	2	9.01	6.04	4.65	6.60
8	2	5.92	6.03	4.65	6.49
9	2	8.92	6.35	4.65	6.84
10	2	6.22	6.34	4.65	6.76
11	2	6.07	6.31	4.65	6.68
12	2	8.80	6.52	4.65	6.90
13	2	5.83	6.47	4.65	6.80
14	2	3.20	6.24	4.65	6.50
15	2	6.02	6.22	4.65	6.46
16	2	8.97	6.39	4.65	6.64
17	2	5.62	6.35	4.65	6.57
18	2	11.87	6.66	4.65	6.91
19	2	11.81	6.93	4.65	7.19
20	2	6.00	6.88	4.65	7.13
21	2	5.98	6.84	4.65	7.07
22	2	6.07	6.80	4.65	7.02
23	2	8.86	6.89	4.65	7.11
24	2	0.39	6.62	4.65	6.80
25	2	6.04	6.60	4.65	6.77
26	2	11.82	6.80	4.65	6.98
27	2	6.03	6.77	4.65	6.94
28	2	15.13	7.07	4.65	7.26
29	2	3.20	6.94	4.65	7.11
30	2	3.20	6.81	4.65	6.97
31	2	11.87	6.98	4.65	7.14
32	2	8.84	7.03	4.65	7.19
33	2	6.22	7.01	4.65	7.16
34	2	6.22	6.99	4.65	7.13
35	2	5.83	6.95	4.65	7.09
36	2	11.82	7.09	4.65	7.23
37	2	5.98	7.06	4.65	7.20
38	2	6.12	7.03	4.65	7.17
39	2	6.07	7.01	4.65	7.14
40	2	6.04	6.99	4.65	7.11
41	2	8.92	7.03	4.65	7.16
42	2	6.12	7.01	4.65	7.13
43	2	5.92	6.99	4.65	7.10
44	2	6.22	6.97	4.65	7.08
45	2	6.04	6.95	4.65	7.06
46	2	9.04	6.99	4.65	7.10
47	2	6.22	6.98	4.65	7.08
48	2	6.22	6.96	4.65	7.06
49	2	9.01	7.00	4.65	7.10
50	2	6.00	6.98	4.65	7.08
51	2	6.00	6.96	4.65	7.06
52	2	6.04	6.95	4.65	7.04
53	2	0.39	6.82	4.65	6.91
54	2	6.22	6.81	4.65	6.90
55	2	-0.02	6.69	4.65	6.77
56	2	6.00	6.68	4.65	6.75
57	2	11.87	6.77	4.65	6.84
58	2	6.07	6.76	4.65	6.83

TABLE D-1b

SIMULATION WITH RULE 1 - SECOND RUN

n	π_j	$_nX_j$	S_n/n	$_n\bar{X}_1$ (μ_1=6.60)	$_n\bar{X}_2$ (μ_2=7.25)
1	1	8.87	8.87	8.87	0.00
2	2	8.80	8.84	8.87	8.80
3	1	0.54	6.07	4.71	8.80
4	2	5.62	5.96	4.71	7.21
5	2	9.04	6.58	4.71	7.82
6	2	8.84	6.96	4.71	8.08
7	2	-0.02	5.96	4.71	6.46
8	2	6.12	5.98	4.71	6.40
9	2	9.01	6.32	4.71	6.78
10	2	5.92	6.28	4.71	6.67
11	2	8.92	6.52	4.71	6.92
12	2	6.22	6.49	4.71	6.85
13	2	6.07	6.46	4.71	6.78
14	2	6.03	6.43	4.71	6.72
15	2	5.83	6.39	4.71	6.65
16	2	3.20	6.19	4.71	6.40
17	2	6.02	6.18	4.71	6.38
18	2	8.97	6.34	4.71	6.54
19	2	15.13	6.80	4.71	7.05
20	2	11.87	7.05	4.71	7.31
21	2	11.81	7.28	4.71	7.55
22	2	6.00	7.22	4.71	7.47
23	2	5.98	7.17	4.71	7.40
24	2	6.07	7.12	4.71	7.34
25	2	8.86	7.19	4.71	7.41
26	2	0.39	6.93	4.71	7.12
27	2	6.04	6.90	4.71	7.07
28	2	11.82	7.07	4.71	7.26
29	2	0.39	6.84	4.71	7.00
30	2	6.00	6.82	4.71	6.97
31	2	11.82	6.98	4.71	7.13
32	2	6.00	6.95	4.71	7.10
33	2	8.80	7.00	4.71	7.15
34	2	3.20	6.89	4.71	7.03
35	2	5.83	6.86	4.71	6.99
36	2	6.02	6.84	4.71	6.96
37	2	5.62	6.81	4.71	6.93
38	2	3.20	6.71	4.71	6.82
39	2	9.01	6.77	4.71	6.88
40	2	8.97	6.83	4.71	6.94
41	2	6.22	6.81	4.71	6.92
42	2	8.80	6.86	4.71	6.97
43	2	5.62	6.83	4.71	6.93
44	2	6.00	6.81	4.71	6.91
45	2	5.98	6.79	4.71	6.89
46	2	8.86	6.84	4.71	6.93
47	2	6.07	6.82	4.71	6.92
48	2	6.12	6.81	4.71	6.90
49	2	6.12	6.79	4.71	6.88
50	[illegible]	5.9[illegible]	6.78	4.71	6.86
51	[illegible]	8.86	6.82	4.71	6.90
52	2	[illegible]	6.86	4.71	6.94
53	2	11.92	6.95	4.71	7.04
54	2	11.82	7.04	4.71	7.13
55	[illegible]	8.97	7.08	4.71	7.17
56	2	0.38	6.96	4.71	7.04
57	2	6.07	6.94	4.71	7.02
58	2	3.20	[illegible].88	4.71	6.95

TABLE D-1c

SIMULATION WITH RULE 1 - THIRD RUN

n	π_j	${}_n X_j$	S_n/n	$\mu_1 = 6.60$ ${}_n\bar{X}_1$	$\mu_2 = 7.25$ ${}_n\bar{X}_2$
1	1	3.09	3.09	3.09	0.00
2	2	6.12	4.61	3.09	6.12
3	2	8.92	6.05	3.09	7.52
4	2	8.80	6.74	3.09	7.95
5	2	5.62	6.51	3.09	7.37
6	2	9.04	6.94	3.09	7.70
7	2	8.84	7.21	3.09	7.89
8	2	-0.02	6.30	3.09	6.76
9	2	6.04	6.28	3.09	6.67
10	2	9.01	6.55	3.09	6.93
11	2	5.92	6.49	3.09	6.83
12	2	11.82	6.94	3.09	7.29
13	2	6.22	6.88	3.09	7.20
14	2	6.07	6.82	3.09	7.11
15	2	6.03	6.77	3.09	7.03
16	2	5.83	6.71	3.09	6.95
17	2	3.20	6.51	3.09	6.72
18	2	6.02	6.48	3.09	6.68
19	2	8.97	6.61	3.09	6.81
20	2	15.13	7.04	3.09	7.25
21	2	11.87	7.27	3.09	7.48
22	2	11.81	7.47	3.09	7.68
23	2	6.00	7.41	3.09	7.61
24	2	5.98	7.35	3.09	7.54
25	2	6.07	7.30	3.09	7.48
26	2	8.86	7.36	3.09	7.53
27	2	0.39	7.10	3.09	7.26
28	2	6.00	7.06	3.09	7.21
29	2	3.20	6.93	3.09	7.07
30	2	9.01	7.00	3.09	7.13
31	2	5.62	6.96	3.09	7.08
32	2	5.62	6.91	3.09	7.04
33	2	6.22	6.89	3.09	7.01
34	2	6.07	6.87	3.09	6.98
35	2	6.07	6.85	3.09	6.96
36	2	6.07	6.82	3.09	6.93
37	2	5.83	6.80	3.09	6.90
38	2	3.20	6.70	3.09	6.80
39	2	6.07	6.69	3.09	6.78
40	2	8.80	6.74	3.09	6.83
41	2	8.97	6.79	3.09	6.89
42	2	6.22	6.78	3.09	6.87
43	2	9.01	6.83	3.09	6.92
44	2	6.12	6.82	3.09	6.90
45	2	8.84	6.86	3.09	6.95
46	2	5.98	6.84	3.09	6.93
47	2	5.98	6.82	3.09	6.91
48	2	6.02	6.81	3.09	6.89
49	2	8.92	6.85	3.09	6.93
50	2	11.87	6.95	3.09	7.03
51	2	5.98	6.93	3.09	7.01
52	2	6.00	6.92	3.09	6.99
53	2	11.87	7.01	3.09	7.08
54	2	9.01	7.05	3.09	7.12
55	2	8.80	7.08	3.09	7.15
56	2	8.97	7.11	3.09	7.18
57	2	3.20	7.04	3.09	7.11
58	2	-0.02	6.92	3.09	6.99

TABLE D-2a

SIMULATION WITH RULE 7 - FIRST RUN

Set B $\begin{cases} \pi_1: 3, 10, 28 \\ \pi_2: 5, 14, 36 \end{cases}$

n	a_j	$_nX_j$	S_n/n	$\mu_1=6.30$ $_n\bar{X}_1$	$\mu_2=7.25$ $_n\bar{X}_2$
1	1	3.09	3.09	3.09	0.00
2	2	8.84	5.97	3.09	8.84
3	2	5.98	5.97	3.09	7.41
4	2	9.01	6.73	3.09	7.95
5	1	5.97	6.58	4.53	7.95
6	2	5.83	6.46	4.53	7.42
7	2	6.07	6.40	4.53	7.15
8	2	5.92	6.34	4.53	6.95
9	2	6.12	6.32	4.53	6.83
10	2	8.92	6.58	4.53	7.09
11	2	8.80	6.78	4.53	7.28
12	2	5.62	6.69	4.53	7.12
13	2	9.04	6.87	4.53	7.29
14	1	8.81	7.01	5.96	7.29
15	2	11.81	7.33	5.96	7.67
16	2	-0.02	6.87	5.96	7.08
17	2	6.04	6.82	5.96	7.00
18	2	6.00	6.77	5.96	6.94
19	2	0.39	6.44	5.96	6.53
20	2	11.82	6.71	5.96	6.84
21	2	6.22	6.68	5.96	6.80
22	2	6.07	6.86	5.96	6.77
23	2	6.03	6.63	5.96	6.73
24	2	8.86	6.72	5.96	6.83
25	2	3.20	6.58	5.96	6.67
26	2	6.02	6.56	5.96	6.64
27	2	8.97	6.65	5.96	6.74
28	2	15.13	6.95	5.96	7.07
29	2	11.87	7.12	5.96	7.26
30	2	3.20	6.99	5.96	7.11
31	2	-0.02	6.77	5.96	6.85
32	2	8.86	6.83	5.96	6.92
33	2	-0.02	6.62	5.96	6.69
34	2	5.83	6.60	5.96	6.66
35	2	9.01	6.67	5.96	6.74
36	1	11.82	6.81	7.43	6.74
37	1	5.78	6.78	7.10	6.74
38	1	6.25	6.77	6.96	6.74
39	1	5.67	6.74	6.77	6.74
40	1	6.12	6.73	6.69	6.74
41	2	15.13	6.93	6.69	6.99
42	2	8.97	6.98	6.69	7.05
43	2	3.20	6.89	6.69	6.94
44	2	0.39	6.75	6.69	6.76
45	2	6.07	6.73	6.69	6.74
46	2	8.92	6.78	6.69	6.80
47	2	5.83	6.76	6.69	6.77
48	2	6.03	6.74	6.69	6.75
49	2	6.04	6.73	6.69	6.74
50	2	11.81	6.83	6.69	6.86
51	2	8.92	6.87	6.69	6.90
52	2	8.92	6.91	6.69	6.95
53	2	9.04	6.95	6.69	7.00
54	2	6.02	6.93	6.69	6.93
55	2	-0.02	6.81	6.69	6.83
56	2	0.39	6.69	6.69	6.69
57	1	8.87	6.73	6.94	6.69
58	1	0.54	6.62	6.30	6.69

TABLE D-2b

SIMULATION WITH RULE 7 - SECOND RUN

Set B $\begin{cases} \pi_1: & 3, 10, 28 \\ \pi_2: & 5, 14, 36 \end{cases}$

n	π_j	$_nX_j$	S_n/n	$\mu_1=6.60$ $_n\bar{X}_1$	$\mu_2=7.25$ $_n\bar{X}_2$
1	1	8.81	8.81	8.81	0.00
2	2	9.01	8.91	8.81	9.01
3	2	5.83	7.89	8.81	7.42
4	1	11.82	8.87	10.32	7.42
5	1	5.78	8.25	8.81	7.42
6	1	6.25	7.92	8.17	7.42
7	1	3.09	7.23	7.15	7.42
8	2	6.07	7.09	7.15	6.97
9	1	6.12	6.98	6.98	6.97
10	2	5.92	6.87	6.98	6.71
11	1	8.87	7.06	7.25	6.71
12	1	0.54	6.51	6.41	6.71
13	2	6.12	6.48	6.41	6.59
14	1	5.98	6.45	6.37	6.59
15	2	8.92	6.61	6.37	6.98
16	2	8.80	6.75	6.37	7.24
17	2	5.62	6.68	6.37	7.04
18	2	9.04	6.82	6.37	7.26
19	2	8.84	6.92	6.37	7.42
20	2	-0.02	6.57	6.37	6.74
21	2	6.04	6.55	6.37	6.69
22	2	5.98	6.52	6.37	6.63
23	2	0.39	6.26	6.37	6.19
24	1	3.31	6.13	6.06	6.19
25	2	11.82	6.36	6.06	6.56
26	2	6.22	6.36	6.06	6.54
27	2	6.07	6.35	6.06	6.51
28	2	6.03	6.34	6.06	6.49
29	2	8.86	6.42	6.06	6.61
30	2	3.20	6.32	6.06	6.44
31	2	6.02	6.31	6.06	6.42
32	2	8.97	6.39	6.06	6.54
33	2	15.13	6.65	6.06	6.91
34	2	11.87	6.81	6.06	7.12
35	2	11.81	6.95	6.06	7.31
36	1	8.86	7.00	6.32	7.31
37	2	6.00	6.98	6.32	7.26
38	2	3.20	6.88	6.32	7.11
39	2	11.82	7.00	6.32	7.28
40	2	0.39	6.84	6.32	7.04
41	2	6.22	6.82	6.32	7.01
42	2	6.12	6.81	6.32	6.98
43	2	11.82	6.92	6.32	7.13
44	2	6.00	6.90	6.32	7.10
45	2	11.87	7.01	6.32	7.24
46	2	5.92	6.99	6.32	7.20
47	2	8.97	7.03	6.32	7.25
48	2	6.03	7.01	6.32	7.22
49	2	6.00	6.99	6.32	7.19
50	2	8.80	7.03	6.32	7.23
51	2	5.83	7.00	6.32	7.19
52	2	6.00	6.99	6.32	7.16
53	2	11.87	7.08	6.32	7.28
54	2	3.20	7.01	6.32	7.18
55	2	8.84	7.04	6.32	7.22
56	2	6.03	7.02	6.32	7.19
57	2	11.82	7.11	6.32	7.29
58	2	6.22	7.09	6.32	7.27

TABLE D-2c

SIMULATION WITH RULE 7 - THIRD RUN

Set B $\begin{cases} \pi_1: & 3, 10, 28 \\ \pi_2: & 5, 14, 36 \end{cases}$

$\mu_1 = 6.60$ $\mu_2 = 7.25$

n	π_j	${}_nX_j$	S_n/n	${}_n\bar{X}_1$	${}_n\bar{X}_2$
1	1	5.78	5.78	5.78	0.00
2	2	5.83	5.81	5.78	5.83
3	2	5.92	5.65	5.78	5.88
4	2	6.12	5.92	5.78	5.96
5	1	6.25	5.98	6.02	5.96
6	1	3.09	5.50	5.04	5.96
7	2	8.92	5.99	5.04	6.70
8	2	8.80	6.34	5.04	7.12
9	2	5.62	6.26	5.04	6.87
10	2	9.04	6.54	5.04	7.18
11	2	8.84	6.75	5.04	7.39
12	2	-0.02	6.19	5.04	6.57
13	2	6.04	6.18	5.04	6.51
14	1	6.12	6.17	5.31	6.51
15	2	9.01	6.36	5.31	6.74
16	2	0.39	5.99	5.31	6.21
17	2	11.82	6.33	5.31	6.64
18	2	6.22	6.33	5.31	6.61
19	2	6.07	6.31	5.31	6.58
20	2	6.03	6.30	5.31	6.54
21	2	8.86	6.42	5.31	6.68
22	2	3.20	6.27	5.31	6.49
23	2	6.02	6.26	5.31	6.46
24	2	8.97	6.38	5.31	6.59
25	2	15.13	6.73	5.31	7.00
26	2	11.87	6.93	5.31	7.22
27	2	11.81	7.11	5.31	7.42
28	2	6.00	7.07	5.31	7.36
29	2	5.98	7.03	5.31	7.30
30	2	6.07	7.00	5.31	7.26
31	2	11.87	7.16	5.31	7.43
32	2	6.00	7.12	5.31	7.38
33	2	-0.02	6.90	5.31	7.12
34	2	8.92	6.96	5.31	7.18
35	2	5.83	6.93	5.31	7.14
36	1	8.87	6.98	6.03	7.14
37	2	5.98	6.96	6.03	7.10
38	2	6.00	6.93	6.03	7.07
39	2	6.04	6.91	6.03	7.04
40	2	6.02	6.89	6.03	7.01
41	2	5.62	6.86	6.03	6.97
42	2	5.83	6.83	6.03	6.94
43	2	6.04	6.81	6.03	6.92
44	2	8.80	6.86	6.03	6.97
45	2	6.03	6.84	6.03	6.94
46	2	9.01	6.89	6.03	6.99
47	2	8.80	6.93	6.03	7.04
48	2	5.62	6.90	6.03	7.00
49	2	8.97	6.94	6.03	7.05
50	2	11.82	7.04	6.03	7.15
51	2	3.20	6.97	6.03	7.07
52	2	0.39	6.84	6.03	6.93
53	2	8.80	6.88	6.03	6.97
54	2	6.00	6.86	6.03	6.95
55	2	5.98	6.84	6.03	6.93
56	2	6.04	6.83	6.03	6.91
57	2	8.80	6.87	6.03	6.95
58	2	9.04	6.90	6.03	6.99

TABLE D-3a

SIMULATION WITH MINIMAX RULE - FIRST RUN

Assumed $\sigma_1 = \sigma_2 = 3.0$

n	π_j	$_nX_j$	S_n/n	$\mu_1 = 6.60$ $_n\bar{X}_1$	$\mu_2 = 7.25$ $_n\bar{X}_2$
1	1	8.81	8.81	8.81	0.00
2	2	11.82	10.32	8.81	11.82
3	1	5.78	8.81	7.30	11.82
4	2	9.01	8.86	7.30	10.42
5	1	6.23	8.33	6.94	10.42
6	2	5.98	7.94	6.94	8.94
7	1	11.81	8.50	8.16	8.94
8	2	6.07	8.19	8.16	8.22
9	1	8.81	8.26	8.29	8.22
10	2	15.13	8.95	8.29	9.61
11	1	11.82	9.21	8.88	9.61
12	2	6.22	8.96	8.88	9.04
13	1	11.77	9.18	9.29	9.04
14	2	3.20	8.75	9.29	8.21
15	1	3.34	8.39	8.55	8.21
16	2	8.92	8.42	8.55	8.30
17	1	3.23	8.12	7.96	8.30
18	2	6.04	8.00	7.96	8.05
19	1	8.74	8.04	8.04	8.05
20	2	6.02	7.94	8.04	7.85
21	1	8.71	7.98	8.10	7.85
22	2	6.12	7.89	8.10	7.69
23	1	8.79	7.93	8.16	7.69
24	2	9.04	7.98	8.16	7.80
25	1	3.09	7.78	7.77	7.80
26	2	5.83	7.71	7.77	7.65
27	1	0.54	7.44	7.25	7.65
28	2	8.80	7.49	7.25	7.73
29	1	5.97	7.44	7.17	7.73
30	2	11.81	7.59	7.17	8.01
31	1	5.76	7.53	7.08	8.01
32	2	5.62	7.47	7.08	7.86
33	1	6.25	7.43	7.03	7.86
34	2	-0.02	7.21	7.03	7.39
35	1	8.81	7.26	7.13	7.39
36	2	8.86	7.30	7.13	7.47
37	1	5.79	7.26	7.06	7.47
38	2	8.84	7.30	7.06	7.55
39	1	5.78	7.26	7.00	7.55
40	2	6.03	7.23	7.00	7.47
41	1	6.23	7.21	6.96	7.47
42	2	11.87	7.32	6.96	7.68
43*	2	5.92	7.29	6.96	7.60
44	2	0.39	7.13	6.96	7.29
45	2	6.07	7.11	6.96	7.24
46	2	8.97	7.15	6.96	7.31
47	2	6.00	7.12	6.96	7.26
48	2	8.86	7.16	6.96	7.32
49	2	5.98	7.14	6.96	7.27
50	2	5.62	7.11	6.96	7.21
51	2	5.92	7.08	6.96	7.17
52	2	3.20	7.01	6.96	7.04
53	2	8.86	7.04	6.96	7.10
54	2	5.83	7.02	6.96	7.06
55	2	6.00	7.00	6.96	7.03
56	2	6.22	6.99	6.96	7.01
57	2	9.04	7.02	6.96	7.06
58	2	3.20	6.96	6.96	6.96

TABLE D-3b

SIMULATION WITH MINIMAX RULE - SECOND RUN

Assumed $\sigma_1 = \sigma_2 = 3.0$

				$\mu_1 = 6.60$	$\mu_2 = 7.25$
n	π_j	$_nX_j$	S_n/n	$_n\bar{X}_1$	$_n\bar{X}_2$
1	1	6.23	6.23	6.23	0.00
2	2	5.98	6.11	6.23	5.98
3	1	11.81	8.01	9.02	5.98
4	2	6.07	7.53	9.02	6.03
5	1	8.81	7.78	8.95	6.03
6	2	15.13	9.01	8.95	9.06
7	1	11.82	9.41	9.67	9.06
8	2	6.22	9.01	9.67	8.35
9	1	11.77	9.32	10.09	8.35
10	2	3.20	8.71	10.09	7.32
11*	1	3.34	8.22	8.97	7.32
12	1	3.23	7.81	8.15	7.32
13	1	8.74	7.88	8.22	7.32
14	1	8.71	7.94	8.28	7.32
15	1	8.79	7.99	8.33	7.32
16	1	3.09	7.69	7.85	7.32
17	1	8.81	7.75	7.93	7.32
18	1	5.97	7.66	7.78	7.32
19	1	5.76	7.56	7.64	7.32
20	1	6.25	7.49	7.55	7.32
21	1	8.81	7.55	7.63	7.32
22	1	5.79	7.47	7.52	7.32
23	1	5.78	7.40	7.42	7.32
24	1	6.23	7.35	7.36	7.32
25	1	3.31	7.19	7.16	7.32
26	1	5.98	7.14	7.10	7.32
27	1	3.08	6.99	6.92	7.32
28	1	3.11	6.86	6.75	7.32
29	1	8.72	6.92	6.84	7.32
30	1	5.67	6.88	6.79	7.32
31	1	2.77	6.75	6.63	7.32
32	1	8.86	6.81	6.72	7.32
33	1	5.78	6.78	6.68	7.32
34	1	8.74	6.84	6.76	7.32
35	1	6.12	6.82	6.73	7.32
36	1	0.54	6.64	6.53	7.32
37	1	8.87	6.70	6.61	7.32
38	1	11.82	6.84	6.77	7.32
39	1	3.11	6.74	6.66	7.32
40	1	8.74	6.79	6.72	7.32
41	1	8.74	6.84	6.77	7.32
42	1	8.74	6.89	6.83	7.32
43	1	3.09	6.80	6.73	7.32
44	1	8.74	6.84	6.78	7.32
45	1	2.77	6.75	6.68	7.32
46	1	5.97	6.74	6.66	7.32
47	1	5.79	6.72	6.64	7.32
48	1	3.34	6.65	6.57	7.32
49	1	5.97	6.63	6.55	7.32
50	1	3.08	6.56	6.48	7.32
51	1	11.82	6.66	6.59	7.32
52	1	5.79	6.65	6.58	7.32
53	1	6.23	6.64	6.57	7.32
54	1	3.08	6.57	6.50	7.32
55	1	8.74	6.61	6.54	7.32
56	1	3.31	6.55	6.48	7.32
57	1	6.23	6.55	6.47	7.32
58	1	8.74	6.59	6.52	7.32

TABLE D-3c

SIMULATION WITH MINIMAX RULE - THIRD RUN

Assumed $\sigma_1 = \sigma_2 = 3.0$

n	π_j	$_nX_j$	S_n/n	$\mu_1=6.60$ $_n\bar{X}_1$	$\mu_2=7.25$ $_n\bar{X}_2$
1	1	8.81	8.81	8.81	0.00
2	2	15.13	11.97	8.81	15.13
3	1	11.82	11.92	10.32	15.13
4	2	6.22	10.50	10.32	10.68
5	1	11.77	10.75	10.80	10.68
6	2	3.20	9.50	10.80	8.19
7	1	3.34	8.62	8.94	8.19
8	2	6.07	8.30	8.94	7.66
9	1	3.23	7.74	7.60	7.66
10	2	6.04	7.57	7.80	7.34
11	1	8.74	7.67	7.96	7.34
12	2	6.02	7.54	7.96	7.12
13	1	8.71	7.63	8.06	7.12
14	2	6.12	7.52	8.06	6.98
15	1	8.79	7.61	8.16	6.98
16	2	9.04	7.70	8.16	7.23
17	1	3.09	7.42	7.59	7.23
18	2	5.83	7.34	7.59	7.08
19	1	11.81	7.57	8.02	7.08
20	2	8.80	7.63	8.02	7.25
21	1	5.97	7.55	7.83	7.25
22	2	9.01	7.62	7.83	7.41
23	1	5.76	7.54	7.66	7.41
24	2	5.62	7.46	7.66	7.26
25	1	6.25	7.41	7.55	7.26
26	2	-0.02	7.13	7.55	6.70
27	1	8.81	7.19	7.64	6.70
28	2	8.86	7.25	7.64	6.86
29	1	5.79	7.20	7.52	6.86
30	2	5.98	7.16	7.52	6.80
31	1	5.78	7.11	7.41	6.80
32	2	11.82	7.26	7.41	7.11
33	1	6.23	7.23	7.34	7.11
34	2	11.87	7.37	7.34	7.39
35	1	3.31	7.25	7.12	7.39
36	2	5.92	7.21	7.12	7.31
37	1	5.98	7.18	7.06	7.31
38	2	0.39	7.00	7.06	6.95
39	1	3.08	6.90	6.86	6.95
40	2	6.07	6.88	6.86	6.90
41	1	3.11	6.79	6.68	6.90
42	2	8.97	6.84	6.68	7.00
43	1	8.72	6.89	6.77	7.00
44	2	6.00	6.87	6.77	6.96
45	1	5.67	6.84	6.73	6.96
46	2	6.03	6.82	6.73	6.92
47	1	2.77	6.74	6.56	6.92
48	2	11.81	6.84	6.56	7.12
49*	2	8.84	6.88	6.56	7.19
50	2	8.92	6.92	6.56	7.26
51	2	6.07	6.91	6.56	7.21
52	2	6.02	6.89	6.56	7.17
53	2	-0.02	6.76	6.56	6.92
54	2	6.02	6.75	6.56	6.89
55	2	0.39	6.63	6.56	6.68
56	2	6.04	6.62	6.56	6.66
57	2	9.04	6.66	6.56	6.74
58	2	6.03	6.65	6.56	6.71

TABLE D-4a

SIMULATION WITH BACKWARDS INDUCTION RULE - FIRST RUN

Assumed $\sigma_1=\sigma_2=3.0$ *Indicates observation from n^b

n	π_j	$_nX_j$	S_n/n	$\mu_1=6.60$ $_n\bar{X}_1$	$\mu_2=7.25$ $_n\bar{X}_2$
1	1	6.23	6.23	6.23	0.00
2	2	5.98	6.11	6.23	5.98
3	1	11.81	8.01	9.02	5.98
4	1	8.81	8.21	8.95	5.98
5	1	11.82	8.93	9.67	5.98
6	1	11.77	9.41	10.09	5.98
7	1	3.34	8.54	8.97	5.98
8	1	3.23	7.88	8.15	5.98
9	1	8.74	7.97	8.22	5.98
10	1	8.71	8.05	8.28	5.98
11	1	8.79	8.12	8.33	5.98
12	1	3.09	7.70	7.85	5.98
13	1	8.81	7.78	7.93	5.98
14	1	5.97	7.65	7.78	5.98
15	1	5.76	7.53	7.64	5.98
16	1	6.25	7.45	7.55	5.98
17	1	8.81	7.53	7.63	5.98
18	1	5.79	7.43	7.52	5.98
19	1	5.78	7.35	7.42	5.98
20	1	6.23	7.29	7.36	5.98
21	1	3.31	7.10	7.16	5.98
22	2*	6.07	7.05	7.16	6.03
23	1	5.98	7.01	7.10	6.03
24	1	3.08	6.84	6.92	6.03
25	1	3.11	6.70	6.75	6.03
26	1	8.72	6.77	6.84	6.03
27	1	5.67	6.73	6.79	6.03
28	1	2.77	6.59	6.63	6.03
29	2*	15.13	6.89	6.63	9.06
30	2	6.22	6.86	6.63	8.35
31	2	3.20	6.75	6.63	7.32
32	2	8.92	6.81	6.63	7.59
33	2	6.04	6.79	6.63	7.37
34	2	6.02	6.77	6.63	7.20
35	2	6.12	6.75	6.63	7.08
36	2	9.04	6.81	6.63	7.28
37	2	5.83	6.79	6.63	7.15
38	2	8.80	6.84	6.63	7.29
39	2	9.01	6.90	6.63	7.42
40	2	5.62	6.86	6.63	7.29
41	2	-0.02	6.70	6.63	6.80
42	2	8.86	6.75	6.63	6.93
43	2	8.84	6.80	6.63	7.04
44	2	11.82	6.91	6.63	7.31
45	2	11.87	7.02	6.63	7.55
46	2	5.92	7.00	6.63	7.47
47	2	0.39	6.86	6.63	7.13
48	2	6.07	6.84	6.63	7.08
49	2	8.97	6.88	6.63	7.17
50	2	6.00	6.87	6.63	7.12
51	2	6.03	6.85	6.63	7.07
52	2	11.81	6.95	6.63	7.26
53	2	15.13	7.10	6.63	7.55
54	2	11.87	7.19	6.63	7.70
55	2	6.04	7.17	6.63	7.65
56	2	8.92	7.20	6.63	7.69
57	2	8.92	7.23	6.63	7.73
58	2	6.12	7.21	6.63	7.68

TABLE D-4b

SIMULATION WITH BACKWARDS INDUCTION RULE - SECOND RUN

Assumed $\sigma_1 = \sigma_2 = 3.0$ *Indicates observation from n^b

n	π_j	$_nX_j$	S_n/n	$\mu_1=6.60$ $_n\bar{X}_1$	$\mu_2=7.25$ $_n\bar{X}_2$
1	1	6.23	6.23	6.23	0.00
2	2	6.07	6.15	6.23	6.07
3	1	3.34	5.22	4.79	6.07
4	2	6.12	5.44	4.79	6.10
5	2	5.62	5.48	4.79	5.94
6	2	5.98	5.56	4.79	5.95
7	2	6.03	5.63	4.79	5.97
8	2	-0.02	4.92	4.79	4.97
9	1*	3.31	4.75	4.30	4.97
10	1*	3.23	4.59	4.03	4.97
11	2	8.92	4.99	4.03	5.53
12	2	8.80	5.31	4.03	5.94
13	2	9.01	5.59	4.03	6.28
14	2	15.13	6.27	4.03	7.17
15	2	8.86	6.45	4.03	7.32
16	2	6.22	6.43	4.03	7.23
17	2	3.20	6.24	4.03	6.92
18	2	0.39	5.92	4.03	6.46
19	2	6.02	5.92	4.03	6.43
20	2	5.92	5.92	4.03	6.40
21	2	6.07	5.93	4.03	6.38
22	2	11.82	6.20	4.03	6.68
23	2	6.04	6.19	4.03	6.65
24	2	9.04	6.31	4.03	6.77
25	2	8.97	6.42	4.03	6.87
26	2	11.87	6.63	4.03	7.10
27	2	5.83	6.60	4.03	7.04
28	2	8.84	6.66	4.03	7.12
29	2	11.81	6.86	4.03	7.31
30	2	6.00	6.83	4.03	7.26
31	2	6.22	6.81	4.03	7.22
32	2	11.82	6.96	4.03	7.38
33	2	0.39	6.77	4.03	7.14
34	2	6.00	6.74	4.03	7.10
35	2	-0.02	6.55	4.03	6.87
36	2	15.13	6.79	4.03	7.13
37	2	6.03	6.77	4.03	7.10
38	2	11.82	6.90	4.03	7.24
39	2	-0.02	6.72	4.03	7.03
40	2	6.12	6.71	4.03	7.01
41	2	5.98	6.69	4.03	6.98
42	2	3.20	6.61	4.03	6.88
43	2	8.97	6.66	4.03	6.93
44	2	11.82	6.78	4.03	7.05
45	2	8.84	6.83	4.03	7.10
46	2	15.13	7.01	4.03	7.29
47	2	11.87	7.11	4.03	7.40
48	2	6.02	7.09	4.03	7.37
49	2	6.12	7.07	4.03	7.34
50	2	6.04	7.05	4.03	7.31
51	2	6.03	7.03	4.03	7.28
52	2	15.13	7.18	4.03	7.45
53	2	6.07	7.16	4.03	7.42
54	2	6.02	7.14	4.03	7.39
55	2	11.82	7.23	4.03	7.48
56	2	0.39	7.10	4.03	7.34
57	2	6.00	7.09	4.03	7.32
58	2	5.92	7.07	4.03	7.29

TABLE D-4c

SIMULATION WITH BACKWARDS INDUCTION RULE - THIRD RUN

n	π_j	${}_nX_j$	S_n/n	$\mu_1=6.60$ ${}_n\bar{X}_1$	$\mu_2=7.25$ ${}_n\bar{X}_2$
1	1	11.81	11.81	11.81	0.00
2	2	8.80	10.31	11.81	8.80
3	1	3.23	7.95	7.52	8.80
4	2	11.87	8.93	7.52	10.34
5	2	5.98	8.34	7.52	8.89
6	2	6.07	7.96	7.52	8.18
7	2	15.13	8.99	7.52	9.57
8	2	6.22	8.64	7.52	9.02
9	2	3.20	8.04	7.52	8.19
10	2	8.92	8.13	7.52	8.28
11	2	6.04	7.94	7.52	8.03
12	1*	6.23	7.80	7.09	8.03
13	2	6.02	7.66	7.09	7.83
14	2	6.12	7.55	7.09	7.67
15	2	9.04	7.65	7.09	7.79
16	2	5.83	7.54	7.09	7.64
17	2	6.03	7.45	7.09	7.52
18	1*	0.54	7.06	5.46	7.52
19	2	9.01	7.17	5.46	7.62
20	2	5.62	7.09	5.46	7.50
21	2	-0.02	6.75	5.46	7.06
22	2	8.86	6.85	5.46	7.16
23	2	8.84	6.93	5.46	7.25
24	2	11.82	7.14	5.46	7.47
25	2	11.81	7.33	5.46	7.68
26	2	5.92	7.27	5.46	7.60
27	2	0.39	7.02	5.46	7.29
28	2	6.07	6.98	5.46	7.24
29	2	8.97	7.05	5.46	7.31
30	2	6.00	7.02	5.46	7.26
31	2	3.20	6.89	5.46	7.11
32	2	6.07	6.87	5.46	7.07
33	2	6.03	6.84	5.46	7.03
34	2	5.98	6.82	5.46	7.00
35	2	5.98	6.79	5.46	6.97
36	2	6.02	6.77	5.46	6.94
37	2	11.81	6.91	5.46	7.08
38	2	6.12	6.89	5.46	7.06
39	2	6.12	6.87	5.46	7.03
40	2	11.82	6.99	5.46	7.16
41	2	6.12	6.97	5.46	7.14
42	2	0.39	6.81	5.46	6.96
43	2	8.80	6.86	5.46	7.01
44	2	15.13	7.05	5.46	7.21
45	2	5.62	7.02	5.46	7.17
46	2	8.86	7.06	5.46	7.21
47	2	5.98	7.03	5.46	7.18
48	2	5.62	7.01	5.46	7.15
49	2	5.92	6.98	5.46	7.12
50	2	3.20	6.91	5.46	7.03
51	2	8.86	6.95	5.46	7.07
52	2	5.83	6.92	5.46	7.05
53	2	6.00	6.91	5.46	7.03
54	2	6.22	6.89	5.46	7.01
55	2	9.04	6.9[illegible]	5.46	7.05
56	2	3.20	6.87	5.46	6.98
57	2	6.07	6.85	5.46	6.96
58	2	6.00	6.84	5.46	6.94