CONTRACT REPORT

C/PERS 303

THE PROGRAMMING OF PSYCHOLOGICAL TESTS ON THE PLATO SYSTEM

Submitted to

CONTROL DATA CORPORATION

NATIONAL INSTITUTE FOR PERSONNEL RESEARCH
COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH

CSIR Contract Report C/PERS 303 (p. i - vii; 1 - 57)
UDC 681.322.063.3 : 159.9.073
Johannesburg, Republic of South Africa, March 1981

CONTROL DATA CORPORATION

by

T.R. Taylor, N. Gerber and M.E. Rendall

NATIONAL .INSTITUTE FOR PERSONNEL RESEARCH
COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH

CSIR Contract Report C/PERS 303 (p. i - vii; 1 - 57)
UDC 681.322.063.3 : 159.9.073
Johannesburg, Republic of South Africa, March 1981

CSIR Contract Report C/PERS 303

Published by
National Institute for Personnel Research
Council for Scientific and Industrial Research
P.O. Box 32410
Braamfontein Republic of South Africa

March 1981

Printed in the Republic of South Africa by
National Institute for Personnel Research

A C K N O W L E D G E M E N T S

This report describes some of the work done to date on project 80/6, "Adaptation of selected NIPR tests for PLATO". The research is sponsored by the Control Data Corporation.

The project was directed by Dr. G.K. Nelson, Director of the National Institute for Personnel Research.

SUMMARY

This report covers three major areas:

1) A survey of the literature on computerized testing. The advantages and disadvantages of computerized testing are discussed.

2) A description of some of the features and facilities of the PLATO system.

3) A description of the tests which have been computerized on the PLATO system. These tests include: the Arithmetic Reasoning Test, the Estimation Test, the Words in Context Test, the Scrambled Sentences Test, the Figure Classification Test and the Continuous Symbol Checking Test.

OPSOMMING

Hierdie verslag dek drie hoofareas:

1) 'n Oorsig van die literatuur wat betrekking het op gerekenariseerde toetsing. Die voordele en nadele van gerekenariseerde toetsing word bespreek.

2) 'n Beskrywing van sommige van die kenmerke en moontlikhede van die PLATO-sisteem.

3) 'n Beskrywing van die toetse wat deur middel van die PLATO-sisteem gerekenariseer is. Hierdie toetse sluit in: Die Rekenkundige Redeneringstoets, die Skattingstoets, die Woorde-in-Kontekstoets, die Geskommelde Sinnetoets, die Figuur-Indelingstoets en die Ononderbroke Simboolnasien-toets.

TABLE OF CONTENTS

Page

TABLE OF FIGURES

1. INTRODUCTION

Over the last two decades, the computer has played an ever greater role in applied and research psychology. During this period the computer has grown in sophistication and shrunk in cost; it is not surprising therefore that the use of this machine in psychology has extended well past its original application in data analysis. New fields of computer assisted instruction (CAI), computer managed instruction (CMI), adaptive control of experimental situations and test administration have developed (Johnson, 1978; Schmidt and Urry, 1978; Kalish, 1980; Cory, 1977; Boyle and Smith, 1976).

There has been a dramatic increase in interest in computerized psychological testing over the last few years, and several studies investigating the feasibility of computerized testing and comparing conventional with computerized tests have been undertaken (Hansen and Ross, 1977a,b ; Johnson et al, 1978; Biskin, 1977; Seguin, 1976; Elwood and Clark, 1978; Cory, 1977). A substantial body of test theory for the new computerized tests has been developed by Lord (1970,1977), Samejima (1977a,b), Mc Bride (1977), Jensema (1977), Krus and Ceurvorst (1978) and others. To date, computerized tests have been used almost exclusively in an experimental context. Johnson et al (1978) describe a commercially available system; this system, however, is used mainly for administering and scoring clinical instruments like the MMPI, and is almost completely lacking in interactive features.

The NIPR's interest in computerized testing is certainly not limited to purely experimental applications. Although many basic research questions will be investigated using computerized tests, a major aim is to produce a commercially viable computerized testing system which will offer a full range of psychological tests for use in selection and placement contexts. These tests will be available to suitably qualified and registered persons who have access to a PLATO terminal. Terminals may be installed anywhere in the country and connected to the system via telephone lines.

As can be seen from the above paragraph, the NIPR's involvement in computerized testing is large-scale and ambitious. To the best of the author's knowledge, there is no other system in the world designed to be used commercially for selection and placement on a national basis. Although the NIPR system is certainly not ready to be brought into service at the present moment, substantial development has been achieved and a number of tests have been programmed onto the system. One of the aims of this report is to describe some of the more noteworthy features of these new tests.

Computerized testing affords many advantages over conventional testing and few disadvantages. These advantages and disadvantages will be discussed in Chapter 3. As the computer can assess a subject's response immediately after it has been made, it is possible to apply adaptive procedures both in instructional and in testing material. In other words, the computer can be programmed to react in a way appropriate to the subject's response. The present array of NIPR computerized tests are adaptive in the presentation of the instructional material, but not in the presentation of items. The development of full-scale adaptive testing will comprise a second phase of test construction. This phase will be a fairly natural development of the first phase which will supply information on item characteristics to be used in the construction of fully adaptive tests.

The NIPR's computerized test program is being supported financially by the Control Data Corporation. Control Data's PLATO system is being used for this purpose. PLATO was designed to be a teaching system and a vast array of didactive programs on a wide spectrum of subjects is available on the system. With the addition of psychometric material, PLATO becomes a medium for both teaching and testing.

2. A BRIEF SURVEY OF THE LITERATURE ON COMPUTERIZED TESTING

One of the most important questions in computerized testing is, "How does the subject react to this new testing medium?" Most people have had minimal experience in viewing and responding to textual and graphic material presented on a cathode ray tube (CRT). Also, few have had the experience of interacting with a computer, of talking to a machine which instantly makes an evaluation of one's response and reacts in an appropriate manner. Would the computer be seen in such circumstances as a sinister onmiscient machine?

Investigations conducted into the responses of subjects to computerized testing have generally found very positive reactions to this type of testing. Schmidt et al (1978) tested 79 males and 84 females on compterized tests and then measured their reaction to the tests using a Likert-type questionnaire. The responses were overwhelmingly positive. Verster (personal communication) has also found very positive subject reactions to computer-interactive testing. Elwood and Clark (1978) successfully administered a computer version of the Peabody Picture Vocabulary to young children. Biskin (1977) applied the MMPI to subjects in conventional and computerized CRT versions. There was little difference in scores, except in the "Cannot say" scale where more people responded positively than with "?" on the CRT version. This finding might indicate that the subjects doing the CRT tests were more involved and interested in the task.

Pencil-and-paper tests have been found to correlate highly with equivalent computerized tests (Cory et al, 1977). Very high test-retest reliability correlations (often greater than 0,9) have been reported for computerized tests (Elwood and Clark, 1978). These very high correlations are probably attributable to the standardized administration conditions, especially with regard to the instructions. This is a major advantage of computerized testing.

The standardized administration capability of computerized testing is complemented by its branching capability. Boyle and Smith (1976) describe an application of computerized testing where this facility was used extensively. CAI was used for second language teaching, then knowledge was tested using computerized tests. The branching facility enabled the testing to be used in a diagnostic manner to identify areas of difficulty for the student.

Immediate scoring is another capability of computerized testing. It is possible to give subjects feedback on their performance, for instance to inform subjects whether they answered each item correctly or incorrectly. Gialluca and Weiss (1980) found that feedback of results after each wrong response did not improve test scores or affect test dimensionality. Betz (1977) divided her sample into high ability and low ability groups. She found that in the high ability group, knowledge of results improved performance. In the low ability group, no significant difference was found. Knowledge of results might have widened the gap between high ability and low ability performers because high ability performers used the feedback to improve their future performance, whereas low ability performers became discouraged.

Computerized testing can be used to measure abilities and skills which would be difficult or impossible to measure using conventional methods. Cory (1977) mentions the advantages of CRT display for the presentation of pictorial material, and animated displays. Cory has used the CRT to measure five abilities which are well suited to administration via CRT. In his battery of Graphic Interactive Processing tests (GRIP), there are measures of short term memory, perceptual speed, closure, movement detection and sequential concept identification. Cory found GRIP tests very useful in predicting on-the-job performance on visual displays, adjusting equipment and working with distraction.

A substantial amount of research has been put into the development and investigation of tests with variable item content, although little clarity exists as to the relative merits of different varieties of these tests. The underlying structural and statistical models have reached a fairly advanced state, without having attained full maturity. The methods go by various names - stradaptive testing (Weiss, 1974), adaptive ability testing (Mc Bride, 1977; Owen, 1975), tailored testing (Samejima, 1977b) and flexilevel testing (Lord, 1970, 1971, 1977). However, all share a common feature that the subject's responses influence what items he is given.

One of the greatest advantages of adaptive tests is that they tend to be substantially shorter than conventional tests. Time is not wasted by giving a subject items which are far above or below his ability level. McKinley and Reckase (1980) state that the standard error of measurement for conventional tests is usually higher at the extremes of the ability domain than is the case with adaptive tests.

Waters (1977) randomly assigned 102 subjects to one of two testing groups. One group was given a conventional version of the School Ability Test (SCAT) and the other group a computerized version of the SCAT. Results showed significantly higher reliability for the adaptive test and equivalent validity indices. Adaptive testing involved 20 - 30 items as against about 50 for the conventional test.

Kalish (1980) found a 75 % timesaving on adaptive tests with equal or better prediction of a criterion (mastery of weapons maintenance). Hansen and Ross (1977a) found a somewhat smaller saving (30 %). Correlations between adaptive tests and their conventional counterparts was very high (of the order of 0,99). Kingsbury and Weiss (1979) state that checking whether a subject has mastered course material as a criterion to decide whether to allow the individual to go on is rather hit-and-miss when a cut-off on a fixed set of questions is used. A fuller knowledge of the subject's level of mastery can be obtained by selecting mastery items from a pool, the selection being based on performance on previous items.

3. ADVANTAGES AND PROBLEMS OF COMPUTERIZED TESTING

Some of the advantages of computerized testing have been mentioned in previous chapters. In this chapter we will present a more formal exposition of the advantages and problems of this type of testing.

3.1 Advantages

3.1.1 Flexibility of testing times

As formalized testing groups are not a feature of computerized testing, testing can be distributed over the whole day; in fact it must be in order to optimize the use of the terminal. With careful attention to the scheduling of terminal time, subjects can be "fitted in" when it suits them without disrupting the testing procedures.

3.1.2 Standardization of testing procedures

Conventional tests have manuals which specify correct testing procedures. However, this method of standardization is inadequate in several ways:

* Testers often inadvertently or deliberately depart from the standard "patter".
* The tone of the tester's voice can either emphasize or de-emphasize a point.
* Gestures and idiosyncracies on the part of the tester can affect the quality of instruction.
* A tester may rush through the instructional material, leaving some of the testees "behind" or he may be so meticulous that he bores and irritates some of the testees.
* The test manual may be inadequate in that it omits to give contingency procedures for certain situations.
* The timing of tests may not be implemented with sufficient care.

Computerized testing overcomes all of these problems. The presentation of the instructional material is absolutely standard. Points cannot be emphasized or de-emphasized. The subject paces himself through the instructions; hence he can move to new material when he feels ready to do so. The timing of the test is controlled highly accurately by the computer itself. The programmer has to think very carefully and logically about the instructional material while programming a test. Usually all the possible courses which the program can take are set out in a flowchart. (Some of these will be presented later.) Hence it is almost certain that all contingencies will have been taken into account.

3.1.3 Comprehension of Instructional Material

Instructions in a computerized test can be presented in such a way as to optimize comprehension. Tests can be built into the instructions at various points to determine whether the subject has understood the material up to that point. A subject who fails a test can be branched to more explicit and simply phrased instructional material and then tested again. If he fails the test a specified number of times, he can be eliminated from the testing procedure before he reaches the items. Hence, well written instructional programs can optimize comprehension, test for complete understanding of instructions before letting the subject into the test itself and eliminate subjects who should not be given the test due to their inability to understand what they are expected to do or their inability to do elementary examples.

Often many of the zero or near-zero scores in a distribution of test scores are contributed by subjects who failed to comprehend what was required of them. Computerized testing is fairer in that it does not permit this to happen. It also permits more satisfactory test development; the psychometrician can be sure that his sample is not made up of two subsamples - the "comprehenders" and the "noncomprehenders". Scores on the test are a pure reflection of ability uncontaminated by differences in comprehension.

3.1.4 Demands on the Tester and Control of Testing

As a computerized test can be designed to run itself and take account of all contingencies in the testing situation, the demands on the tester are light, and he may be able to do other tasks while testing is in progress. All he has to know is how to call up programs, to set up batteries and what various termination messages mean. The tester would also have to take action in the event of system failure.

The composition of batteries might be specified for the tester by a superior. The whole testing procedure can be controlled and monitored by a highly qualified expert who is not physically at the place of testing. This person can monitor and control events at a number of sites: test results can be called up by him immediately after a test has been completed. These results might cause him to modify the rest of the test battery. Control by one expert over a number of distributed testing operations, and the flexibility to modify test batteries on the basis of previous results are two of the great advantages of computerized testing.

3.1.5 Distribution of Test Sites

Terminals may be sited anywhere where there is access to a telephone line. For psychological testing, further requirements would be that the terminal be situated in a quiet room with the correct lighting levels and a minimum of distracting stimuli. These requirements are fairly easy to meet. Points mentioned in this section and in the previous one should have made it clear that greater decentralization is possible with computerized testing. This is a cost-saving feature. The expense of transporting either testees or testers long distances can be reduced substantially.

3.1.6 Control over who uses the Test Programs

In conventional testing, abuses can occur in test administration. Tests are categorized into three levels: A, B and C. Certain minimum qualifications are specified for administration of each of these categories, and registration with the Test Commission is required before a person is empowered to use tests for which he is qualified.

Effective monitoring of test administration is extremely difficult if not impossible due to the widely dispersed use of tests. Abuses no doubt occur, with unqualified and unregistered persons applying tests, possibly in an incompetent manner.

With computerized testing, much tighter control over test usage is possible and the likelihood of incompetent test administration reduced to a minimum. The security of test results can also be brought to a very high level with computerized testing. More will be said about these topics in a report by M.A. Coulter.

3.1.7 Elimination of the Human Factor in Scoring

Computerized tests are generally programmed to score themselves. Hence the clerical tasks of scoring and referring scores to norm tables are eliminated; savings in the time of personnel and the elimination of clerical error can be effected. The scoring of some pencil-and-paper tests is time-consuming and laborious, and the tests are often not used for this reason. Computerization eliminates these problems.

3.1.8 Range of Scores Available

In computerized testing, one can keep track of many more aspects of the subject's test behaviour than is possible in conventional testing. It is possible, for instance, to measure the exact amount of time which the subject spent on each item and what answers he selected before giving his final answer.

3.1.9 Updating Test Material and Information

Updating "hard copy" tests and manuals involves the publication and distribution of new material and the withdrawal of old material. This is a laborious and time-consuming procedure. Necessary updates might sometimes not be done because of the expense and effort involved. Like tests, manuals can also be computerized and made available to users via CRT display. Updating both tests and manuals is then a simple task, involving in many cases no more than a few minutes work by a programmer. Warning could be given to users via the computer that a new version of a test or manual would be available from a particular date. Overnight, the old version could be withdrawn and the new version substituted.

Norms can also be kept up to date very easily in computerized testing. Smaller organizations could use norms developed by the test supplier, and larger organizations could develop their own norms. These norms could be kept right up to date by basing them on the most recent N (say 300) testees.

3.1.10 Royalties

Test suppliers have limited control over how test material is used after sale. The only continued source of income is for answer sheets, and unscrupulous users might reproduce these (and possibly even the tests) privately. These abuses are eliminated with computerized testing. Control can be kept over test use to the extent that the test supplier is aware of how many times each of his programs has been run. Users can be charged a royalty for each use of a program. This seems to be a much more equitable arrangement for the test supplier.

3.1.11 Construction of New Types of Tests

This is one of the greatest advantages of computerized testing. In computerized CRT testing, it is possible to achieve the following:

* Animation (on certain systems)
* Accurate measurement of reaction times
* Branching capability - the material presented to the subject can be dependent upon previous responses
* Adaptive item presentation - this is a special case of branching, although a very important one

We have already seen in the previous chapter that Cory (1977) has capitalized on some of the capabilities of computerized testing by measuring abilities which cannot be measured easily using conventional methods (e.g. short term memory, perceptual speed, sequential concept formation). Computerization affords great potential for test development in the areas of problem solving and decision making. Conventional test scores are usually "static" in that they describe a measurable feature of a psychological construct on a scale of intensity. This intensity is in some ways analogous to the mechanical index of <u>efficiency</u> which is an overall measure of a machine's performance but one which tells nothing about how the machine works, how different parts of the machine interact with one another. Computerized testing, on the other hand, can be dynamic in that underlying processes of mental functioning, not just overall levels, can be investigated. What makes computerized testing capable of being dynamic is the interactive nature of the machine-man relationship. The testee may ask for more information; the computer may comment on the testee's performance or offer alternatives. The testee and the computer form a closed interacting system. For instance, the testee may make a certain decision and the computer may feed back to him the results of his decision. This can act as the basis for further decision-making on the part of the testee. Tests such as the In-Basket, which are beset with problems in conventional form, could be transformed in a computer-interactive version.

3.2 The Problems of Computerized Testing

3.2.1 Literacy

At present, only literate people can be tested on computerized tests, as the instructions are presented in written form. This is not a great drawback, as most test users would probably envisage using computerized testing at the middle and upper levels of the testing population. In any event, the literacy problem is only temporary. Some of the new generation computer terminals will have synthesized speaking facilities, and methods exist for recording responses without the subject having to type or write anything.

3.2.2 Testing is Individual

Unless a large number of terminals are available, it is not possible to test a large number of people in a very short period of time. However, with effective scheduling of terminal time, it is usually possible to minimize this problem.

3.2.3 Terminal Cost-Effectiveness

Cost per testing run can be high if the terminal is underutilized. Computerized testing becomes much more economic if the terminal is in use most of the day. If this is done and terminals are decentralized (i.e. placed in areas where there are concentrations of testees), then travel savings can be effected. Also, the relatively low demands which computerized testing places on testers means that economies in staff time and in the level of staff required can be realised.

3.2.4 Making Computer Versions of a Pencil, Rubber, Answer Sheet and Finger to Turn the Page

In conventional testing, we take all of the above for granted, but in computerized testing, equivalents for them have to be devised. In conventional testing, the subject is usually free to skip items, rub out his answer, turn any number of pages back to previous items. In order to give subjects a comparable amount of freedom in a computer test, computer versions of the above have to be programmed.

This has been achieved to a very large degree in the NIPR tests programmed on PLATO. The subject has almost as much freedom as he would have in an equivalent paper-and-pencil test.

3.2.5 System Failures

What happens if the system "crashes"? The subject might be interrrupted in the middle of a test. Nothing can be done about the immediate consequences of a "crash"; the current display may remain frozen on the screen, or the screen may be overwritten with a system failure message, or the screen may go blank. It is possible, however, to prevent any further disruption by "crash-proofing" tests. This makes it possible to restart the testee, once the system comes up again, at the correct place in the test and with no loss of information about his performance up to the point when the system went down. More will be said about this in Chapter 5.

3.3 Conclusion

This chapter should have made it clear that there are many very worthwhile advantages to compterized testing. It should also be apparent that there are relatively few disadvantages and that most of these can be minimized by judicious programming and efficient use of terminals.

4. SOME FEATURES OF PLATO

PLATO is a Control Data system which was designed originally for didactive purposes. The PLATO terminal consists of a keyboard and a 22 x 22 cm screen which can be illuminated at any of 256 x 256 points. The screen is 64 characters wide and 32 lines deep. The screen is also touch-sensitive. It can record the position of a touch by a finger or pencil on a 16 x 16 grid.

PLATO has a sophisticated graphic facility which makes it possible to construct detailed pictorial material. Some degree of animation is also possible. A wide range of typefonts is available and new fonts can be constructed by the programmer. Verbal material can be written at any angle on the screen. A character library can be consulted if some representational figure is desired (e.g. a hand pointing). Programmers can also construct their own characters de novo.

The system has a built-in clock which makes it possible to perform various timing operations.

It is possible to communicate with other PLATO users through what is called a "term talk" facility. This facility could be used by supervisors of psychological testing operations to implement any urgent changes in testing procedures.

It is possible to devise sophisticated security systems which can be used to control access to programs and information.

The PLATO system has its own programming language called TUTOR. This language is designed to be effective in interactive applications. The language has two modes: judging and normal. Numerous judging commands exist in the language. This enables the programmer to develop a flexible interactive interface between machine and respondent. A further advantage of the language is that it can be mastered fairly easily by individuals who have previously not been involved in programming. Useful pairings of skills can be embodied in a single person (e.g. psychology and programming, teaching and programming).

As was mentioned earlier, PLATO was originally designed as a testing medium. A vast library of teaching programmes covering a wide range of subjects is available on the system. However, up to the point when the NIPR began to program psychometric material onto PLATO, no tests of psychological abilities were available on it. Although the system was not explicitly designed for this purpose, it proved to be highly amenable to this type of application. The testing of knowledge and understanding of material is a regular feature of PLATO lessons. The machine can be programmed to judge the correctness of responses and, on the basis of this, route a person to material designed to fill in lacunae in his knowledge. The usefulness of this in psychological testing, especially in the creation of instructional material, should be immediately apparent.

In each lesson (or test), 150 student variables are available for whatever program-relevant purpose the programmer wishes to put them to. These variables are called student variables because they are used for controlling the program and recording information in accordance with the events of that particular program run. The student variables "belong" to the testee or student for the duration of the program run. This does not mean that only one person at a time can use a particular program. Each testee gets his own copy of the program and hence his own set of student variables. Many people can be using the same program simultaneously and they can all be at different places in the program. Each subject's "own" student variables are controlling the program and recording information about his performance.

An example might help to illustrate how student variables are used. Suppose that you wish to present certain material (e.g. a section of the instructions for a psychological measuring instrument) and then to test whether the subject has understood this material adequately. To do this, you present a few questions after the subject has seen the instructional material. You want to route the subject to new material if he gets all the questions right and to revision material if he fails to get them all correct. If he goes to the revision material, you want to route him back to be retested on the questions (after he has gone through this material); if he fails again, you want to return him to the revision material. If he fails to answer the questions correctly after three attempts, you want to eliminate him from the test.

All this can be done using a single variable and suitable TUTOR commands. Let this variable be called "ques". Figure 1 shows how ques can be used to perform the above procedure.

Student variables can be used for a variety of purposes, e.g. recording subject responses, controlling branches in the program (as in Figure 1), timing the test. The values in student variables are lost once a program has been terminated. It is necessary to save some of these values on disc so that they will still be available after the program has terminated. The following values might be written to disc before the subject exits from a test: the subject's response to each item, how long he spent on each item, how much time (if any) was left over after he exited from the test, what item he reached (if he was exited from the test because the time was up).

The structures and procedures involved in saving, storing and accessing subject information will be discussed in another report by M.A. Coulter.

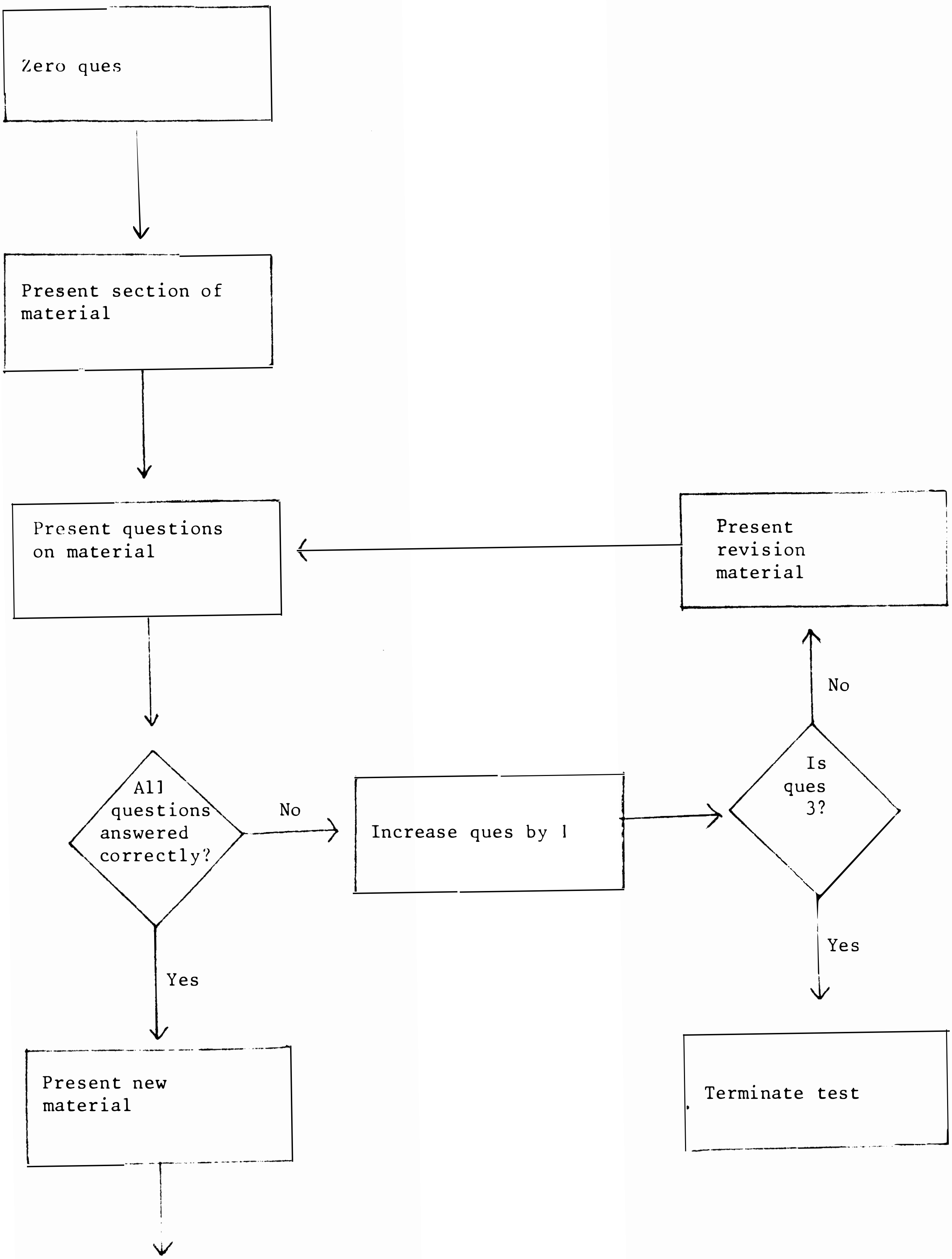


Figure 1   Example of variable controlled branching in TUTOR

5. THE PROGRAMMING OF TESTS ON PLATO

This chapter will describe the actual tests which have been programmed on the PLATO system and will discuss and illustrate some of the major features of the programs. The tests mentioned in this chapter are those which have been programmed up to a stage of completion or near-completion. Others which are at a more preliminary stage will not be described. (These include the tests from the NIPR High Level and Intermediate Batteries and two new perceptual tests.) The aim is ultimately to offer a complete array of tests, ranging across all the major ability domains and at levels which make them suitable for application to the main testee populations in South Africa.

It is necessary before going on to discuss the details of the programmed tests, to mention some of the features which are common to all or most of these tests.

5.1 Some General Features of the Programmed Tests

5.1.1 Exclusive use of the Touch-sensitive Screen

In all the tests described in this chapter, the testee never touches the keyboard. All responses, both in the instructions and in the test itself, are entered by touching the screen in an appropriate place. This mode of response entry has been selected because it is simpler and clearer. The keyboard has many keys which would not be used for any given response and which could confuse the testee.

Suppose, for instance, you gave the subject the following choice/

Do you want to

a) review the instructions, or
b) go straight on to the test?

Press the appropriate key.

If the subject is required to use the keyboard, he has to find the appropriate keys on a fairly crowded keyboard.

With touch processing, the same choice would be presented on the screen in the following way:

Touch the appropriate box to indicate what you want to do next:

Review the instructions	Go straight on to the test

The most commonly used touch processing controls the progression of the program. On most displays, there is a box marked "NEXT". The subject touches this when he is ready to go on to new material. Almost all aspects of the instructions and tests are self-paced.

5.1.2 The Back sequence

In most of the programmed tests, the subject is allowed to skip items which he does not want to attempt at that moment. He is then given the option at regular intervals in the test to return to the items which he skipped. This option is also offered after the last item in the test. When a subject elects to go back to a previous item, he is said to be on a Back sequence. The sequence works as follows:

1) The subject is offered the opportunity to go back or go on.
2) If he elects to go back, he is shown a list of the items which he has skipped. He is also allowed to change his mind and go to a new item if he so wishes.
3) The subject can go back to a given item by touching the appropriate part of the screen (a box with the item number in it).
4) He is then returned to the requested item. He can answer the item, or skip it a second time.
5) When he is finished with the item, he is returned to the display which lists his skipped items. He can again go

back to a previous item or he can go on to a new item. He is not limited in the number of times he is allowed to go back to a previous item. (Obviously, if the subject attempts all his previous uncompleted items whilst on a Back sequence, he is returned automatically to the main sequence.)

6) Whenever a subject emerges from a Back sequence, he is shown the next new item in the test. Hence, if he went into a Back sequence on item 15, he would emerge from the sequence at item 16.
7) Back sequences are only offered if there are unattempted previous items.

5.1.3 Changing Responses

All the programmed tests allow the subject to change his answer. In some tests, the subject touches a box labelled "CLEAR" and then enters his new answer. In other tests, touching another answer automatically clears the answer box and enters the new answer. In all tests (except Continuous Symbol Checking), the subject can change his answer only while he is doing the item in question. As soon as he touches "NEXT" to go on to the next item, his final answer is stored and he cannot change it. He is allowed only to return to unattempted items (using the Back sequence).

5.1.4 "Crash-proofing"

Precautions have been taken against the eventuality of system failure during testing. All relevant information on the subject's performance is saved after each item. If the system fails, this information is not lost; once the system comes up again, the subject can be restarted at the appropriate place in the test. (If, for instance, the system crashes while the subject is on item 10, the testee will be brought back into the test at this point. All information on his responses and skips up to item 9 will have been saved and fed into the program.)

A further refinement is that when the system comes back up after a failure, the subject is offered the opportunity either to see the instructions again and then resume at the appropriate place in the test, or to go straight to the item which he reached when the system went down.

5.1.5 Timing

All the programmed tests described in this chapter are timed using an internal system clock. For all tests except the Continuous Symbol Checking Test, the remaining testing time is displayed for the benefit of the testee. (The Continuous Symbol Checking Test does not have this feature because the subject is deliberately kept in ignorance about the test duration.)

The computerized Estimation Test has an additional timing feature: there is a time limit for each item and the subject is automatically stepped on to the next item if he remains on the same item for 60 seconds. He is given a warning five seconds before the time limit is up.

In all the tests, the length of time spent by the subject on each item is saved. This information can be used in item analysis and is also used to reset the testing time correctly after a system failure.

5.2 Description of the Programmed Tests

5.2.1 The Figure Classification Test (program fcts) and High Level Figure Classification Test (program fct)

The above two programs are complemented by two Afrikaans version tests (programs fctas and fcta), but as these are directly comparable with their English counterparts, only fcts and fct will be discussed.

The computerized test fcts is based on the pencil-and-paper NIPR test FCT. The items are virtually identical; only small changes have been introduced occasionally in order to make the item more suitable for CRT display. There are no instances of basic change in the nature of items. All the concepts embedded in the items of FCT are also to be found in fcts.

Test fct is also based on a pencil-and-paper test - High Level FCT - which, although not officially an NIPR test yet, is at the point where it is ready for submission to the Test Commission for registration. The High Level FCT is intended for administration to Blacks with Standard 9 or higher qualifications and to Coloureds, Indians and Whites in the range Standard 6 - Standard 9. (The standard FCT is intended for Blacks with Standard 5 - Standard 8).

A few words should be said about what these tests measure. Taylor (1977) describes the FCT as a measure of concept identification. This construct is closely related to certain other constructs which have appeared in the psychological literature, e.g. abstraction, ability to categorize, stimulus generalization, concept formation. In both the FCT and High Level FCT, the subject has to identify a wide range of concepts which are essential to effective functioning in a Western-technological society (e.g. rotation, symmetry, repetition, analogous relationships).

These concepts are presented graphically in both FCT tests. Each item consists of six diagrams embodying two complementary concepts. Three of the diagrams express one concept and three the other. Various embedding and distracting material is present in the diagrams. The subject therefore has to be able to filter out the specific "noise" in order to identify conceptual similarities between diagrams. Because of the limitations imposed on the item structure (always six diagrams, two concepts and three diagrams representing each concept) it is possible to impose a unique right answer on each item.

Figure 2 presents an illustration of an item of fct and Figure 3 a flowchart of the fct and fcts programs. On the screen, fct and fcts appear identical up to the end of the instructions; both fct and fcts have 36 items, but the items of fct are more difficult.

The main interactive features of fct and fcts are illustrated in the flowchart. Note for instance that the subject has to demonstrate his understanding of the basic requirements of the task by answering three questions correctly. If he fails to achieve this, he is sent to revision material which presents the rules of the test more simply. Only after the subject has answered all questions correctly is he allowed to see an example of an actual item; subjects who fail the questions three times are eliminated from the test. The reasoning behind this is that an individual who is incapable of understanding the basic instructions of the test after three exposures to them is not likely to benefit from additional attempts at clarification and hence should not be doing the test.

There is a further criterion which a subject has to pass before being allowed into the test. He has to answer two consecutive practice items correctly. If he fails a practice item, he is informed that his answer was wrong and he is told which concepts he should have identified. He is then sent back to do the item again. This procedure is repeated if necessary until he has answered the item correctly. The subject then has to start "from scratch" and answer two consecutive items correctly; answering a failed item correctly on a second or third try does not count towards the criterion. There are five practice items. If the subject runs out of practice items before ataining the criterion, he is eliminated from the test.

There are many other feedback features of the fct and fcts programs which have not been incorporated in the flow diagram. For instance, when the subject is being taken through the answering of the example item, messages confirming correct behaviour are written on the screen at appropriate times.

Nearly all the instructions are given in a display in which the example item appears. As the instructions refer to different aspects of the item display, it is of obvious advantage to have both instructions and an item on the screen simultaneously. Flashing arrows are used to point out the item features to which the instructions refer.

All instructions are self-paced. The subject moves on to the next section of the instructions by touching a relevant area of the screen ("NEXT").

The PLATO system allows part-erasures of the screen to be effected. In a large section of the fct and fcts instructions, the example item remains on-screen, but new instructional material replaces the old when the subject moves the program on by touching "NEXT". If the subject spends an inordinately long period at any one place in the instructions, a message appears on the screen urging him to respond, e.g. "Waiting for you to touch diagram A...".

One of the disadvantages of printed instructions is that a page full of printing may not be read carefully. The subject might feel himself to be under time pressure and only skim the material. This problem can be overcome in PLATO by sequentially building up a page of instructions. The subject touches the screen and instruction 1 appears. He is asked to touch the screen again, and instruction 2 appears below instruction 1. The subject has to touch the screen yet again for instruction 3, and so on. In this way a whole page of information can be built up gradually.

This technique can be supplemented by using slow writing. Verbal material can be made to appear on the screen at about the rate of one word a second. Although the extensive use of slow writing can make for tedious reading, limited use of this technique can be effective for emphasizing points.

Both the page build-up and slow writing techniques are used in fct and fcts for giving certain final information before the start of the test.

We now come to the test itself. Figure 2 should be consulted in order to give context to the discussion below.

There are nine places on the screen where the subject may touch during an item display. These are: the boxes marked "NEXT", "CLEAR" and "SKIP" and the six diagrams marked A - F. The test has been programmed so that the computer will "know" which of these areas has been touched.

The subject commences an item by touching three diagrams which belong together in the same concept. The letters which label these diagrams appear in the top answer box. This is the point that has been reached in Figure 2. The subject then completes the item by identifying the second concept in the same way that he identified the first. Both answer boxes are then filled. If the subject is satisfied with his answer, he can touch the "NEXT" box. The next item will then appear on the screen (if, after every fifth item, the option to go onto a Back sequence). If the subject wishes to change his answer, he can touch "CLEAR". The answer boxes will then clear and he will have an opportunity to do the item again. "CLEAR" can actually be touched at any time while doing an item. If a subject cannot do an item, or if he wishes to be able to come back to it later, he can touch "SKIP".

The program does not permit the use of the same diagram twice in the same group. If, for instance, the subject touches B, E, B, the second B will not be accepted and the subject will be instructed to select another diagram, or clear his answer boxes and start afresh. The program does not object, however, if the same symbol is used in different groups. The reason for this is that a replication in a single group is likely to be a clerical error, whereas a replication in different groups might be a conceptual error.

The subjects can fill in their answers in any order.

| A E F | | B C D | , | B C D | | A E F | and

| F E A | | C B D | are all regarded by the program as equivalent. This makes the computer version of the FCT somewhat more flexible than the pencil-and-paper version, which requires the subject to begin with diagram A.

The following information on the subject's performance is available immediately after testing:

1) His score on the test
2) His actual response for each item (this makes it possible to determine what actual answer he gave if he answered incorrectly, and whether he omitted the item).
3) How long (in seconds) he spent on each item, including time spent on returns to the item via Back sequence.
4) How much time he spent on the test. The program only counts the time while items are on the screen. All other time is "dead time".
5) What item he reached.
6) Whether or not the system crashed during testing.

Although the average test user is unlikely to want detailed information on each item ((2) and (3)), this information can be very useful in test development and in monitoring individual items over protracted time periods or across different samples.

5.2.2 Words in Context Test (program wict) and High Level Words in Context Test (program wicts)

The Words in Context Test measures verbal comprehension, with particular emphasis on the ability to discriminate between the meanings of different words. The testee is required to read and understand a sentence from which a word has been omitted.

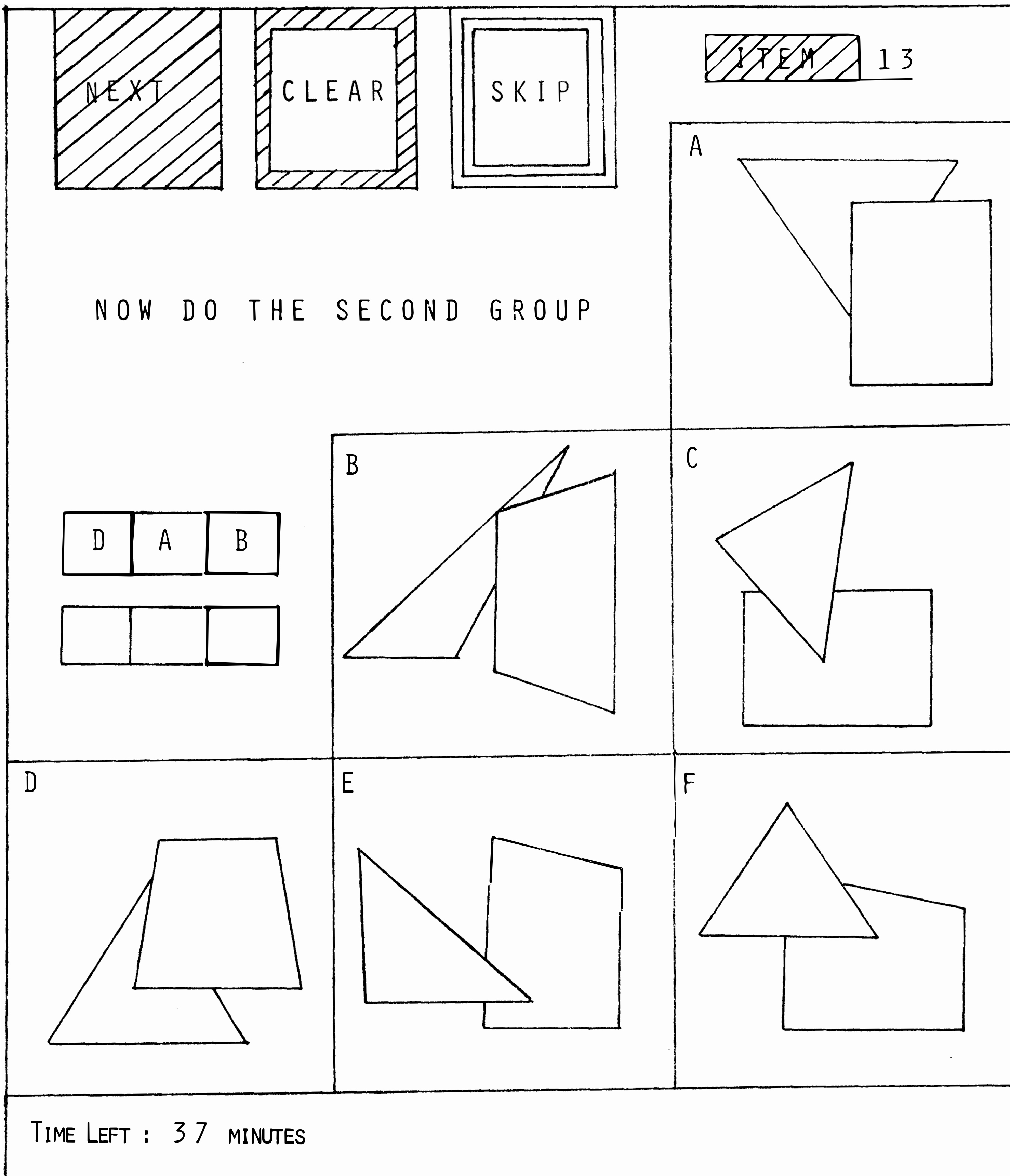


Figure 2 An example of an item from fct, the program of the High Level English FCT. The subject has just identified the first concept. Note that the relevant letters have been filled in the top answer box and a message has been written on the screen instructing him to do the second concept.

Figure 3 Flowchart for fct, fcts, fcta and fctas

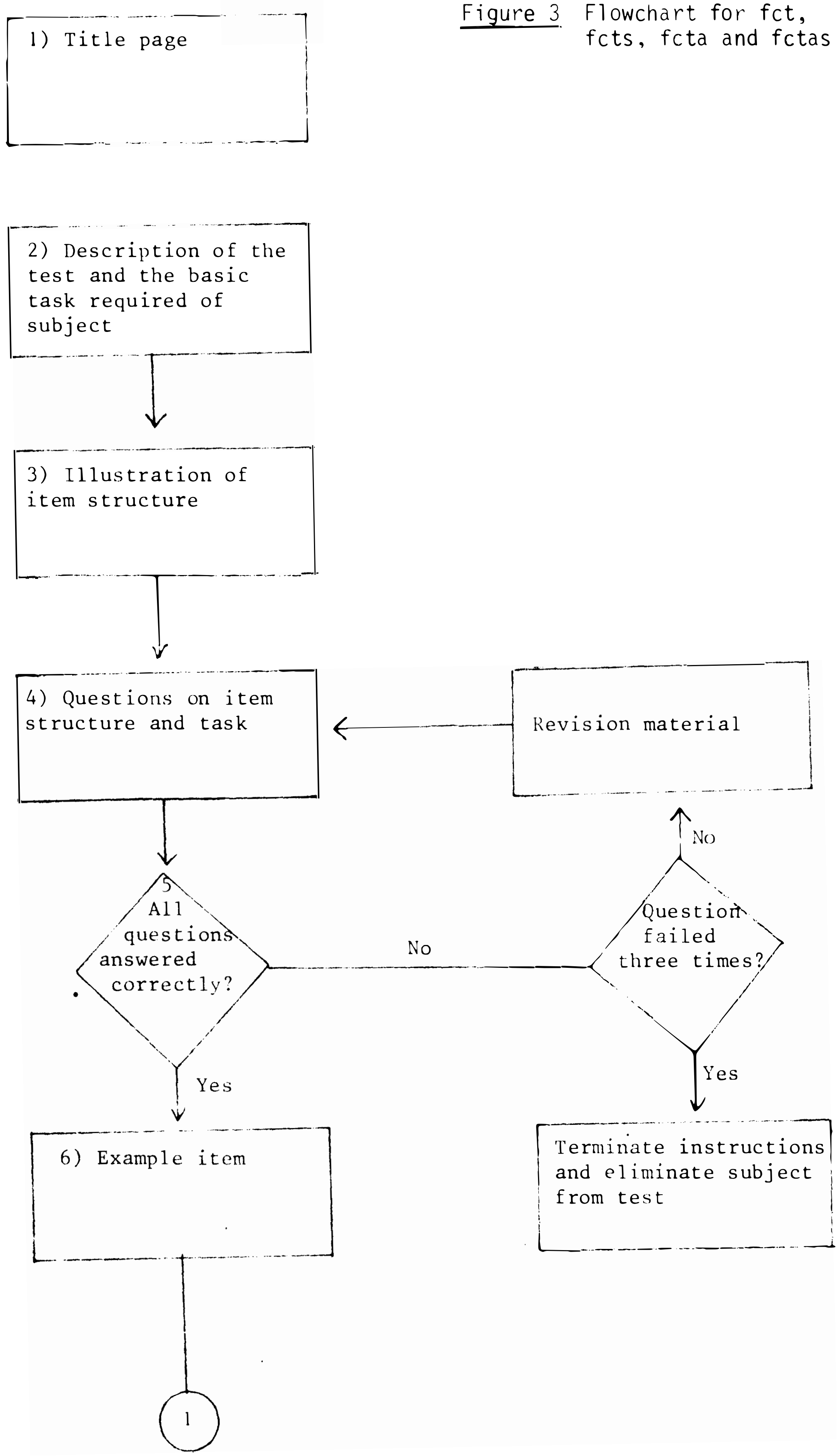


Figure 3 Flowchart for fct, fcts, fcta and fctas

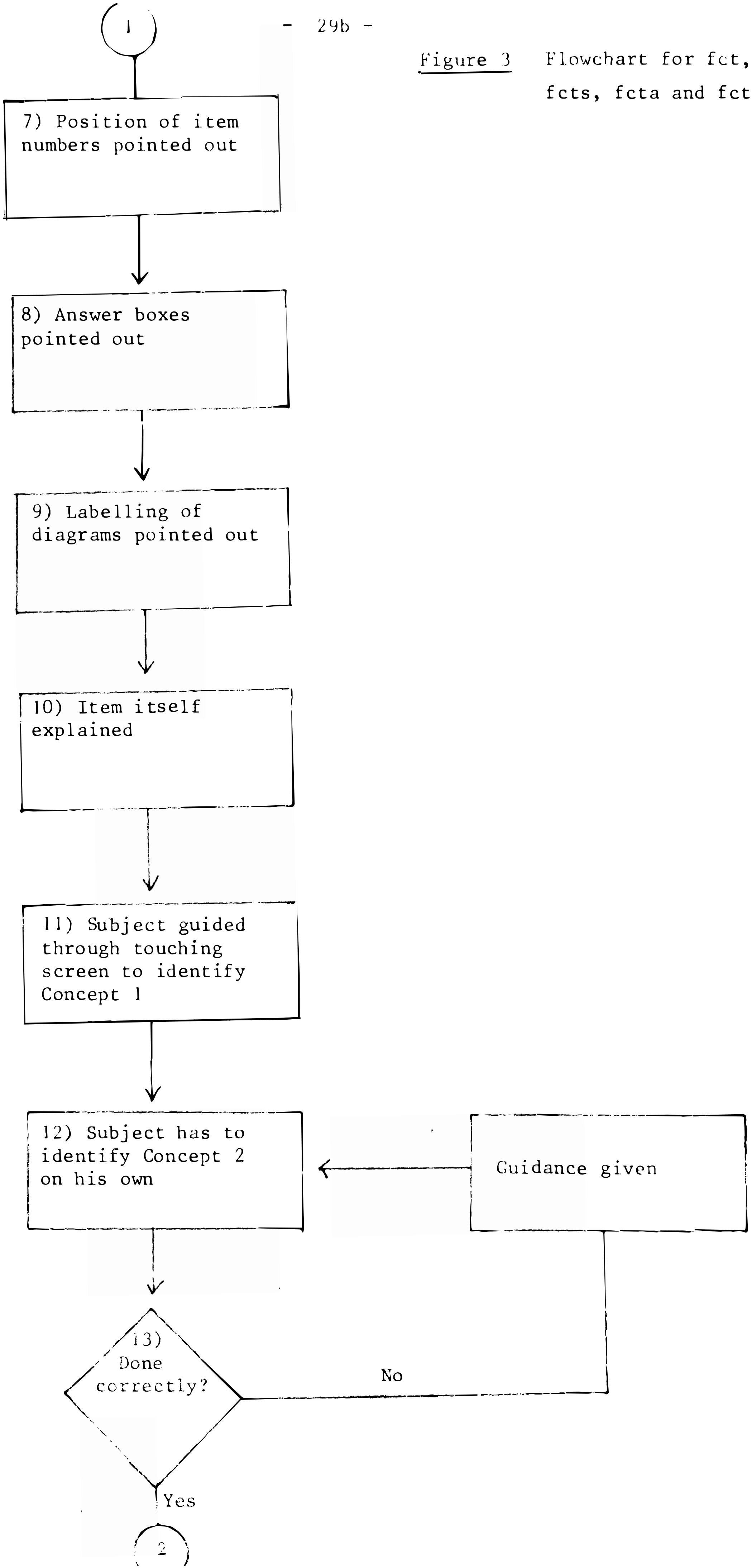


Figure 3 (continued)

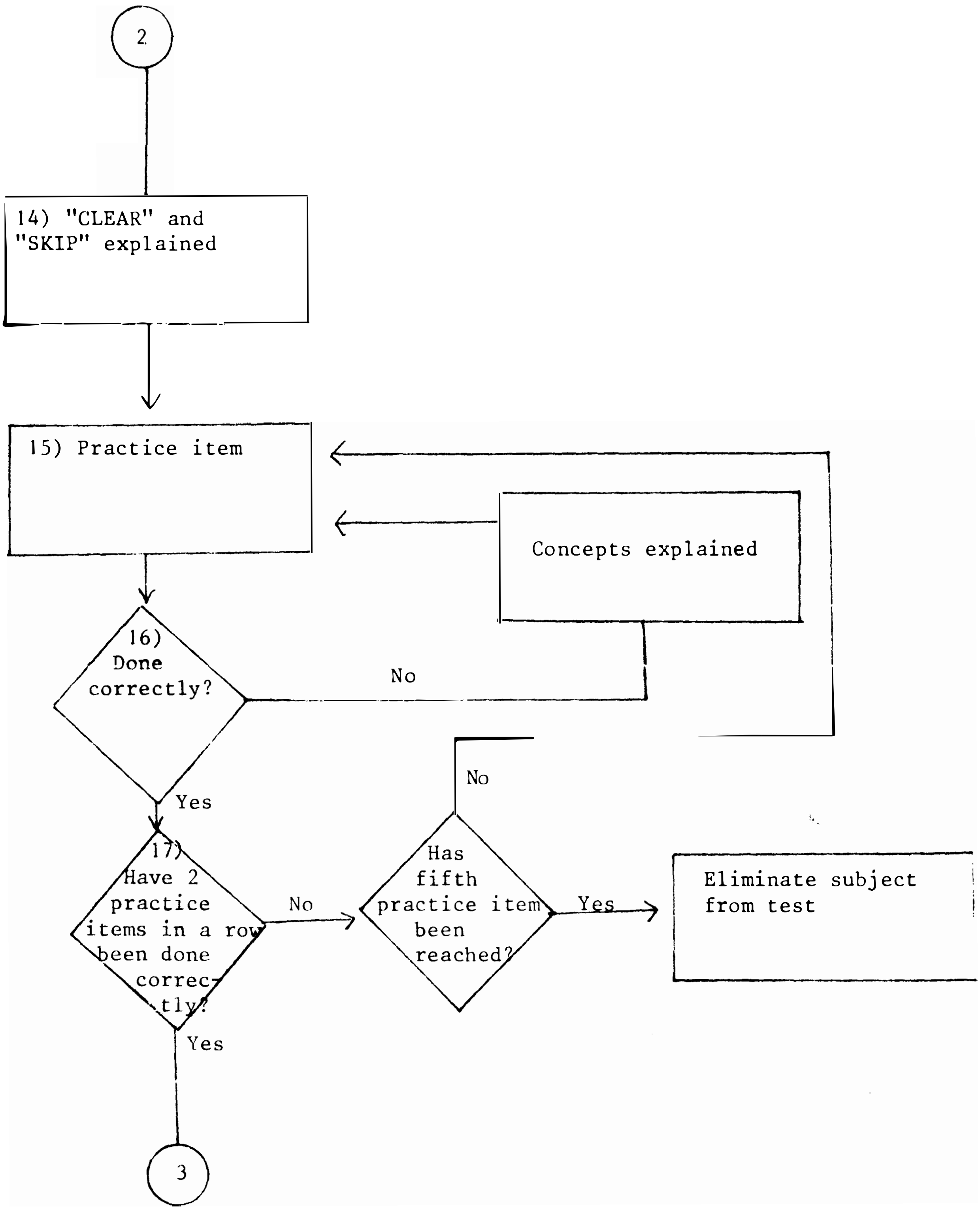
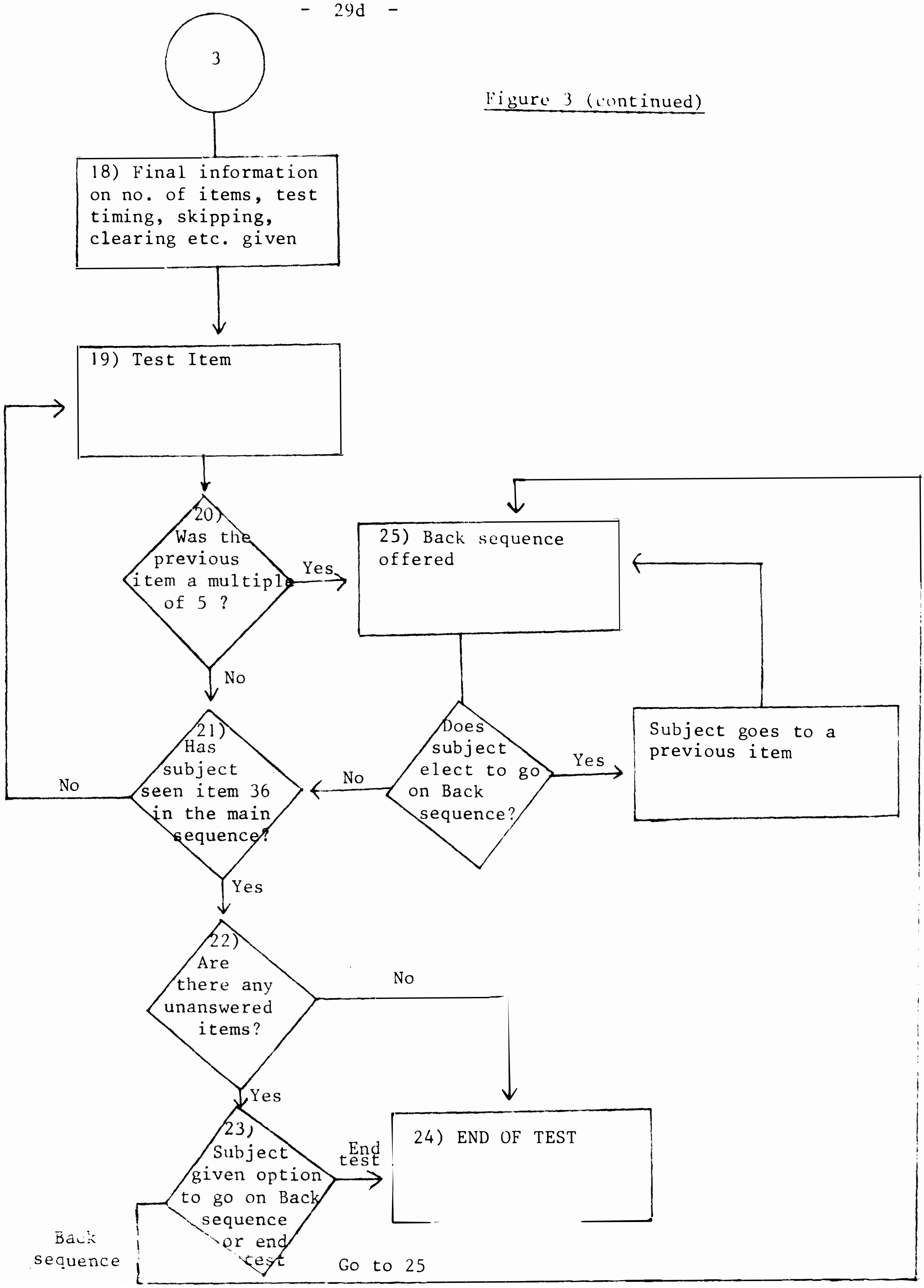


Figure 3 (continued)



The item answering procedure is as follows. As in all the programmed tests mentioned in this report, the testee enters his responses by touching the screen. In each item, a sentence is displayed at the top of the screen with an empty box where the missing word should be (see Figure 4). Five possible answers appear in boxes arranged vertically. The testee touches the word he thinks most suitable to complete the sentence. This word then appears in the answer-box. The testee can change his answer before finally entering it by touching other words, which will replace one another in the answer-box. Feedback is provided when touching is inaccurate i.e. when a subject touches a part of the screen which is not in one of the boxes. Summarised instructions can always be accessed by touching the box marked "HELP". Progression to the next item is achieved by touching the box marked "NEXT ITEM". The word in the answer-box at this time is taken as the testee's final answer to the item. Items can be left unanswered by touching the box marked "CLEAR" before going on to the next item. Every fifth item, testees are given the option of returning to unanswered items via Back or of continuing with a new item. The following information can be accessed immediately after the subject has completed the test.

1) Number of items answered correctly.
2) Response to each item (that is, the alternative chosen).
3) Total time spent on the test.
4) Time spent on each item.

5.2.3 The Arithmetic Reasoning Test (program ariths) and the High Level Arithmetic Reasoning Test (program arith)

The Arithmetic Reasoning Test measures the individual's grasp of the fundamentals of arithmetic and the effective manipulation of these fundamentals in logical sequences. An attempt has been made in this test to design the material so that solution of the problems can be achieved only if the basic principles of computation are understood deeply; the application of standard arithmetic procedures which are "programmed" into people at

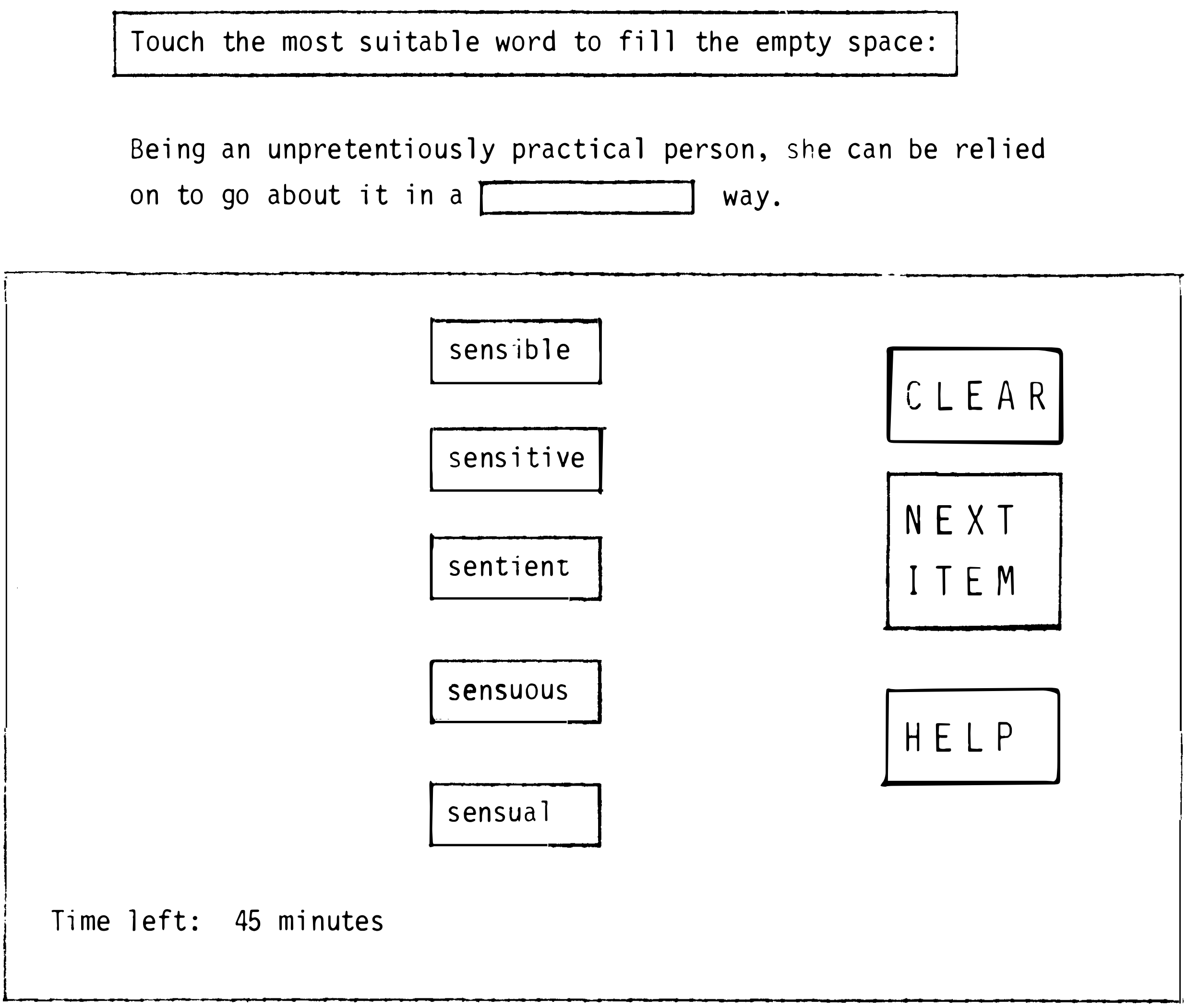


Figure 4: An example of an item from wict, the program for the High Level Words in Context Test.

<u>Figure 5</u> Flowchart for wict, wicts, wicta and wictas

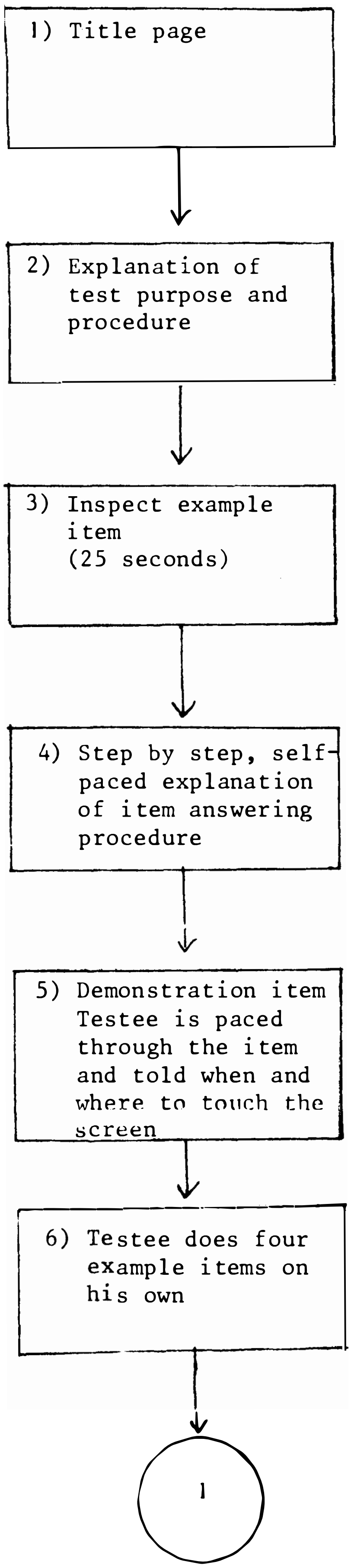
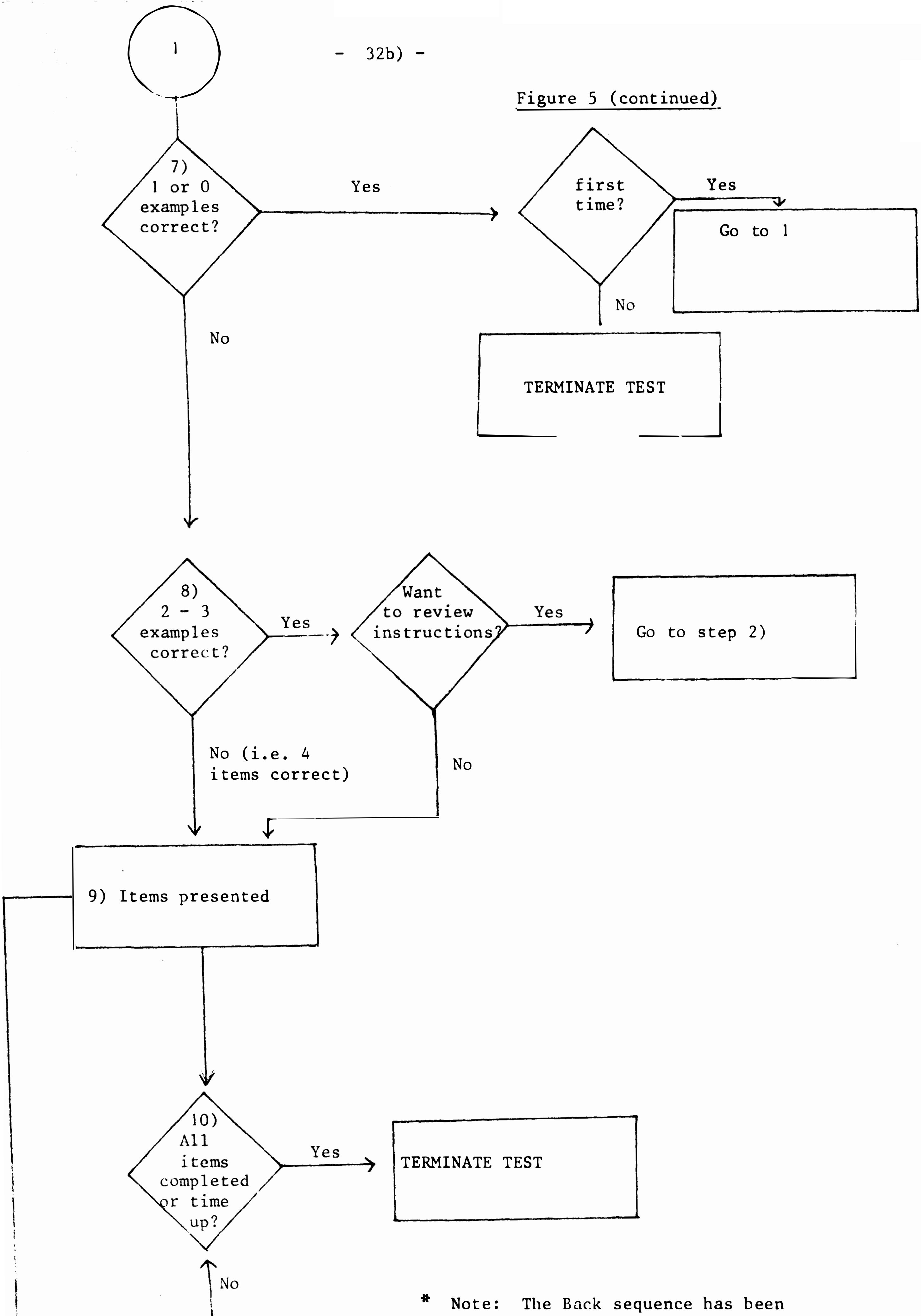


Figure 5 (continued)



* Note: The Back sequence has been omitted from the flowchart

school is generally ineffective. More innovative procedures or strategies are required, and the subject has to draw on his understanding of the fundamental rules of arithmetic in order to do the items. It should be mentioned that no large or complicated computations are involved. The aim is <u>not</u> to measure how well the subject can perform massive multiplications, divisions, subtractions etc. without making clerical errors. Each item requires a different strategy, which might involve up to about six logical steps. If an appropriate strategy is applied, an item is soluble in a short period of time (less than 30 seconds); all the steps can be executed without the need to set pencil to paper.

Let us look briefly at the rules of the test.

There are three "unknown" digits which may appear in arithmetic equations:

1) ⟦⟧ which is any positive digit in the range 1 - 9

2) ⟦ which is a digit equal to half ⟦⟧
(Hence, when both ⟦⟧ and ⟦ appear in an equation, ⟦⟧ must be even.)

3) ⟦H⟧ which is always equal to ⟦⟧ + 1

These symbols can also appear as digits in larger numbers. For instance, if ⟦⟧ = 3, then ⟦⟧7 = 37. Because the symbols can appear as digits in multi-digit numbers, solution of the items using standard algebraic procedures is much too time-consuming to be a viable strategy.

In many items, an arithmetic operator is also omitted and replaced by the symbol ^. Part of the subject's task is to discover what ^ is (i.e., whether it is +, -, x or ÷).

The subject's job is to find the value of ⟦⟧ . Here is an example of an item (one of the more difficult ones from the High Level version of the test):

▯ 4 ^ ⊞ = $9\frac{1}{4}$

These are the steps involved in its solution:

1) As the number on the right hand side of the equation contains a fraction, ^ must be ÷.

2) Hence $\frac{▯\,4}{⊞} = 9\frac{1}{4} = \frac{37}{4}$

3) Try equating numerators and denominators. But in the numerator, ▯ 4 cannot be equal to 37 for obvious reasons.

4) Hence ▯ 4 must be a multiple of 37. The obvious multiple is 2; 37 doubled will produce the digit 4 in the units column and this tallies with the 4 in ▯ 4.

5) Double 37 to produce 74. Hence ▯ = 7.

6) An unnecessary but reassuring step would be to check the whole equation, substituting 7 + 1 = 8 for ⊞ :

$$\frac{74}{8} = 9\frac{1}{4}$$

As the equation balances, all is well.

There is a fair amount of material to be presented in the instructions, and the interactive facility of PLATO is ideal for this purpose. The following are the main steps in the presentation of the arith and ariths instructions:

1) Basic nature of the task is explained.

2) Rules of the symbols are given.

3) Knowledge of rules are tested. Feedback is given for wrong answers. If the test is failed, the subject is returned to the relevant instructions. He is then retested on a similar rules test. If he fails again, he is eliminated from the test. If he passes, he goes back onto the mainline of the program.

4) Information about ^ is given.

5) Information on how to answer the test is given.

6) Subject is given simple example items to do. He must get two in a row right in order to be allowed into the test. If, after five examples, he has not succeeded in achieving this criterion, he is eliminated from the test. Whenever a subject answers a practice item incorrectly, he is given feedback; hence, he can learn from his mistakes.

7) If the subject passes the criterion in 6), he goes to some final instructions and then begins the test. Figure 6 presents a typical item of arith and Figure 7 presents the program's flowchart. The flowcharts of ariths, and of the Afrikaans programs aritha and arithas are identical to that of arith.

It will be noticed in Figure 6 that the boxes "NEXT" and "CLEAR" are in dotted lines. This is to indicate that these boxes are not actually on the screen at this point. When the item is initially written on the screen, only the "SKIP" box is visible. This gives the subject the opportunity to skip the item if he does not wish to attempt it. If he does attempt it, however, the "SKIP" box disappears and "NEXT" and "CLEAR" come on the screen. The subject attempts an item by touching one of the numbered boxes. As soon as he touches a numbered box, it "illuminates", that is, its outline becomes thicker to indicate that it has registered his touching. If the subject is satisfied with his answers, he touches "NEXT" and the following item (or a display offering the opportunity to use Back) will appear on the screen. If, on the other hand, the subject is dissatisfied with his answer, he touches "CLEAR". The illuminated box then returns to normal, "CLEAR" and "NEXT" disappear, and "SKIP" reappears. The item now appears as it did at the beginning of the display. The program can be cycled through these two types of item display as many times as desired (i.e., as many times as the subject wishes to clear his answer).

The ability of computerized tests to teach fairly complex material step by step, test knowledge, and loop back subjects

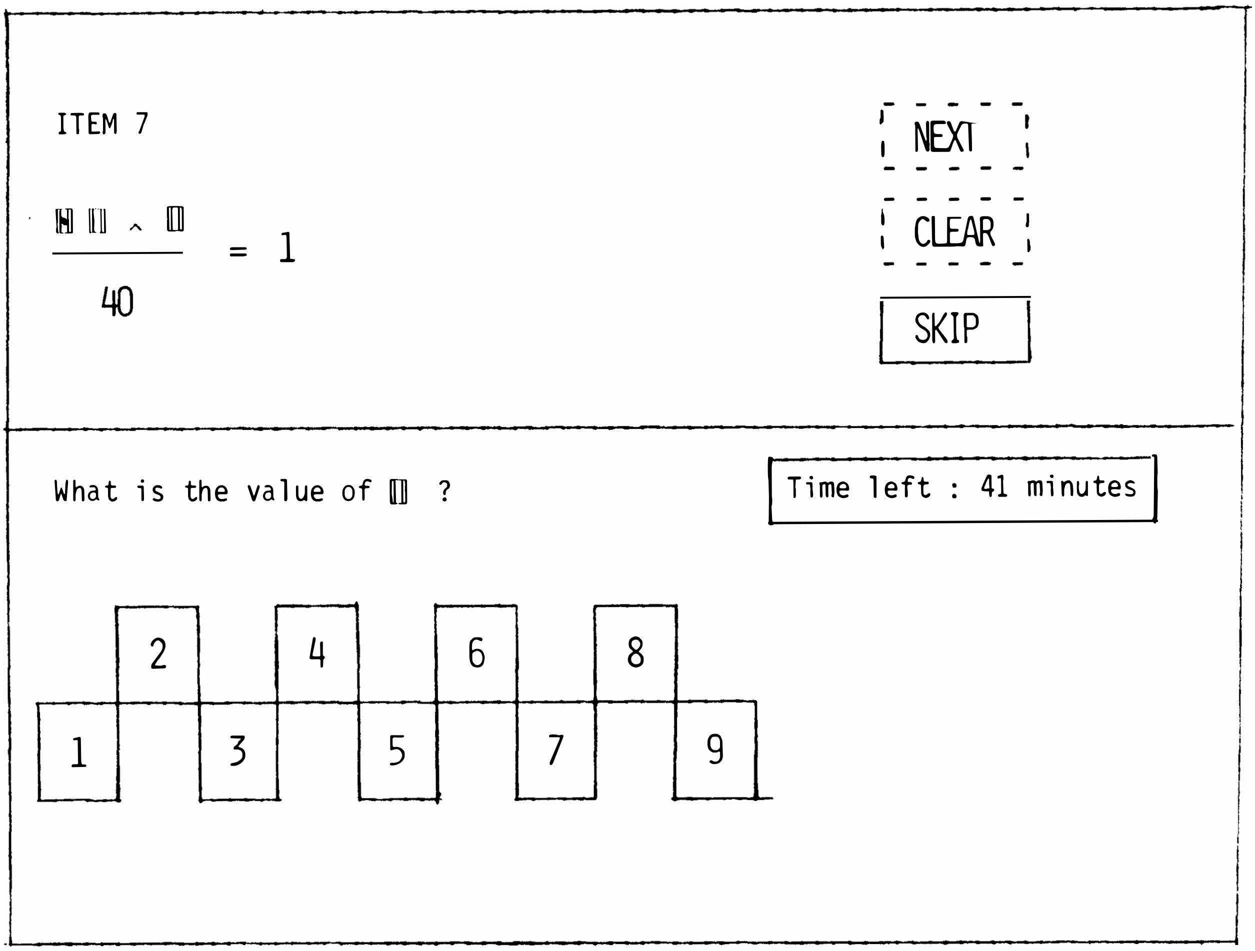


Figure 6 An example item from arith

5) Information on how to answer the test is given.

6) Subject is given simple example items to do. He must get two in a row right in order to be allowed into the test. If, after five examples, he has not succeeded in achieving this criterion, he is eliminated from the test. Whenever a subject answers a practice item incorrectly, he is given feedback; hence, he can learn from his mistakes.

7) If the subject passes the criterion in 6), he goes to some final instructions and then begins the test. Figure 6 presents a typical item of arith and Figure 7 presents the program's flowchart. The flowcharts of ariths, and of the Afrikaans programs aritha and arithas are identical to that of arith.

It will be noticed in Figure 6 that the boxes "NEXT" and "CLEAR" are in dotted lines. This is to indicate that these boxes are not actually on the screen at this point. When the item is initially written on the screen, only the "SKIP" box is visible. This gives the subject the opportunity to skip the item if he does not wish to attempt it. If he does attempt it, however, the "SKIP" box disappears and "NEXT" and "CLEAR" come on the screen. The subject attempts an item by touching one of the numbered boxes. As soon as he touches a numbered box, it "illuminates", that is, its outline becomes thicker to indicate that it has registered his touching. If the subject is satisfied with his answers, he touches "NEXT" and the following item (or a display offering the opportunity to use Back) will appear on the screen. If, on the other hand, the subject is dissatisfied with his answer, he touches "CLEAR". The illuminated box then returns to normal, "CLEAR" and "NEXT" disappear, and "SKIP" reappears. The item now appears as it did at the beginning of the display. The program can be cycled through these two types of item display as many times as desired (i.e., as many times as the subject wishes to clear his answer).

The ability of computerized tests to teach fairly complex material step by step, test knowledge, and loop back subjects

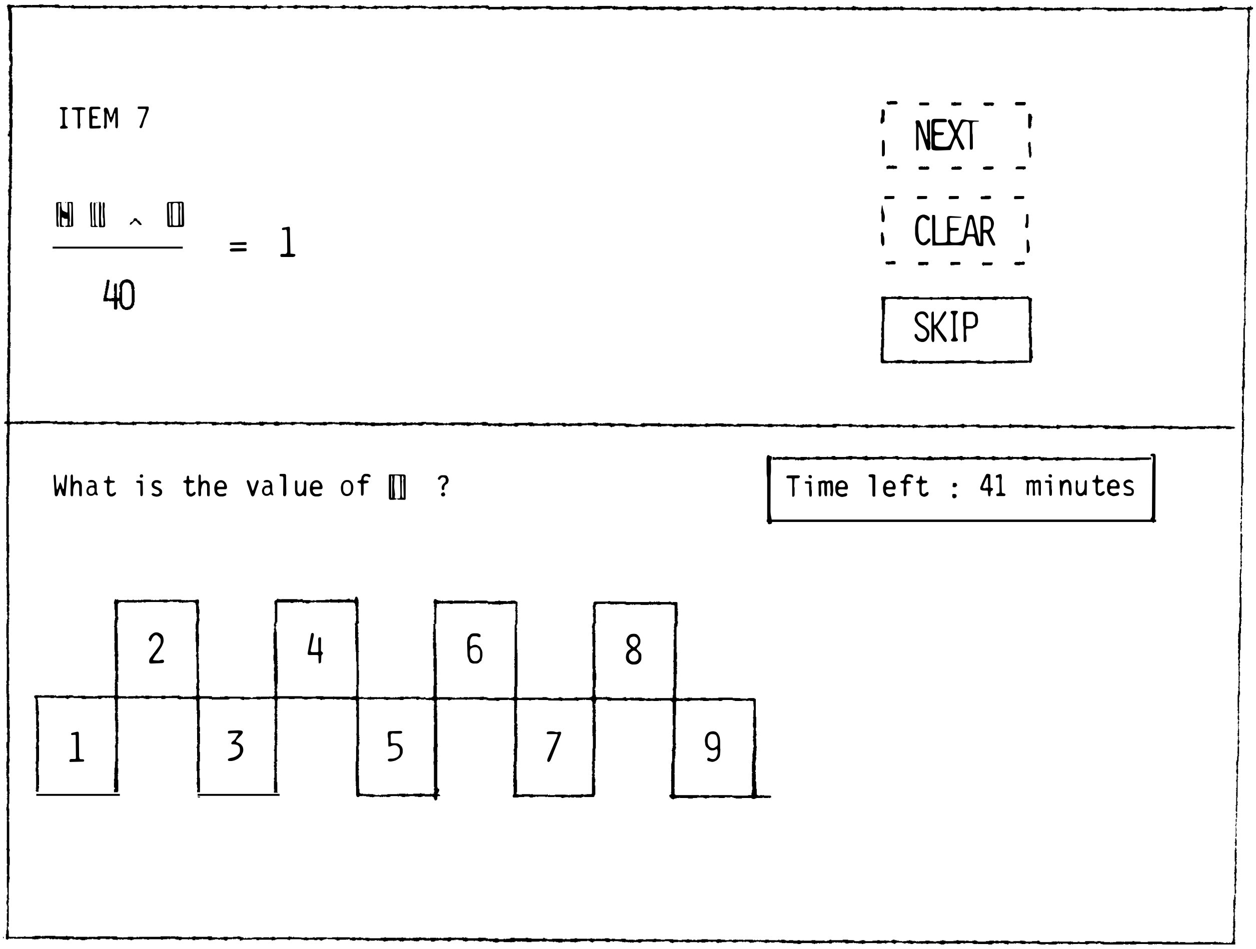


Figure 6 An example item from arith

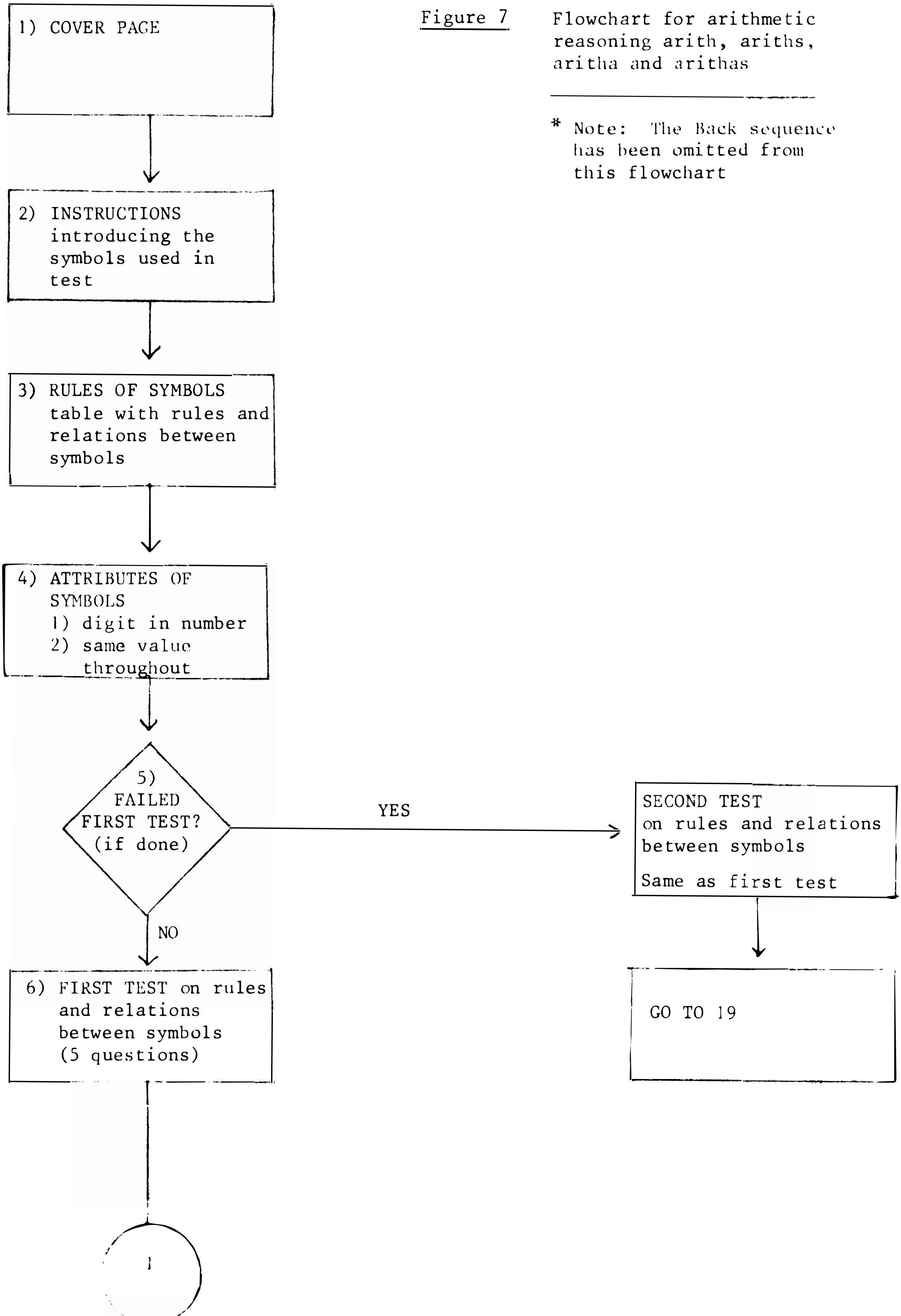


Figure 7 Flowchart for arithmetic reasoning arith, ariths, aritha and arithas

* Note: The Back sequence has been omitted from this flowchart

Figure 7 (continued)

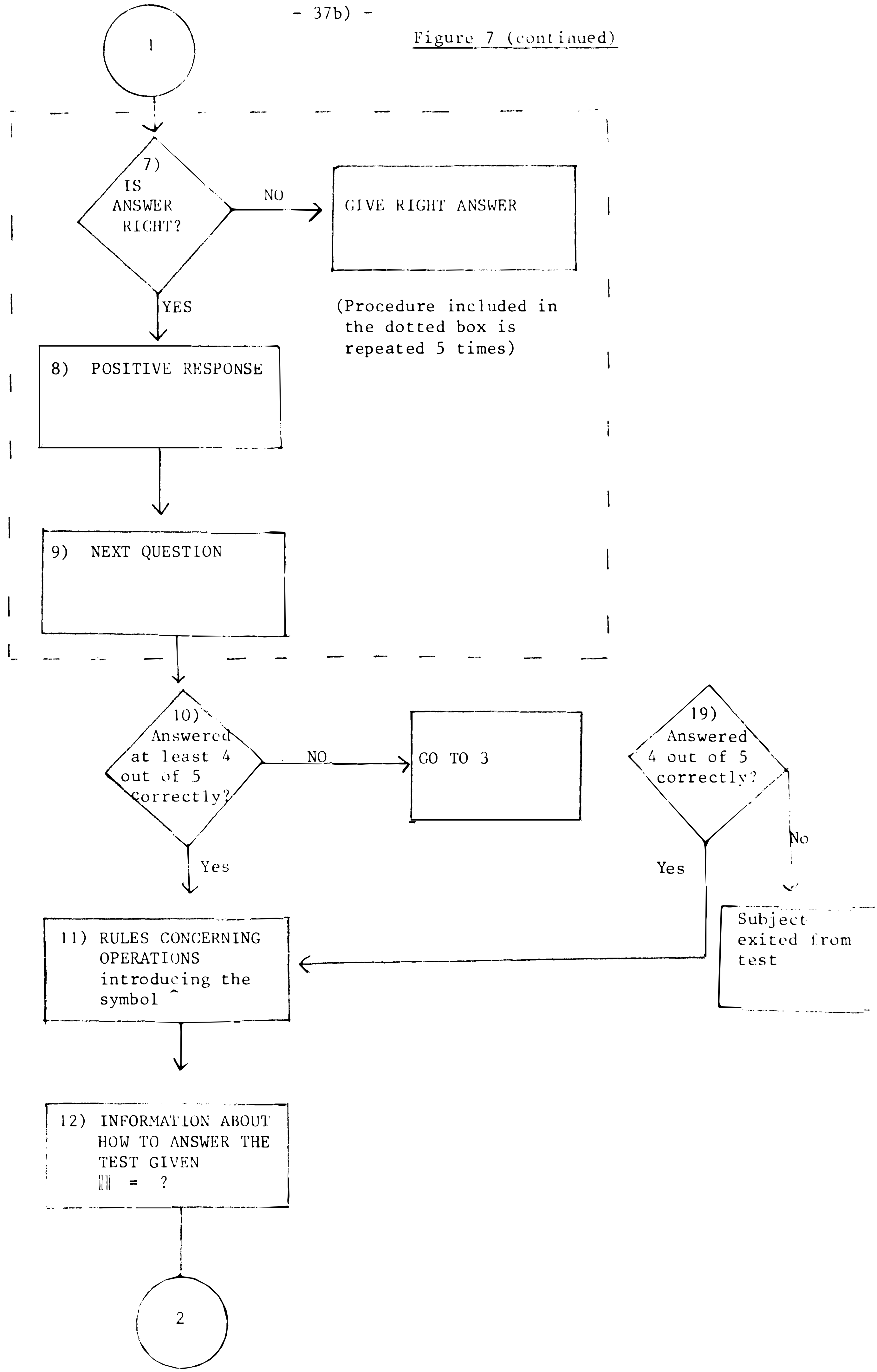
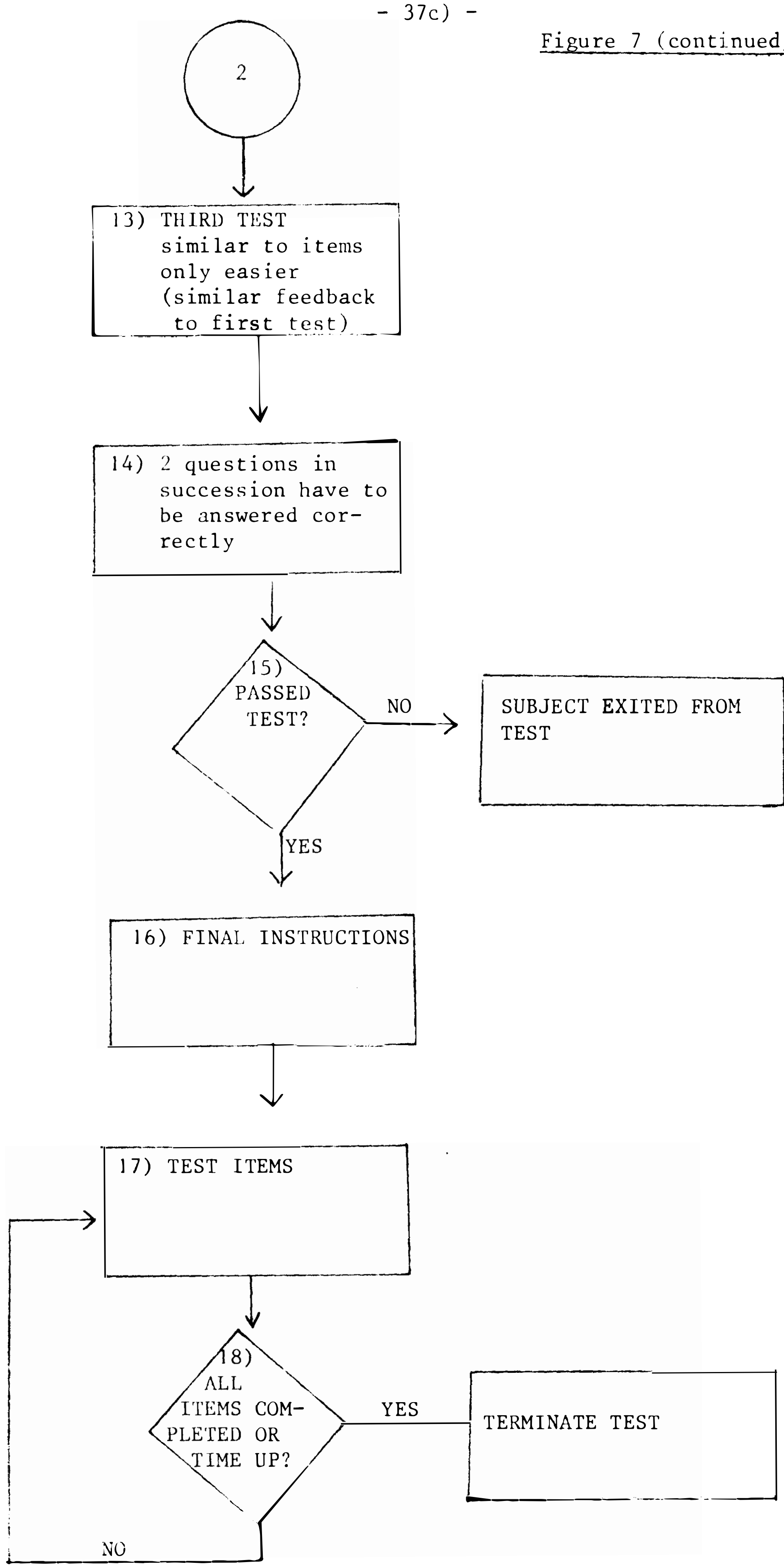


Figure 7 (continued)



who do not display an adequate grasp of the material is very useful in tests like this where the rules and procedures are fairly involved. One is likely to find a higher proportion of people minunderstanding or failing to grasp what is required of them than one would in a test where the rules and procedures are simpler. The ability of the arith and ariths programs to eliminate subjects who fail to meet certain criteria is hence very valuable.

5.2.4 The Estimation Test (program estims) and the High Level Estimation Test (program estim)

The growing importance of computers and electronic calculators in modern society is making the ability to perform complex arithmetical calculations less important. However, the ability to make a quick and accurate estimate of the answers to arithmetical problems and to recognise obvious errors in calculations done by electronic means is becoming more important. The Estimation Test measures this ability. There are two levels of the Estimation Test - the higher level, which is intended for people with Standard 10 and upwards; and the standard level, which is intended for people with Standards 7 to 10. Afrikaans and English versions of both levels have been programmed. The programs for the standard and High Level Afrikaans versions are estimas and estima respectively.

The testee goes through the following instructional procedure. Firstly, he is informed of the aim of the test, namely to measure his ability to estimate the value of arithmetic expressions quickly. The conventions used in the test are explained to him: commas always indicate decimal commas, and spaces are used to indicate thousands, millions, etc. An example item then appears on the screen. As each part of the item appears, its function is explained in a specially provided area. The testee is then shown, step by step, how to answer an item:

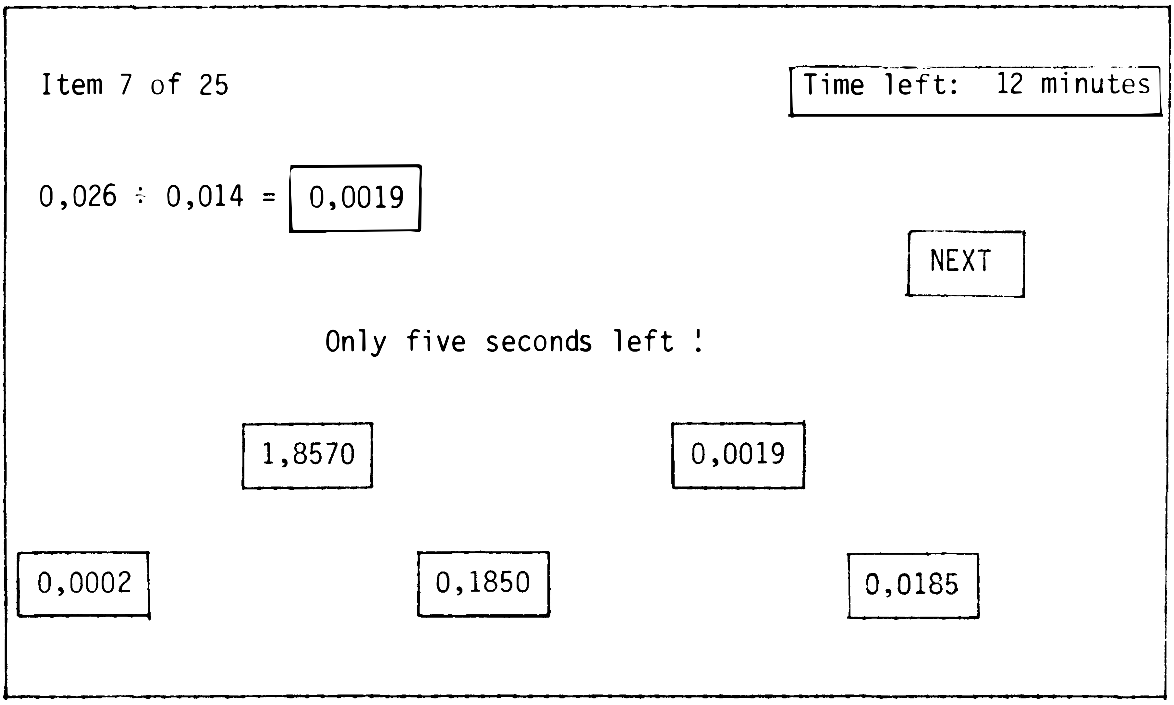


Figure 8: An example of an item from estim, the program for the standard version of the Estimation Test.

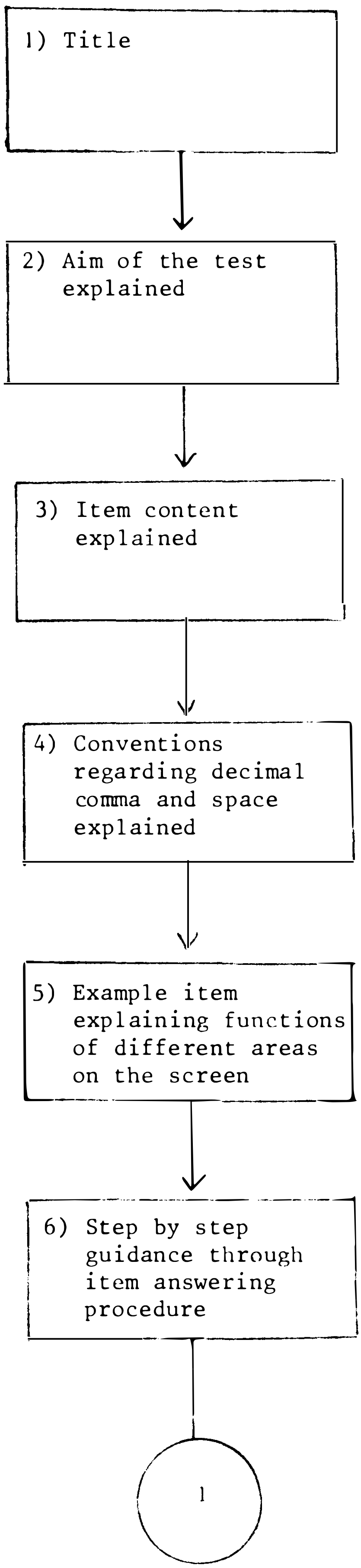
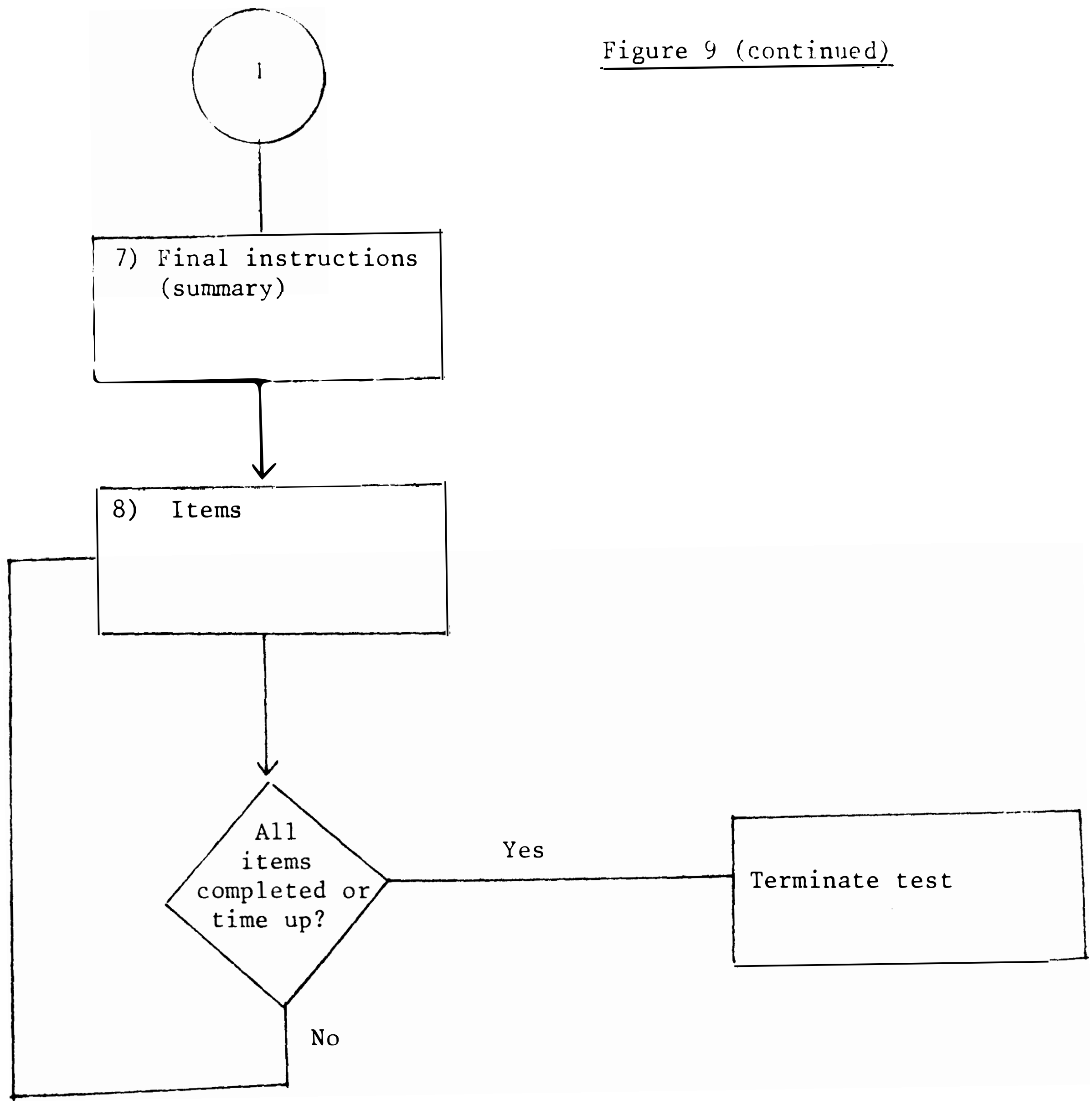


Figure 9 Flowchart for estim, estims, estima and estimas

Figure 9 (continued)



a) There are five possible answers in each item. The testee can endorse an alternative by touching it. It then appears in the answer box (see Figure 7).

b) He can change his answer by touching another alternative, which replaces the previous answer.

c) He enters his answer and moves on to the next item by touching "NEXT".

d) He can leave an item unanswered by touching "NEXT" before touching an answer.

e) He cannot return to unanswered items. (In other words, Back is not a feature of this test.)

f) He is not allowed to spend more than a minute on each item. After 55 seconds a warning appears - five seconds later the item is replaced by another one. This is done to stop the subject from calculating out the answers.

The testee is then given a self-paced set of final instructions before going on to the items. In these final instructions, the importance of working quickly is re-emphasized. It is pointed out that although there is a minute time limit on each item, there are 25 items to do in 15 minutes. Hence, on average, only 36 seconds can be spent on each item.

The following information is available after testing:

1) Number of items correct
2) Total time spent on test
3) Time spent on each item
4) Actual responses to each item

5.2.5 The Scrambled Sentences Test (program scrams)

The Scrambled Sentences Test is a measure of a subject's ability to order verbal material both logically and grammatically. The subject is presented with a disordered set of sentences or parts of sentences. His task is to re-order the sentences so that the resultant material makes both logical and grammatical sense. The test is aimed at subjects who have at least Standard 10. It is expected that the Scrambled Sentences Test will be an effective predictor of performance in jobs which place a premium on verbal skills.

Program scrams is not complete at the time of writing. This program will at a later stage be joined by an Afrikaans counterpart, scramsa. Each test will consist of five paragraphs, and each paragraph will comprise approximately ten sentences or parts of sentences.

The instructions for the test are fairly lengthy because so many options are available to the testee. A flowchart of the instructions and test is shown in Figure 12.

The first display explains the aim of the test. The second display identifies the task and this is followed by an example of an item. This first example consists of numbers that have to be rearranged into ascending order. The subject is stepped through this task so that he can learn how to move sentences around, insert and remove punctuation, etc. A second example is presented. Again, the subject is stepped through the item, combining the sentences in different way so that they are:-

a) Logically correct but grammatically incorrect

b) Grammatically correct but logically incorrect

c) Logically and grammatically correct.

This is done to emphasize that logic and grammar are both important. The subject is told to make mistakes at certain stages so that he can see the messages written on the screen. This is to reassure him that should he forget what to touch, or touch the wrong area, guidance will always be available.

The subject is then presented with a third example which he must do by himself. The program judges the final ordering of the sentences. If it is incorrect, feedback is provided and the subject is allowed to try again. If the subject fails to order the paragraph correctly after three tries, he is eliminated from the test. If the subject passes this criterion, some final information is given to him and he is then allowed to start the test.

After working through the instructions, the subject is presented with his first paragraph. A representation of this initial screen display is shown on page 44 (Figure 10).

The screen has been divided into three sections.

Area A: Should the subject wish to insert punctuation, he can touch one of the three punctuation options. By touching "NEXT PARAGRAPH", various scores are recorded and the following paragraph is presented. Space has been left at the top of the screen so that a message can be flashed at the top of the screen should the subject touch an inappropriate area. The time remaining for the paragraph is displayed on the left.

Area B: Here the sentences and their identifiying letters are displayed. The letters only are touch sensitive. They have been presented in a zigzag manner so as to maximise the separation between the touch areas.

Area C: This is the main work area.

The initial screen display allows the subject three options:

a) If he decides to rearrange the sentences, he must touch the letter associated with the sentence he wishes to move. This will cause Area A to be erased. The question mark in Area C Will be replaced by the letter which he touched. Two blocks, "BEFORE" and "AFTER" will appear after the letter, and "CLEAR" will appear on the lower right hand side of the screen (see Figure 11).

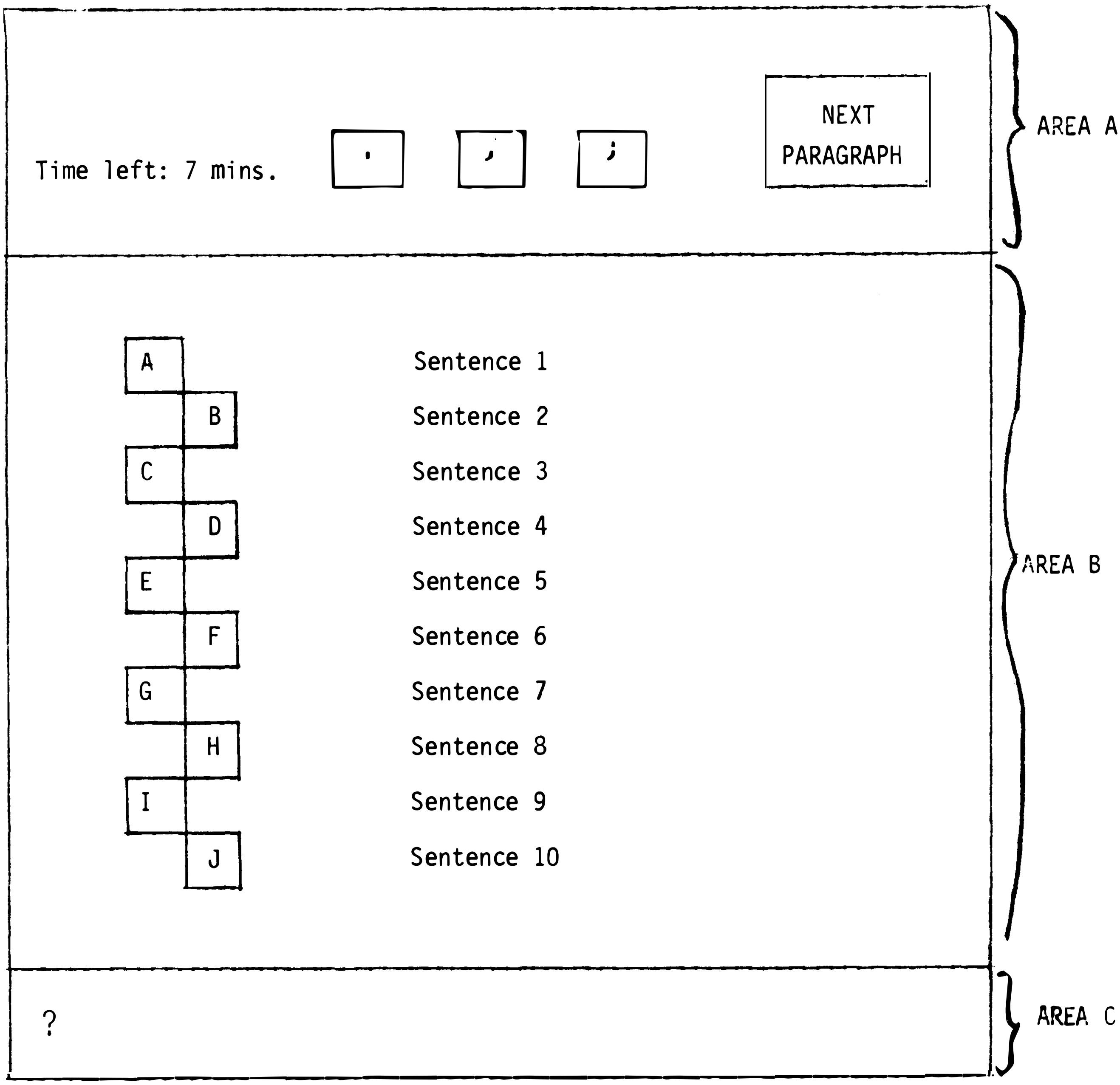


Figure 10. Initial Screen Display of scrams

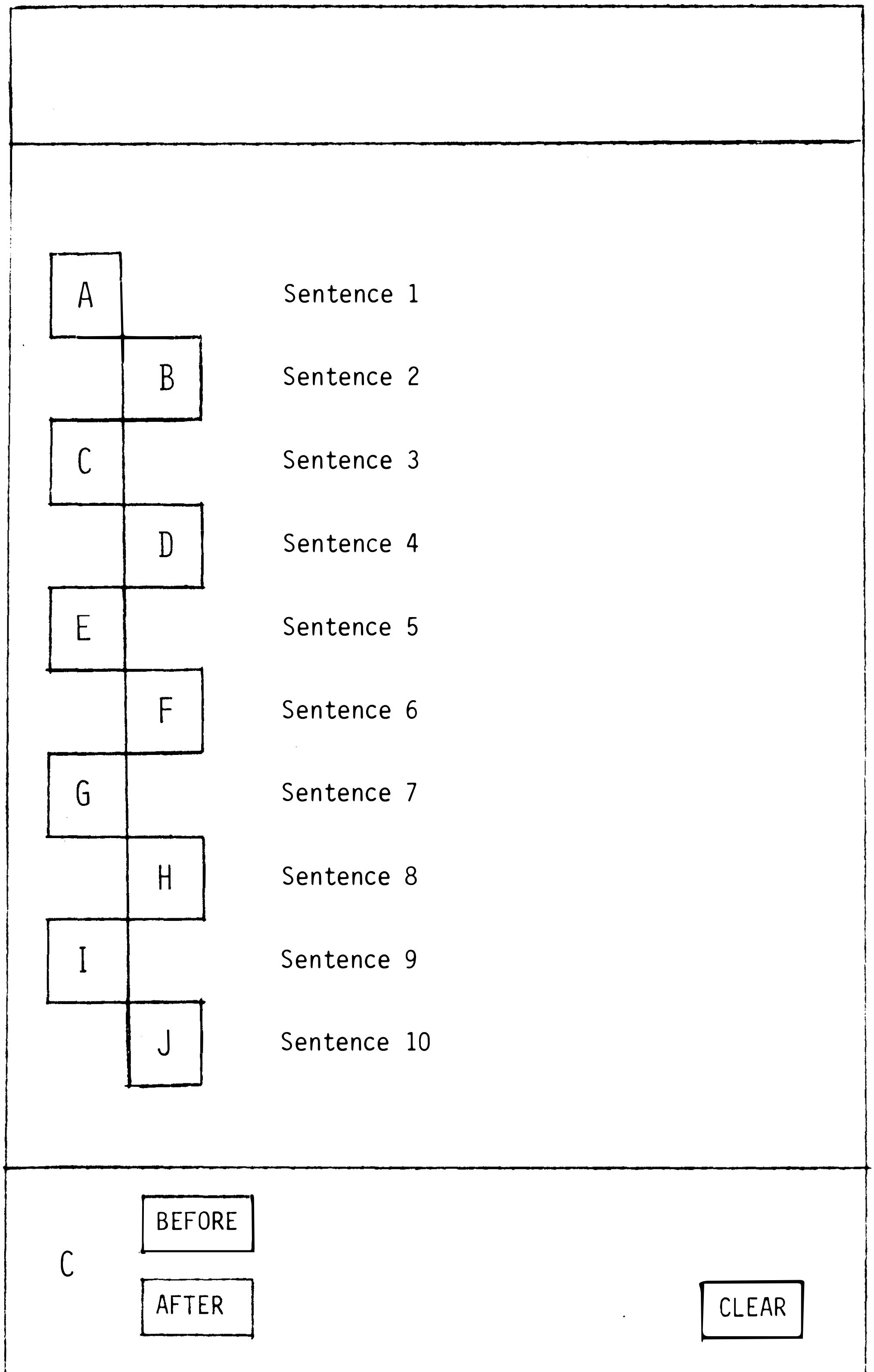


Figure 11 Second Screen Display of scrams

Assuming that the subject does not touch "CLEAR", he can now touch another letter. The second question mark is replaced by the letter and a new box, "CONFIRM", appears on the screen above "CLEAR". Figure 12 shows how Area C now appears to the testee.

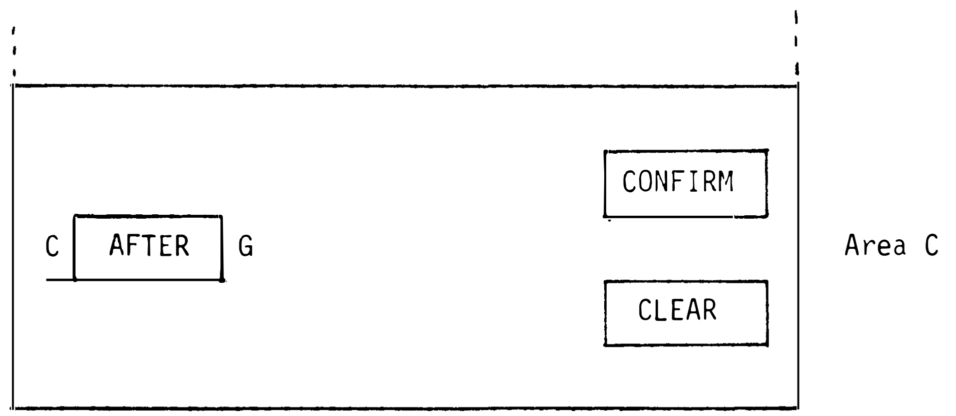


Figure 12. Confirmatory Screen Display

If the subject touches "CONFIRM", the sentences are reordered as requested and the screen is reset as in Figure 10. If he should press "CLEAR", the screen is reset as in Figure 10 but the sentence order remains as it was.

b) The subject's second option is to insert punctuation. This is done by touching one of the punctuation squares. For example, suppose a full stop is touched. Area A is then erased and left blank. In Area C the question mark is replaced by a full stop. Since the subject is never required to insert punctuation in the middle of a line, "AFTER" is automatically written after the full stop and is followed by a second question mark. The option to "CLEAR" appears. The subject specifies the sentence or sentence segment that is to receive the punctuation by touching the appropriate letter. This letter replaces the second question mark and "CONFIRM" is written above "CLEAR". If "CONFIRM" is touched, the punctuation is inserted at the end of the sentence and the screen is reset to the initial screen display. A new option, "CLEAR PUNCTUATION", is now available to the subject. This option will only appear if one of the sentences contains punctuation. It works in the same way as the punctuation option.

If it is touched, the first question mark is erased and in its place is written "CLEAR PUNCTUATION AFTER ?". After the appropriate letter is entered, the subject can either "CONFIRM" or "CLEAR".

c) The third option available from the initial screen display is "NEXT PARAGRAPH". Touching this box will cause the sentences to be erased and the new paragraph to be placed on the screen.

Every effort has been made to guide the subject through the test. Only the options which are logically available to him are shown on the screen, e.g. it is impossible for him to press "CONFIRM" until all the information required has been given. If he touches the screen in the wrong place, a message will appear on the screen telling him what options are available, e.g. if he touches a sentence instead of the letter associated with it whilst on the initial screen display, a message saying, "touch a letter, punctuation or 'NEXT PARAGRAPH'" will appear.

A final scoring system has not yet been devised for this test. Because there is so much information that could be recorded, it was felt that it would be better to observe a number of people doing the test, and from that, decide what scores to record. A few examples of scores that could be recorded are listed below. Only the first five of these could be easily obtained from a conventional paper-and-pencil test.

a) Final ordering of items

b) Minimum number of moves to change the subject's answer to the correct answer.

c) Grammar score

d) Logic score

e) Length of longest correct run

f) Time taken to order a paragraph

g) Time between moves

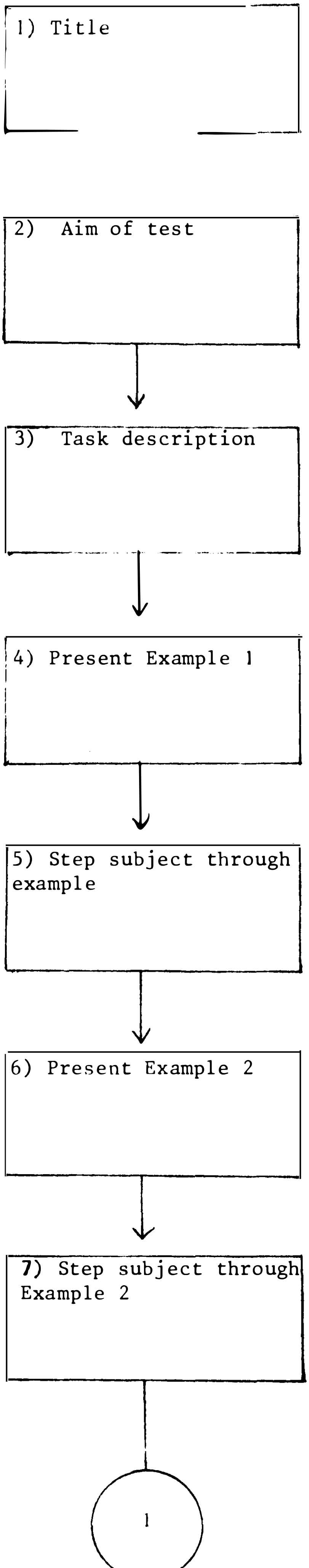
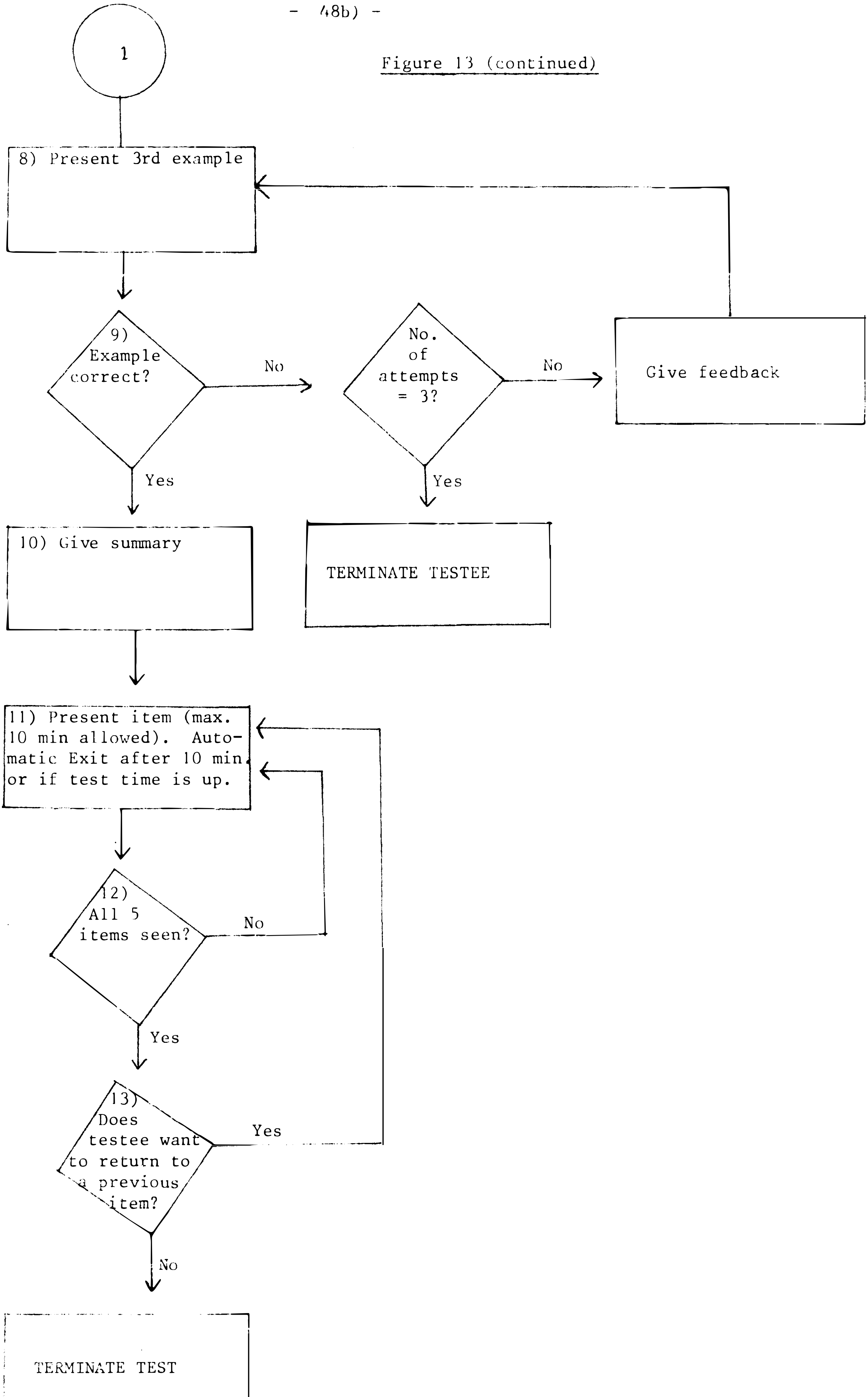


Figure 13 Flowchart for scrams

Figure 13 (continued)



The final three scores give useful information on the subject's response strategy. One can see whether he is a quick, impulsive worker, a slow accurate worker, whether he works by trial and error, etc.

Program scrams could be easily adapted for use as a measure of personality attributes. In this application, there would be no inherent logical or grammatical order to the verbal segments. The ordering imposed by the testee would then reflect personality rather than logical factors. Naturally, it would be necessary to select the verbal segments very carefully to be relevant to the particular personality dimension under study.

5.2.6 The Continuous Symbol Checking Test (program consym)

The Continuous Symbol Checking Test is a continuous work test. it requires the testee to maintain a high level of concentration during a long period of repetitive mental work. Because the repetitive task that the testee is expected to do (checking rows of symbols for differences) is so simple, the test is considered to be suitable for application to persons of any culture and level of education. The test yields information on the following characteristics of work performance:

a) Accuracy and Neatness - the number of errors and corrections made are used as indications.

b) Speed of work is indicated by the total quantity of the testee's output.

c) Increase or decrease of performance over time.

d) The relative position in the work process at which the increase or decrease occurs and the sharpness of any upward or downward curve in performance over time.

e) Steadiness or variability of performance as indicated by fluctuations in speed and accuracy.

The computerized version of this test is much simpler to administer and score than the original paper-and-pencil version. It is no longer necessary to use special timing devices and to rely on the testee to draw a line after every three minute period. The test user also does not have to struggle with complicated formulae any more, as the scoring calculations are done by the computer and scores are available virtually instantly.

There is an Afrikaans version of consym, called consyma.

The testee sees the following displays in consym. After the title has appeared on the screen, the testee is presented with a short description of what the test will involve - namely, a lengthy repetitive task. Then follows an introduction, illustrated with a part of the test material. The following aspects are explained, and the elements under discussion are highlighted by displays.

a) The rows of symbols are pointed out using an animated technique.

b) The answer boxes, which indicate the number of differences between consecutive rows of symbols, are pointed out.

c) The testee is instructed that he must count the differences between the rows by comparing pairs of symbols directly above each other, and indicate his answer by touching the answer box with the correct number of dots.

d) The two rows to be compared and their answer box are indicated by a frame which moves down one row after each response.

The testee is then allowed to practice, and is not allowed to proceed until he has done the practice examples correctly. It is then explained to the testee that he can change his answer after the frame has moved on by touching the correct answer box in the previous row of answer boxes. He can change only his most recent answer. A short summary of the answering procedure is then presented before the testee is given some more practice material to exercise changing answers. The testee is then given a choice to either revise the instructions (in which case he is taken back to the first set of practice items), or to start the test. He is not told how long he will have to work on the test material. (The test actually lasts 60 minutes.) Indices of the following are available immediately after completion of the test:

a) Accuracy

b) Neatness (Number of corrections)

c) Total amount of test material covered

d) Convexity of the work curve (Thie index can be positive or negative)

e) Fluctuation - the stability of the work performance

f) The degree of the initial drop in performance (This index can be positive or negative.)

g) The time interval during which the initial drop occurred.

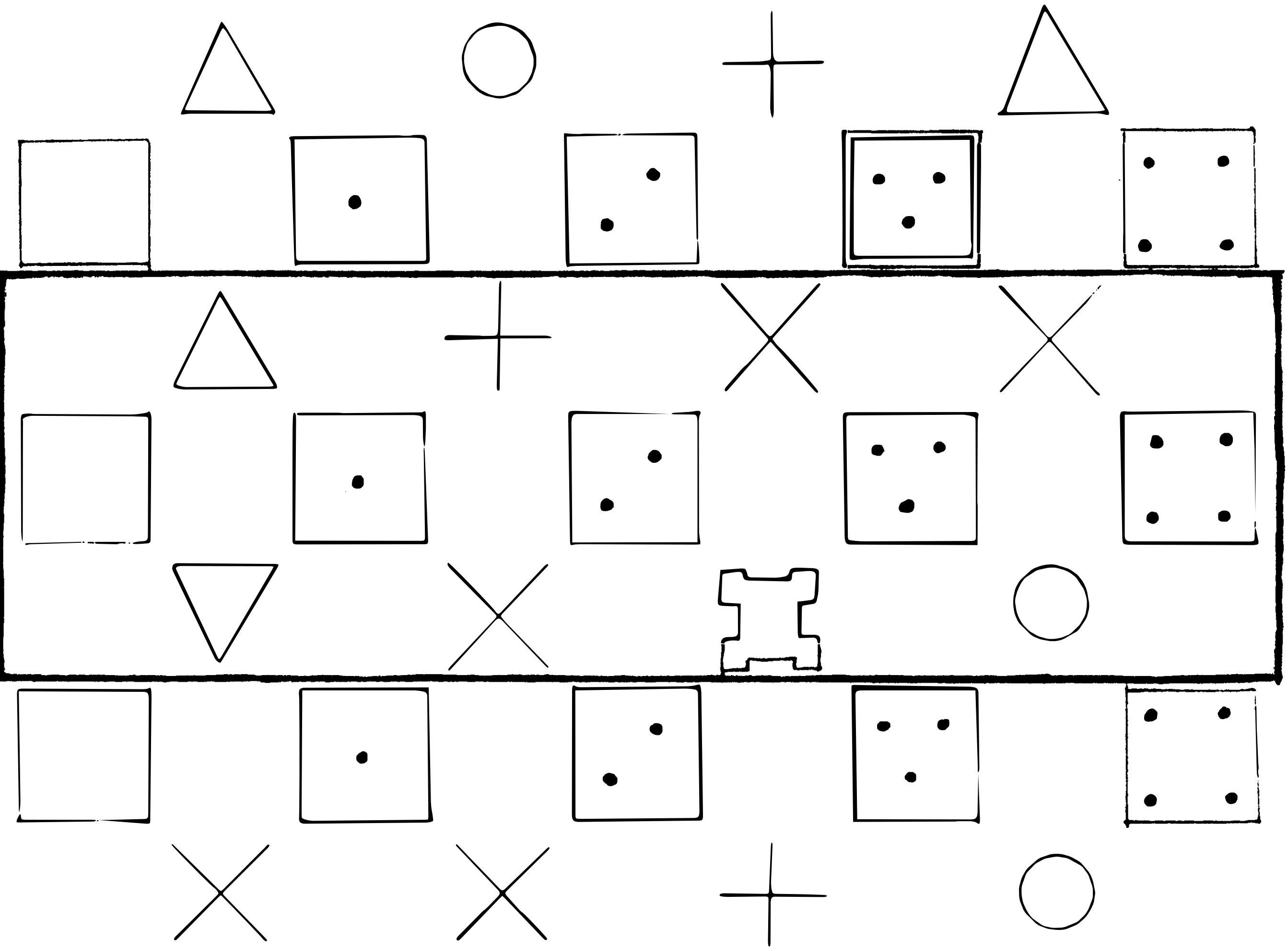


Figure 14: A sample of test material for the Continuous Symbols Checking Test, showing the rows of symbols and the rows of answer boxes between them.

The large frame indicates the two rows that are to be compared next. The small frame around the answer box with three dots indicates the testee's previous answer: there are three differences between the first and second rows of symbols.

Figure 15 Flowchart for consym and consyma

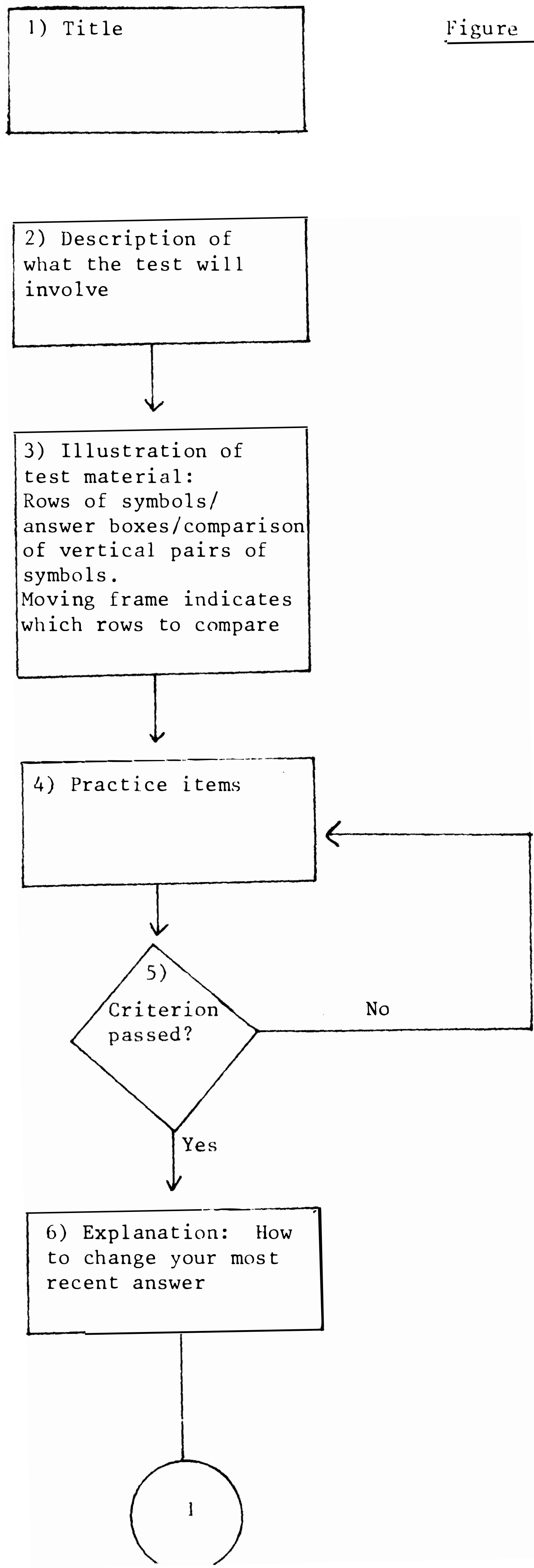
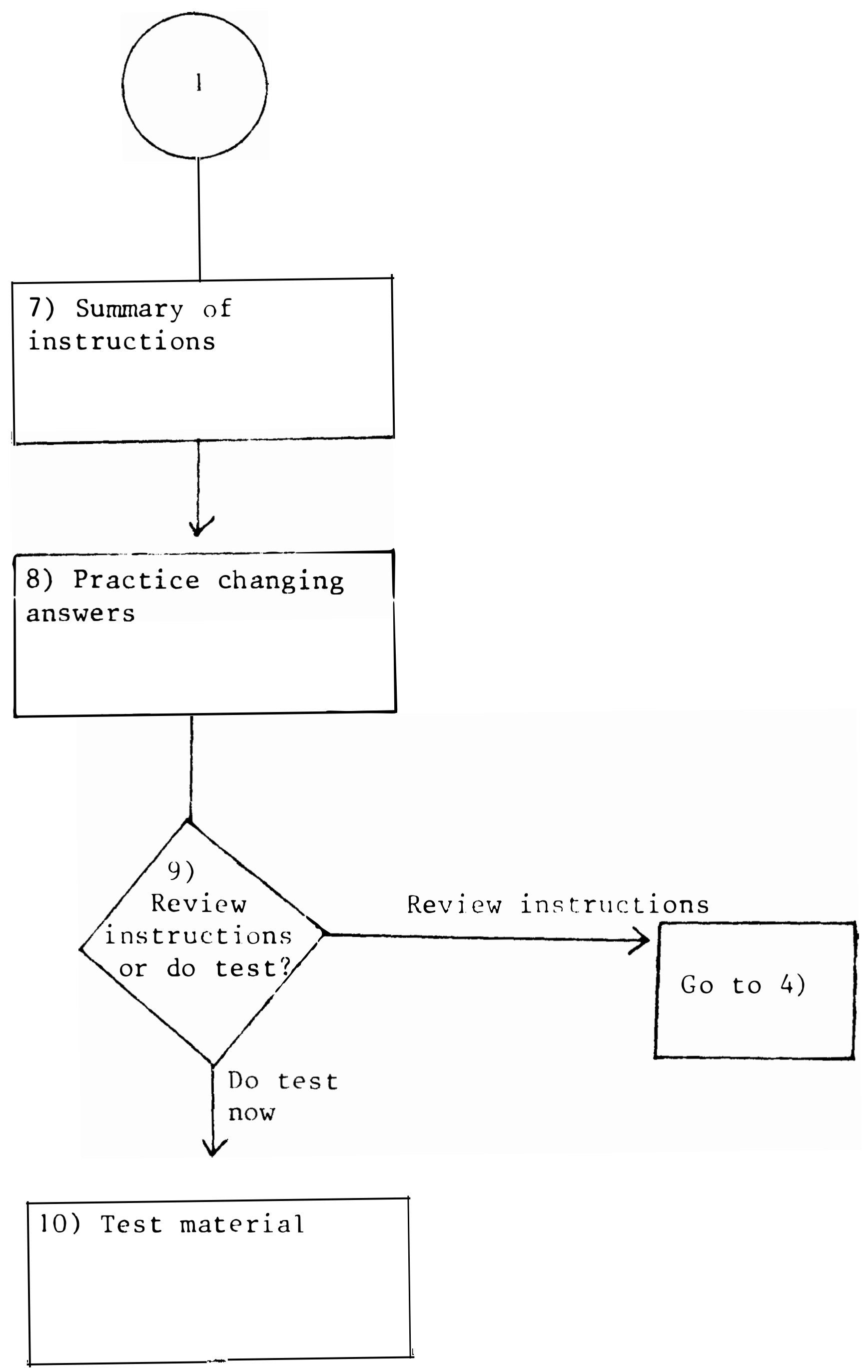


Figure 15 (continued)



7. REFERENCES

BETZ, NANCY E. (1977) Effects of immediate knowledge of results and adaptive testing on ability test performance. Applied Psychological Measurement, 1, 259-266

BIRNBAUM, A. (1968) Some latent trait models and their use in inferring an examiner's ability. In F.M. Lord and M.R. Novick (Eds). Statistical theories of mental test scores, Reading, Massachusetts: Addison-Wesley

BISKIN, B.H. (1977) Effects of computerized administration on scores on the Minnesota Multiphasic Personality Inventory Applied Psychological Measurement, 1, 543-549

BOYLE, T.A. and W.F. SMITH (1976) Computer mediated testing: a branched program achievement test. Modern Language Journal, 60, 428-440

BROWN, J.M. and D.J. WEISS (1977) An adaptive Testing strategy for achievement batteries. Research Report 77-6, Psychometric Methods Program, Department of Psychology, University of Minnesota, Minneapolis

CORY, C.H. (1977) Relative utility of computerized versus paper-and-pencil tests for predicting job performance. Applied psychological Measurement, 1, 551-564

CORY, C.H., B. RIMLAND and R.A. BRYSON (1977) Using computerized tests to measure new dimensions of abilities: an exploratory study. Applied Psychological Measurement, 1, 101-110

ELWOOD, D.L. and CAROLYN CLARK, (1978) Computer technology. Computer administration of the Peabody Picture Vocabulary Test to young children. Behaviour Research Methods and Instrumentation, 10, 43-46

GIALLUCA, KATHLEEN A. and D.J. WEISS (1980) Effects of immediate knowledge of results on achievement test performance and test dimensionality. Research Report 80-1, Computerized Adaptive Testing Laboratory, Psychometric Methods Program, Department of Psychology, University of Minnesota, Minneapolis

HANSEN, D.N. and S. ROSS (1977a) Flexilevel adaptive testing paradigm: hierarchical concept structures. Air Force Human Resources Laboratory Report AFHRL-TR-77-35(11)

HANSEN, D.N. and S. ROSS (1977b) Flexilevel adaptive testing paradigm: validation in technical training. Air Force Human Resources Laboratory Report AFHRL-TR-77-35(1)

JENSEMA, C.J. (1977) Bayesian tailored testing and the influence of item bank characteristics. Applied psychological Measurement, 1, 111-120

JOHNSON, J.H., R.A. GIANETTI and T.A. WILLIAMS (1978) A self-contained microcomputer system for psychological testing. Behavior research Methods and Instrumentation, 10, 579-581

KALISH, S.J. (1980) Computerized instructional adaptive testing model: formulation and validation. Air Force Human Resources Laboratory Report AFHRL-TR-79-33

KINGSBURY, G.G. and D.J. WEISS (1979) Adpative testing strategy for mastery decisions. Research Report 79-5, Psychometric Methods Program, University of Minnesota

KRUS, D.J. and R.W. CEURVORST (1978) Computer assisted construction of variable norms. Educational and psychological Measurement, 38, 815-818

LORD (1970) Some test theory for tailored testing. In W.H. Holtzman (ed.) Computer-assessed instruction testing and guidance. New York: Harper and Row.

LORD, F.M. (1971) The self-scoring flexilevel test. Journal of educational measurement, 8, 147-151

LORD, F.M. (1977) A broad-range tailored test of verbal ability. Applied psychological Measurement, 1977, 95-100

MC BRIDE, J.R. (1977) Some properties of a Bayesian adaptive ability testing strategy. Applied psychological Measurement, 1, 121-140

MC KINLEY, R.L. and M.D. RECKASE (1980) A successful application of latent tract theory to tailored achievement testing. Research Report 80-1, Tailored Testing Research Laboratory, University of Missouri, Columbia

OWEN, R.J. (1975) A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American statistical Association, 70, 351-356

SAMEJIMA, F. (1977a) Effects of individual optimization in setting the boundaries of dichotomous items on accuracy of estimation. Applied psychological Measurement, 1, 77-94

SAMEJIMA, F. (1977b) A use of the information function in tailored testing. Applied psychological Measurement, 1, 223-248

SCHMIDT, F.L.V.W. URRY and J.F. GUGEL (1978) Computer assisted tailored testing: examinee reactions and evaluations. Educational and psychological Measurement, 38, 265-273

SEGUIN, S.P. (1976) An exploration study of the efficiency of the flexi-level testing procedure. Dissertation Abstracts International, September, 1196-1197

TAYLOR, T.R. (1977) The measurement of concept identification differences among literate Blacks. Psychologia Africana, 17, 31-47

URRY, V.W. (1974) Approximations to item parameters of mental test models and their uses. Educational and psychological Measurement, 34, 253-269

VERSTER, J.M. (1979) The development of a computer-interactive testing system and the measurement of speed and error in different cognitive processes in different cultures. Unpublished document, National Institute for Personnel Research.

WATERS, B.K. (1977) An empirical investigation of the stratified adaptive computerized testing model. Applied psychological Measurement, 1, 141-152

WEISS, D.J. (1974) Strategies of adaptive ability measurement. (Research Report 74-5) Minneapolis: University of Minnesota, Department of Psychology Psychometric Methods Program (AD 004270)