

R-1721-ARPA

March 1976

CAAPM: Computer-Aided Admissible Probability Measurement on PLATO IV

Suzanne Landa

A Report prepared for
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Rand
SANTA MONICA, CA. 90406

The research described in this Report was sponsored by the Defense Advanced Research Projects Agency under contract No. DAHC15-73-C-0181. Reports of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

PREFACE

This report was prepared as part of Rand's DoD Training and Manpower Management Program, sponsored by the Human Resources Research Office of the Defense Advanced Research Projects Agency (ARPA). With manpower issues assuming an ever greater importance in defense planning and budgeting, it is the purpose of this research program to develop broad strategies and specific solutions for dealing with present and future military manpower problems. This includes the development of new research methodologies for examining broad classes of manpower problems, as well as specific problem-oriented research.

The report describes the use of a computer program based on "admissible probability measurement," a testing procedure that permits a user to express his (or her) degree of certainty as to the correctness of alternative answers to multiple-choice questions. The program is implemented on PLATO IV, an interactive computer-based education system available at many military sites.

Two previous Rand reports prepared under the DoD Training and Manpower Management Program provide additional information on admissible probability testing. *Rationale of Computer-Administered Admissible Probability Measurement*, R-1371-ARPA, by E. H. Shuford, Jr., and T. A. Brown, presents the theory behind admissible probability testing. An experimental computer implementation of admissible probability testing techniques at The Rand Corporation is described in *A Prototype Computer Program for Interactive Computer-Administered Admissible Probability Measurement*, R-1258-ARPA, by W. L. Sibley. Also for the purpose of testing the theory, a version available on a portable, stand-alone terminal was developed based on the prototype implementation.

This report is a guide for instructors and students interested in using computer-aided admissible probability measurement (CAAPM) on PLATO IV. The instructional text in Secs. III and IV is also available online: the tutorial for students is found in lesson rand3 on PLATO; the author guide in lesson rand10.

SUMMARY

This report provides instructions on how to create and take computer-aided admissible probability measurement (CAAPM) tests using programs available on PLATO IV. Admissible probability measurement, as discussed in Sec. I, is a testing procedure that permits a user to express a degree of uncertainty as to the correctness of alternative answers.

Section II describes PLATO IV, an interactive computer-based education system on which CAAPM resides. Section III presents a tutorial for students on how to take CAAPM tests, demonstrating how the standard test-taking approach may force the student to guess an answer, whereas CAAPM lets the student express his (or her) uncertainty. Section IV tells instructors how to create CAAPM tests easily and quickly on the PLATO IV system,^{*} and Sec. V discusses recommended extensions to CAAPM. For example, the use of slides, touch panels, and random-access audio devices could enhance the testing capabilities of the system. Student data collection and analysis should also be added. The external validity graph is illustrated and explained in App. A, and standard tests available in CAAPM format on Plato IV are included in App. B.

^{*}Sections III and IV are also available online in lessons rand3 and rand10, respectively.

ACKNOWLEDGMENTS

The author would like to thank Robert H. Anderson, William L. Sibley, and Thomas A. Brown (consultant) of The Rand Corporation for their continued guidance and support throughout the project. In addition, the generous contributions in time and effort of Eileen Sweeney, Craig Burson, Alec Hahwich, Larry Francis, and Robert Bohn of the Computer-Based Education Research Laboratory (CERL), University of Illinois, are gratefully acknowledged.

CONTENTS

PREFACE	iii
SUMMARY	v
ACKNOWLEDGMENTS	vii
Section	
I. INTRODUCTION	1
II. COMPUTER SYSTEM DESCRIPTION	3
PLATO IV System	3
PLATO Terminal and Keyboard	3
Getting to CAAPM	7
III. STUDENT GUIDE	8
IV. AUTHOR GUIDE	20
V. RECOMMENDED EXTENSIONS TO CAAPM	26
Input/Output Devices	26
Instruction Capability	27
Test Effectiveness Analysis	27
Student Data Collection and Analysis	27
Appendix	
A. THE EXTERNAL VALIDITY GRAPH	29
B. SAMPLE TESTS	32
REFERENCES	42

I. INTRODUCTION

When a student submits an essay answer to a question he is graded as if he believed 100 percent in that answer. On a multiple-choice type of test he is asked to select one right answer. In either case the student may guess, and thus his response will not reflect his true state of knowledge on the subject. Admissible probability measurement (APM) is a testing procedure that permits a user to express his degree of certainty or uncertainty as to the correctness of his answer.

The multiple choice type of test is used in the current implementation of APM. Three possible answers are displayed at the corners of a triangle. The student moves a cursor around this space toward or away from the possible answers. The spatial relationship of the cursor position to each corner reflects the user's beliefs as to the correctness of each answer. This approach permits use of APM without extensive explanation on subjective probability.

Effective use of APM dictates its implementation on an interactive computer system. It was first made available at Rand on the Video Graphics System [1]. This version of APM, however, could only be demonstrated at Rand. We next developed APM on an IMLAC, a stand-alone portable terminal using cassettes for input and output. This enabled us to demonstrate APM both here and at other sites. The final version, described in this report, resides on the PLATO IV system, an interactive computer-based education system available at many military sites [2]. In addition to being widely available for demonstration, APM may now be used by instructors and students in actual courses.

It is the purpose of this report to provide instructions on how to create and take admissible probability tests using programs available on PLATO IV. The report includes:

- o A description of the computer system, PLATO IV, on which the CAAPM programs reside and an explanation of how to use these programs.
- o A tutorial for students on how to take CAAPM tests.
- o A guide for authors on how to create CAAPM tests.

- o A discussion of recommended extensions to the current CAAPM design.
- o An illustration of the external validity graph.
- o Sample tests using CAAPM available on PLATO IV.

II. COMPUTER SYSTEM DESCRIPTION

PLATO IV SYSTEM

The PLATO system is an interactive computer-based education system developed over the last decade by the Computer-Based Education Research Laboratory (CERL) at the University of Illinois. In its fourth version the system handles more than 700 terminals at over 100 locations throughout the United States, Canada, and Europe. There are presently more than 2000 lessons available in 65 fields of study [3]. Students have logged over 25,000 contact hours of instruction at all school levels on PLATO IV [2]. This version of PLATO is specifically designed to lower the terminal/computer costs to 25-30 cents per student contact hour. For additional information on the history and design of the system, see Ref. 4.

The major hardware components and the communications options for linking the components of the PLATO IV system are identified in Fig. 1 [2].

PLATO TERMINAL AND KEYBOARD

Whether the user is a student or an author of CAAPM, he communicates to PLATO via the keyboard (Fig.2). Students use the six arrow



keys to move the cursor. These keys may be pressed directly without the use of -SHIFT-. The -NEXT- and -DATA- keys are pressed to continue through the lesson. Authors use the standard keyboard characters to enter questions and answers; the -NEXT- key to continue; the SHIFT-STOP to try a test; and -BACK- to back out of a lesson when editing.

The user's communication with PLATO is displayed at the terminal. The parts of the terminal that are used with CAAPM are identified and explained in Fig. 3 [2].

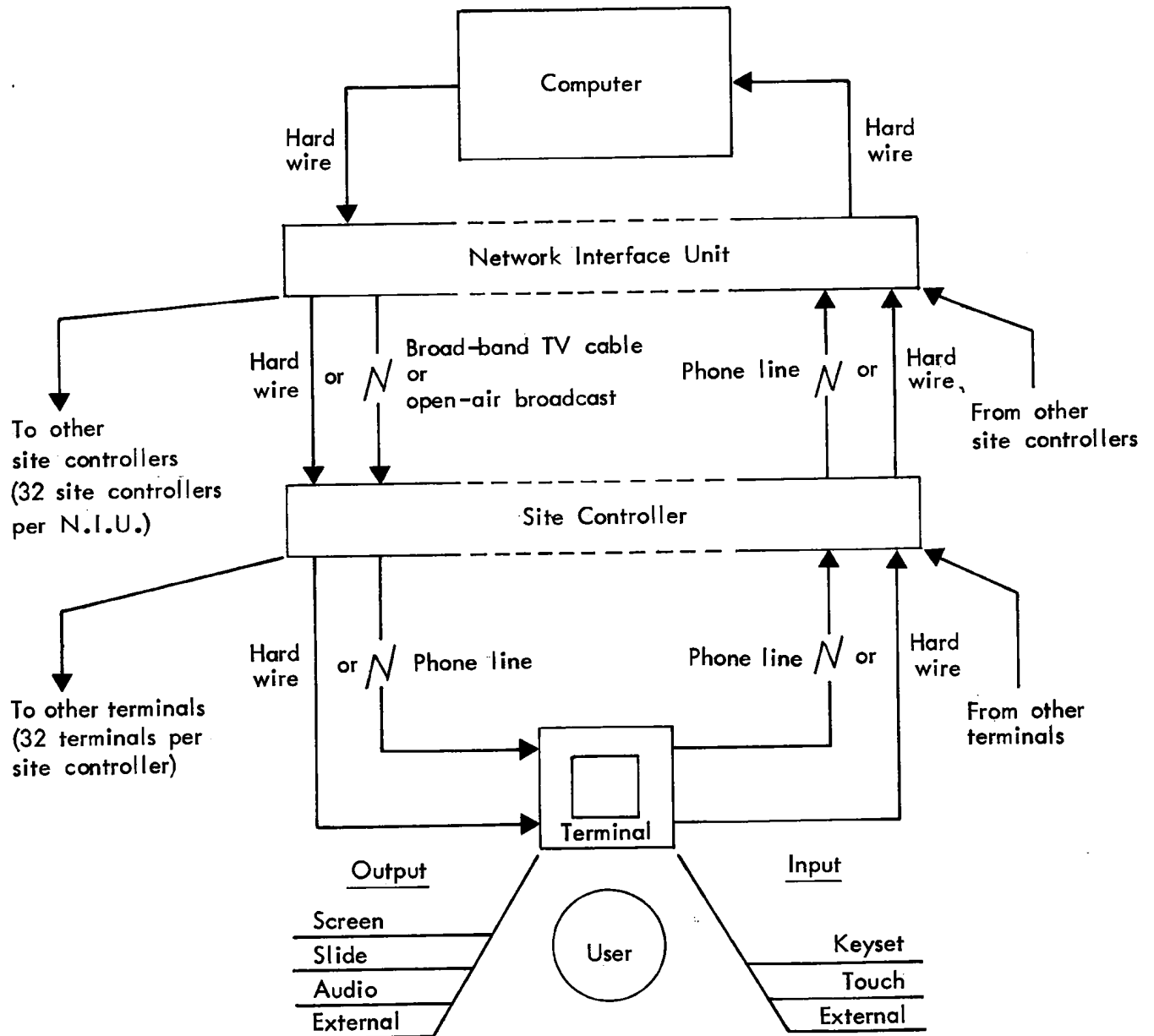


Fig. 1 — Components of Plato IV system

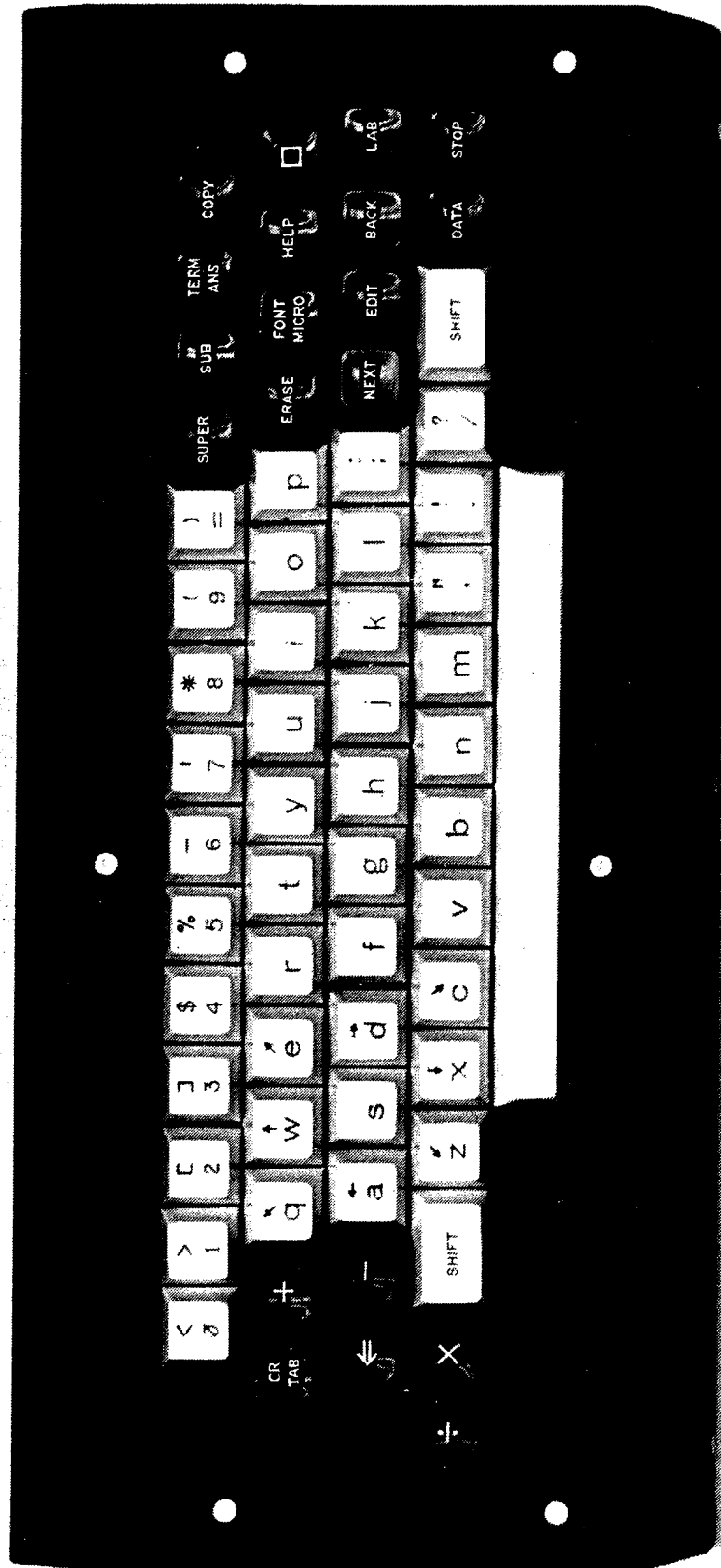
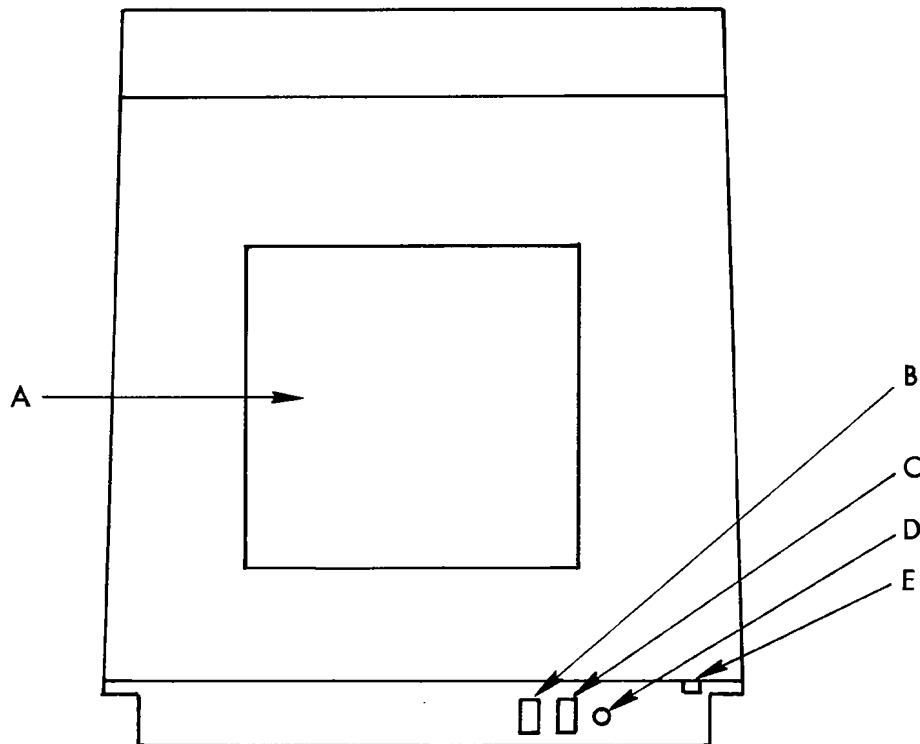


Fig. 2—PLATO IV keyboard



- A - Screen (plasma panel): Displays text and graphics.
- B - Terminal clear switch: Resets the circuitry in the terminal and erases the screen. Use when screen has been cluttered with garbage and/or the error reset switch has not corrected the communication error.
- Error reset switch: Resets the circuitry that indicates a non-correctable communication error has occurred.
- D - Error light: Turns on (red) when communication errors occur.
- E - Power switch: Turns terminal on and off.

Fig. 3—PLATO IV terminal

GETTING TO CAAPM

PLATO may be used only if it recognizes the user's name, course, and password. To enter this information into the system and to request lesson space for new CAAPM tests, authors should contact CERL (217) 333-2374. Students should contact authors for permission to use PLATO and for the name of CAAPM test lessons. Once signed onto the system, a lesson name must be entered. Sections III and IV direct the reader to existing example CAAPM lessons.

III. STUDENT GUIDE

How to take tests using CAAPM will be discussed in this section. An online version of this tutorial is available in lesson rand3 on PLATO.

CAAPM represents a new concept in test-taking. You will have an opportunity to use your knowledge in a novel way to answer questions.

Consider the following analogy question:

Angle is to degree as (ANGLE:DEGREE::)

1. area:square inch
2. time:minute
3. milk:quart

Standard test-taking procedures ask you to respond with *one* choice. If you are certain that the first answer, for example, is the correct answer, then your response accurately reflects your knowledge. However, if you think the correct answer might be the second or third choice, or if you have no idea which is the correct answer, the standard approach forces you to guess and your response no longer reflects your knowledge of the subject. CAAPM is a test-taking procedure that lets you express your uncertainties as well as your certainties.

For each question you will see a triangle with the possible answers displayed near each corner (see Fig. 4). At the beginning of each question a point will be displayed in the center of the triangle. By moving this point within the triangle or along the lines toward or away from a possible answer you will indicate your belief as to the correctness of that answer.

Question: ANGLE: DEGREE::

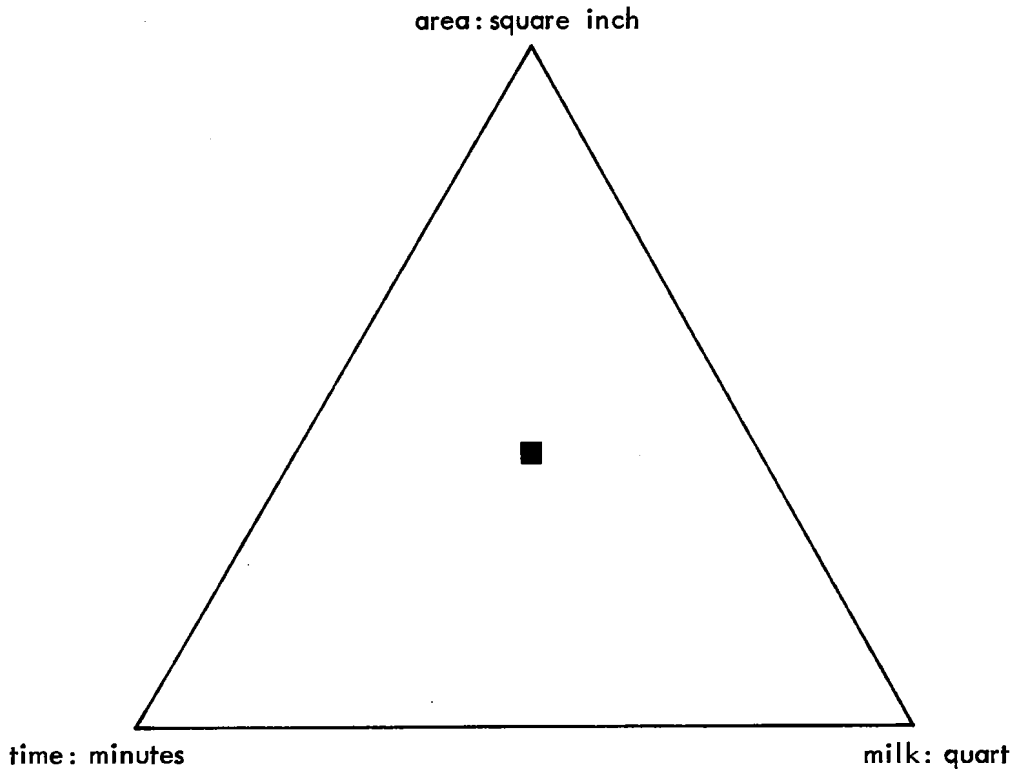


Fig. 4

To move the point, use the six arrow keys on the keyboard (see Fig. 2). The direction in which each arrow will move the point is demonstrated in Fig. 5. Because responses must be given within the triangle, you will not be able to move the point outside the triangle.

Question: ANGLE:DEGREE::

area:square inch

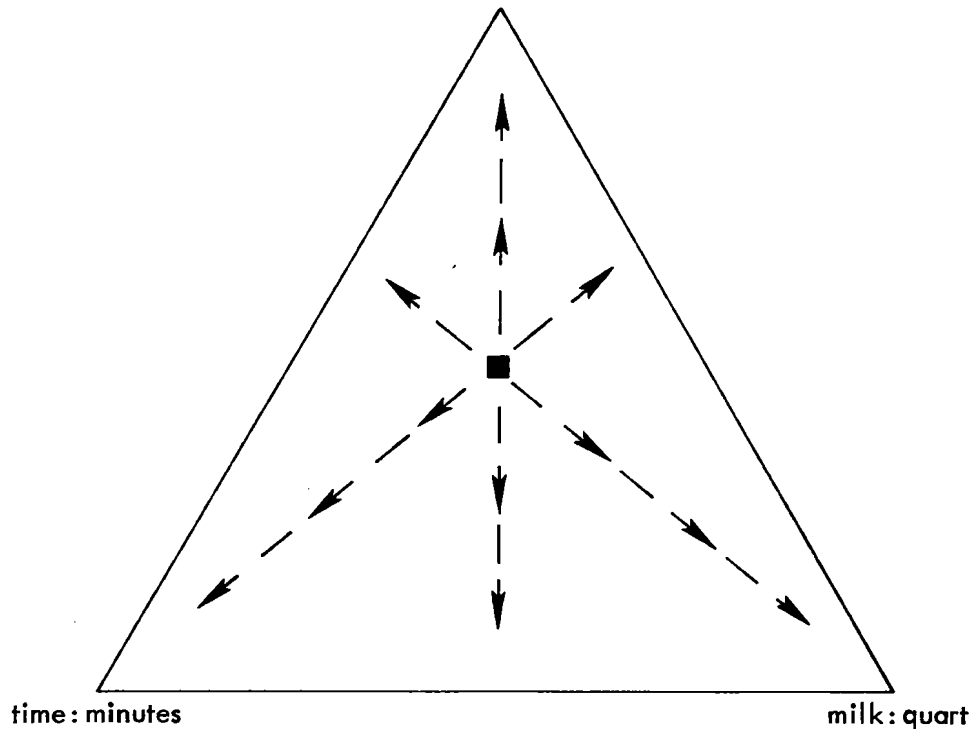


Fig. 5

As you move the point toward the triangle your potential score is displayed at each corner. The scores displayed at the beginning of each question when the point is located in the center of the triangle are 0,0,0 as shown in Fig. 6.

Question: ANGLE:DEGREE::

area : square inch

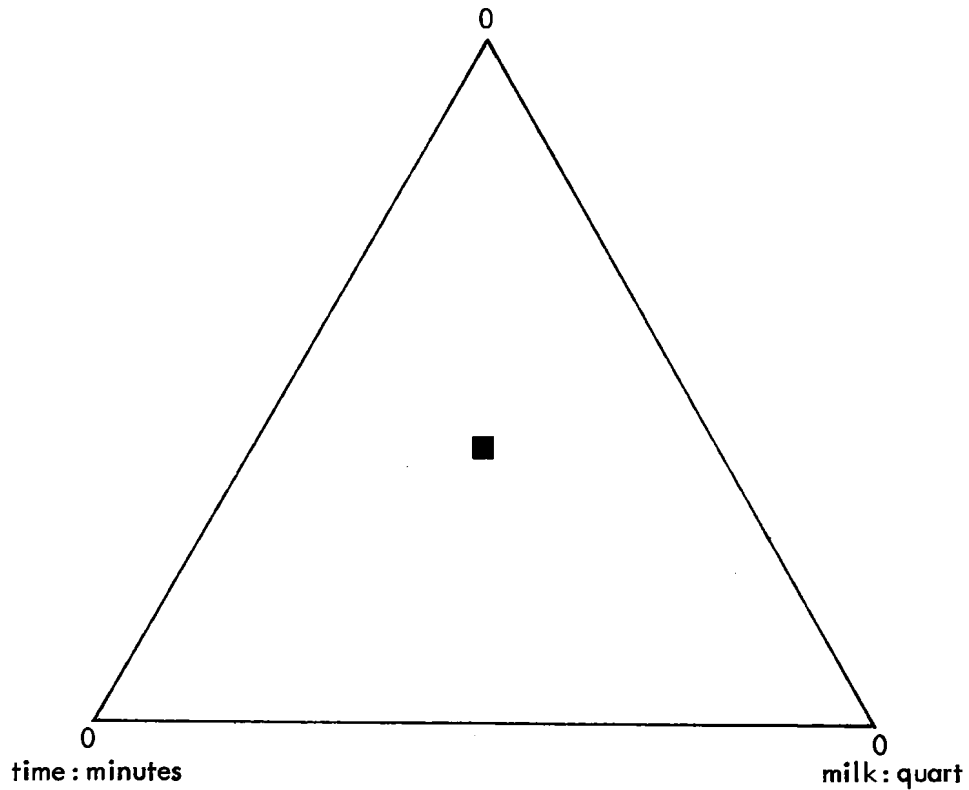


Fig. 6

The maximum potential gain you can make is 23, while the maximum potential loss is 76. These scores are achieved by moving the point to a specific corner. For example, in Fig. 7 the cursor has been moved to the "area:square inch" corner. The 23 means you will gain 23 points *if* the answer "area:square inch" is the correct one. But if you have selected "area:square inch" and it is *not* the correct answer, you will lose 76 points.

Scores do not reveal the correct answer. The number at a corner only indicates your potential gain or loss.

Question: ANGLE:DEGREE::

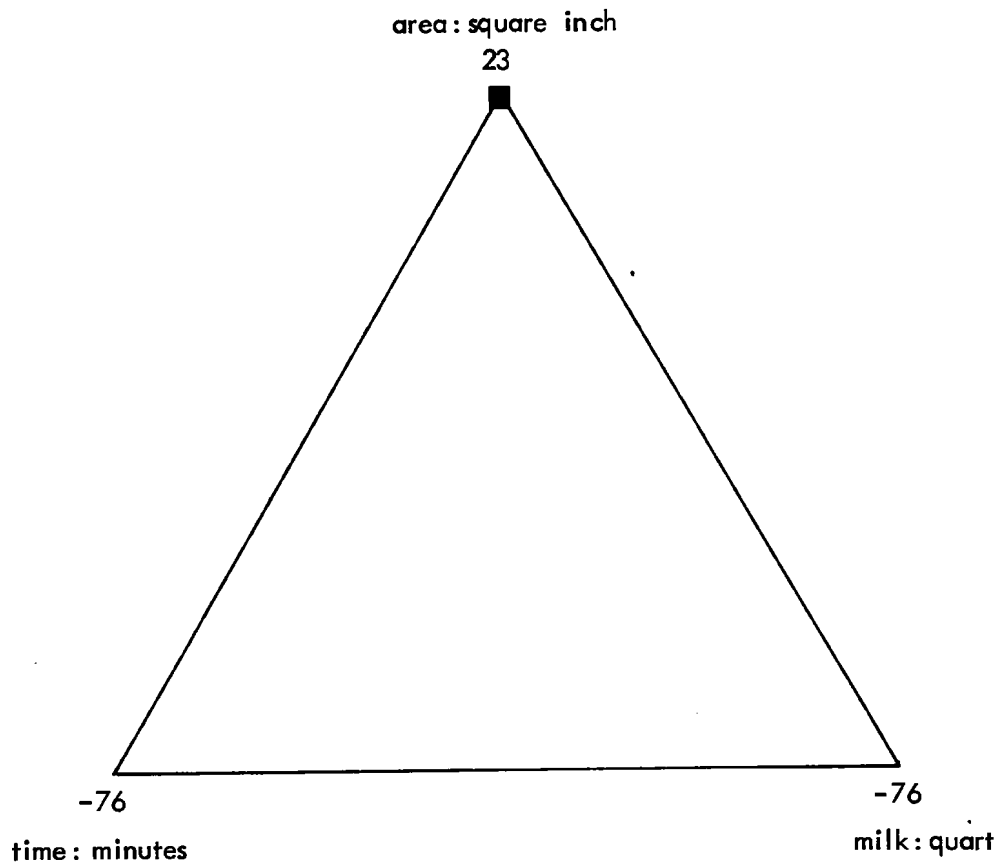


Fig. 7

As you can see, you can lose many points if you choose one answer and it is wrong. Moving the point all the way to one corner is appropriate only if you are certain of the correct answer. Thus if you are not certain which is the correct answer you may wish to reduce your potential loss (as well as gain) by moving the point away from a specific corner. Compare Figs. 8 and 9. Notice that by moving away from the corner just a little you reduce your potential loss by 50 percent (76 to 37) while reducing your potential gain by only 10 percent (23 to 21).

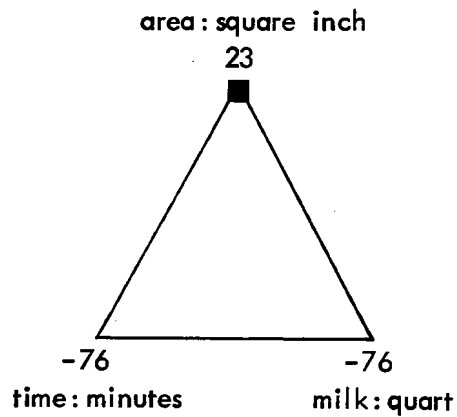


Fig. 8

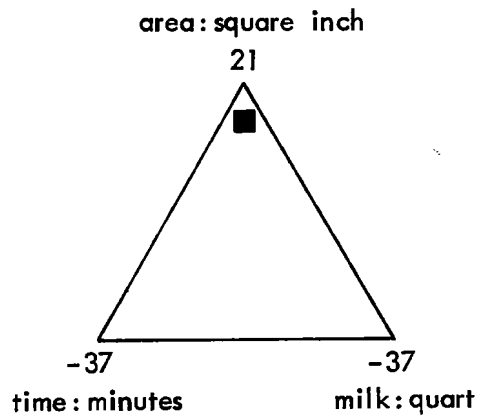


Fig. 9

If you knew, however, that milk:quart was not the correct answer, but you were unsure as to which of the other two were correct, you would move the point to the line between time:minutes and area:square inch as shown in Fig. 10.

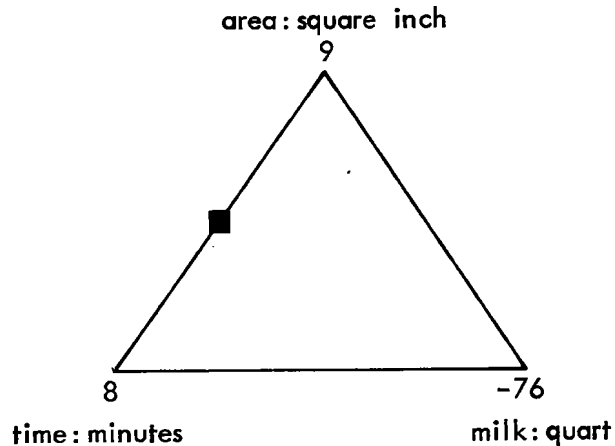


Fig. 10

If you have no idea what the correct answer might be, you may choose to receive no points for the question by positioning the point in the center of the triangle (as in Fig. 5).

You should not guess if you do not know for sure what the answer is. The scoring system is designed to heavily penalize a wrong answer and to encourage you to express your degree of certainty or uncertainty.*

To continue to the next question, you press the -NEXT- key. This instruction will appear at the bottom of each display. If the instruction "press -DATA- for score" also appears, you may check your score on individual questions by pressing -DATA-. This is an option set by your instructor.

*See Ref. 5 for an explanation of the theory behind the scoring systems.

Your response is not recorded until you press -NEXT- or -DATA-; therefore you may adjust the point until you are satisfied with the set of scores.

If you press -DATA- and your instructor has set this option, a display like Fig. 11 will appear. Your score on the current question is flashed next to the correct answer. If, for example, you pressed -DATA- after moving the cursor to the point shown in Fig. 9, the score "21" would flash as shown in Fig. 11 because "area:square inch" is the correct answer. If, however, "time:minutes" were the correct answer, the "-37" would flash. Your cumulative score, the sum of scores achieved on all questions up to and including the current one, is plotted on the graph in relation to the maximum and minimum cumulative score that could be attained on the test as each question is answered.

Question: angle:degree::

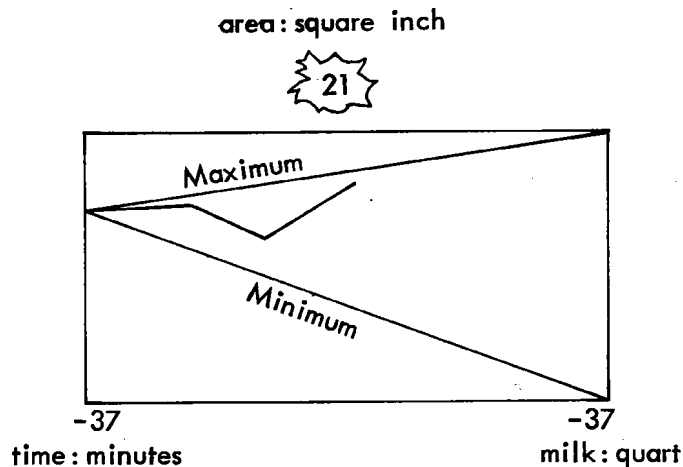
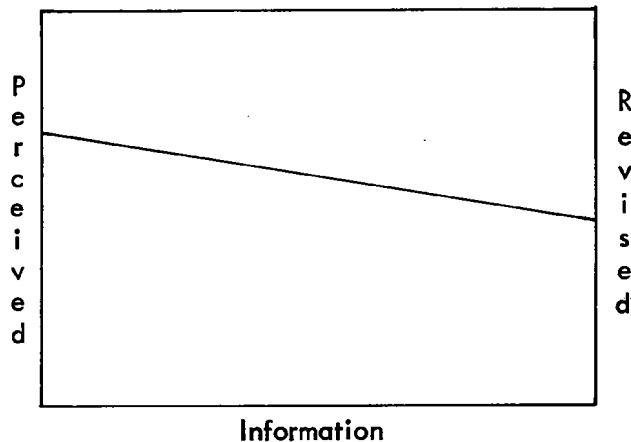


Fig. 11

If you press -DATA- and your instructor has not set this option, the display will not change. To continue, press -NEXT-.

At the completion of the test you will see an assessment of your test performance. An example of a test performance assessment is given in Fig. 12. In the following paragraphs the information presented in a test performance assessment is interpreted. For an explanation of the derivation of the performance diagram and scores, see App. A.



You tend to overvalue your information.

You can improve your score by 37 points
by more realistic use of your knowledge.

You can improve your score by 224 points
by more study.

You can improve your score by 261 overall.

Fig. 12

The graph that appears compares how much you thought you knew (perceived) with an estimate of how much you really knew (revised) on the test. The difference between these two measures indicates your bias. Most people tend to be biased in the use of their information; that is, they tend to overestimate or underestimate the validity of their facts and reasons. You should try to learn to correctly assess the value of your facts and reasons, and thus eliminate all bias from your responses.

If you are successful at this, the line of the information graph will tend to be horizontal for any test you take (Fig. 13). If, however, you overvalue your information the line will slope downward (Fig. 14); if you undervalue your information, the line will slope upward (Fig. 15).

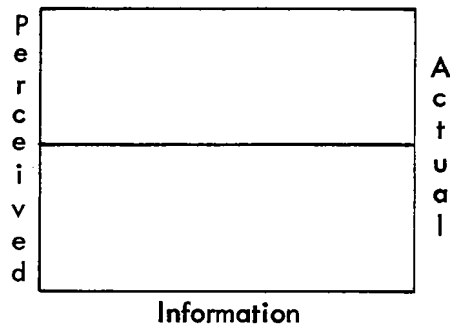


Fig. 13—Correctly valued

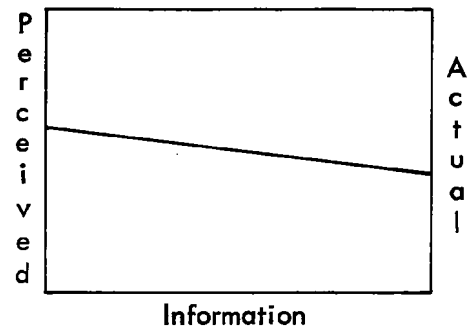


Fig. 14—Overvalued

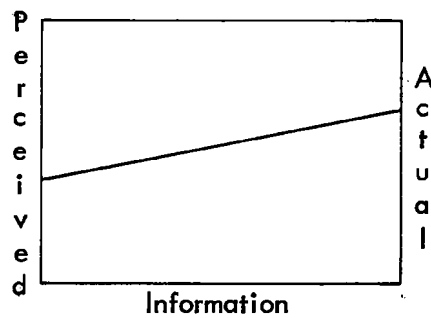


Fig. 15—Undervalued

The height of the line on the information graph indicates your knowledge of the subject matter. In Fig. 16, the student knew very little about the test subject and would gain considerably by more study. In Fig. 17, the student knew the subject matter well. In both cases he correctly valued his information.

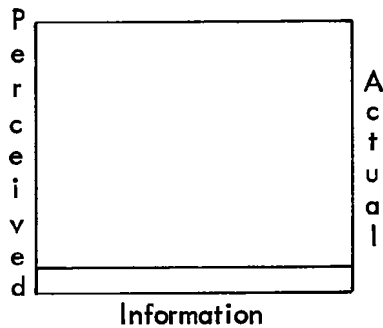


Fig. 16—Subject matter difficult

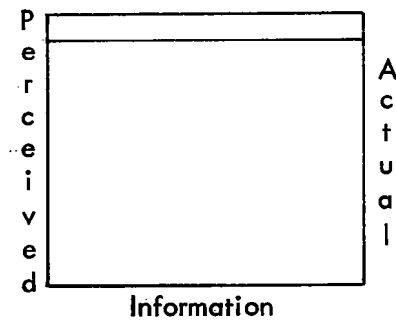


Fig. 17—Subject matter easy

One reason for striving toward removing bias from your response is so that you can accurately communicate degrees of certainty or uncertainty not only to yourself but also to others. Another reason is that you lose points by being biased. In addition to a graphic display of your performance, CAAPM provides a numeric evaluation of this loss by:

1. Estimating a revised test score that you would have made if your responses had been unbiased.
2. Subtracting your test score from this revised test score.
3. Reporting the result as "You can improve your score by _____ points by more realistic use of your information."

When you can improve your score in this manner by less than 10 points, CAAPM displays under the graph, "You tend to correctly value your information." When this potential improvement is 10 or more points, CAAPM displays, "You tend to overvalue your information" or "You tend to undervalue your information," as the case may be.

You can and probably will lose some points as a result of lack of complete information about the subject matter of the test or about the intent of the writer of the test questions. CAAPM evaluates this loss by:

1. Computing the best possible score for the test (in this case, 23 points per question).
2. Subtracting your revised test score from this maximum score.
3. Reporting the result as, "You can improve your score by ____ points by gaining more information about the subject matter."

Your overall improvement score, equal to the sum of the other two improvement scores, is reported by, "You can improve your score by ____ points overall." You can do best by striving to maximize your total test score.

To gain experience with CAAPM try taking one of the following practice tests:

Lesson

- rand4: International Relations
- rand6: Statistics
- rand7: Art
- rand8: Music
- rand9: Literature

IV. AUTHOR GUIDE

How to create CAAPM tests will be discussed in this section; instructions are also available on PLATO in lesson rand10.

Your new test will be written in the TUTOR language used by the PLATO system [6]. Although you will not have to write extensively in the language, you should note the format of TUTOR statements.

write A Test on International Relations

The first part of statement (write) is called the command. The remainder (A Test on International Relations) is called the tag. You may tab over to enter a tag by pressing the key

CR TAB

 on the left side of the keyboard.

To begin, type all statements as they appear inside the box in Fig. 18 into your lesson. The universal delimiter "◄►" is created by pressing the micro key and then the comma key (see Fig. 2). There are seven items you add to this list of statements to create a new test:

1. The number of questions in the test.
2. A score flag.
3. The title of the test.
4. The test questions.
5. The correct answers to the questions.
6. One set of wrong answers.
7. A second set of wrong answers.

These items are entered using the edit command INSERT in the positions indicated by the numbers outside the box in Fig. 18. For instructions on the use of INSERT and other edit commands, press -HELP- while in your lesson. An example of a completed test with these seven items included is given in Fig. 19.

```
(1) { unit          begin
(2) } goto          one
      use          randl, caapt
      use          caapt1
      use          caapt2
      use          caapt3
      use          caapt4
      *
(3) > unit          title
      at          3025
      write       press -NEXT-
      *
      unit          ques
(4) > writec       num ◀▶▶▶▶▶
      *
      unit          corr
(5) > writec       num ◀▶▶▶▶▶
      *
      unit          wrgl
(6) > writec       num
      *
      unit          wrg2
(7) > writec       num ◀▶▶▶▶▶
      *
```

Fig. 18--The skeleton test lesson

```

unit      begin
define    nq=5
          score=1

goto      one
use       rand1,caapt
use       caapt1
use       caapt2
use       caapt3
use       caapt4
*
unit      title
write     A Test on International Relations
at        3025
write     press -NEXT-
*
unit      ques
writec    num<><><><>
          Selling goods for export at prices lower than
          those charged domestic buyers is an example of<>
          The communist theorists expound today's world
          movements as a triumph for the revolutionary concept of<>
          Identify the most important policy advisory body that
          works directly with the Secretary of Defense:<>
          The decision as to which countries the U.S. shall
          recognize diplomatically lies with the<>
          What do we call the strategy of buying goods to prevent
          a rival nation from getting them

*
unit      corr
writec    num<><><><>
          dumping<>Lenin<>Armed Forces
          Policy Council<>President<>preemptive buying

*
unit      wrg1
writec    num<><><><>
          subsidy<>Marx<>National Security
          Council<>Senate<>loading

*
unit      wrg2
writec    num<><><><>
          embargo<>Stalin<>Defense Research and
          Engineering Group<>Secretary of State<>embargo

```

Fig. 19--A completed test lesson

To specify the number of questions in your test you insert the instruction

command	tag
	
define	nq=

entering after the equal sign the appropriate number. In Fig. 19, the number of questions was set equal to 5. In this implementation of CAAPM, the maximum number of questions per test lesson is 9.

As the author, you may choose to reveal or not to reveal to the student the correct answer to each question. You do this by defining the variable score, equal to 1 for yes (display score) and 0 for no (do not display):

$$\text{define score} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

In Fig. 18, score=1, and therefore the correct answer to a question will be displayed when the student presses the -DATA- key.

To specify the title of the test you must insert a write command after the "unit title" instruction and then type in the title. The title may be up to 20 lines in length, each line containing a maximum of 50 characters. The write command is entered only on the first line of the title. For each subsequent line, press -NEXT- and tab directly over to the tag field and continue the title. In Fig. 19, the title "A Test on International Relations" was entered.

The next four items are concerned with the questions and answers for your test. The first question in the question list is associated with: the first correct answer in the correct answer unit "corr"; the first wrong answer in the first wrong answer unit "wrg1"; the first wrong in the second wrong answer unit "wrg2". The fifth, then, in each list is associated with the fifth in the other three lists. If you write out your test as shown in Fig. 20, you will have created the appropriate tags for each of these units.

QUESTIONS		ANSWERS		
QUES	Corr	Wrg1	Wrg2	
1. Selling goods for export at prices lower than those charged domestic buyers is an example of◀▶	dumping◀▶	subsidy◀▶	embargo◀▶	
2. The communist theorists expound today's world movements as a triumph for the revolutionary concept of◀▶	Lenin◀▶	Marx◀▶	Stalin◀▶	
3. Identify the most important policy advisory body that works directly with the Secretary of Defense:◀▶	Armed Forces Policy Council◀▶	National Security Council◀▶	Defense Research and Engineering Group◀▶	
4. The decision as to which countries the U.S. shall recognize diplomatically lies with the◀▶	President◀▶	Senate◀▶	Secretary of State◀▶	
5. What do we call the strategy of buying goods to prevent a rival nation from getting them	preemptive buying	loading	embargo	

Fig. 20--Example question and answer tags

All questions are inserted in the unit "ques" sequentially. The questions must be separated by the universal delimiter "◄►" as shown in Fig. 19. Each question may be up to six lines in length, 60 characters per line.

Answers are inserted in units "corr", "wrg1", and "wrg2". Enter answers in an order to match the questions. As with the questions, each answer must be separated from the next by the universal delimiter "◄►". Each answer may be up to two lines in length, 20 characters per line. For attractive display of answers taking two lines, press -NEXT- at an appropriate midpoint in your answer and continue on the next line. In the first example below, a break occurs between Policy and Council; in the second, between Forces and Policy. The latter alternative is preferred for display.

<u>Tag</u>	<u>Display</u>
dumping ◄► Lenin ◄► Armed Forces Policy Council ◄► President ◄► preemptive buying	Armed Forces Policy Council
dumping ◄► Lenin ◄► Armed Forces Policy Council ◄► President ◄► preemptive buying	Armed Forces Policy Council

After all questions and answers have been inserted, you are ready to try your new test (press -SHIFT- and -STOP- simultaneously). Each time you try your test, note that CAAPM randomizes the order of questions and the location of answers in order to minimize fallacious associations on the part of test repeaters.

V. RECOMMENDED EXTENSIONS TO CAAPM

INPUT/OUTPUT DEVICES

CAAPM uses the keyboard for input and the plasma panel for output. The addition of several other input and output devices available on the PLATO system could enhance the use of CAAPM.

In the current implementation, the student moves the cursor on the screen by pressing the arrow keys on the keyboard. A touch panel is available that enables a point to be indicated on the screen by simply touching, with pencil or finger, that point. The touch panel works on the photoelectric eye principle. It divides the screen area into $16 \times 16 = 256$ half-inch squares [2]. For CAAPM use, however, only approximately 50 squares would be active--that is, those squares that fall within the boundaries of the equilateral triangle. This resolution, similar to the current arrow-key version which recognizes approximately 50 locations within the triangle, is sufficiently accurate. Touch response is especially effective with young children.

In Sec. IV we described how an author could enter a text-type question into CAAPM. But what if the author needs to display a picture (perhaps complex, perhaps even color!) with, or instead of, a text question? For example, a pathology instructor may want to quiz students on visual recognition of diseases by showing them photographs of various growths. Available on PLATO is a slide selector which can select any one of 256 full-color slides rapidly and project the image onto the terminal screen. Overlaying or beneath this picture on the screen, then, would be the CAAPM-generated interactive response triangle.

Effective testing in some areas--for example, in phonics and conversational French--requires the use of audio questioning. PLATO has available a random-access audio device that can select in a half second or less one of up to 4000 messages stored on a 15-inch flat disc. The total message time on one disc is about 23 minutes [2].

Use of these devices with CAAPM would require modifications to the existing program.

INSTRUCTION CAPABILITY

CAAPM is currently designed to administer tests only. Addition of selective feedback of remedial information based on user's assessment probabilities would allow CAAPM to be used for instruction, as well as testing. This would involve expansion of the author mode to support remedial information input as well as changes to the program to permit selective branching.

TEST EFFECTIVENESS ANALYSIS

Creating effective tests--that is, tests that measure what one wants them to measure--demands systematic analysis of questions and answers. This analysis might include, for example, application of the Ease Index (a measure of the ease of a question) and the Differentiation Index (a measure of the effectiveness of a question in differentiating among students) to the test [7]. However, the influence of guessing on these measurements has limited their usefulness. Since the CAAPM approach eliminates guessing, adding these types of measurements to CAAPM should enable authors to create better tests and, as a result, to collect more accurate and informative data on student performance.

STUDENT DATA COLLECTION AND ANALYSIS

PLATO IV provides a means for collecting specific types of data on students, for example, the number of wrong answers provided by the student per question. Since a student is not precisely wrong or right in CAAPM, these data are not collected. A capability needs to be developed in CAAPM to collect and analyze relevant data--for example, scores--on students.

Appendix A
THE EXTERNAL VALIDITY GRAPH*

Even though the student may learn to express his own probabilities accurately, the question remains as to how those probabilities relate to reality. That question can be answered in part by his external validity graph [8].

The external validity graph can be used to determine how accurately the student assesses the predictive validity of his facts and reasons. It can be thought of as his "track record" as an assessor. We may collect all those events that he claims have a probability of, say, 0.80 of occurring and determine what percentage of them did in fact occur. If he were a perfect assessor, we would expect roughly 80 percent of them to occur. Figure A.1 illustrates such a graph when the number of observations is unlimited. The diagonal AB represents a perfect assessor. If we collect, say for 20 questions, all the probabilities used by a student and compute the relative frequency with which each particular probability is associated with a correct alternative, we may construct an external validity graph for that student taking that test. We then assume that the relationship between student response and relative frequency can be approximated satisfactorily by a straight-line simple linear regression.

We call the resulting line the student's realism line. The lines A'B' and A''B'' in Fig. A.1 illustrate possible results. The line A'B' has the following interpretation: About 60 percent of the events that were given a probability of 1.0 of occurring did in fact occur. We interpret this to mean that the student tends to overvalue his facts and reasons. Similarly, interpretation of the line A''B'' shows that the events that were given a probability of 0.60 (or greater) of occurring did in fact occur 100 percent of the time. In other words, the student tends to undervalue his facts and reasons. We may further interpret the realism line in the following way: "When he responded

* This appendix is taken from Ref. 1, pp. 9-12.

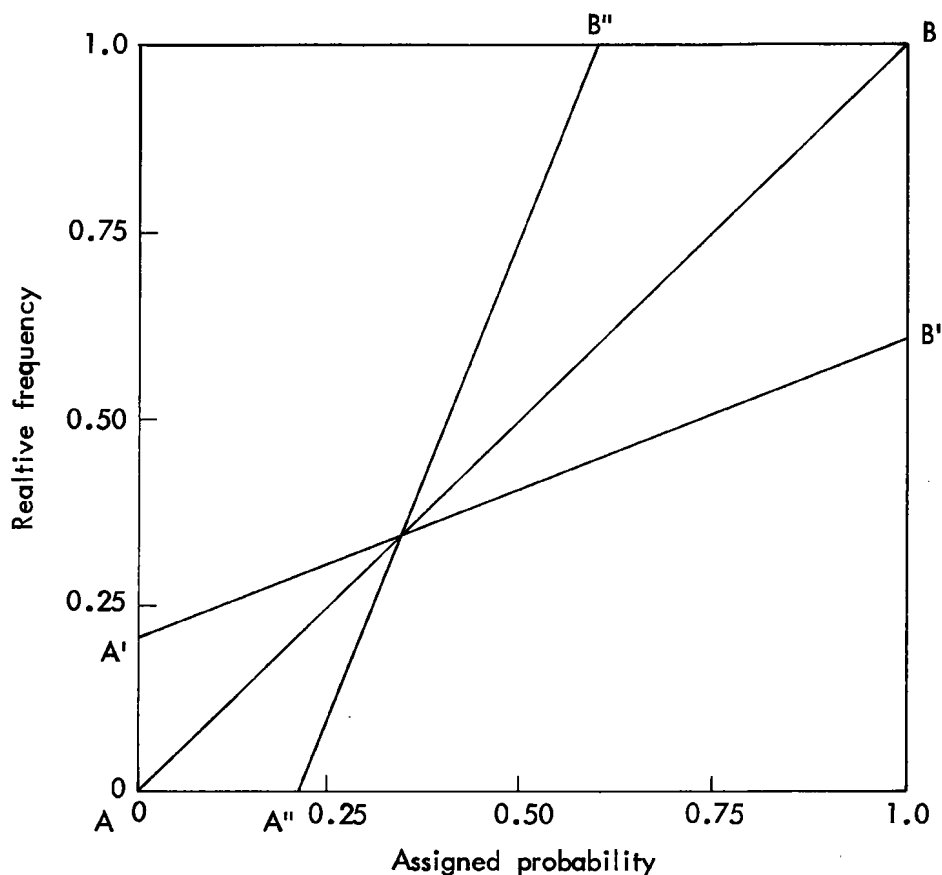
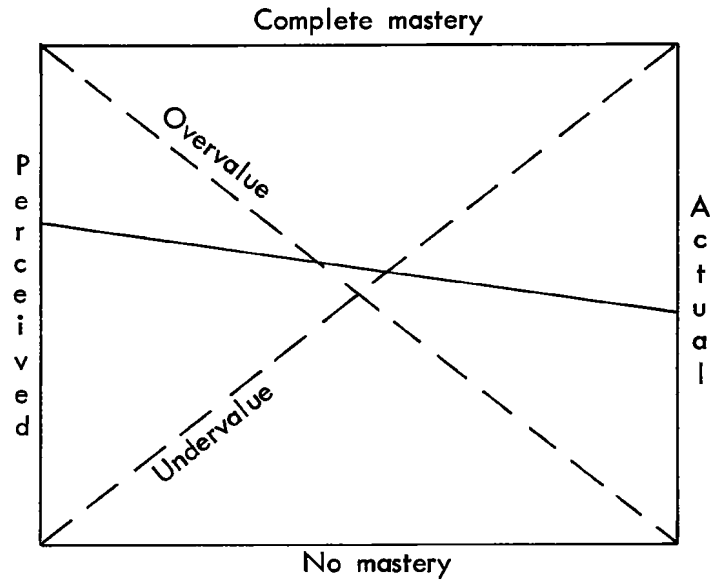


Fig. A.1 — External validity graph and realism lines .

x he should have responded with $Ax + B$," where $Ax + B$ is the equation of his realism line. We may then transform the student's probabilities, using his realism line, and recalculate his score on the basis of the new probability assignments. This revised score can then be used to decompose his original score into the portion attributable to bias in his assignment of probabilities and the portion attributable to lack of information about the subject matter itself. The lower section of Fig. A.2 illustrates such a score decomposition.

The scores are actually reported as the difference between the student's scores and the maximum attainable for a test with a specified length. The overall gain is that maximum less the original score. The gain attainable from more information is that maximum less the revised score. The gain from more "realistic" use of probabilities is simply the revised score less the original score.



You tend to overvalue your information.

You can improve your score by 37 points
by more realistic use of your knowledge.

You can improve your score by 224 points
by more study.

You can improve your score by 261 points
overall.

Fig. A.2—Score interpretation display

Appendix B

SAMPLE TESTS

rand4: A Test on International Relations

rand6: A Test on Statistics

rand7: A Test on Art

rand8: A Test on Music

rand9: A Test on Literature

Lesson: rand4

A TEST ON INTERNATIONAL RELATIONS

- * correct answer
- 1. Selling goods for export at prices lower than those charged domestic buyers is an example of:
 - *a. dumping
 - b. subsidy
 - c. embargo
- 2. The communist theorists expound today's world movements as a triumph for the revolutionary concept of:
 - a. Marx
 - *b. Lenin
 - c. Stalin
- 3. Identify the most important policy advisory body that works directly with the Secretary of Defense:
 - *a. Armed Forces Policy Council
 - b. National Security Council
 - c. Defense Research Engineering Group
- 4. The decision as to which countries the U.S. shall recognize diplomatically lies with the:
 - a. Senate
 - b. Secretary of State
 - *c. President
- 5. What do we call the strategy of buying goods to prevent a rival nation from getting them?
 - a. loading
 - *b. preemptive buying
 - c. embargo

Lesson: rand6

A TEST ON STATISTICS

- * correct answer
- 1. The least square estimator for the model $Y=XB+U$ is best since it:
 - a. is unbiased
 - b. is consistent
 - *c. has smallest variance
- 2. In matrix algebra, the transpose of a sum is just the:
 - *a. sum of the transposes
 - b. product of the transposes
 - c. sum of the products
- 3. An F (R,S) distribution (i.e., an F with R and S degrees of freedom is the ratio of):
 - a. chi-squared with R degrees of freedom and a normal with S degrees of freedom
 - *b. two chi-squareds with R and S degrees of freedom
 - c. two T-distributions with R and S degrees of freedom
- 4. A matrix is singular if it:
 - a. has no linearly dependent rows or columns
 - *b. has at least two linearly dependent columns
 - c. has a nonzero determinant
- 5. In testing whether or not there are increasing returns to scale in an estimated Cobb-Douglas production function, you would want to use a:
 - a. two tailed F test
 - *b. one tailed T test
 - c. one tailed chi-squared test
- 6. In the model $Y=XB+U$ under conditions of heteroscedasticity, the OLS estimator of B is not:
 - a. unbiased
 - b. consistent
 - *c. efficient

7. In the model $Y=XB+U$ with N observations and G explanatory variables plus a constant term, S -squared (i.e., sum of squared residuals divided by degrees of freedom) is distributed as a:
 - *a. chi-squared with $N-G-1$ degrees of freedom
 - b. T-distribution with $N-G-1$ degrees of freedom
 - c. chi-squared with $N-G$ degrees of freedom
8. In the case of heteroscedasticity, you would want to estimate the coefficients by:
 - a. ordinary least squares
 - *b. weighted least squares
 - c. two-stage least squares
9. It is possible to obtain a large F value (i.e., all the coefficients are significantly different from zero) and a low R -squared if you have:
 - *a. many observations and few parameters
 - b. few observations and few parameters
 - c. few observations and many parameters

Lesson: rand7

A TEST ON ART

* correct answer

1. Of the following categories, the one in which visual perspective was most thoroughly explored is:
 - *a. late Renaissance paintings
 - b. Byzantine mosaics
 - c. gothic stained glass
2. Of the following, the watercolorist noted for seascapes is:
 - a. Theodore Gericault
 - *b. Winslow Homer
 - c. Jacob Epstein
3. Which of the following Dutch artists is noted for his genre paintings?
 - *a. Jan Steen
 - b. Jan Vermeer
 - c. Jacob van Ruisdael
4. Of the following painters, the one who is best known for paintings of children is:
 - *a. Mary Cassatt
 - b. Rosa Bonheur
 - c. Georges Braque
5. The drawings of William Blake were based on:
 - *a. imaginative scenes
 - b. contemporary scenes
 - c. urban life
6. A technique in which no colors are mixed but tones are built up by a stipple of pure brilliant color is called:
 - *a. pointillistic
 - b. expressionistic
 - c. impressionistic
7. "The Night Watch" was painted by:
 - a. Franz Hals
 - b. Titian
 - *c. Rembrandt van Rijn

ART

2

8. The French artist who used the entertainment world almost exclusively as subject matter for his paintings was:
 - *a. Henri de Toulouse-Lautrec
 - b. Francois Boucher
 - c. Georges Seurat
9. Foremost among the so-called action painters was:
 - a. Salvador Dali
 - b. Paul Klee
 - *c. Jackson Pollock

Lesson: rand8

A TEST ON MUSIC

- * correct answers
- 1. An unaccompanied solo passage in a concerto, usually brilliant and virtuoso in style, is called a:
 - *a. cadenza
 - b. coda
 - c. counterpoint
- 2. The "Emperor Concerto" for piano and orchestra was composed by:
 - a. Frederic Chopin
 - b. Franz Liszt
 - *c. Ludwig van Beethoven
- 3. The romance of Rodolfo and Mimi is a prominent story thread in the opera:
 - *a. "La Boheme"
 - b. "La Traviata"
 - c. "Il Trovatore"
- 4. The "book" detailing the sequence of events in an opera is called the:
 - a. synopsis
 - b. score
 - *c. libretto
- 5. Among the following, the instrument that does not use a reed is the:
 - a. oboe
 - *b. flute
 - c. bassoon
- 6. The composer who wrote a famous "Ave Maria," using a prelude from Bach's "Well-Tempered Clavier" as an accompaniment, was:
 - a. Franz Schubert
 - *b. Charles Gounod
 - c. Gabriel Faure
- 7. Which composer wrote a symphony depicting the emotional stress of a city under siege?
 - a. Sergei Rachmaninoff
 - *b. Dmitri Shostakovich
 - c. Jean Sibelius

MUSIC

2

8. Violetta is a character in the opera:
- *a. "La Traviata"
 - b. "Il Trovatore"
 - c. "Rigoletto"
9. According to current expert opinion, which one of the following is regarded as an important jazz composer and arranger?
- a. Count Basie
 - *b. Duke Ellington
 - c. Benny Goodman

Lesson: rand9

A TEST ON LITERATURE

* correct answers

1. Which of the following poets reached his highest attainment in satire:
 - *a. Alexander Pope
 - b. William Blake
 - c. John Milton
2. Niccolo Machiavelli, in his political treatise "The Prince," advocates, as the basic policy of efficient government:
 - a. aggressive expansion
 - b. rigid economy
 - *c. diplomatic expediency
3. The poem "Lapis Lazuli" was written by:
 - a. T. S. Eliot
 - b. Dylan Thomas
 - *c. William Butler Yeats
4. "My head is bloody, but unbowed" epitomizes the mood of the poem:
 - *a. "Invictus"
 - b. "Lay of the Last Minstrel"
 - c. "The Soldier"
5. A literary work that tells about a hero who slays a dragon is:
 - *a. "Beowulf"
 - b. "The Battle of Brunanburh"
 - c. "The Wanderer"
6. Macbeth's tragic flaw can best be described as:
 - *a. lust for power
 - b. jealousy
 - c. overbearing pride
7. "A Valediction Forbidding Mourning" is a poem by which of the following metaphysical poets?
 - *a. John Donne
 - b. Andrew Marvell
 - c. George Herbert

LITERATURE

2

8. The name of Upton Sinclair's novel "Dragon Harvest" calls to mind an incident in the story of:
 - a. Hercules
 - b. Perseus
 - *c. Jason

9. The greatest of the Greek comic dramatists, noted for his satires of the life of his time, was:
 - a. Menander
 - *b. Aristophanes
 - c. Euripides

REFERENCES

1. Sibley, William L., *A Prototype Computer Program for Interactive Computer-Administered Admissible Probability Measurement*, The Rand Corporation, R-1258-ARPA, April 1974.
2. Melles, David V., *Using PLATO IV*, Computer-Based Education Research Laboratory, University of Illinois, Urbana, Illinois, July 1974.
3. Lyman, Elisabeth, *PLATO Curricular Material*, Computer-Based Education Research Laboratory, University of Illinois, Urbana, Illinois, July 1974.
4. Bitzer, D., and D. Skaperdas, *The Design of an Economically Viable Large-Scale Computer-Based Education System*, Computer-Based Education Research Laboratory, University of Illinois, Urbana, Illinois, January 1972.
5. Brown, T. A., Shuford, H. E., *Rationale of Computer-Administered Admissible Probability Measurement*, The Rand Corporation, R-1371-ARPA, July 1974.
6. Sherwood, Bruce Arne, *The TUTOR Language*, Computer-Based Education Research Laboratory, University of Illinois, Urbana, Illinois, June 1974.
7. Gardner, Willie C., *The Use of Confidence Testing in the Academic Instructor Course*, The Shuford-Massengill Corporation, Lexington, Massachusetts, SMC R-10, February 1970.
8. Brown, Thomas A., *Probabilistic Forecasts and Reproducing Scoring Systems*, The Rand Corporation, RM-6299-ARPA, June 1970.
9. Thorndike, Robert L., ed., *Educational Measurement*, American Council on Education, Washington, D.C., 1971.