ESD-TR-67-229

CYBERNETIC TESTING - PROJECT 2806 - TASK 280609
for the period November 1963 thru May 1965

Emir H. Shuford, Jr.

March 1967

DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts 01730

Distribution of this document is Unlimited.

ESD-TR-67-229

CYBERNETIC TESTING - PROJECT 2806 - TASK 280609
for the period November 1963 thru May 1965

Emir H. Shuford, Jr.

March 1967

DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts 01730

FOREWORD

The study upon which this report is based was conducted in support of Project 2806, Task 280609, during the period of November, 1963 through May, 1965.

This report consists of a paper presented as part of the symposium, "Current Trends in Computer-Based Instructional Systems," at the National Society for Programed Instruction convention in Philadelphia, Pennsylvania on 8 May 1965.

In part, it describes a computer-based subjective probability response technique developed by this Laboratory. Based upon the mathematical concepts defined by Toda (in ESD-TDR-63-407), the prototype design of this measurement technique was created jointly by the author and W. E. Organist. The technique shows promise of getting more information per response for use in computer-assisted instruction, testing, and psychological experimentation.

Subsequent to this report, this measurement approach evolved into a system that serviced four subject stations at the same time and has been used in experiments which will be reported on separately.

Robert T. Rizzo, of the Arcon Corporation, was responsible for the computer program design and programming of the prototype. James D. Baker and Ira Goldstein contributed to the final design and implementation.

This Technical Report has been reviewed and is approved.

Walter E. Organist

WALTER E. ORGANIST
Project Officer
Decision Sciences Laboratory

James S. Duva

JAMES S. DUVA
Technical Director
Decision Sciences Laboratory

ABSTRACT

This report presents a concrete realization of an admissible probability measurement procedure utilizing a computer-driven scope and light pen. This particular technique is appropriate for all multiple-choice type of testing.

Empirical results are reported from an analogous pencil-and-paper realization of the same admissible probability measurement procedure. These results indicate a marked superiority for admissible probability measurement over traditional multiple-choice testing.

It is suggested that further gains can be obtained by using admissible probability measurement procedures to sequentially test the scope of knowledge of a student.

TABLE OF CONTENTS

CYBERNETIC TESTING

Emir H. Shuford, Jr.

A computer is essentially a factory for the very rapid processing of information. Computers can be used effectively to reduce the cost of information processing whenever the speed of the computer can be applied to an information processing problem which is of a very highly repetitive nature. This allows the cost of programming to be amortized over many instances of application, each justifying a part of the total cost.

Most applications of computers have involved just the substitution of automatic information processing for some part of a more complex, already existing enterprise. Such a direct substitution can dramatically reduce the cost and increase the capacity for information processing, but the full potential of this change can generally not be realized unless other changes are made in the operation of the enterprise. For example, it is sometimes necessary to reduce also the cost of obtaining and reacting to information by introducing techniques for the automatic sensing of and responding to information. This type of effective application of a computer is well represented by the computer-based instructional systems just described by Professors Hansen and Stolurow.

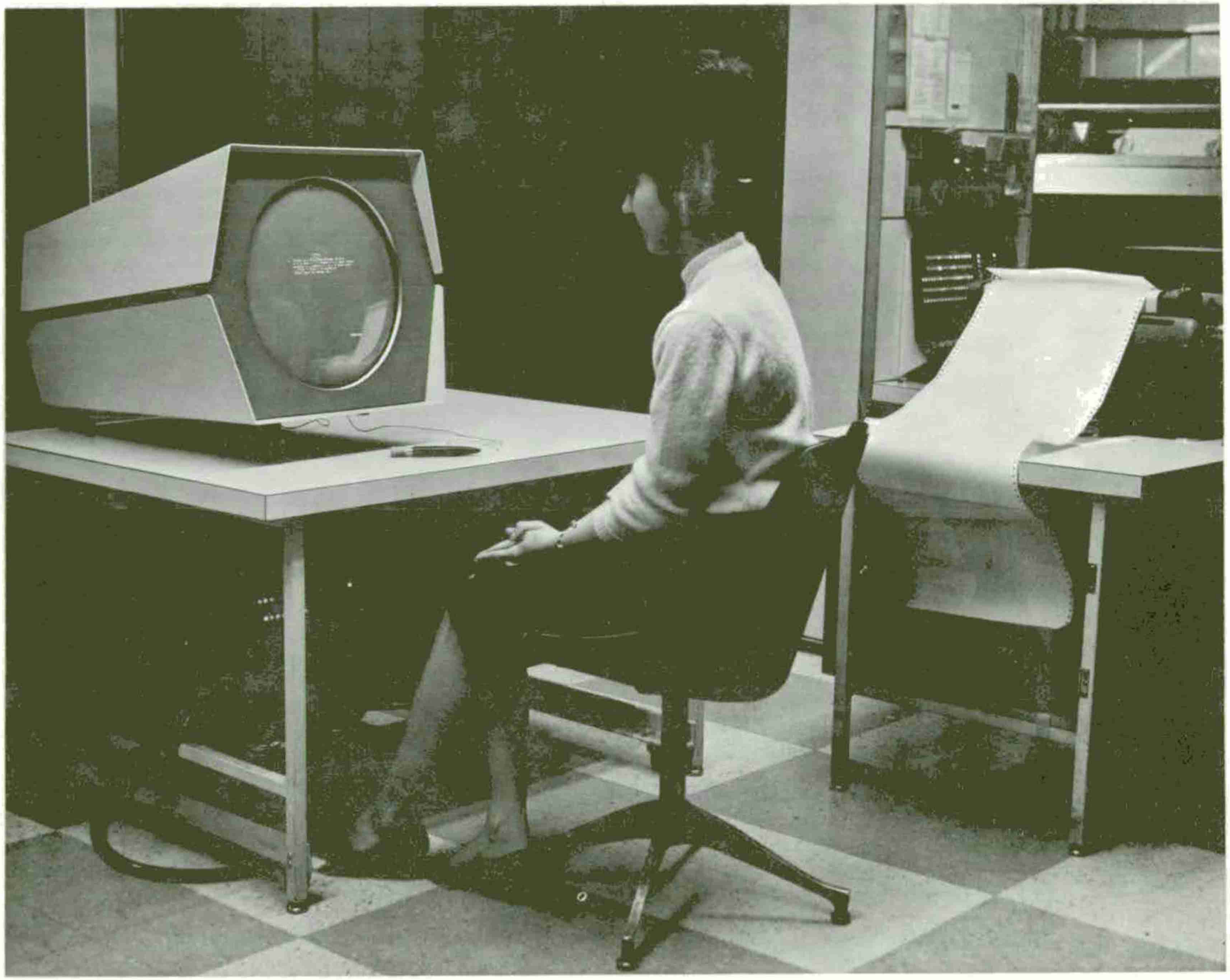
In these applications, the computer systems (a) measure the current state of the student's knowledge, (b) process this information to determine what instructional material must be presented next in order to improve the student's knowledge, and (c) effect the presentation of the material. This is quite clearly a cybernetic control process with the computer used as a <u>controller</u> which senses the state of a <u>controlled system</u> and then takes <u>corrective action</u> to move the controlled system to a more <u>desirable state</u>. Notice, however, that complete automation is not essential to the nature of the cybernetic process. A teacher conducting a course in a classroom, a school system promoting students to the next grade level, and a student guiding the course of his own study are also examples of the cybernetic control process as applied to the development of knowledge. (Shuford & Massengill, 1965).

Now, when the educational process is looked at from this point of view, it is not difficult to see that the effectiveness of the educational process depends, in part, upon how well we can observe the present state of the student's knowledge. This observational process, in turn, determines the sensitivity with which we can follow the educational development of the student. Indeed, the recent emergence of admissible probability measurement procedures (Shuford, Albert, & Massengill, 1965) which yield much more information about the current state of a student's knowledge than do the multiple-choice and constructed-response test procedures (Massengill & Shuford, 1965), suggests that it may be possible to achieve even greater increases in effectiveness over and above that resulting solely from the introduction of computer-based instructional systems based on traditional measurement techniques.

In order to distinguish these new applications based on probability measurement from the other currently used applications based on choice procedures, I would like to introduce two new terms. First, <u>cybernetic</u>

instruction refers to any computer-based instructional system utilizing probability measurement to follow the development of a student's knowledge. Second, *cybernetic testing* refers to the use of probability measurement where the computer may be used to control the testing or to analyze the results, but not to control the complete course of instruction. Thus, cybernetic testing may be used to aid any instructional procedure or in association with any instructional media.

Now, what is an admissible probability measurement procedure? First, let me define it and then we will get down to cases. An admissible probability measurement procedure has a scoring system which guarantees that any student, at whatever level of knowledge or skill, can maximize his expected score if and only if he follows instructions and honestly reflects his degree-of-belief probability as to the correctness of each possible answer to the test item. These degree-of-belief probabilities contain all of the information that can be made available about the student's knowledge structure as a consequence of asking the particular question under consideration. By way of contrast, multiple-choice and constructed-response test procedures can yield only partial information as to whether or not these probabilities exceed certain values or lie within a very broad range. It is probably best at this point to consider a concrete example of an admissible probability measurement procedure used in conjunction with a computer-based system, i.e., cybernetic testing. Let's look at some pictures which illustrate multiple-choice testing on a computer-driven scope and light pen.



<u>Figure 1</u> shows a student seated before a computer driven scope and light pen ready to begin taking the multiple-choice test.

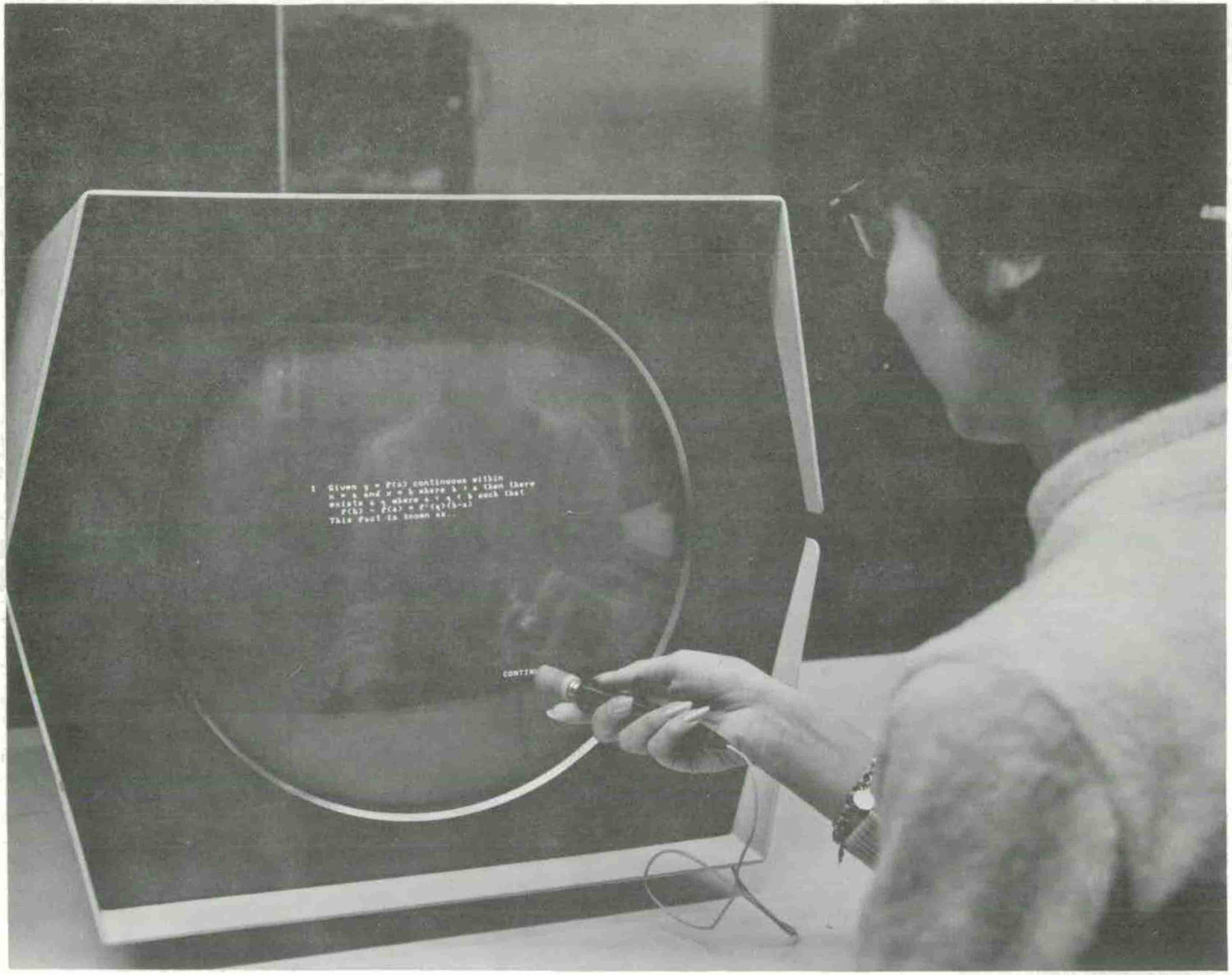


Figure 2 shows the first question on the test. The student has read the question and feels that she is ready to answer it. So, she points the light pen at the CONTINUE sign.

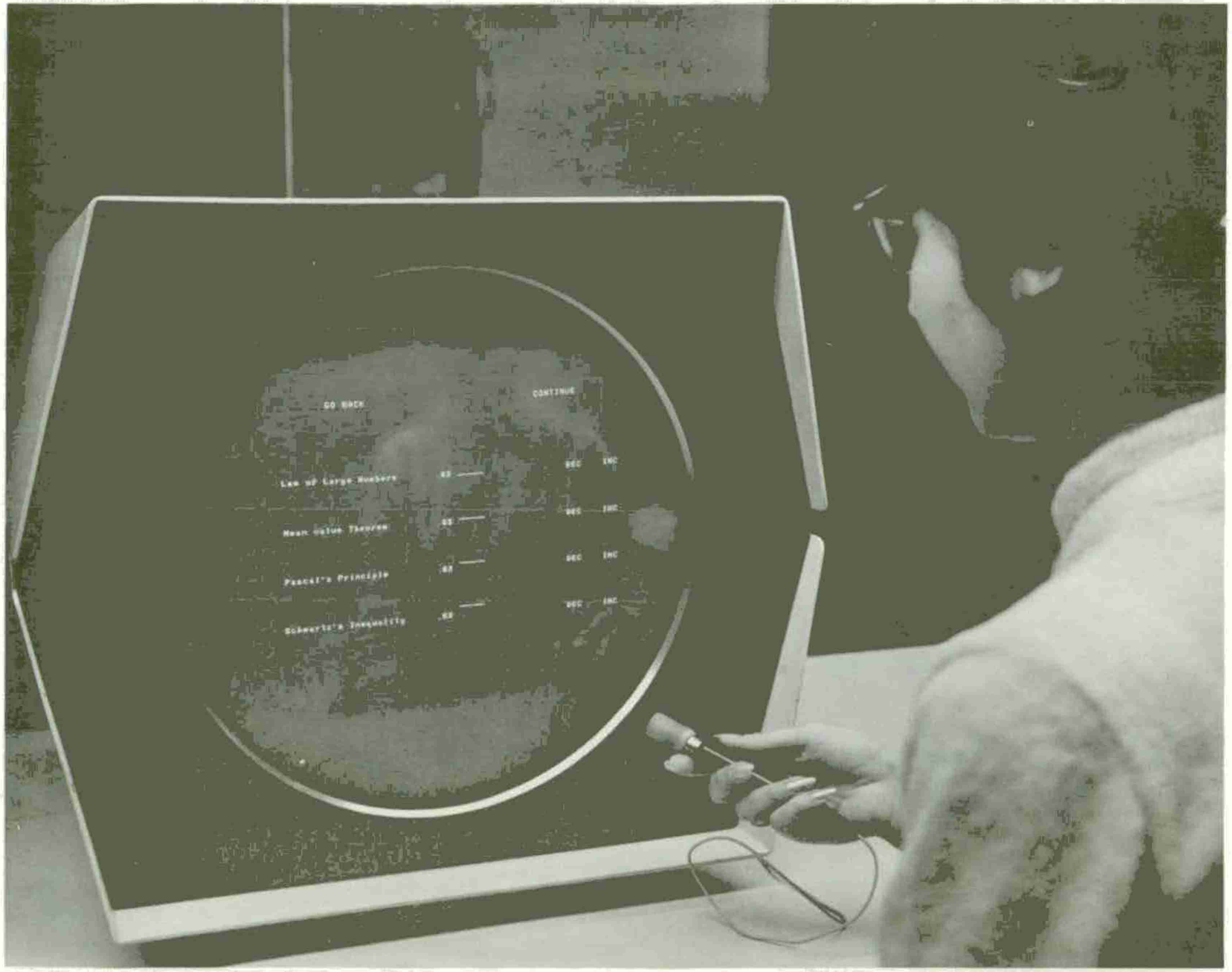
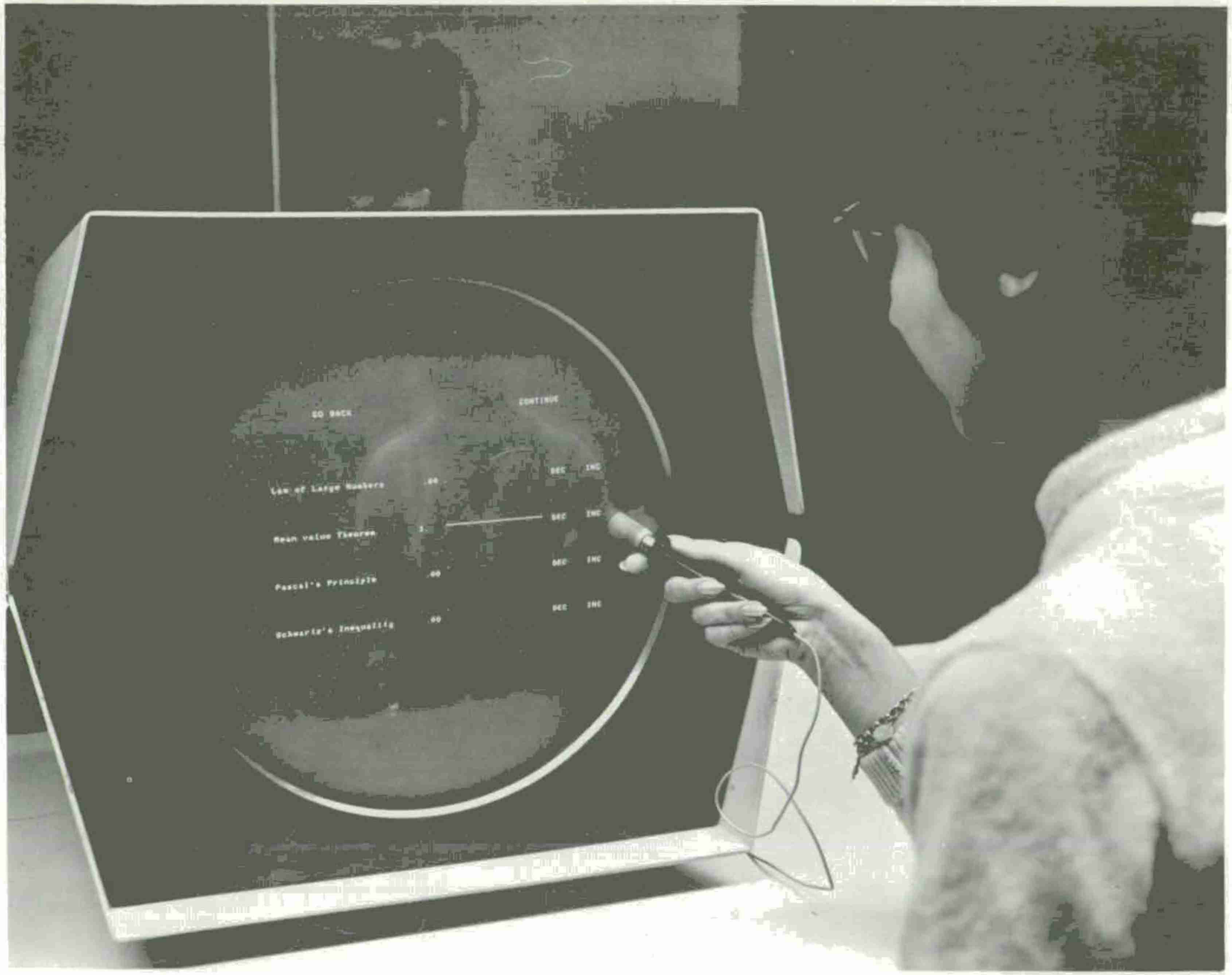


Figure 3 shows the four mutually exclusive and exhaustive possible answers to Question 1. The horizontal line by each answer represents the probability currently assigned to the correctness of that answer. The number to the left of the line represents the score that the student would receive if, in fact, that answer were correct. The score ranges between zero and one instead of being limited to just the extreme values of zero and one as is the current scoring practice.



<u>Figure 4</u> shows the student pointing the light pen to adjust the probabilities and possible scores. The student has no doubt that the correct answer is the "Mean Value Theorem" so she points the light pen at the INC sign. The probability assigned to the "Mean Value Theorem" increases at a constant rate while total probability is conserved by the automatic decrease of the remaining probabilities. Now the student will receive a score of 1.0 if this is the correct answer, but nothing if any of the other answers is correct.

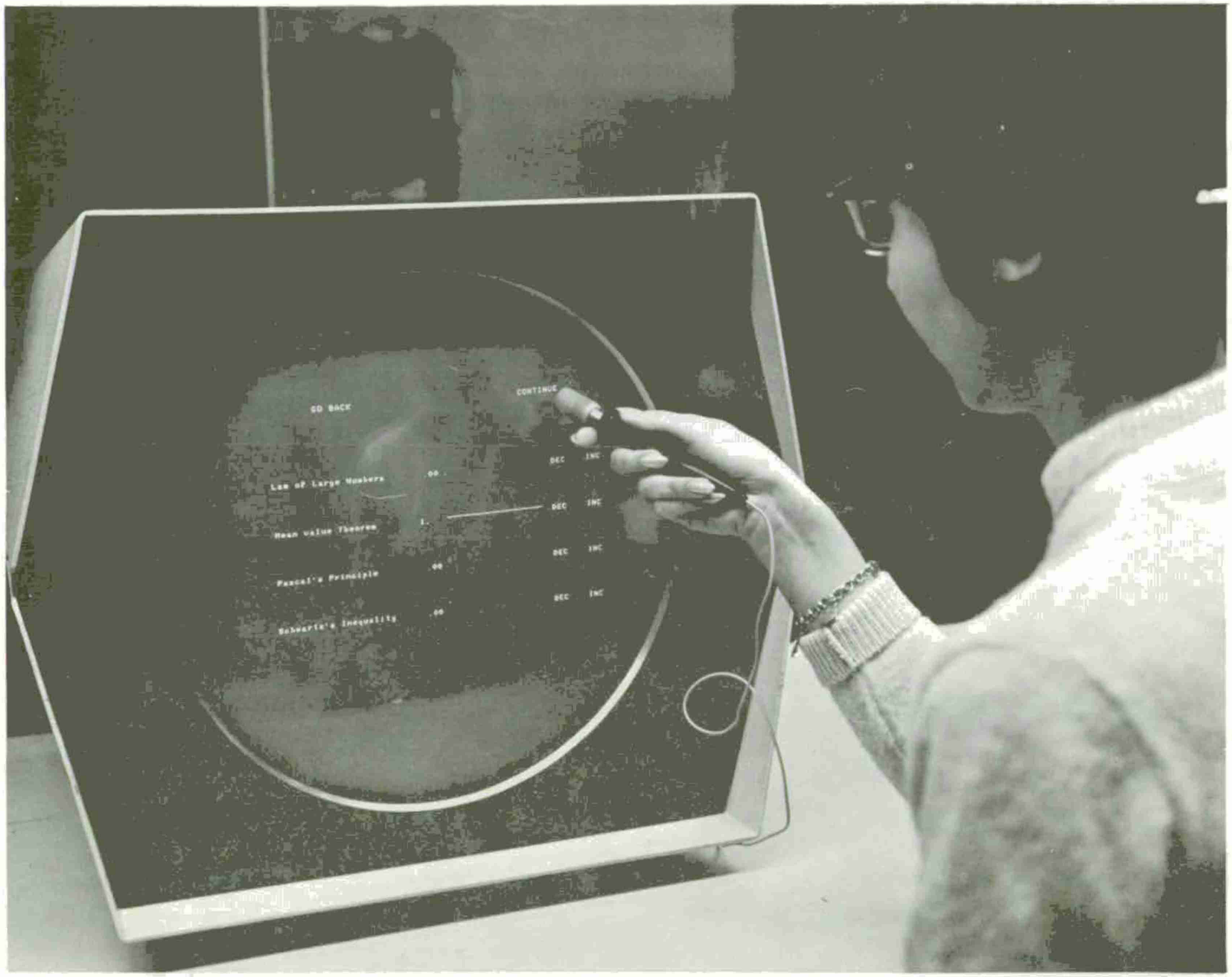


Figure 5 shows the student pointing the light pen at CONTINUE. She is satisfied with her probability assignment and wants to find out how well she scored.

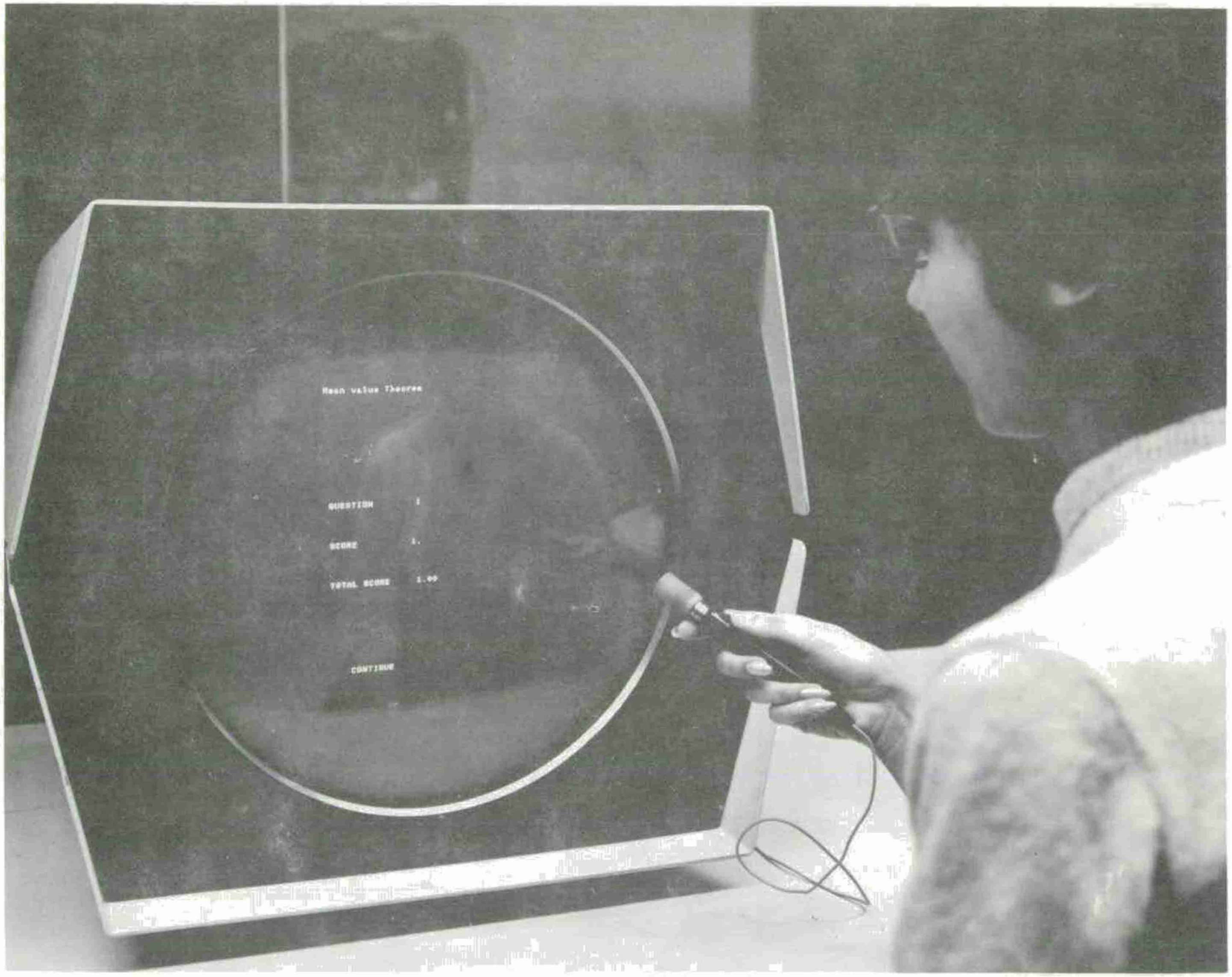
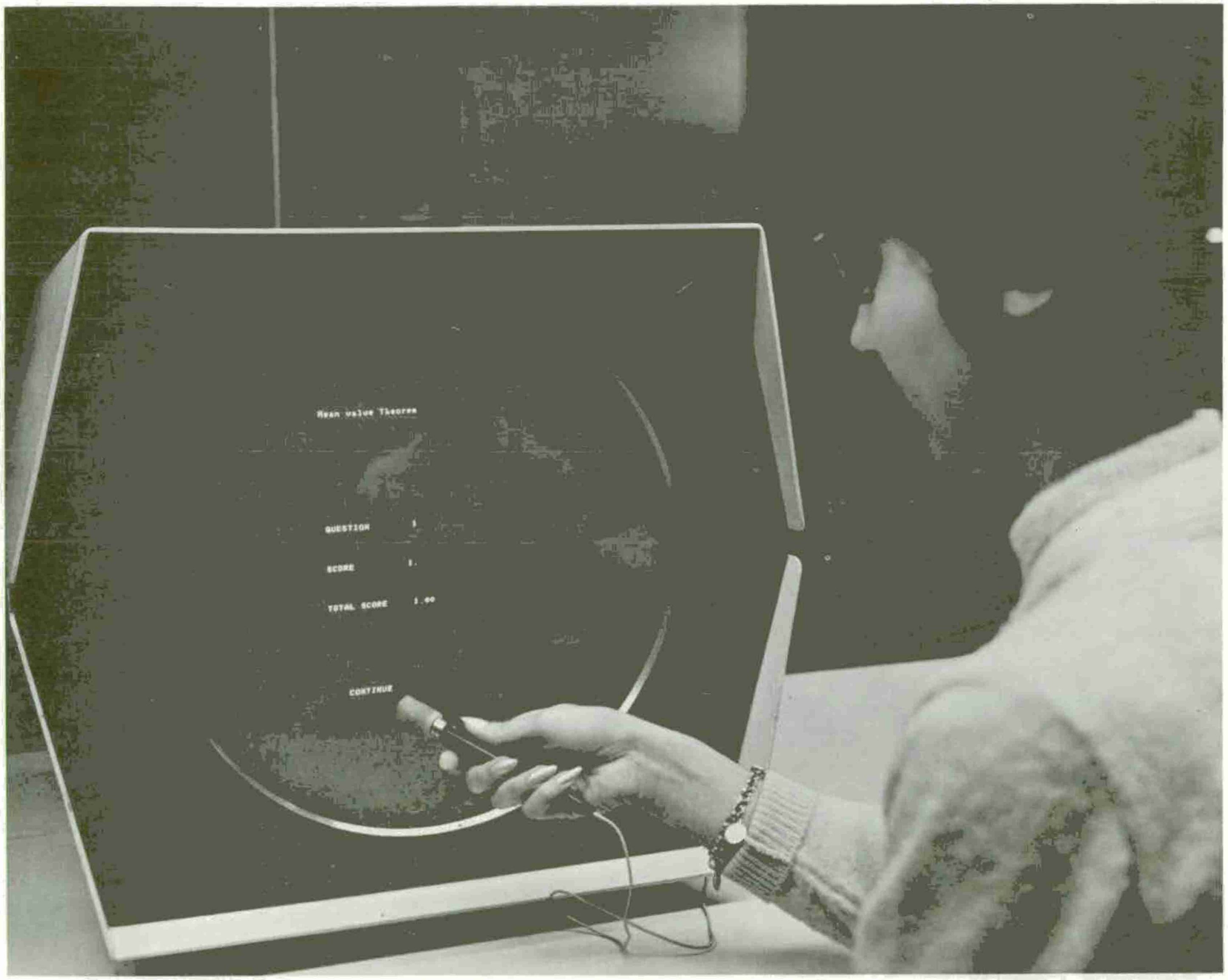
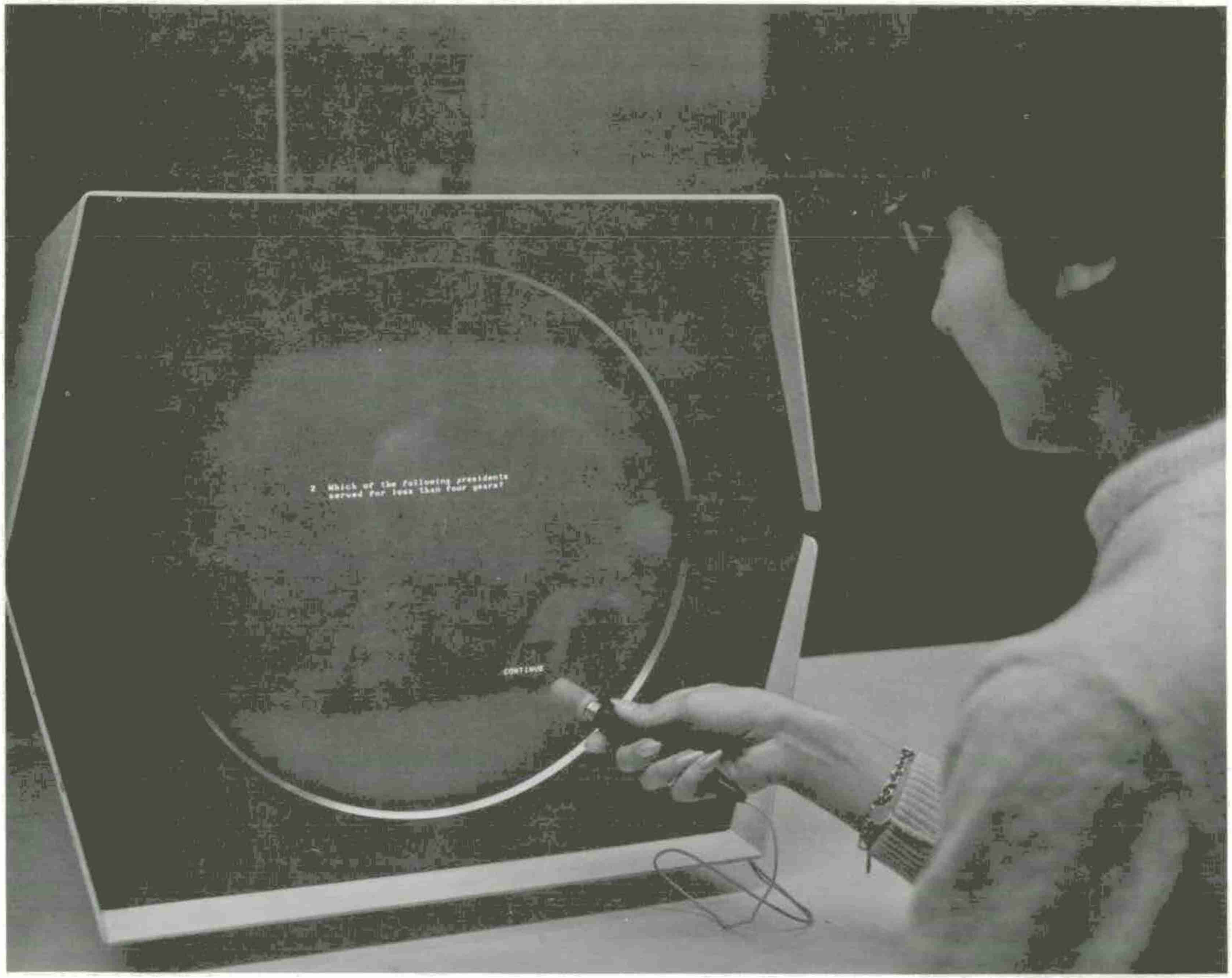


Figure 6 shows that the "Mean Value Theorem" is the correct answer to the first question, that the student received a score of 1.0 on the question, and that her total score to this point in the test is 1.0.



<u>Figure 7</u> shows the student pointing at CONTINUE. This will cause the next question to be displayed.



<u>Figure 8</u> shows the second question. The student has read the question and points to CONTINUE.

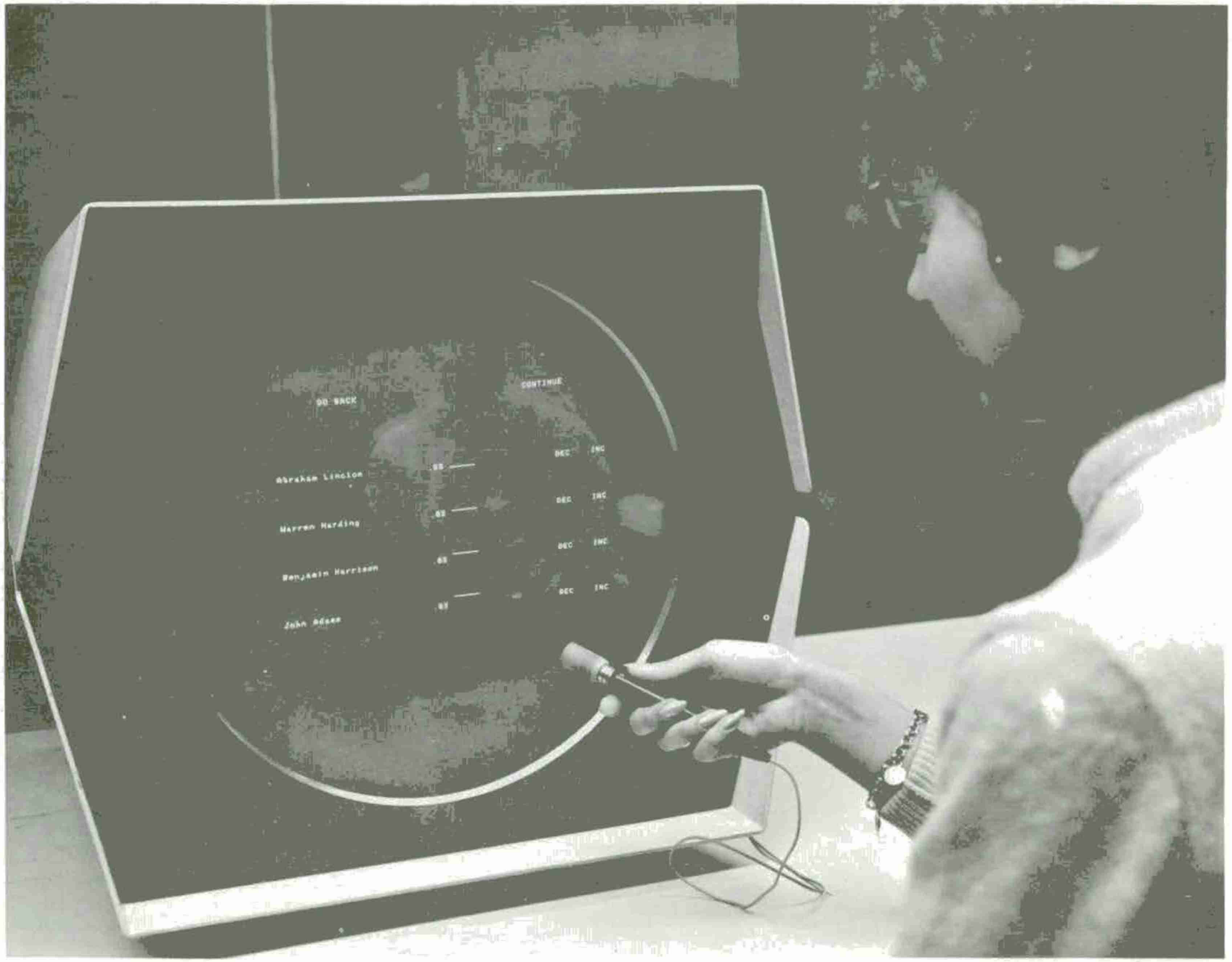
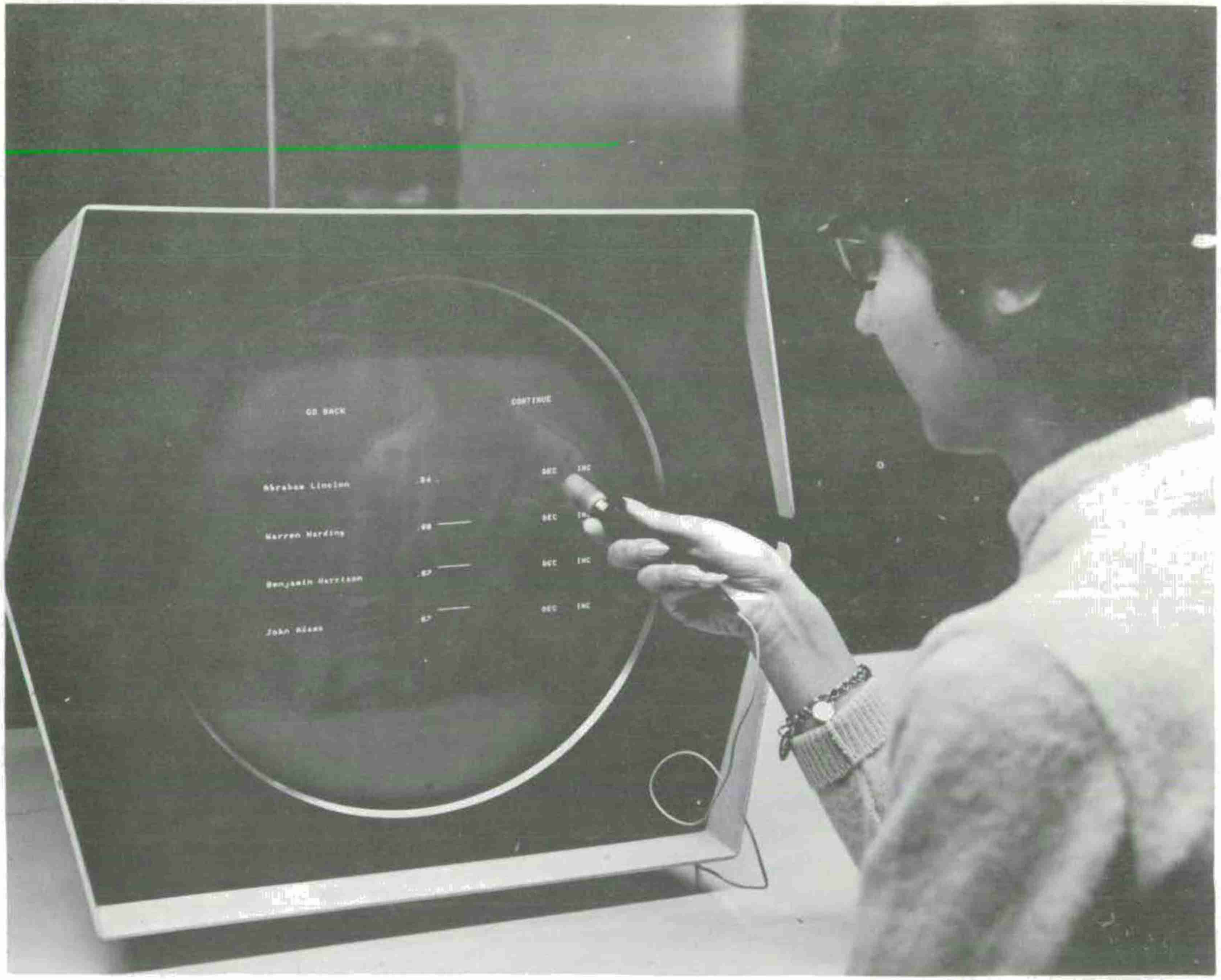


Figure 9 shows the four possible answers to the second question.



<u>Figure 10</u> shows the student pointing at the DEC sign associated with "Abraham Lincoln." After reading the four answers, she is quite certain that "Abraham Lincoln" is not correct so she points at the DEC sign until the probability associated with that answer is reduced to zero.

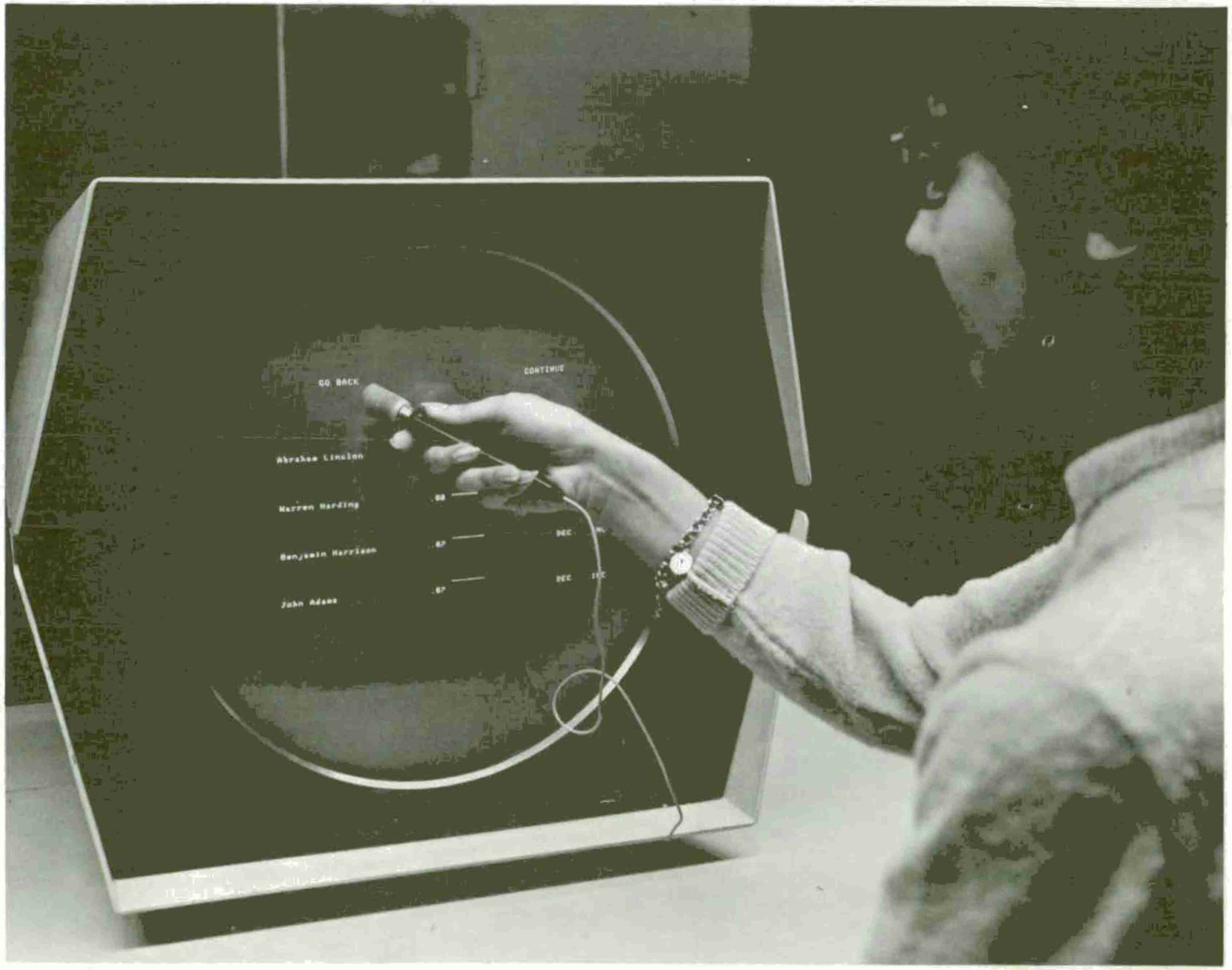
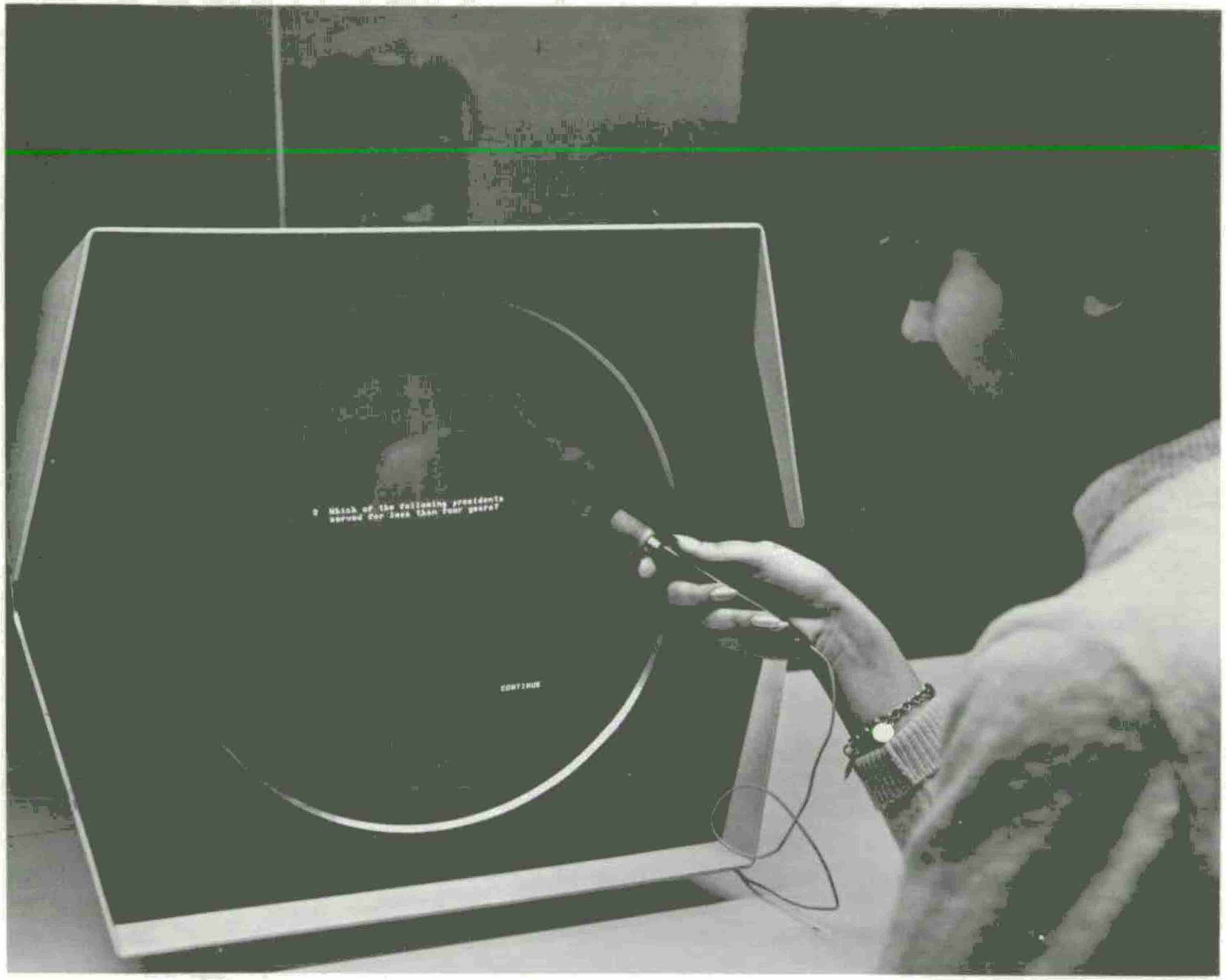


Figure 11 shows the student pointing at the GO BACK sign. She would like to review the question and does so by pointing at the GO BACK sign.



<u>Figure 12</u> shows the second question being displayed again. After the student has read it, she will point at CONTINUE.

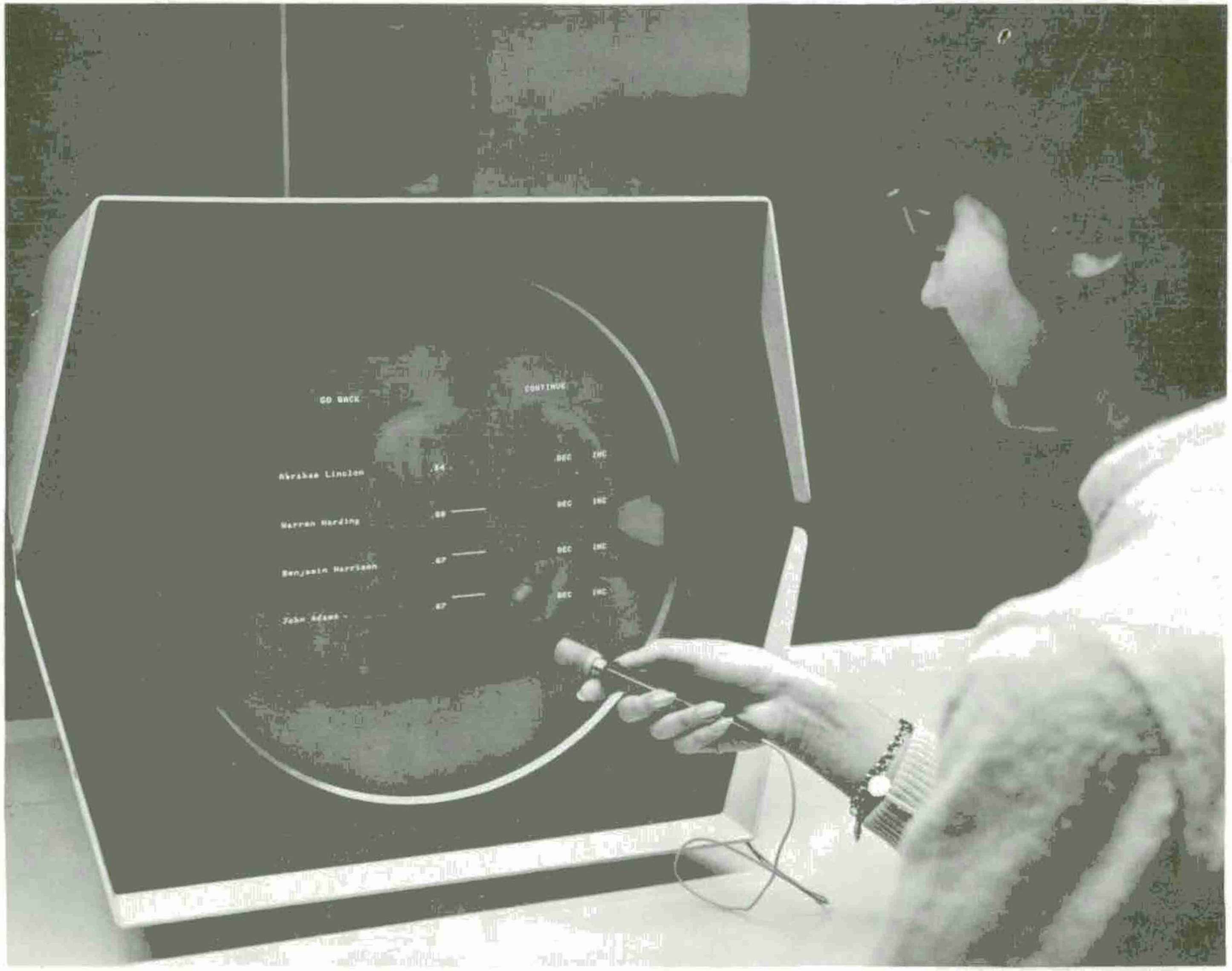


Figure 13 shows the response frame of the second question being redisplayed. Note that the frame appears exactly as it did when the student pointed at the GO BACK sign.

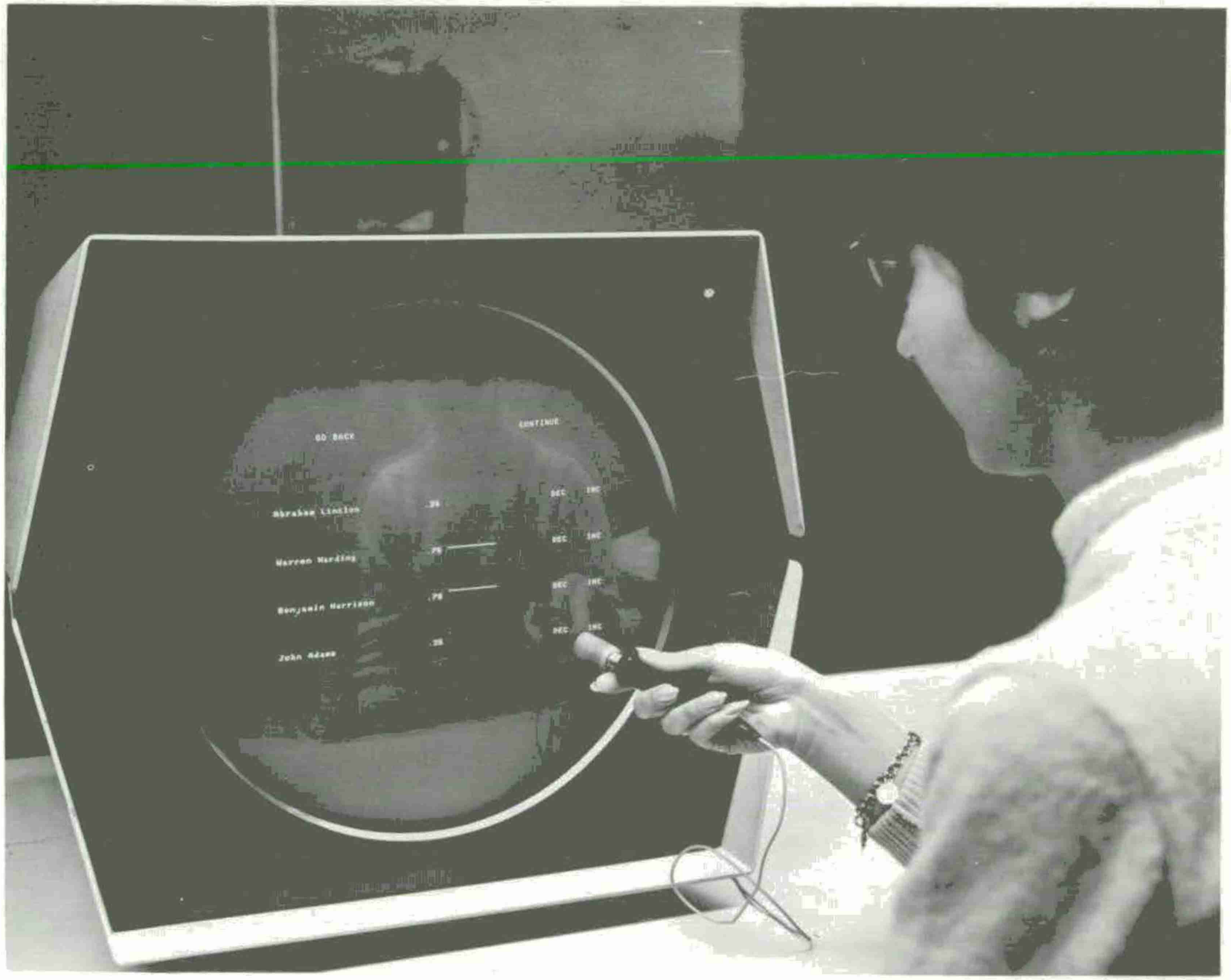


Figure 14 shows the student pointing at the DEC sign associated with "John Adams." She has decided that the fourth answer, "John Adams," is certainly not the correct one and so reduces the probability assigned to this answer to zero.

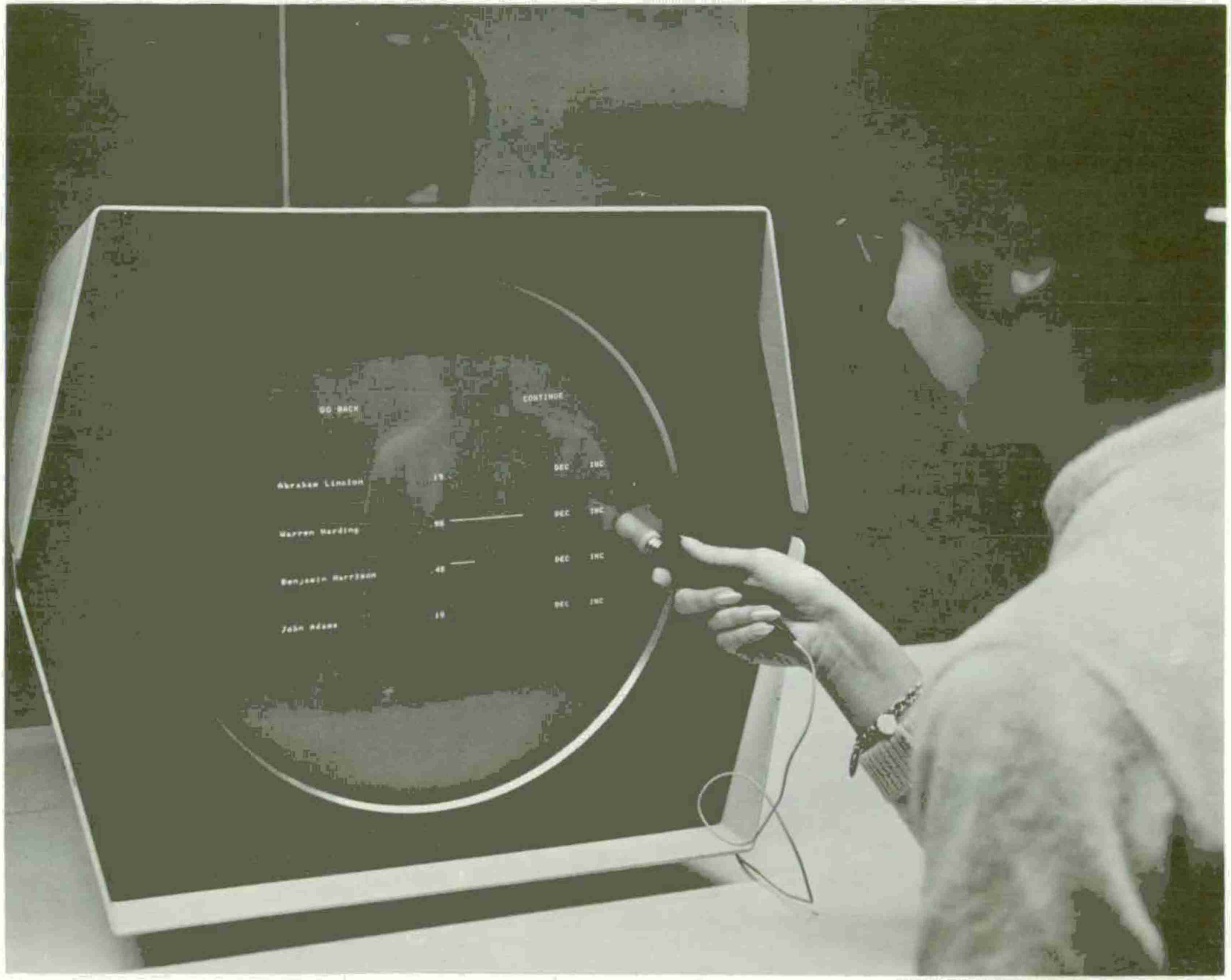


Figure 15 shows the student pointing at the INC sign beside "Warren Harding." She is sure that neither the first nor the fourth answers are correct, but she is not completely certain which of the remaining two answers is correct. She is, however, fairly certain that the second answer, "Warren Harding," is the correct one, so she points at the INC sign to divide the probability between these two answers to reflect this feeling. Notice that she does not feel that she can exclude the third answer, "Benjamin Harrison."

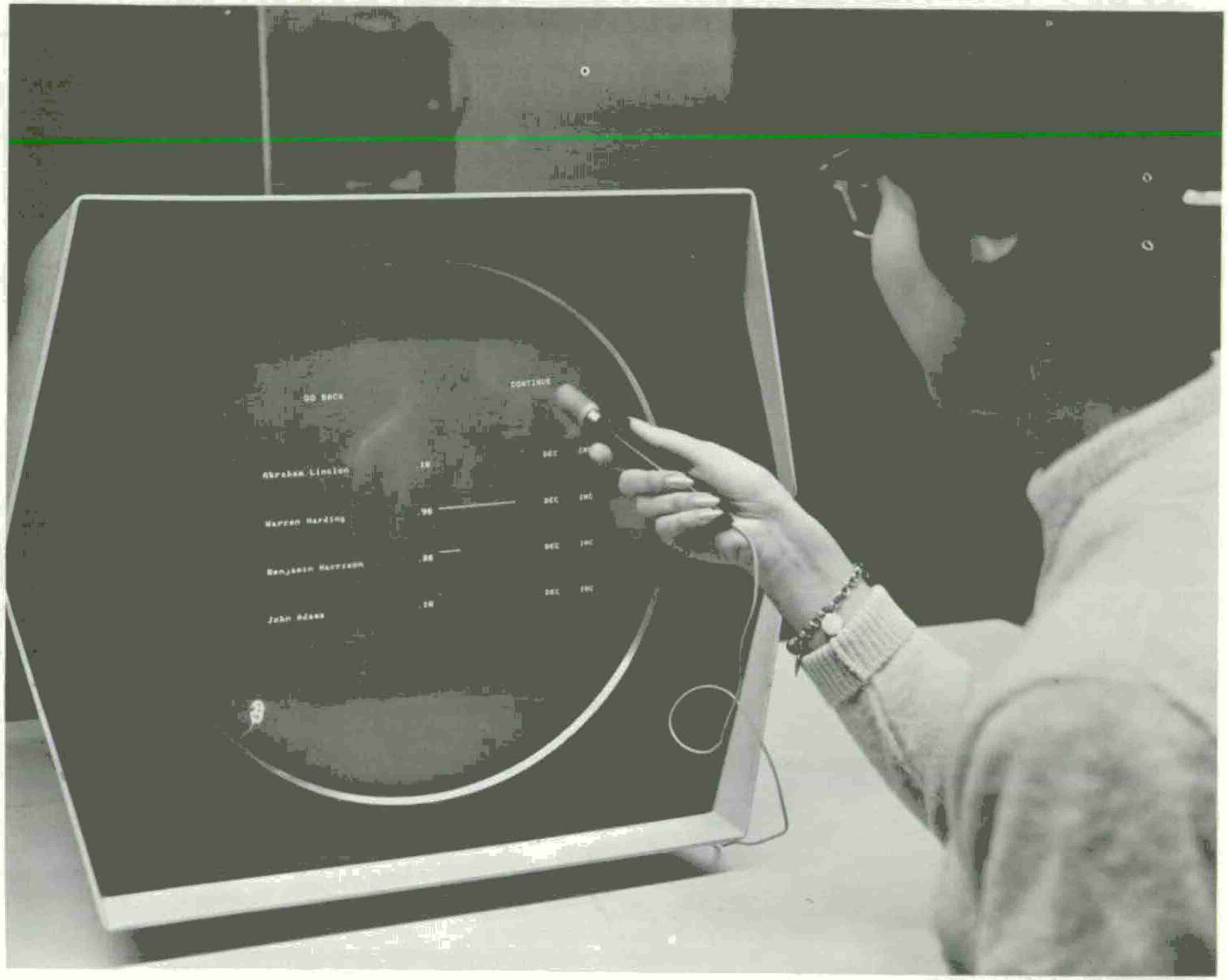
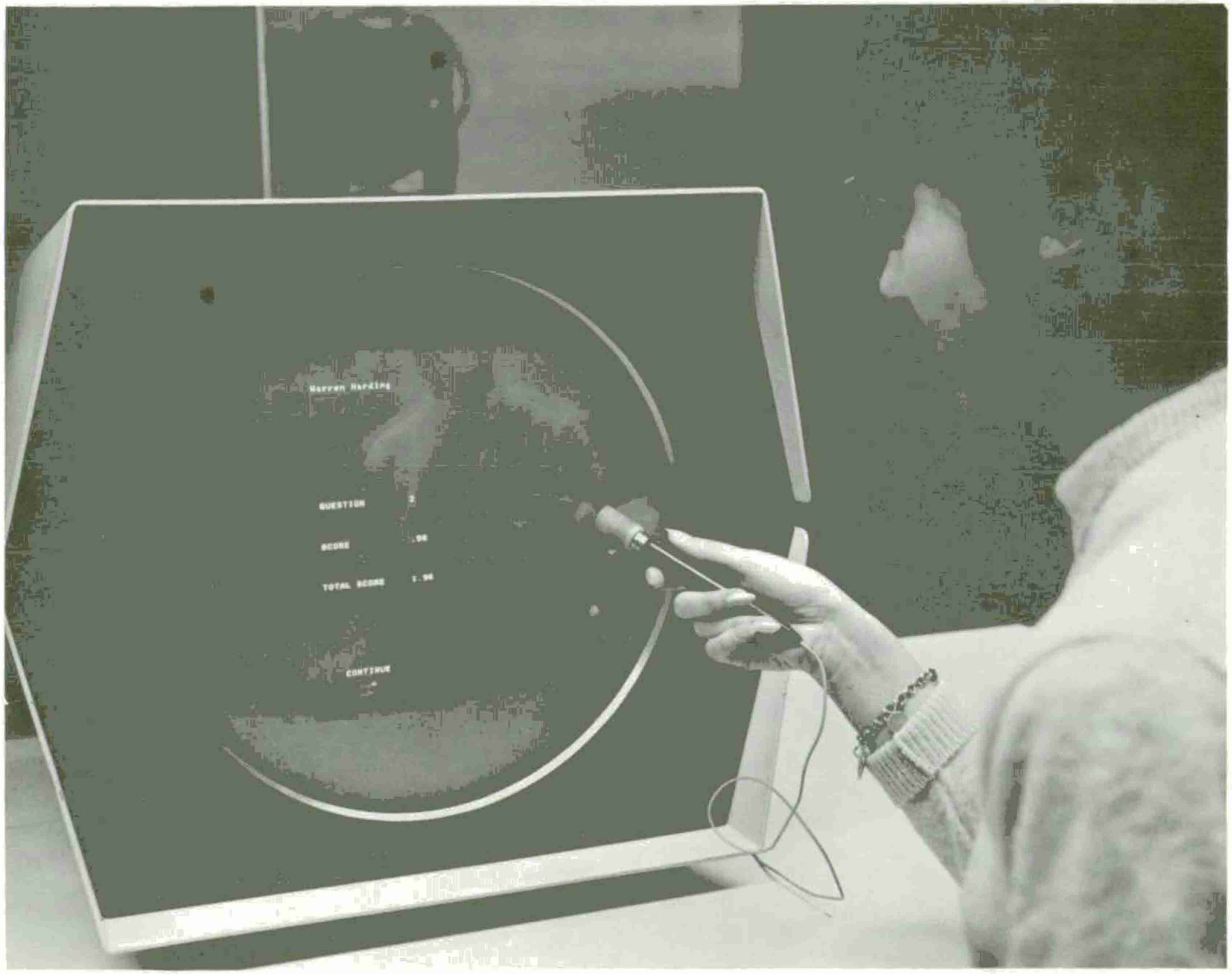


Figure 16 shows the student pointing to CONTINUE, indicating that she is finished in her selection of answers.



<u>Figure 17</u> shows that "Warren Harding" is the correct answer to the second question, that the student received a score of .96 on this question, and that now her total test score is 1.96. When the student points at CONTINUE, she will move on to the next item, and so on.

Though these pictures have illustrated admissible probability measurement only for multiple-choice type items, it is important to note that procedures exist for having the student supply his own answer (Shuford, Albert, & Massengill, 1965). Thus, it is now possible to measure a student's degree-of-belief probabilities for almost all objective test and programmed instructional material. Realize that no information is lost by substituting admissible probability measurement procedures for the choice procedures currently in use since a student's choices can be reconstructed from knowledge of his probabilities, i.e., the student would be expected to choose the most likely answer if given the opportunity (Shuford & Massengill, 1965).

The guarantee that no information is lost would be sufficient to justify the use of admissible probability measurement procedures and high-speed digital computers only if this substitution were a cheaper way of doing what was done before. It is not. It generally takes a student a little longer to express his probabilities and the purchase of a computer system is, at present, not a trivial economic decision. Therefore, cybernetic testing is going to have to be able to do things somewhat better, either more effectively or more cheaply. So, let's consider the gains that can result from cybernetic testing.

Using college students and a pencil-and-paper test form of probability measurement, ~~Walt~~ Organist ~~and I~~ found that multiple-choice tests yielding split-half reliabilities in the range .6 to .7 for the number of items correct, i.e., scored in the usual way yielded split-half reliabilities in the vicinity of .9 for total test scores and other measures obtained through probability measurement. In addition, theoretical arguments can be given which indicate that these increased reliabilities will be found in almost all testing situations encountered in practice. Therefore, a teacher using cybernetic testing can reasonably expect to more accurately and precisely grade her students and, of course, since correlations and validities are limited by test reliabilities, she can expect her tests to give better predictions and to have higher validity.

To consider another result, first realize that using cybernetic testing there is no longer any need to average over test items or over different students since reliable information can be obtained from each individual query. There is, however, something interesting that can be done by examining the pattern of probabilities given to the answers of one item by all students in a class. In most cases, it can be determined with great precision both how well the subject matter has been taught and how well the test items and answers have been written. This cannot be done with currently used testing techniques, but by having a computer examine the pattern of probabilities, a teacher can obtain information that would enable her to improve her teaching of the course and the quality of the items that she uses to test for understanding of the subject matter. She can also, of course, by examining the pattern of probabilities for each student, gain diagnostic information useful in giving individual attention to her students and in understanding the teaching-learning process.

Now, I could proceed gradually through many levels of increasing sophistication of application, each level promising a further increase in the effectiveness of the educational process, and finally arrive at the level of adaptive programmed instruction with branching decisions based on the student's probabilities rather than on his choices. In fact, Jim Baker is experimenting with this type of cybernetic instruction at the present time and I think that he is finding it quite exciting. However, due to lack of time, I would like to skip these intermediate levels of application and, instead, briefly introduce the notion of sequential testing where the next item to be presented to the student depends upon the previous items and his responses to these items. Choice methods leave too much ambiguity about the student's knowledge in order to be used this way, but the existence of admissible probability measurement procedures make this type of testing appear to be highly promising. The promise resides in the possibility that by utilizing information about the structure of the subject matter material and about the way the student learns, the scope of a student's knowledge about a content area can be determined by asking only a minimal number of questions. The test would be tailored to each student.

For example, in some cases test items can be written with different degrees of difficulty so that if a student knows a particular item, he is almost certain to know the easier items. Thus, if a student indicates almost complete certainty in the correct answer to this particular item, a much more difficult item could be presented next while if he indicated almost complete uncertainty, a much easier item could be presented next. Such a testing strategy could determine his level of knowledge very quickly by asking very few questions.

For another example and in other cases, items can be written so that knowledge of the correct answer depends jointly on knowledge of several different, less complex items. Proofs in mathematics and the translation of sentences or phrases provide concrete examples of this type of structure. In this case, if a student expresses a great deal of confidence in the correct answer to one of these complex items, he could then be tested on a different topic represented by another complex item while if he expresses considerable uncertainty, he could then be tested on one of the less complex items to determine the source of his uncertainty. This is another testing strategy which would, as before, determine the scope of a student's knowledge with great efficiency.

The usefulness of sequential testing could be further increased by associating with the questions at different levels references to chapters and to sections in textbooks and, where appropriate, additional problems and examples. This would allow the diagnostic information provided by sequential testing to be used to recommend remedial or supplementary study for the individual student according to the scope of his knowledge.

Clearly, sequential testing would be a more efficient and a more enjoyable form of testing. More enjoyable to the student, I should hasten to add. Writing these sequential tests would require much too much time of a classroom teacher operating under typical conditions. Therefore, we should expect that textbook publishers will make available sets of sequential test materials to accompany their texts and that test publishers will develop sequential tests for much improved diagnostic and achievement testing.

Since sequential testing requires flexibility in the presentation of items and considerable information processing, it should be conducted under the control of a computer and possibly with computer-driven scopes and light pens. Thus, we should expect that computer manufacturers will make available to the schools completely pre-programmed computer systems ready to accept the sequential testing materials provided by the publishers and to give the tests to students both for evaluating their progress through the course of instruction and as a means of guiding their study of textbooks and other materials.

Finally, in what other ways can the combination of computers with admissible probability measurement procedures improve instruction? I don't know, but I do know that our ability to improve education depends, in part, upon our knowledge of the teaching-learning process which in turn depends upon our being able to observe the effect of instructional procedures upon the knowledge structures of individual students. And this observational process is accomplished with exquisite sensitivity and precision by cybernetic testing.

REFERENCES

Massengill, H. Edward & Shuford, Emir H., Jr. (1965) Direct vs. indirect assessment of simple knowledge structures. ESD-TR-65-542 ,Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

Shuford, Emir H., Jr., Albert, Arthur, & Massengill, H. Edward (1965) Admissible probability measurement procedures. Psychometrika, (In Press).

Shuford, Emir H., Jr. & Massengill, H. Edward (1965) On communication and control in the educational process. ESD-TR-65- , Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Decision Sciences Laboratory, ESD L.G. Hanscom Field, Bedford, Mass. 01730	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A

3. REPORT TITLE

Cybernetic Testing - Project 2806

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

November 1963 thru May 1965

5. AUTHOR(S) (Last name, first name, initial)

Shuford, Emir H., Jr.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1967	29	3

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. 2806, Task 280609	ESD-TR-67-229
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	None

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

This report presents a concrete realization of an admissible probability measurement procedure utilizing a computer-driven scope and light pen. This particular technique is appropriate for all multiple-choice type of testing.

Empirical results are reported from an analogous pencil-and-paper realization of the same admissible probability measurement procedure. These results indicate a marked superiority for admissible probability measurement over traditional multiple-choice testing.

It is suggested that further gains can be obtained by using admissible probability measurement procedures to sequentially test the scope of knowledge of a student.

DD FORM 1 JAN 64 1473

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

GPO 886-551

Printed by
United States Air Force
L. G. Hanscom Field
Bedford, Massachusetts