THE UNIVERSITY OF ALBERTA

INDIVIDUALIZED INTELLIGENCE TESTING BY COMPUTER

by

© JOHN ERNEST BOYLE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

THE DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1973

ABSTRACT

This study explored the feasibility of using the computer as a "substitute" examiner in place of a "human" examiner for the administration and evaluation of an individualized intelligence test. The feasibility of this idea was investigated through the construction of a computer model of the Verbal Section of the Wechsler Intelligence Scale for Children (WISC).

The model was developed and implemented on the IBM 1500 Instructional System at the University of Alberta as a Coursewriter II computer program. Its purpose was to supervise the administration and evaluation of the test to clients.

In order to develop the model, and the resulting computer program, a modeling process common to industrial, commercial and scientific sectors was used. It was based upon the concept that the real phenomenon being modeled could be transformed stage-by-stage, into a conceptual model, then into the computer program. The stage-by-stage transformation process, coupled with very careful documentation, error checking and correction procedures, established the validity of the developed model by ensuring that all the criteria being modeled were considered and implemented into the model.

The criteria in the WISC manual deemed as being necessary to the administration and evaluation of the test were taken as the real phenomenon. These were transformed into a conceptual model, then into the computer program.

The computer model was evaluated in three stages. First, it was used to administer the Verbal Section to graduate students, then to a school psychologist whose response was taken from protocols, and then to legitimate clients.

The Computer-Examiner's performance was exceptional, taking into consideration a number of constraints which limited the model's development and that it was a first-generation model. The computer model made no administrative errors and its capability to correctly evaluate responses was encouraging. In its administration to the legitimate clients, the computer model was able to correctly categorize 73.4% of the 1,043 responses it examined, as compared to a human examiner's ability to evaluate the same responses.

The results of this study clearly suggest there is a feasibility in using computer-examiners to administer and evaluate individualized intelligence tests. However, in order to be able to develop a practical computer examiner several technical problems will have to be solved and considerable research and development must ensue.

ACKNOWLEDGEMENTS

This writer is indebted to his supervisor, Dr. S. M. Hunka, Coordinator of the Division of Education Research, for his guidance. He spent many hours listening to, and commenting on, numerous ideas and problems, and offered many helpful suggestions. I deeply appreciate the enthusiasm, encouragement and patience he gave throughout the course of the study.

I would like to thank Dr. A. Scott, Department of Educational Psychology, whom I consulted many times on matters relating to the correct way of administering and evaluating the Wechsler Intelligence Scale for Children.

I appreciate the help of Norman McGinnis, Systems Analyst for the Division of Educational Research, who solved a number of technical problems relating to the implementation of the computer model on the IBM 1500 Instructional System, and also the System's Computer Operators who lent a great deal of their time and assistance to implementing the model.

Gratitude is also to be extended to my fellow Graduate Students who enthusiastically received the computer model and lent assistance in testing it out.

Finally, I would like to acknowledge the tremendous assistance given to development of the computer model by my wife, Jeanette, who spent many hours assisting in testing the program, analyzing performance and recordings and preparing the manuscript.

TABLE OF CONTENTS

CHAPTER PAGE

CHAPTER PAGE

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (CONTINUED)

CHAPTER 1

INTRODUCTION

1. COMPUTER APPLICATIONS IN GENERAL

The evolution of the electronic computer and its application to our technologically oriented society has been phenomenal. Since the first all electronic computer ENAIC (Electronic Numerical Integrator and Calculator) was put into operation in 1946 as a purely laboratory device (Borko, 1962, p. 42), its application to the ever expanding needs of man has accelerated consistently. In the 1960's the computer had undergone an unprecedented revolution. Developments in computer hardware and the increasing sophistication of software support systems have paved the way for potential employment of the computer to almost every conceivable corner of our society. Business, industry, government, service agencies, the military; professions such as medicine, law, and education have all found uses for the computer.

2. COMPUTER APPLICATIONS IN EDUCATION

In comparison with personnel in such areas as business, industry and government, the personnel in education have been relatively slow in realizing that the computer has great potential in their field. To a large degree, this has been because of the relatively high costs involved in the acquisition and maintenance of computer hardware and software support systems. In education the computer has found some application, mainly in clerical activities, with minor use in scheduling, planning and

educational research. Its use in the main activity of education, the learning process, is still very experimental. Ever since the first publication on the use of the computer as an instructional tool appeared in 1959 (Rath et al.) research has been confined almost entirely to the campuses of universities, to military agencies and to computer business corporations.

As an instructional tool, the computer has been applied to a number of subject areas. Hickey (1968) indicates that courses have been prepared in the social sciences, psychology, economics, languages, science, mathematics, statistics, the health professions, business administration, engineering, computer technology, vocational guidance and counseling, besides others. In three of these areas, psychology, vocational guidance and counseling, experimentation and application has still been relatively scarce, and within these areas there has been particularly little use and experimentation in test administration - specifically in the area of psychological testing concerned with the general measurement of intellectual development.

This study is concerned with the area of psychological testing. Because of the frequent use of tests which measure general intellectual development, the results obtained from this study have implications in such fields as clinical and research psychology and psychiatry, industrial and educational psychology, vocational guidance, clinical and educational counseling services and all other areas which make use of individually administered tests.

3. COMPUTER APPLICATIONS IN PSYCHOLOGICAL TESTING

The use of computers in the general area of psychological testing has been mainly limited to aptitude testing and vocational guidance interviews. Studies in which the computer assisted the vocational guidance counselor as the interviewer were conducted by Cogswell and Estavan(1965), and Romaniuk (1969). Computer-assisted instructional programs for assisting and assessing the potential job placement of students have been written by Loughary, Tandow and Yabroff (Lekan, 1970, p. 79) and by Kostenbauder (Lekan, 1970, p. 8). As yet, there has been no descriptive literature published for these programs.

Aptitude testing and personality inventories seem to have been the main interest for application of the computer in psychological testing. Lekan's Index to Computer Assisted Instruction (1970, pp. 78 - 82) indicates that the following programs have been prepared:

(1) Minnesota Multi-Phasic Personality Inventory: This program is still in the developmental stage at Columbia University. No descriptive literature is available.[1]

(2) Kuder: This program is operational. It administers and scores the Kuder Preference Record. The developers are Benham and Lippert at Florida State University. No descriptive literature is available.

(3) Cavit; Scant; Vast; and Compan: These programs are concerned with academic aptitude testing. They were developed by Knoll, French, and Tandibuono at Pennsylvania State University. Descriptive literature

[1]In the CAI area it is common to find that programs are not adequately documented or not documented at all.

is reported in the Pennsylvania State Reports (1967, pp. 51 - 53; 1968, pp. 16 - 32).

(4) A Personality and Aptitude Testing Program: This program is operational but no descriptive literature has been issued. The purpose of the program is to provide for the following personality aptitude measures; Trait and State Anxiety Scale (STAI); Impulsivity Scale (BIS); Social Desirability Scale; I.Q. tests; Math Aptitude. The authors are O'Neill and Aurbauch at Florida State University.

Two programs most related to the area of intelligence testing have been developed by Tiegs and Clark (Lekan, 1970, p. 79) called URLTOO and URLTO1. These programs are computer versions of the California Achievement Tests, Upper Primary forms W and X and Elementary forms W and X. The programs were developed at Harvard University using an IBM 360/65 computer and an IBM 1050 Audio Visual Terminal. Descriptive literature is contained in the Program Abstracts of the Harvard University Computer-Aided Instructional Laboratory.

To the author's knowledge no study has been performed investigating the adequacy of the computer with regard to administering and evaluating an individualized intelligence test.

4. COMPUTER APPLICATIONS TO PSYCHOLOGICAL TESTING

In our contemporary society professional bodies which offer services of counseling, psychiatry, personnel assessment and guidance have been undergoing changes in the nature of and in the methods of performing these services. Consider, for example, the counseling services rendered to educational institutions. Blocher (1969, p. 18)

observes that the one-to-one counseling relationship is no longer the primary tool of the counselor, but only one of a whole new bag which has been developed as the result of the application of technology. Included are such techniques as group counseling and laboratory education. He also observes that the counseling and guidance function is developing to include many areas. It is expanding further into the heart of the educational process, and in this respect, the school psychologist is becoming an important consultant in the educational process. His duties and activities, in addition to the traditional support of the individual, will increasingly expand to the family, to groups, and ultimately to the educational and business communities.

An immediate and obvious implication of such an expansion of counseling services is that counselors are increasingly being taxed for time. Not only do they have their traditional responsibilities to carry out but they now have to face increased demands on their time and skills.

One element of the guidance and counseling process which uses a considerable portion of the school psychologist's day is the one-to-one interview situation with the client. As an instrument for gathering information about the client, or giving information back to him, it has played an extremely important role, but in terms of time its use has been expensive. The psychologist might be able to make use of the computer to help maintain the interpersonal relationship as a primary instrument in the guidance and counseling process. It could be used as a "substitute interviewer or examiner" in certain kinds of interviews either to gather information, dispense it, or interact with the client in predefined ways. Such an interview might satisfy the client's individual needs, and yet save the psychologist the time required to

conduct the interview. As additional benefits, a computer interviewer would with ease and without hesitation record a complete and accurate account of the interview for the psychologist's record, thereby assisting the psychologist in post-evaluations. Going one step further, with a computer interviewer, it might be possible to make a more accurate assessment of the client's behavior or even to dispense information directly to the client. This latter possibility has been demonstrated as being feasible by the experimental studies of Cogswell and Estavan (1965), and Romaniuk (1969). These investigations have shown that the decision making behavior exhibited by the guidance counselor or psychologist while helping clients make vocational decisions can be very effectively simulated by computer models.

If the vocational guidance counselor's decision making behavior can be simulated by a computer model in precisely defined areas it means that the skills involved are no longer uniquely human in quality. One might define a uniquely human skill as being one which cannot, as yet, be duplicated by machine or computer. If this is the case, then why have psychologists been performing such tasks when there is a more pressing need for skills which are uniquely human in the areas of guidance, counseling and assessment? Let the computer perform such tasks, and free the psychologist for other responsibilities.

The problems of determining how to effectively use time and other related types of problems are not exclusively limited to the school psychologist. Similar problems are experienced in other areas of psychology and psychiatry. These have common activities such as guidance interviews, personality assessment and psychological testing. They depend heavily on the one-to-one interview as the primary

instrument for gathering data from the client. Therefore, all to some degree, make considerable use of the psychologist's time. If it can be demonstrated as being feasible that the computer is able to model the behavior required by the interviewer to conduct these interviews, and if it can be demonstrated that clients do relate favorably to computer interviewers, the psychologist could be freed from such tasks to make more efficient use of his skills and time. The computer has a number of characteristics that could be utilized to facilitate such substitutions. It has capabilities for rapid and accurate computation, an infallible memory, information retrieval and display capabilities, it is free from human foibles of mood, bias, and subjectivity and most important, can be controlled by man.

CHAPTER 2

THE PROBLEM

1. THE PURPOSE OF THE STUDY

The purpose of this study was to determine the feasibility of using the computer as a substitute examiner in place of a human examiner for the administration and evaluation of an individualized intelligence test. This was done by specifying objectives for the design, construction, implementation and evaluation of a computer model of an individualized intelligence test on the IBM 1500 Instructional System. The model is called the Computer-Examiner. It was not a simulation of the human examiner's characteristics per se, but more a simulation of the standardized administration and evaluation criteria, i.e., the criteria which the test manual sets forth as being necessary for the proper administration and evaluation of the test.

The test selected for this study was the Wechsler Intelligence Scale for Children (WISC). It was chosen for basically two reasons. First, the WISC, along with the Stanford Binet Intelligence Scale, are the two most generally used measures of intellectual development for children (Chauney and Dobbin, 1963, p. 23). Secondly, of these two tests, it would appear that the WISC was the better test to model because it is conveniently structured into 12 subtests, each of which attempts to measure a different intellectual skill. The Stanford Binet does not make this differentiation, instead derives an I.Q. score by measuring various skills characteristically possessed by children at

each age level as an inseparable group of test items. By using the WISC it was possible to delineate with more precision those areas in which the Computer-Examiner was a success or failure. Also, the subtest structure of the WISC facilitated an easier development of the computer model.

This investigation was only concerned with the administration and evaluation of the Verbal Section of the WISC because the task of generating a computer model of a test like the WISC was predicted to be highly complex. By taking advantage of the subtest structure of the WISC, it was possible to conveniently scale this task down to a realistic size. Secondly, some of the Performance subtests of the WISC, such as Block Design and Object Assembly, could not be readily or easily implemented in a computer model which uses the facilities of the IBM 1500 Instructional System. The Verbal Section did not appear to present the same problems at this initial stage of study. Third, the Verbal subtests formed an integral part from which the Verbal I.Q. score was derived. If necessary, this score could be used as a basic unit of comparison between human and computer examiners.

2. THE NEED FOR THE STUDY

THE CONFLICT OF SIMULATING HUMAN BEHAVIOR

A number of reasons exist for justifying an investigation into the need for computer examiners. First, the problem of person to computer interactions, their form and outcome is contemporary. One point of view maintains that the computer is not only not capable of simulating such sophisticated human skills as the administration and

evaluation of intelligence tests, but there is no place for it in this kind of activity because the human element is absolutely essential to the situation. It is argued that the computer would only provide for a generally dehumanizing kind of experience. The opposite view maintains that computers could duplicate human skills and provide for a very close approximation to the examiner's behavior. To date, the author knows of no existing evidence to support either view, and therefore there is a need for studies which will provide cumulative empirical evidence as to the degree of success the computer would have as an examiner. Preliminary investigations, in all probability, will be of the kind which will attempt to outline in greater detail and with greater clarity some of the problems of designing and constructing computer models, their magnitude and possible solutions. This was the purpose of this study.

COMPUTERS AS EXAMINERS

A second reason for substituting the human examiner with a computer is a practical one - it would assist the professional to better manage his time. Loughary and Tandow (1969) indicate "the demand for counselor time is greater than the foreseeable supply (p. 32)". Assuming this prediction is correct, it would appear that all possible solutions to this problem, which might help professionals who use intelligence tests to more efficiently use their time should be investigated. If the computer is able to successfully fulfill the examiner's role, the saving in time which would result could be used for tasks more demanding of the professional's skill and experience. An assumption is being

made that the computer would only be able to perform this substitution in real life situations if hardware and software development costs, computer rentals, terminal accessibility, and other such factors did not outweigh the costs of employing a psychologist and any additional fringe benefits of having a psychologist perform the examination function.

PERFORMING ROUTINE TASKS

The possibility of replacing or assisting human examiners with computer examiners becomes apparent when some of the characteristics of the administration of intelligence tests and the examiner's role in this administration are examined. The test situation is naturally very routine, very rigidly standardized in its administration and evaluation procedures, and time consuming. The examiner is required to be analytic, objective, accurate, observant, alert, and in order to perform his job with expertise, must often draw on a large reservoir of experience and accumulated knowledge. In the scientific, industrial and commercial fields there are similar kinds of job situations where the personnel are required to exhibit similar behavior characteristics. For example, in many industrial processing and manufacturing plants computers have displaced, or have been put in to assist console operators to make observations on complex processes, perform complex analytical calculations and make objective decisions based on past and current data as to how plant operations are to be optimized. Such operators were under pressure to be accurate, alert, responsive to alarm situations and capable of making consistently correct decisions. In many cases, computers now relieve operators in such tasks, and have proved to be

superior to the human operator by monitoring more data, detecting and reacting faster and more accurately to changes in processing and being more consistent and correct in making decisions.

In carrying the analogy back to the psychologist's situation it could be speculated that the computer might be capable of a similar kind of substitution if an adequate computer model can be developed. It could also be speculated that if similar approaches to the task modeling were used as those that have been employed in industrial, scientific or commercial fields a successful model might begin to emerge.

CHARACTERISTICS NOT OPERATIONALLY DEFINED

Undoubtedly many characteristics of the individualized intelligence test and the examiner's role will be easily specified in the form of operational or behavioral definitions and these will be readily modeled on computers. But, there will also be many other characteristics that will not so readily submit to operational specification and as a result cannot be simulated by computers, for example, concepts such as love, friendliness, trust, confidence, compassion, understanding, empathy, insight and so on. Undoubtedly some of these may make a significant contribution to the assessment of a client's intelligence and may have to be a necessary part of any practical computer examiner. It is the responsibility of the modeler to attempt to specify such characteristics, not in terms of other abstract concepts, but operationally or behaviorally. He must specify what action or behavior is being exhibited by the persons involved. For example it would be an impossibility to model the concept of love by considering it as a deep

devotion, affection or a strong sexual passion. In order to simulate love the modeler must know how a person acts or reacts when expressing or receiving love. The basic specifications for the model must be in this format because it is necessary to instruct the computer as to what its actions or reactions would be in any given situation which would involve love. One must tell it how to behave when simulating love.

Even one step before the modeling stage, it must be determined what concepts are important to the administration and evaluation of intelligence tests. For example, the WISC manual emphatically states that "The administration of a standardized test requires ....a friendly relationship between examiner and subject, ...satisfying to the examiner and subject (p. 17)". What does this statement really mean? What is a friendly relationship; when is it satisfying to both the examiner and the subject? How can such a relationship be operationally defined? To what degree must the examiner exhibit the necessary behavior upon which the concept of a friendly relationship is based? In quantitative terms, just how much is necessary in order to meet the objectives of the test? Such a quantitative specification is not currently stated in the WISC manual. The examiner must determine how much friendliness is required based on his own personal and highly subjective past experience. In other words, there is no specific and consistently measurable operational specification for the concept of friendliness between any two or more examiners.

SUMMARY

In summary, it is necessary for the modeler to recognize the importance of the specific characteristics of an individualized intelligence test and of the role which the examiner plays while administering and evaluating the test. The modeler must be able to determine objectively whether or not these characteristics are important to the achievement of the test's objectives. He must also determine to what degree they are important, and before constructing the model, must translate these characteristics into operational or behavioral terms. The success of a computer examiner depends upon the isolation, selection, and definition of such factors. If these steps cannot be completed is the utility of the computer-examiner jeopardized? The answer can only be determined through the process of experimentation. It is hoped that the ultimate result will be the development of a computer examiner that will become a valuable extension of the human examiner's data gathering and assessment capabilities.

Because of the complexity in isolating, defining and modeling many of the undefined characteristics of the WISC this study concentrated on the more clearly defined administration and evaluation characteristics stated in the WISC manual. Since this is one of the first studies in the area of computer intelligence testing, it was considered justifiable to attempt to model only the very basic characteristics of the test. There was no justifiable reason to attempt to model the friendly relationship characteristics at this stage of development. Such problems should be left for later experimentation when computer examiners will have reached greater degrees of sophistication.

3. POTENTIAL ADVANTAGES OF COMPUTER EXAMINERS

There could be a number of potential advantages to using computer examiners. These are:

(1) an increased impetus to further development of intelligence tests,

(2) group administrations,

(3) reduced administration costs,

(4) increased standardization,

(5) increased reliability,

(6) reduced examiner bias,

(7) increased repertoire of skill and experience, and

(8) more comprehensive data collection.

IMPETUS TO DEVELOPMENT

The study and development of computer examiners will force intelligence test developers to look more closely at various aspects in the development of general measures of intelligence. In order to develop a computer examiner, the test developer is required to spell out in operational detail the nature of the test. No longer could the developer say without further specification, what has been repeatedly stated in the WISC manual (1949) "The administration of a standardized test requires ....a friendly relationship between examiner and subject .... (p. 17)." He must now examine this kind of concept and analytically break it down into precise quantitative statements which can be used as the basis for determining how a computer will model a friendly relationship. This information could be used to determine how a human

examiner should carry on a friendly relationship. The application of the computer in other fields, in this same fashion, has forced researchers to look more closely at the concept or operation they are trying to model, and in turn, the results of such research have given the researchers a much clearer picture of, and has led to improvements in the operational specification of the original concept or operation. Applying this observation to the field of intelligence testing means that not only is there the possibility of developing successful computer examiners, but the knowledge gained in the process of modeling could possibly lead to improvement in the original procedures of administering and evaluating intelligence tests.

GROUP ADMINISTRATION

Another reason for developing a computer examiner is that such an examiner could administer an intelligence test to large groups of clients at the same time. No human examiner can justifiably administer a test such as the WISC to more than one client at a time. The IBM 1500 Instructional System, as it is presently configured for instructional stations at the University of Alberta could be used to administer the WISC to as many as 16 clients at the same time. Such a method of administering the test might prove to be a convenient way of gathering data in research studies which require tests such as the WISC. Looking into the future, computer based intelligence tests could be administered to whole classes of students, much like the current group intelligence tests, but there would be the added advantage that the tests would be individualized.

REDUCED ADMINISTRATION COSTS

Group administration would eliminate the expense of employing a human examiner for each client. This would result in a considerable saving in the total time of tester hours. If it takes an examiner approximately 1.5 hours to administer and evaluate the whole WISC, it would require 24 hours of tester time to evaluate the intelligence of 16 clients. Using the present configuration of the IBM 1500 Instructional System, a computer examiner would probably require approximately 1.5 hours of time to evaluate the same 16 clients.

Assuming that an examiner's time is worth approximately $20 per test, it would require $320 to test 16 clients. Using a cost estimate of $1.95 per student terminal hour, derived by Stolurow (Hickey, 1968, p. 95), it would require $48, plus an additional $20 for a human examiner to supervise the test, to test the same clients by computer examiner. Although these estimates are very rough, it is obvious that there are substantial savings to be realized through the use of computer examiners. However, it must be kept in mind, that prospective cost reductions cannot be immediately realized but certainly remain a future possibility.

INCREASED CONTROL OVER STANDARDIZATION

Another advantage of using a computer examiner is increased control over the standardized administration and evaluation criteria. This would again result in an increase in the test's reliability as a measuring instrument. Wechsler (1949) states "It is necessary for the examiner to adhere to certain prescribed procedures which have

been selected as superior to others on the basis of research (p. 17)". Intelligence tests such as the WISC are the end products of rigorous standardization programs. Deviations from prescribed testing procedures increase the possibility of making errors in the assignment of the client's relative I.Q. score. If not controlled, such deviations will lower the test's reliability and validity as a measure of intelligence.

A computer examiner would be of considerable advantage in maintaining, or even increasing, the reliability of a test. Computers do only as they are instructed. If the administration and evaluation procedures of a test can be operationally defined and incorporated into the program of a computer model, the resulting computer examiner would perform to the exact expectations of the test developers. As a result each client would be tested in exactly the same manner specified by the test developer. The computer examiner would not deviate from the prescribed testing path for that client. This would have the advantage of adding additional assurance to the tests' reliability as a measuring instrument. The real advantage to be gained is that each client would receive exactly the same test form, administration procedures, and evaluation process and therefore, would receive the same opportunity to be rated on the relative scale of intelligence. Such conditions are ideal for the standardization of a test.

REDUCED EXAMINER BIAS

A computer examiner would assist in overcoming the examiner's biases influencing the results of the test. In this regard, Gathecole (1968) states "....it is shown that testers can bias the results they

get from tests and it is suggested that automation may be positively useful in overcoming such distortion (p. 139)". Frequently such distortions are due to the personal appearance or personality of the examiner, his current mood, personal biases or prejudices. The behavior of some examiners may be very inconsistent with the behavior of others. The entire climate of an interview can be determined by the mood of the examiner. He may have predetermined expectations of the client's intelligence. The personal appearance and behavior of the examiner may influence how the client will respond in the interview situation. Such problems would be avoided if a computer examiner were employed because the developers could very carefully ensure that it was free from such biases, or at least they were controlled for.

Also, a computer examiner would not be influenced by the personality of the client. The pretty, bright-eyed, well-dressed, blonde girl would receive exactly the same considerations as the unkempt, unruly, dull-eyed boy. Of course an assumption is being made that the computer examiner would be capable of handling both clients. Similarly, a computer examiner would not be subject to varying fits of mood. Its behavior would remain invariant from interview to interview. The immediate gain from this invariance is that the test's reliability as a measuring instrument would increase.

Criticism could be leveled at the invariance of a computer examiner in that it would not be able to meet the individual needs of the clients. Each one would require a slightly different way of being examined, to which the computer examiner would not be able to adapt. How would it handle the more obvious physical motor differences in

clients such as varying abilities to see, to hear, to talk or to do various motor skills with their hands? These must be taken into consideration. Human examiners are capable of noting such differences and taking them into account while administering and evaluating the test. Computer examiners might be able to note the same differences but this will depend upon the ability of the test developers to isolate and operationally define how such discriminations are being made by human testers. The essential characteristic of the computer remains, it would make no uncalled for departures from the planned testing procedure for the population to which the test is normed. Therefore, it would be the responsibility of the developers to know how to account for the various individual differences of clients.

INCREASED REPERTOIRE OF EXPERIENCE AND SKILL

Another advantage in the use of a computer examiner lies in the computer's ability to store and retrieve large amounts of information. Tests such as the WISC require competent examiners who have not only the skill to administer and evaluate the test, but have considerable real experience administering it. In this case experience means knowledge of atypical situations. If the skills and experiences of the best human examiners could be pooled and incorporated into the computer examiner, then the resulting examiner would vie with or excel the best human examiner in terms of administration and evaluation capabilities. In fact, it would be the most competent examiner. At present this concept is theoretical, and a long way from being a practical reality.

For example, in this investigation a scoring supplement for the

WISC manual, the WISC Scoring Criteria (Massey, 1969), was utilized to help establish the criterion for evaluation. This scoring supplement is a collection of scoring responses made possible by the efforts of a number of psychologists and testers. It is expected that the major portion of the Verbal Section of this scoring supplement will be incorporated in the Computer-Examiner's evaluation section. The clients to be tested will benefit from this experience because the Computer-Examiner will have access to this information at all times - it will not forget as might be the case with human examiners. A computer examiner also has the additional advantage of being updated whenever the test developers deem it necessary. They would simply issue a new program, just as an author would issue a new edition of a book and everyone concerned with administration of the test, the developers and testers, would be assured the computer examiner's experience would be the state-of-the-art.

Not all clients are fortunate enough to be tested by competent examiners let alone the best human examiner. Assuming that a computer examiner can be developed, the problem would be alleviated by the fact that each client would be tested by the best possible examiner. The result of having a more consistent examiner would mean that the test's reliability as a measure of intelligence would increase. There would be no variance on the client's I.Q. score which could be attributed to varying degrees of examiner competence.

COMPREHENSIVE DATA COLLECTION

By using the storage capabilities of the computer it would be possible to record verbatim all the client's responses. Such a

complete and accurate record could be very useful for the purpose of a more detailed post-examination or as part of a data file for future research. Tape recorders are often used for the same purpose, but the computer has the advantage of storing the data in a format which readily lends itself to data search, sort and merge techniques, and statistical analysis. Different kinds of variables could be measured and stored in the client's data file, such as response latency, as well as type and frequency of certain responses. Intelligence might possibly be measured by such variables.

4. POTENTIAL DISADVANTAGES OF COMPUTER EXAMINERS

It is also certain, at this stage of development, that there could be disadvantages to the use of computer examiners. For example, hardware and software support systems costs presently make the use of computer examiners uneconomical. It appears this will gradually disappear as general time sharing systems come into more general use and the costs for input/output terminal equipment decrease.

Another disadvantage to the use of computer examiners would be the fact that the client is required to go to the computer, rather than having the computer come to the client. As time sharing networks come into more general use and terminals more portable, computer examiners will eventually test clients in their environment, whether it be the classroom, the clinic or the home.

Computers cannot, as yet, aurally or visually interact with people. This disadvantage is only due to the fact that the current state-of-the-art of computer development has not reached a level of sophistication which would permit this. These possibilities are on

the horizon and equipment prototypes may be a few years away.

In the opinion of some persons the use of computer examiners has one distinct disadvantage. They contend that the use of such machines provides for another dehumanizing kind of experience. However, it must be kept in mind that this is an opinion based on the point of view that the individualized intelligence test requires certain seemingly human qualities. If there are unique human qualities, it must first be determined that they are essential to the administration and evaluation of the test. It is quite conceivable that a computer could simulate most, or even all, of those seemingly unique human qualities. At this stage of development it cannot be stated that the computer has "no place" in the intelligence testing process. It can only be stated, with certainty, that the current state of the development of computer examiners, and the current state of the art of computer technology, have not reached the level of sophistication which would permit computer examiners to administer and evaluate intelligence tests in place of human examiners.

5. STATEMENT OF THE PROBLEM

The problem in this investigation was primarily exploratory in nature. Using the IBM 1500 Instructional System a computer examiner was developed to administer and evaluate the Verbal subtests of the Wechsler Intelligence Scale for Children (WISC) and to derive a Verbal I.Q. score comparable to the Verbal I.Q. score derived by a human examiner.

6. DELIMITATION OF THE STUDY

Potentially there are as many different kinds of person-to-computer interviews as there are kinds of person-to-person interviews. For example, a computer simulating the role of a psychotherapist would have many characteristics which would distinguish it from one simulating the role of an intelligence test examiner. Even within the more limited role of a test examiner, a computer would simulate different characteristics, depending on the kind of role being modeled and the kind of data being collected.

Two types of information are generally collected from an intelligence test - the client's intelligence (the quantitative measure) and how the client responds to questioning (the qualitative measure). The WISC was originally designed to measure only the client's intelligence. However, in the years which have passed since the WISC was developed, it has become common practice to use the WISC as a vehicle to gather qualitative information about the client's psychomotor and psychological characteristics. From the various subtest scores, clinical inferences have been made not only about the client's intelligence, but his behavioral characteristics as well. Littell (1960) states "there is a tendency to attempt to predict a wide variety of types of behavior from scores derived of the WISC (p. 149)". Such qualitative descriptions, although they are accepted in practice, are not legitimately part of the administration and evaluation of the WISC. The WISC, however, has become a vehicle for gathering such information. In this study, it is to be distinctly clear that the immediate concern was not with collecting associated qualitative information, but only with the quantitative measurement of the client's intelligence. Clinical inference was not

attempted.

It is obvious that it was not possible to simulate all of the characteristics indicated in the test manual as being necessary for the proper administration and evaluation of the test. Some of the physical and operational characteristics of the IBM 1500 System imposed constraints and limitations on the development of these characteristics. One of the major limitations of the 1500 System is that it cannot accept aural responses from the client or analyze them. Thus, instead of vocalizing responses, the clients were required to make responses via the Instructional Display Keyboard. Two questions are immediately obvious from such a change. First, does the change violate the standardization characteristics of the WISC? Obviously, it does because the test is now limited to only those clients who can interact with the Instructional Display Unit and who can type to some degree. Secondly, does use of the keyboard change the general factors of intelligence which the test purports to measure? It might do this, and it might not; to what degree is not known. If such problems were to have been given detailed consideration the primary objective of the study could not have been achieved.

Finally, no attempt was made to provide for statistical evaluation of test reliability or external validity. These will be left to future investigations. There was not sufficient time to measure them and at this stage such an effort would be premature.

CHAPTER 3

THE MODELING PROCESS

1. INTRODUCTION

A major portion of this study was devoted to the development of a computer model which would be capable of simulating the administrative and evaluative criteria of the WISC. Earlier in this study it was suggested that likely the most logical approach to developing such a model would be to borrow the procedures and techniques used to develop models in the industrial, scientific and commercial communities. This was precisely the approach taken. The objective of this chapter is to define modeling and modeling terms, state some general modeling principles, develop the modeling concept and breakdown and list stages of the process.

2. MODEL DEFINED

A computer model is a logical and/or mathematical representation of a concept, or system of concepts, or an operation which has been programmed for implementation on a computer (Martin, 1968, p. 5; Pritsker and Kiviat, 1969, p.1).

3. DEFINITIONS

(1) Concept: A thought, idea or mental image of an action or thing.

(2) Conceptual Model: A thought, idea or mental image of a concept, system, or operation in terms of a logical and/or mathematical representation.

(3) Documentation: A written record of the completion of a phase of work. It is a descriptive record of a system, its parameters and variables, operating characteristics and performance.

(4) First-Generation; Second-Generation Model: A first-generation model is the first complete approximation to the real phenomenon. It is generally the most gross approximation. A second-generation model is based on the results and conclusions drawn from the first-generation and is generally a finer and more sophisticated approximation to the real phenomenon.

(5) Flow Chart: A flow chart is an orderly representation of a process; a graphic illustration in which interrelationships are illustrated and activities are defined.

(6) Logical Flow Chart: A logical flow chart illustrates the logical flow of a system or operation.

(7) Program: A sequence of step-by-step operations performed by a computer in order to solve a given problem.

(8) Simulation: An artificial representation of a phenomenon, concept, system, or operation programmed for solution on a computer.

(9) System: A series of processes, organic or organized whole, or complex of men and machines.

(10) Validity: The degree to which the model approximates the real phenomenon.

(11) Subroutine: A sequence of computer instructions which can be repeatedly called and sequenced by the main computer program to perform a specified function.

MODEL PRINCIPLES

A model is basically a representation of reality. Computer models are obviously not the only possible kinds of models that represent reality. There are symbolic models (mathematical and logical equations), iconic models (maps, globes and scale models) and logical models (flow charts and computer programs) (Pritsker and Kiviat, 1969, p. 1). Models can be concrete, having substance like model airplanes, or they can be abstract as an equation. All models have two things in common. First, they are abstractions of reality and secondly, no model can completely duplicate reality; at best they are only approximations. A computer model is the embodiment of a theory about the operation of a specific observable phenomenon.

It is important to note that in any given model only those characteristics of the real situation which can be considered important or essential are included. In the process of building a model, the modeler must determine and consider these characteristics at the formulation stage.

It is also important for the modeler to realize that the ultimate solution to a modeling problem cannot be achieved in the first model. It can only be approximated. Modeling is an iterative process which requires the construction of more than one generation of solutions. The initial or first-generation model is simple and straightforward and is usually implemented more as an exploratory adventure (Martin, 1968, p. 161). The experience and knowledge gained from such a model is used as the basis for a second-generation which is generally more sophisticated and a closer approximation of the real situation. The iterative

approach works best whenever there is generally a lack of information to successfully implement an acceptable first-generation model.

In summary the modeler must recognize that:

(1) A model is an abstraction of reality.

(2) At best models can only be approximations to the real situation.

(3) Models are based only on those characteristics which are deemed important or essential to the situation.

(4) Models are usually constructed by an iterative approach, especially when there is a lack of information about the real phenomenon.

Each of these principles was used to construct the Computer-Examiner. Only the most easily implemented characteristics of the WISC were modeled, obviously due to the lack of operational definitions. This justified using the iterative modeling approach, but because of the size of the modeling task and limitations of time, only one-generation of the Computer-Examiner was created. To compensate for this, the concepts and detail necessary for the second-generation are adequately documented.

5. THE MODELING PROCESS

The number of ways in which the modeling process can be broken down into discrete steps are numerous and, equally, there is a copious use of terminology to identify these steps. The breakdown may depend upon the field in which the problem originated or upon who is doing the modeling. In this study the process and terminology of Martin (1968) was used because it is oriented toward the construction of computer models. But of more importance, the approach attempts to emphasize one

aspect of the process - model validation. This will be discussed shortly.

The outline of the modeling process is as follows:

(1) Model Conceptualization

(A) Definition of the problem.

(B) Analysis of the problem.

(C) Adoption of hypotheses and assumptions.

(D) Establishment of model rationale.

(E) Definition of parameters and variables.

(F) Description of the conceptual model in abstract terms and concepts.

(G) Check validity of conceptual model.

(H) Documentation of conceptual model phase.

(2) Model implementation

(A) Develop logical flow chart.

(B) Check validity of logical model.

(3) Computer program implementation

(A) Design program flow chart.

(B) Check validity of program flow chart.

(C) Program computer model.

(D) Check validity of program.

(E) Debug computer program.

(F) Check validity of debugging.

(G) Document computer program.

(4) Model results

(A) Finalize experimental design.

(B) Execute computer runs.

(C) Analyze and evaluate results.

(D) Draw conclusions.

(E) Make recommendations.

(F) Document results.

Note that the outline is arranged in a sequential order, and each step progressively moves the operational specifications of the real phenomenon toward an abstract model. Also note that several stages throughout this process are used for validity checks.

6. MODEL VALIDATION

The process by which the modeler establishes the accuracy of the approximation or the model to the real phenomenon is called model validation. Martin (1968, p. 153) states that model validation is the critical problem faced in the process and it cannot be overlooked. The valid model logically or mathematically approximates the concept, system, or operation and with the phenomenon being modeled.

Validation, according to Martin, takes place in two stages: validation of the conceptual model and validation of the implemented model. The first validity check is to check the transformation from the real world phenomena to the conceptual model. At this check the question is asked: Is the conceptual model a valid representation of the real world? According to Martin, this question must be asked with objectivity; however, the task of validating a conceptual model often

results in nothin more than a subjective opinion. The key to model validation is that validation is carried out in each stage by the process of comparing the current stage of development with the previous. For example, the program flow chart is compared to the logical flow chart which is compared to the conceptual model and in turn this is compared to the real world.

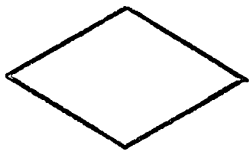
The final stage of model validation constitutes a comparison of the model against real world inputs. This method is unquestionably the best check of model validity. It brings the modeler into close contact with the real world and shows the model's congruence to it.

7. FLOW CHART SYMBOLS

Development of the Computer-Examiner was based upon the use of a large number of flow charts. The symbols used in these flow charts are defined as follows:

The start, the name, or the end of a sequence of activity.

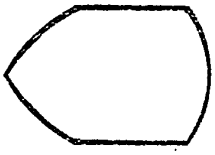
A single, or group of activity (a module).



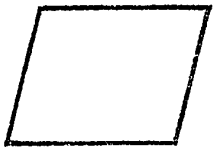
A decision point in the format of a question.

A documentation activity.

An audio message to be played on the 1505 Audio Tape Unit.



A message displayed on the 1510 Instructional Display.



A response to be accepted from the client via the 1510 Instructional Keyboard or Light Pen.

An "off page" connector. The letter "A" designates the on page connector to which the connector is attached.

The direction of the flow of activity.

10 

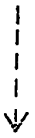
The number of the module. Refers to module 10.

(Exp: Figure 3A)

Refers to an expansion of the module in Figure 3A.

(Doc: Appendix B, List 5)

Refers to the documentation of the module in Appendix B, List 5.



Optional flow of activity depending upon conditions.

Optional single, or group of activity depending upon conditions.

Page: 32

Refers to the page upon which description of the flow chart begins.

CHAPTER 4

DEVELOPMENT OF THE COMPUTER EXAMINER

1. INTRODUCTION

The Computer Examiner was developed by utilizing the concepts and the modeling procedure outlined in Chapter 3. The objective of Chapter 4 is to document this development. In this respect, emphasis will be on description of:

(1) the problem,

(2) development of the conceptual model,

(3) development and implementation of the computer model, and

(4) evaluation procedures.

Evaluation, results, observations and conclusions will be presented in Chapter 5.

THE PROBLEM

The problem, as formulated in Chapter 2 was to develop a computer model capable of administering and evaluating the Verbal Section of the WISC and implementing it on the 1500 Instructional System.

DELIMITATIONS

A number of delimitations were made concerning how the problem was to be handled. These are:

(1) The study was primarily exploratory in scope. Perfection in

terms of the Computer-Examiner's ability to model all of the administration procedures and all of the human examiner's behavior was not the terminal goal. It was to develop a model which could:

(A) administer all questions of the Verbal Section in correct sequence,

(B) evaluate responses as accurately as possible,

(C) correctly award raw score points for each question,

(D) accurately total raw score points,

(E) correctly make scale score conversions,

(F) accurately total subtest scale scores, and

(G) correctly convert the scale score total into the Verbal I.Q.

(2) Deviations are often made in administration of the WISC. These are due to:

(A) many differences which exist from one testing site to the next,

(B) individual differences of clients, and

(C) individual differences in personalities and styles of administration in examiners.

Examiners, despite precautions about making deviations from the prescribed administration procedures so strongly emphasized in the manual, will often consciously or unconsciously, deviate from the specified way of administering the test. No attempt was made to explore, or even consider, such situations. Only the administration procedures set forth in the manual and the supplement as being necessary for the proper administration and evaluation of the test were considered for

the proper administration and evaluation of the test were considered for inclusion in the computer model.

(3) A number of criteria in the manual and supplement stated as being necessary for the proper administration and evaluation of the WISC could not be modeled. The concepts involved were either too vague or lacked operational specificity. No attempt was made to model such criteria. Instead they were documented as not being part of the model.

(4) Sophistication of the Computer-Examiner, as a computer model, was largely limited by the hardware and software capabilities of the 1500 Instructional System. No resources were available to make improvements on these in order to develop a more effective computer model.

(5) A great deal of latitude exists in the approach to the problem and the subsequent development of almost any model. This latitude could be due to:

(A) the fact that there are many kinds of modeling problems,

(B) the constraints in terms of time and resources placed on a modeling task,

(C) the modeler's particular bias in perception of the problem,

(D) the modeler's characteristic way, or style, of developing models,

(E) the amount of data available concerning the problem, or

(F) the precision with which definitions are given operational specificity.

Taking such factors into account, it was realized that the modeling approach outlined in Chapter 3 could not be followed verbatim. The essential tasks of conceptualizing the model, implementation, programming, evaluation and documentation were followed, but beyond this

variation is evident. Because the objective of this study was not to develop the modeling process but to use it as a tool in developing the computer model, the variations made in the modeling process are not accounted for or justified on any other ground than recognition of the fact that latitude can exist because of the factors cited.

DEFINITIONS

(1) Audio Message: An audio message played to the client from an audio tape cartridge loaded on the 1505 Audio Tape Unit.

(2) Behavior: The behavior or behaviors (more than two different actions) which are exhibited by the examiner while administering or evaluating the WISC.

(3) Buffer: Refers to a 100 alphanumeric character storage area of the 1500 Instructional System. The client's response is stored in the response buffer. There are five additional buffers numbered 1 to 5, each with a storage capacity of 100 characters.

(4) Category: Refers to 1 of the 7 categories into which a client's answer can be classified.

(5) Client: The person or subject to whom the WISC test is being administered.

(6) Computer-Examiner: Is the formal name for the computer model developed to administer and evaluate the WISC Verbal Section.

(7) Computer Model: The end product of the modeling process - the computer program of the conceptual model as implemented on the 1500 Instructional System.

(8) Computer Program: The entire sequence of Coursewriter II language instructions which describes the conceptual model to the 1500

Instructional System.

(9) Criteria: Refers collectively to the terms behavior and procedure.

(10) Enter Key: When depressed by client it indicates the response is compete and is to be processed.

(11) Examiner: Refers generally to the administrator of the WISC.

(12) Debug: The process of correcting errors in the computer program.

(13) Display Screen: The television screen of the 1510 Instructional Display, which is used to display messages to the client.

(14) Graphic Message: The visual messages displayed on the 1510 Instructional Display. See Figures 4 to 7 as examples (pp. 89 to 92).

(15) Human Examiner: Refers specifically to a human administration of the WISC.

(16) Index Key: Has the function of advancing the cursor to the next line of the Instructional Display.

(17) Key: A typewriter key of the 1510 Instructional Display Unit.

(18) Keywords: Words selected from the source answers as cues for the computer model in its search for the source answer in the client's response.

(19) Manual: The WISC instruction manual.

(20) Modeling Process: Refers collectively to all the activities involved in development of the Computer-Examiner.

(21) Module X: A single definable group of activity numbered X in development of the Computer-Examiner.

(22) Objectives: Those statments in Appendix B which specify in operational detail how the Computer-Examiner is to administer and evaluate the test.

(23) Procedures: The steps an examiner follows in order to administer and evaluate the Verbal Section of the WISC.

(24) Program Flow Chart: A logical description of the sequence of operations performed by the computer.

(25) Protocol: The record form of a previously administered WISC.

(26) Raw Score: The totaled score of a subtest before scale score conversion.

(27) Response: The information input to the 1510 Instructional Display either by keyboard or light pen.

(28) Response Analysis: The process by which the Computer-Examiner searches the client's response for an answer.

(29) Response Buffer: See Buffer.

(30) Response Category: See Category.

(31) Response Hierarchy: The defined order of priority for the response categories.

(32) Response Repertoire: Collectively all those responses which the Computer-Examiner is programmed to recognize.

(33) Return Key: Has the function of returning the cursor to the right hand side of the Instructional Display.

(34) Source Answer: An answer taken from either the WISC manual or the Massey Supplement.

(35) Skeleton Answer: A keyword answer which has been modified to compensate for spelling or typing errors. This is the answer which the computer model uses as the reference.

(36) Subject: See Client.

(37) Supplement: The Massey WISC Scoring Criteria supplement.

(38) Verbal Section: Refers collectively to the General Information, General Comprehension, Arithmetic, Similarities and Vocabulary subtests.

2. THE MODELING PROCESS

THE STAGES

Development of the computer model was carried out in three major stages of activity:

(1) transformation of the criteria in the manual and supplement into the conceptual model,

(2) transformation of the conceptual model into a computer program capable of running on the 1500 Instructional System, and

(3) experimentation and evaluation of the computer model.

GENERAL DESCRIPTION

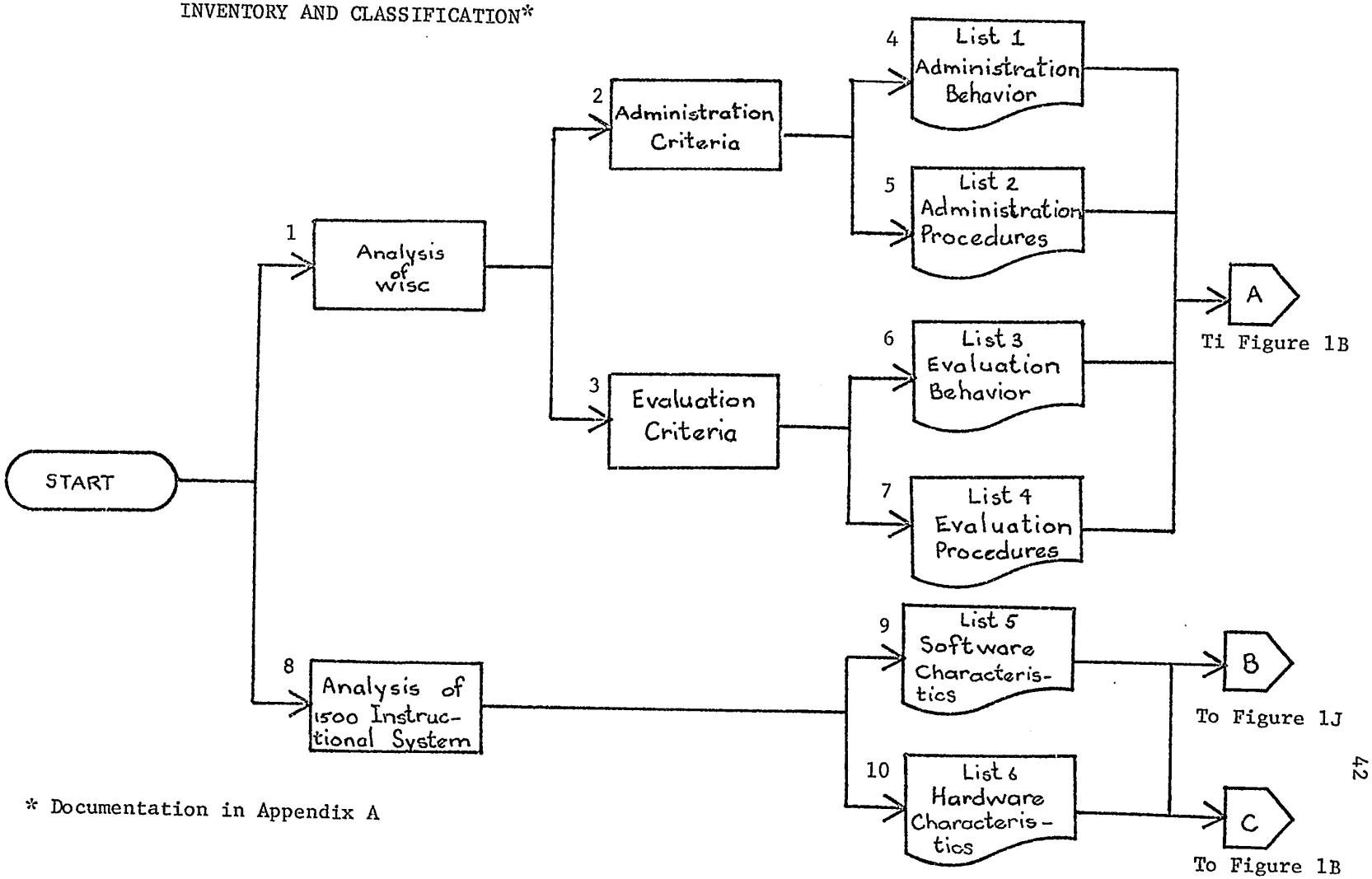
Development of the Computer-Examiner can best be described by making reference to Figures 1A through to 1J (pp. 42 to 51) the details of which will be explained in the sections to follow. These figures diagram the stage-by-stage development of the computer model in a block flow chart form. Figures 1A, 1B, and 1C outline the activity involved in transforming the administration and evaluation criteria of the manual and supplement into the conceptual model. Figures 1D, 1E, 1F, and 1G lay-out the activity followed in transforming the conceptual model into the computer model, while Figure 1H diagrams the initial testing stage for the computer model. Figures 1I and 1J outline the final evaluation stages.

Within most stages of development there were essentially four

FIGURE 1A: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

TRANSFORMATION TO CONCEPTUAL MODEL

INVENTORY AND CLASSIFICATION*

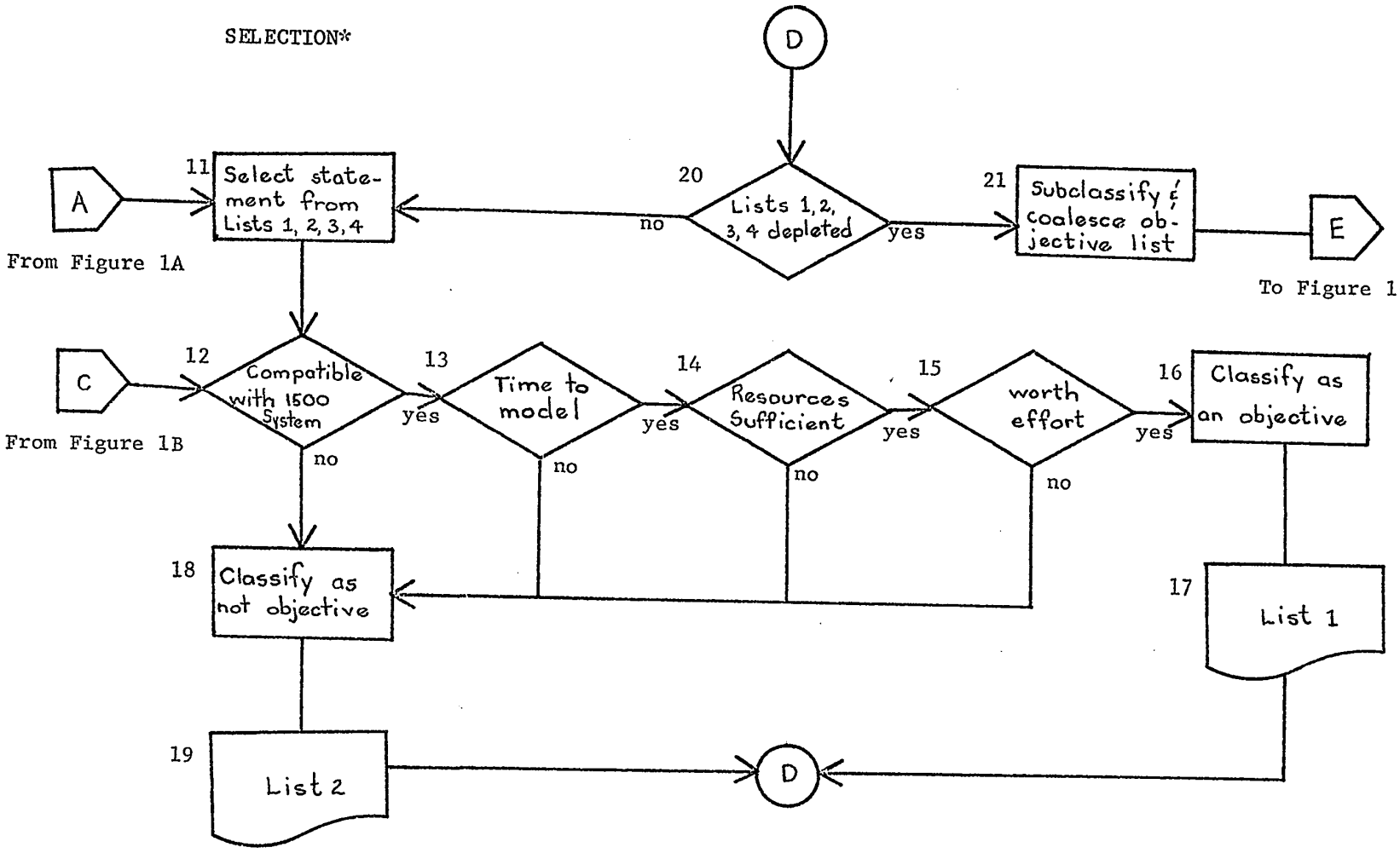


* Documentation in Appendix A

FIGURE 1B: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

TRANSFORMATION TO CONCEPTUAL MODEL

SELECTION*



* Documentation in Appendix B

FIGURE 1C: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

TRANSFORMATION TO CONCEPTUAL MODEL

CONCEPTUALIZATION

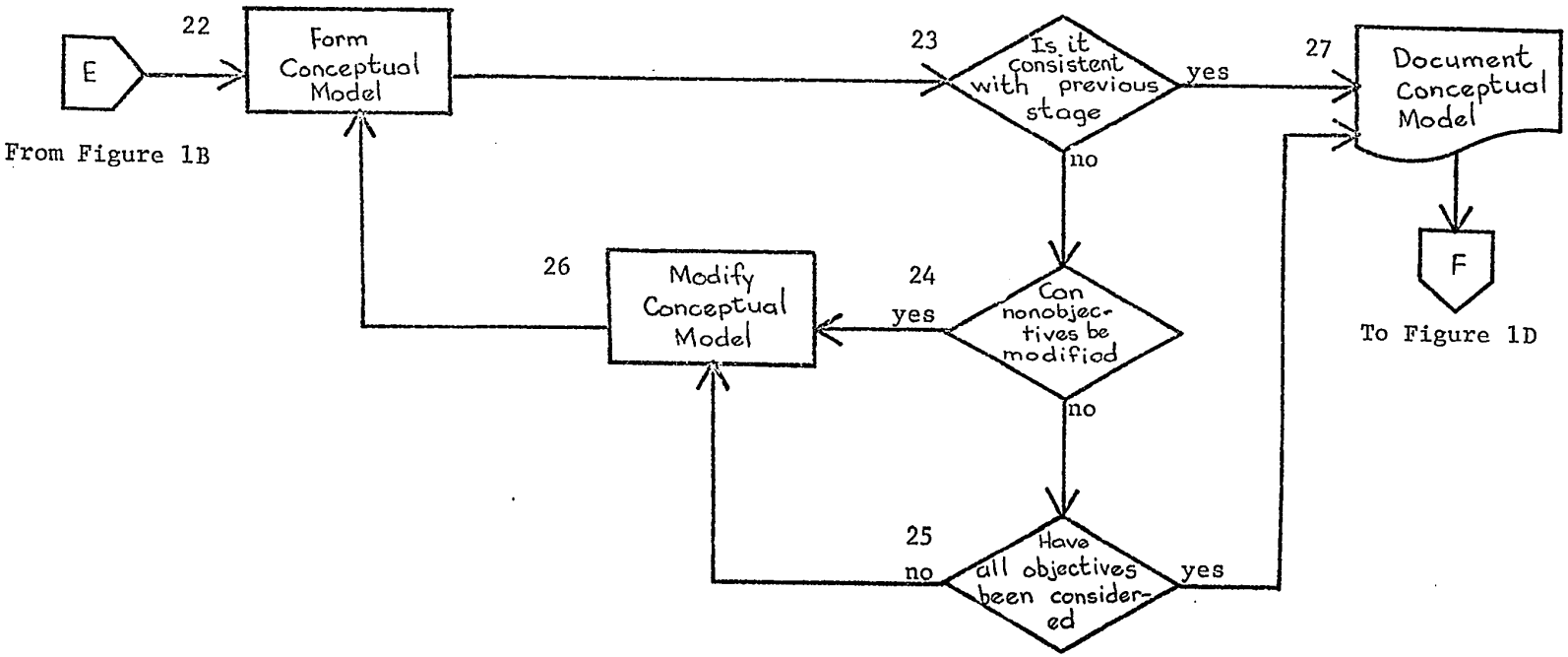


FIGURE 1D: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

TRANSFORMATION TO COMPUTER MODEL

LOGICAL FLOW CHARTING

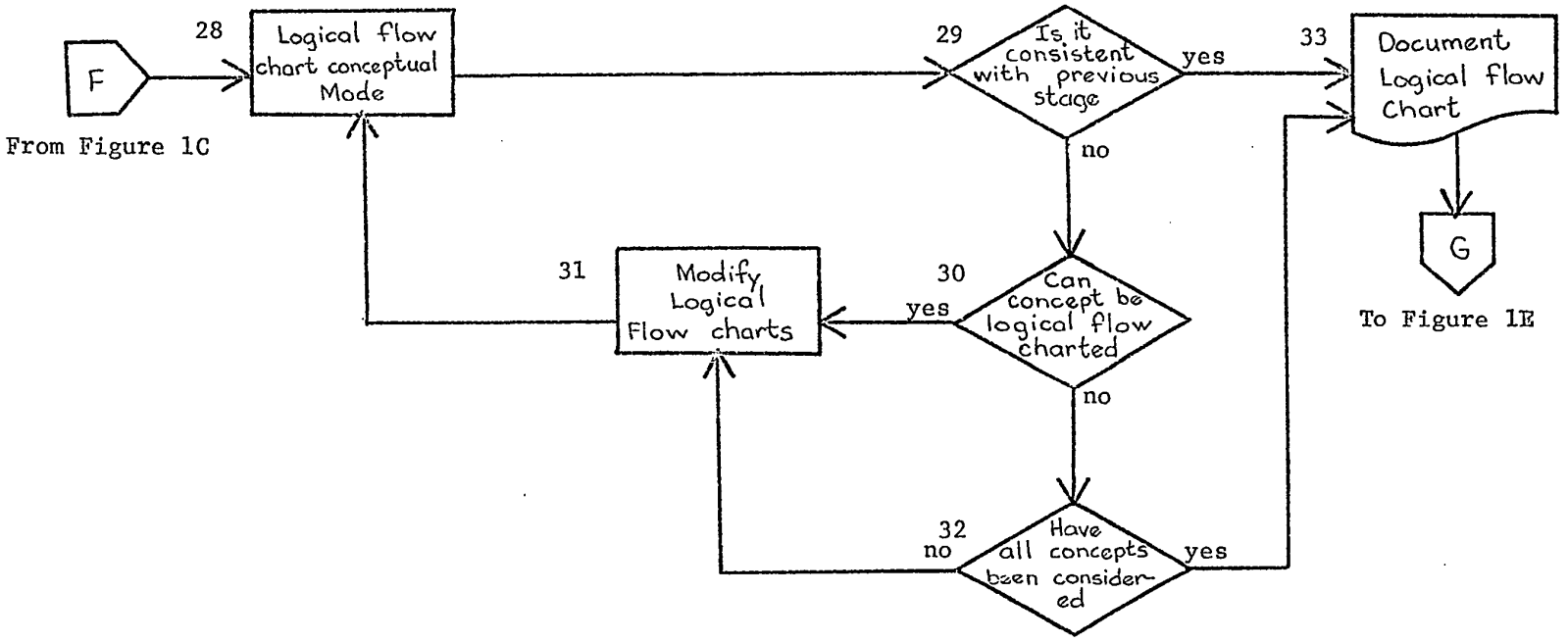


FIGURE 1E: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

TRANSFORMATION TO COMPUTER MODEL

PROGRAM FLOW CHARTING

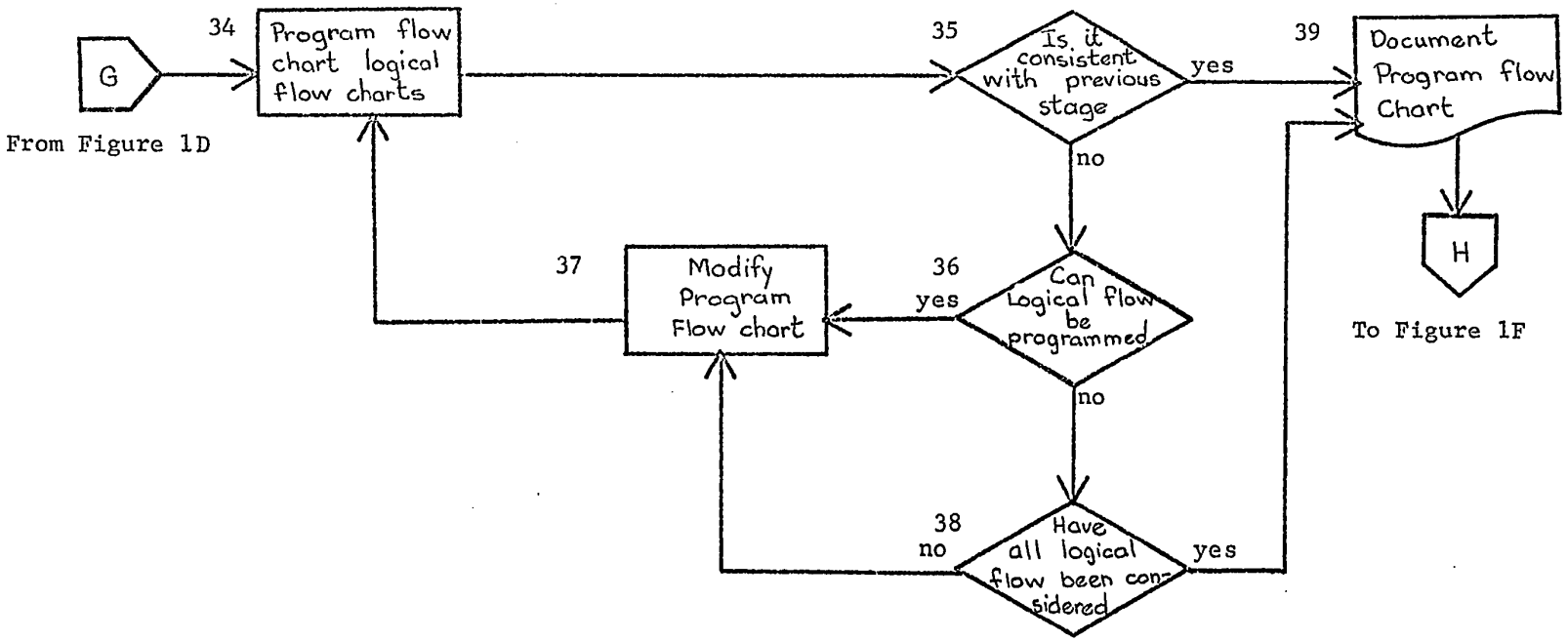


FIGURE 1F: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

TRANSFORMATION TO COMPUTER MODEL

COMPUTER PROGRAMMING

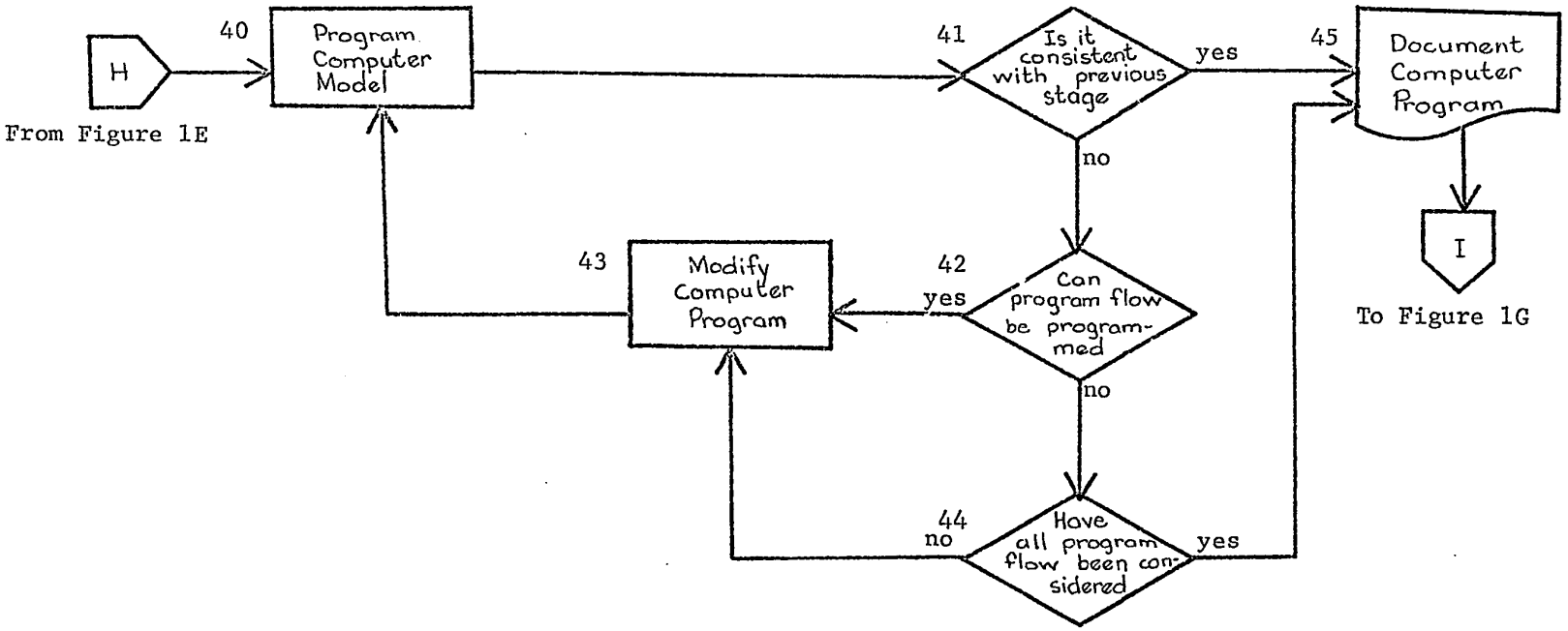


FIGURE 1G: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

TRANSFORMATION TO COMPUTER MODEL

DEBUGGING

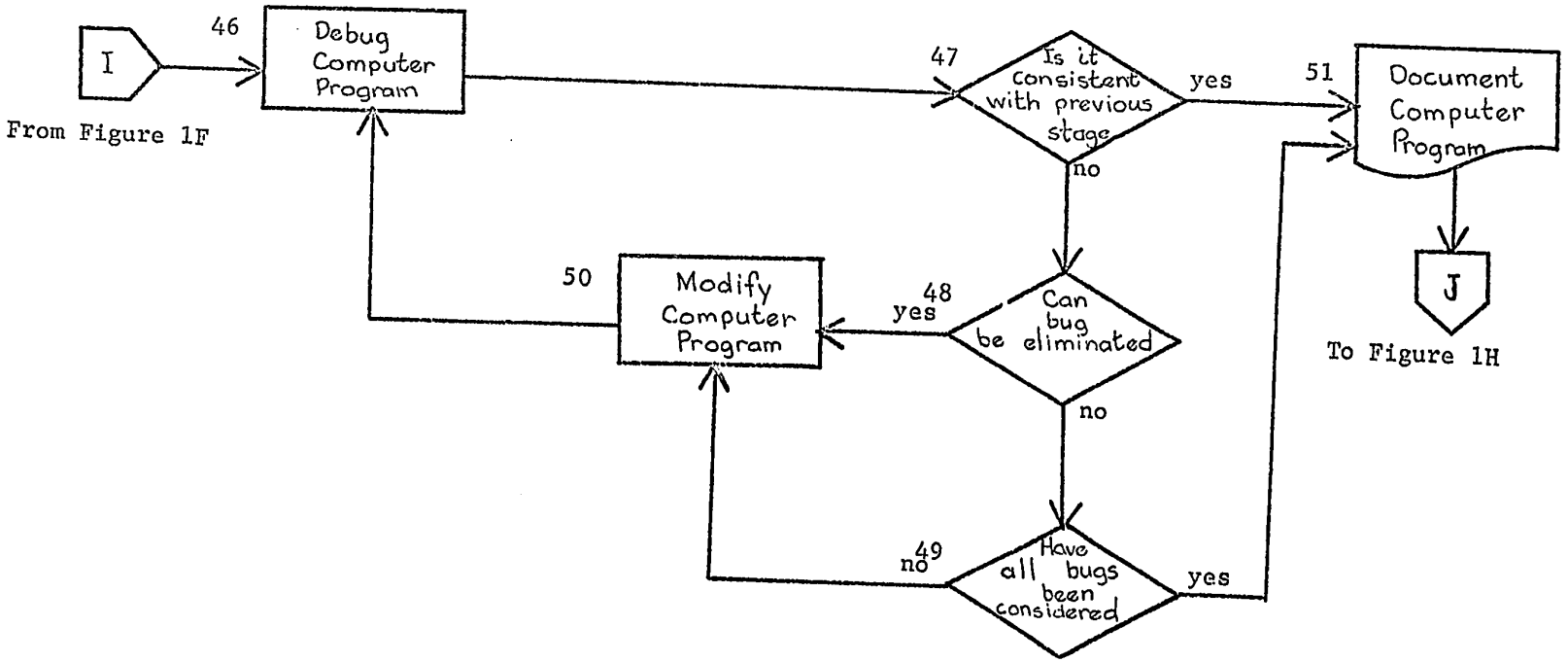


FIGURE 1H: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

COMPUTER MODEL IMPLEMENTATION

VOLUNTEER ADMINISTRATION

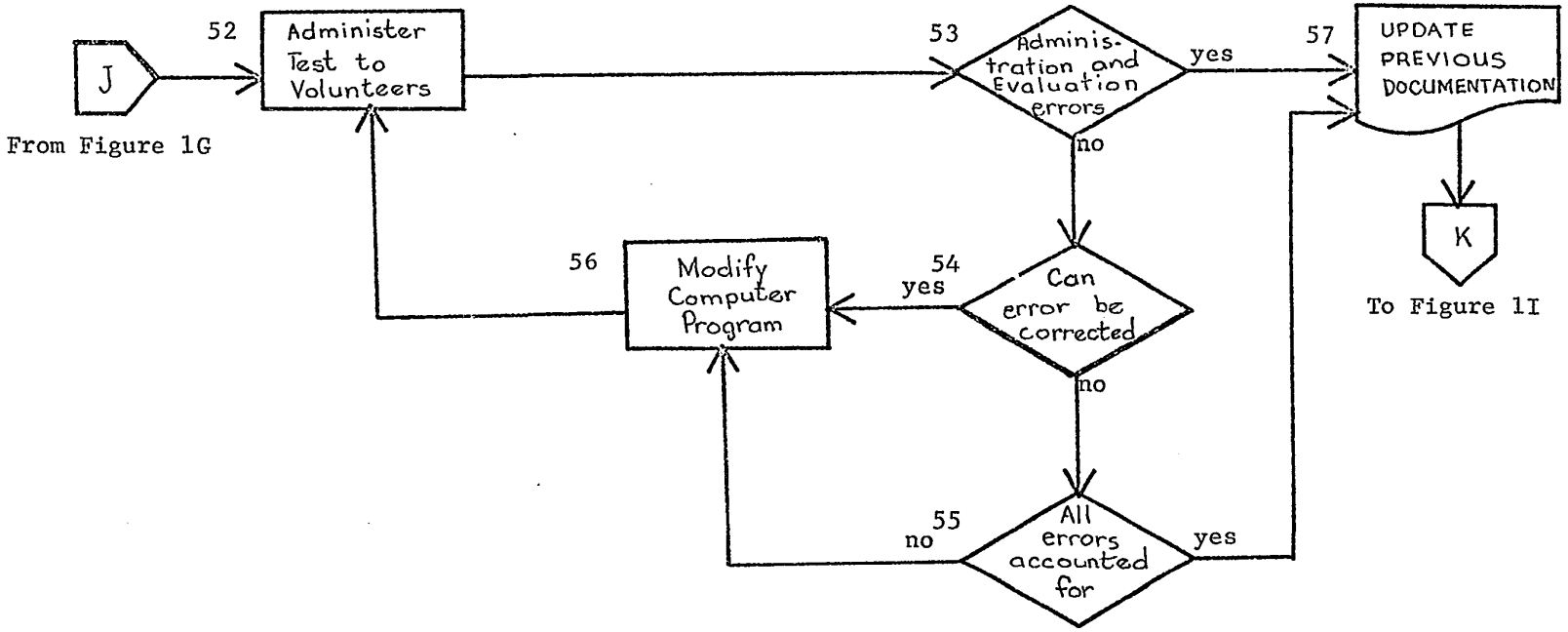


FIGURE 1I: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

COMPUTER MODEL IMPLEMENTATION

PROTOCOL ADMINISTRATION

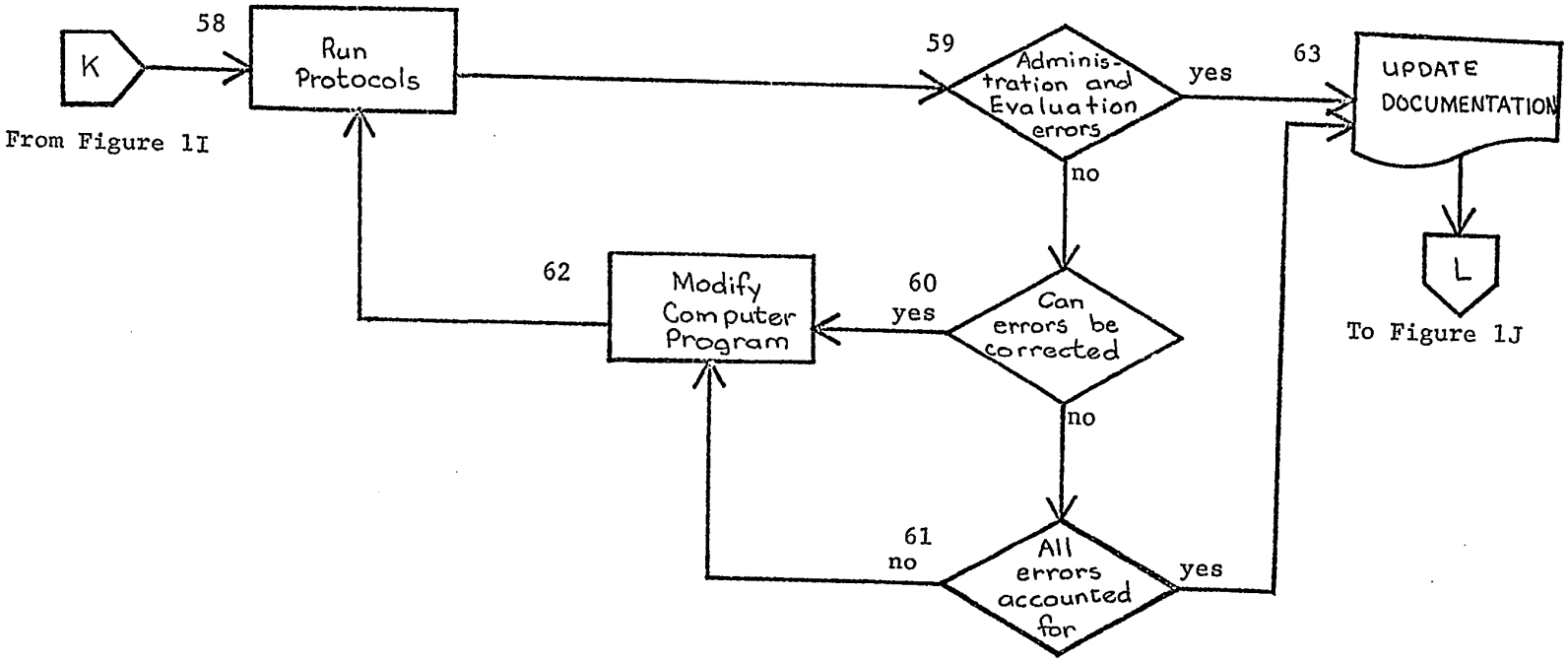
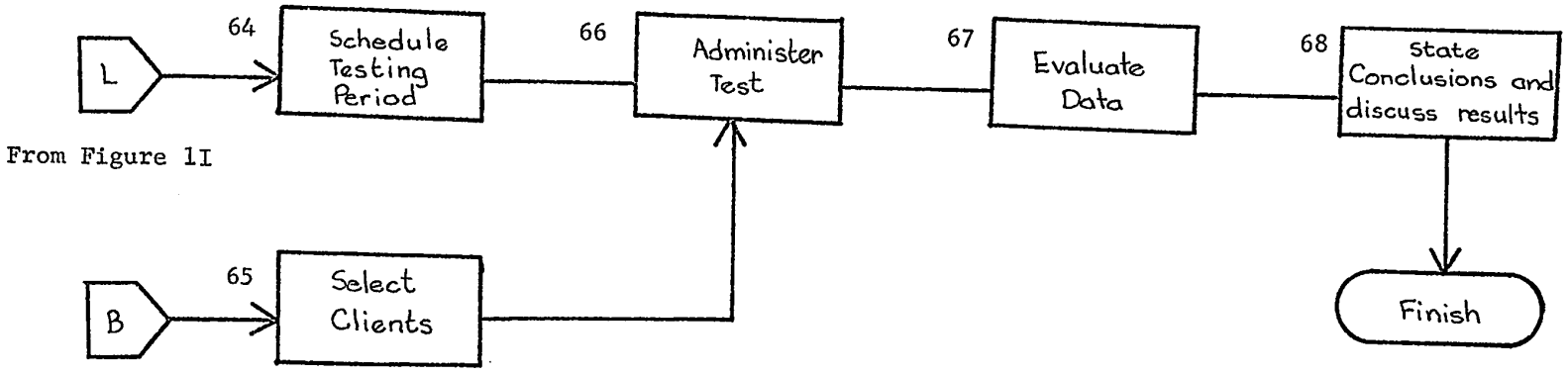


FIGURE 1J: COMPUTER-EXAMINER DEVELOPMENTAL PROCESS

COMPUTER MODEL IMPLEMENTATION

CLIENT ADMINISTRATION



From Figure 1I

From Figure 1A

activities:

(1) performance of the function,

(2) validation,

(3) modification, and

(4) documentation.

The first activity performed in each stage of development was to carry out the function or purpose of the stage; for example, to form the conceptual model (module 22, Figure 1C), to generate the logical flow chart (module 28, Figure 1D), or to program the computer (module 40, Figure 1F).

The second activity performed in most stages was to validate the function. This was in keeping with the concept that a model can achieve some degree of validity by progressively validating each stage of development. This step was accomplished by asking the question: Are the results of the function performed in this stage in congruence with the documentation of the previous stage (module 41, Figure 1F)?

The third activity performed in most stages was to make modifications on the function in order to reduce to a minimum the number of incongruences between that stage and the previous stage. This activity was recursive until all the known incongruences were either eliminated or documented as discrepancies (modules 42, 43 and 44, Figure 1F).

Documentation, the last activity performed in each stage was essential to the task of developing the Computer-Examiner for a number of reasons. First, it was essential to delimit exactly what behaviors and procedures were being modeled and which were not. The task of developing a computer model of the administrative and evaluative functions of the WISC is basically one of paralleling the behavior

required or prescribed of the examiner as the test is administered to the client. To say the least, this is a difficult and complex task, because human behavior - even the restrictive behavior prescribed for the examiner - is extremely complicated and variable. Much of this behavior has yet to be expressed operationally, and the resources needed for this were definitely not available to this study.

In this respect, documentation played the important role of explicitly separating the behaviors being modeled from those which were not. By doing so it was assumed that final evaluation of the Computer-Examiner could be carried out on two discrete planes - first, on an external plane in which the behavior exhibited by a human examiner, and secondly, on an internal plane in which the behavior exhibited by the Comput -Examiner would be compared to the behavior it was designed to exhibit. For the former, the question would be asked, How did the Computer-Examiner differ from human examiners? and for the latter, Did the Computer-Examiner perform to expectation?

Secondly, the documentation activity performed at the end of each stage of development establishes the validity of the function performed in that stage in reference to the documentation activity performed at the end of the previous stage (module 45, Figure 1F).

Third, to some the task of modeling the WISC would be considered impossible. They would not accept any level of sophistication in behavior less than that of the human examiner as evidence of the Computer-Examiner's competence. When evaluating the model they would cite behaviors which could not or were not modeled as evidence of incompetence. In order to prevent this type of ciriticism it is necessary to explicitly

delimit that behavior which could not be modeled from that which could. As previously indicated, this forces evaluation of the Computer-Examiner into either an external or an internal plane. Critics of the model would then be forced to address either one of the planes but not both at the same time. This will prevent them from using legitimate criticism from the external plane. i.e. those behaviors which were not modeled, as grounds for condemnation on the internal plane, i.e. the behaviors which were modeled.

Last, documentation will help to ensure that a data base is available for a successive generation of the computer model.

3. TRANSFORMATION TO CONCEPTUAL MODEL

The following activities were undertaken in order to transform the WISC test into the conceptual model:

(1) INVENTORY AND CLASSIFICATION of administration and evaluation criteria of the WISC, and hardware and software characteristics of the 1500 Instructional System (Figure 1A).

(2) SELECTION of the criteria to be included in the computer model (Figure 1B), and

(3) CONCEPTUALIZING the computer model (Figure 1C).

INVENTORY AND CLASSIFICATION (Figure 1A, p. 42)

The WISC manual and Massey supplement were inventoried for all administration and evaluation procedures which were indicated to be necessary for the proper use of the test (module 1). The criteria for being necessary was taken to be those statements in the manual

and supplement which were either imperative or declarative in meaning.

Statements in the list were then sorted into the following classifications (modules 2 through 7):

(1) EXAMINER'S ADMINISTRATION BEHAVIOR,

(2) EXAMINER'S ADMINISTRATION PROCEDURES,

(3) EXAMINER'S EVALUATION BEHAVIOR,

(4) EXAMINER'S EVALUATION PROCEDURES.

At the same time, within each of these classifications, statements were further subclassified according to the area of concern. These were:

(1) GENERAL TO WHOLE WISC.

(2) GENERAL INFORMATION,

(3) GENERAL COMPREHENSION,

(4) ARITHMETIC,

(5) SIMILARITIES,

(6) VOCABULARY, and

(7) DIGIT SPAN.

The purpose of classifying and subclassifying was to group similar procedures and behaviors. Many statments were found to be redundant; that is, the same procedure or behavior was implied in more than one statment. These statments were consolidated. The resulting lists are documented in Appendix A as lists 1 (p. 179), 2(p. 180), 3(p. 185), and 4(p. 186).

Next, the operating manuals of the 1500 Instructional System were analyzed to obtain a list of its operating characteristics (module 8). Statments in this list were then sorted into the following classifications (modules 9 and 10):

(1) SOFTWARE CHARACTERISTICS, and

(2) HARDWARE CHARACTERISTICS.

The resulting lists are documented in Appendix A as Lists 5 (p. 192) and 6 (p. 195)

SELECTION (Figure 1B, p. 43)

In the second stage of the activity in transforming the WISC into a conceptual model each statement within lists 1, 2, 3 and 4 (Appendix A, p. 179) was evaluated in light of the following questions (module 11).

(1) Considering the hardware and software characteristics of the 1500 System, as outlined in lists 5 and 6 (Appendix A, p. 192) could the criteria specified in the statement be included in the conceptual model (module 12)?

(2) Would there be enough time to implement the criteria specified in the statement in the computer model (module 13)?

(3) Are the resources (manpower and financial) required to develop the specified criteria in the computer model available (module 14)?

(4) Would it be worth the effort to implement the criteria in the computer model (module 5)?

If during examination of the statement, the answer to any one of the questions were no the statement was classified as not an objective to be met in the development of the computer model (module 18) and documented (module 19). If the response to all four questions were yes the statement was classified as an objective to be met in the development of the computer model (module 16) and documented (module 17).

Statements from lists 1, 2, 3 and 4 were selected individually and examined in light of the four questions (module 11). Selection continued until all statements had been examined (module 20). Throughout the selection activity the original classification and subclassifications of lists 1, 2, 3 and 4 (modules 4, 5, 6 and

7) were retained in the objective list (module 17). In addition, statements falling in the category General to the Whole WISC (under Examiner's Administration Procedures and Examiner's Evaluation Procedures) were subclassified once more in order to further common group procedures and behaviors (module 21). Only the major title headings were used in the nonobjective list because the majority of statements fell into the objective category. Appendix B contains the documentation of the SELECTION stage. The objective list starts on page 202 and the nonobjective list on page 215. The classification and subclassification scheme used in the appendix for grouping common procedures and behaviors was as follows:

1. LIST OF COMPUTER MODEL OBJECTIVES
   - A. EXAMINER'S ADMINISTRATION BEHAVIOR
     1. General to Whole WISC
        - A. Questioning
        - B. Requestioning
        - C. Order of Test Administration
        - D. Client Management
        - E. Environmental Conditions for Testing
     2. General Information
     3. General Comprehension
     4. Arithmetic
     5. Similarities
     6. Vocabulary
   - B. EXAMINER'S EVALUATION BEHAVIOR
   - C. EXAMINER'S EVALUATION PROCEDURE
     1. General to Whole WISC
        - A. Initial Single Response Evaluation

B. Multiple Subject Responses

C. Questioning

D. Scorekeeping

2. General Information

3. General Comprehension

4. Arithmetic

5. Similarities

6. Vocabulary

2. LIST OF COMPUTER MODEL NONOBJECTIVES

A. EXAMINER'S ADMINISTRATION BEHAVIOR

B. EXAMINER'S ADMINISTRATION PROCEDURE

C. EXAMINER'S EVALUATION BEHAVIOR

D. EXAMINER'S EVALUATION PROCEDURE

CONCEPTUALIZING (Figure 1C, p. 44)

Conceptualization was the third stage of activity involved in the transformation of the WISC test into a conceptual model. In short, this was the thinking and planning stage in which the task of developing the computer model was defined and analyzed, the objectives of Appendix B were evaluated as information and data essential to the design of the model, a number of assumptions were made, and the parameters and variables of the model were defined.

The conceptualizing activity is represented by module 22 (Figure 1C, p. 44). The validation of this activity is represented by modules 23 to 26. Documentation (which is described in this section) is represented by module 27.

Description of the thinking and planning aspects was a difficult task due to abstractness of the conceptual model. To make this description as clear and as concise as possible, frequent references have been made to the next stage of development - the logical flow charts (Figure 1D, p. 45). By doing this it was possible to describe some of the more abstract concepts.

Development of the conceptual model took place in a number of definable areas. As an overview to the description of this development, these areas are:

1. ADMINISTRATION
2. MANAGEMENT OF THE CLIENT
3. TESTING ENVIRONMENT PREPARATIONS
4. EVALUATION OF THE CLIENT'S RESPONSE
    - A. Answer Categorization
    - B. Answer Hierarchy
        - (1) Decision Tables
        - (2) Decision Table Symbols
        - (3) How to Use the Decision Tables
            - (a) Decision Table 1
            - (b) Decision Table 2
            - (c) Decision Table 3
    - C. Answer Analysis
        - (1) Rationale
        - (2) Example: General Information Question 23
        - (3) Processing Categories
        - (4) Processing Source Answers
        - (5) Generation of Keywords

(6) Generation of Skeleton Answers

5. COMMUNICATION BETWEEN CLIENT AND COMPUTER-EXAMINER

(A). Communication From Computer-Examiner to Client

(1) Messages for Questioning

(2) Graphic Messages

(B) Communication From Client to Computer-Examiner

(1) Response Acceptance Via Keyboard

(2) Client Typing Procedure

6. SCOREKEEPING

7. RESPONSE RECORDING

8. MULTIPLE TESTING PERIODS

9. MISCELLANEOUS OBJECTIVES

1. ADMINISTRATION

Objectives 1 and 5 (Appendix B, p. 202) suggest that the examiner must behave in a manner which is skillful and careful. To behave this way, the examiner must consistently make decisions which progress the test in the correct direction. With exception of those objectives which deal with how the client is to be questioned, and the manner in which the client is questioned, the majority of objectives in the Examiner's Administration Procedures section of Appendix B are concerned with operational specification of such decisions and with outlining the resulting courses of action. These objectives are: 19, 26, 27, 29, 30, 35 through to 54 and 56 (pp. 204 to 208). Conceptually they fit into that section of the model which is concerned with progression of the subtests and questions within each subtest. In relation to the Logical Flow Charts this section is referenced as module 3, Figure 10 (p. 100) and in greater detail as modules 8, 9, 10, 11 and 12, Figure 11C(p. 103).

No description of the conceptualization of these objectives will be given because it is difficult to relate specific detail about the individual conceptualization of each objective in the Examiner's Administration Procedures section, beyond the fact they are identifiable as a group at this stage of development. This is because the majority do not come into effect in the development of the model until either the logical flow chart stage of development (module 28, Figure 1D, p. 45) or more so, the program flow chart stage (module 34, Figure 1E, p. 46). In addition the description would be extremely lengthy.

For example, objective 19 (p. 204) deals with the order of the subtests and the conceptualization of this objective surfaces in the logical flow chart stage. This is diagrammed as the order of modules 8, 9, 10, 11 and 12 (Figure 11C, p. 103). Similarly, objective 26 (p. 204) suggests the order of the questions. This is diagrammed as the order of administration within each of the previously cited modules. While objective 29 (p.205), which deals with the limit of consecutive failures for the General Information subtest, cannot be conceptually identified until the program flow chart stage (module 40, Figure 13F, p. 117)

2. MANAGEMENT OF THE CLIENT

Objective 20 to 22 (Appendix B, p. 204) outline behavior of the client which cannot be tolerated during administration of the test. These objectives suggest that, if necessary, the examiner must manage the client's behavior in order to complete the administration of the test. This task could not be incorporated into the program of a computer model which uses the 1500 Instructional System as its computing base.

Its input/output hardware and software systems are not sophisticated enough to permit supervision of a client's behavior or to take the necessary action to correct deviant or hostile behavior.

It was conceptualized that the realistic approach would be to have a human, called either the supervisor, the tester or the human examiner to assist the Computer-Examiner as it administered the test. One of the functions of this person would be to supervise the client's overall behavior in respect to the kind of concerns expressed in objectives 20 to 22 (p. 204).

The supervisor assists with the administration of the test through all sections except calculations of the Verbal I.Q. This is diagrammed as encompassing modules 1, 2 and 3 (Figure 10, p. 100).

3. TESTING ENVIRONMENT PREPARATIONS

Objectives 23 to 25 (Appendix B, p. 204) outline some of the necessary preparations which must be made in order to insure that the client has a minimum amount of discomfort as the test is administered. The nature of the tasks to be performed during this prepartion are in the same category as those classified under the objectives for Management of the Client. The behaviours required are too difficult, or seemingly impossible to model, and therefore, the assistance of a human supervisor is indicated in order to help prepare the client for testing. Conceptually, client preparation takes place before the Computer-Examiner begins administration of the subtests as shown by module 2, (Figure 10, p. 100). To insure that the necessary preparations are carried out, the Computer-Examiner presents the human supervisor a check list, to be completed and verified one item at a time, as

diagrammed in the expanded logical flow chart of module 2 (Figure 10), by modules 29 to 36 (Figure 12F, p. 110). The questions asked in this check list are as follows:

(1) Are you and the clients the only persons in the room? (module 29)

(2) Is the terminal room door locked and the testing sign hung up? (module 30)

(3) Is the client settled? (module 31)

(4) Is the client's seat adjusted for his/her height and position? (module 32)

(5) Is the keyboard properly positioned for the client's reach? (module 33)

(6) Show the client the keyboard, space bar, and light pen. (module 34)

(7) Show the earphones and the volume control to the client, place on the phones, adjust the bands and position until comfortable to the client. (module 35)

(8) Examiner is the client ready to begin? (module 36)

4. EVALUATION OF THE CLIENT'S RESPONSE

The examiner's evaluation behavior is outlined in objectives 57 and 58 (p.208). Evaluation procedures are specified in objectives 59 through to 95 (pp. 209 to 215). The entire group of objectives, 57 through to 95 has led to a conceptualized model of the evaluation section which consists of three distinct parts:

(A) Response Categorization,

(B) Response Hierarchy, and

(C) Response Analysis.

A. ANSWER CATEGORIZATION

Objectives 59 through 64 and 71 (Appendix B, p. 209) suggest that the client's responses can be classified into one of the following categories:

(1) C2 - The client's response to a two point question is totally correct and worth the full two points. Further questioning on the same question is to be terminated.

(2) Q2 - The client's response to a two point question is partially correct (or is a correct but inferior response) and worth a partial one point. Further questioning is to be continued because there is evidence to indicate that requestioning may clarify the client's response sufficiently to gain the two points.

(3) C1 - The client's response to a two point question is partially correct (or is a correct but inferior response) and worth one point, or the client's response to a one point question is totally correct and worth the full one point. Further questioning is to be terminated. In the case of a partially correct two point response evidence does not exist to indicate that the client will clarify the response sufficiently to be awarded the full two points.

(4) Q1 - The client's response to a one or two point question is wrong and worth zero points. However, further questioning is indicated because there is evidence to suggest that if requestioned the client may clarify his response enough to be awarded one point.

(5) H - The client has given the homonym of the correct response and is to be awarded zero points. However, the client is to be questioned further in order to be allowed to give the correct response and an opportunity to be awarded with the appropriate points.

(6) Z - The client's response to a one or two point question is wrong and worth zero points. Further questioning on the same question is to be terminated because no evidence exists to suggest that the client could improve the response.

Objective 66, (p. 210) part C suggests the following classification.

(7) DZ - The client's response to a one or two point question is wrong because the client has negated the correct response (contradicted the correct response) and is to be awarded zero points. Further questioning is to be terminated.

It is important to realize that those responses, or parts of respon , which are not classifiable in one of the seven specified categories are to be ignored. This is outlined in objective 70 (p. 210).

B. ANSWER HIERARCHY

Objectives 65 through to 69 (Appendix B pages 209 to 210) suggest the following:

(1) that the client's response can be a multiple response (a response which contains more than one classifiable response),

(2) that the classifiable responses are ordered into a hierarchy in which one classification response takes precedence over other classifications in the hierarchy,

(3) that the order of precedence using the designations of the last page are:

(a) C2

(b) Q2

(c) C1

(d) Q1

(e) H

(f) Z

TABLE 1

DECISION TABLE 1; SINGLE RESPONSE EVALUATION

CLASS	
C2	A = 6 S = 2
Q2	A = 2 S = 1
C1	A = 6 S = 1
Q1	A = 2 S = 0
H	A = 4 S = 0
Z	A = 5 S = 0
DZ	A = 5 S = 0

TABLE 2

DECISION TABLE 2; MULTIPLE RESPONSE;

LAST RESPONSE > PREVIOUS RESPONSES

C2	A = 6 S 2						
Q2	6 2	2 1					
C1	6 2	2 1	6 1				
Q1	6 2	2 1	6 1	2 0			
H	3 0	3 0	3 0	3 0	4 0		
Z	6 2	2 1	6 1	2 0	5 0	5 0	
DZ	6 2	2 1	6 1	2 0	5 0	5 0	5 0
	C2	Q2	C1	Q1	H	Z	DZ

TABLE 3

DECISION TABLE 3; MULTIPLE RESPONSE;

LAST RESPONSE < PREVIOUS RESPONSES

DZ	A = 5 S = 0						
Z	5 0	5 0					
H	5 0	5 0	4 0				
Q1	5 0	2 1	3 0	2 0			
C1	5 0	6 1	3 0	6 1	6 1		
Q2	5 0	2 1	3 0	2 1	2 1	2 1	
C2	5 0	6 2	3 0	6 2	6 2	6 2	6 2
	DZ	Z	H	Q1	C1	Q2	C2

(g) DZ

with the C2 response taking precedence over all responses, and the DZ being the lowest response in the hierarchy,

(4) that the last classifiable response in a multiple response reverses the order of the hierarchy, if this response is lower in classification than all other responses in the multiple response, and

(5) if the client is given the opportunity to be requestioned, and his new response is lower in score value than the previous, he should be given the original score award.

(1) DECISION TABLES

Objectives 59 through to 64 (pp. 208 to 209) suggest the need for a set of response categories, while objectives 65 through to 69 (pp. 209 to 210) suggest the conditions of a response hierarchy. Consideration of objectives 50 through to 69 together suggests that evaluation of a client's response can be done by classifying its parts, then consulting one of the three decision tables, which give the following information:

(1) what points are to be awarded to the client for his response, and

(2) what action must the Computer-Examiner take next in relation to the question asked.

The criteria for deciding which decision table to select was determined as follows:

(1) DECISION TABLE 1 - this table is consulted if the client's response is a single response.

(2) DECISION TABLE 2 - this table is consulted if the client's response is a multiple response and the last classifiable response is

greater or equal in classification to any other response in the multiple response.

(3) DECISION TABLE 3 - this table is consulted if the client's response is a multiple response and the last classifiable response is less in classification to any other response in the multiple response.

The score points for each cell in the decision tables are specified by objectives 78, 80, 85, 87 and 88 (pp. 210 to 213).

(2) DECISION TABLE SYMBOLS

The various symbols used in the decision tables are defined as follows:

(1) C2, Q2, C1, Q1, H, Z and DZ - refer to the possible classification categories for the client's response as previously outlined.

(2) A - refers to the next action the Computer-Examiner is required to take as a result of evaluation of the client's response. There are six possible values A may have and these are:

(i) 6 - terminate further questioning, award score points (possible values of either 1 or 2), and proceed to the next question. It is important to realize that this level of action is recognition of the fact that the client has answered the question correctly and Computer-Examiner must decrement the consecutive question failure count for the subtest to zero.

(ii) 5 - terminate further questioning, award the score point (only possible value is 0), and proceed to the next question. It is important to realize that this level of action is recognition of the fact that the client has incorrectly answered the question and the Computer-Examiner must increment the consecutive question failure count

for the subtest by one.

(iii) 4 - requestion the client by asking would you give me THE OTHER meaning please? This level of action indicates the response was a homonym.

(iv) 3 - requestion the client by asking Which answer is it? This level of action indicates the response was a multiple response of differences. The client must now decide which single response in the multiple response is the correct response.

(v) 2 - requestion the client by asking Please tell me more about your answer. This level of action indicates the response was not totally correct, but there is sufficient evidence to indicate that it could be improved if the client is given another chance.

(vi) 1 - repeat the question. This level of action indicates that the client requested that the question be repeated.

(3) S - refers to the score point the client is to receive for the evaluated response. The possible values are:

(A) 0 or 1 for a one point question, or

(B) 0, 1 or 2 for a two point question.

(3) HOW TO USE THE DECISION TABLES

(a) DECISION TABLE 1

When the client's response has been evaluated to a single classifiable response, Decision Table 1 is entered on the left hand side of the table at that level of classification. The table values in the right hand cell next to this entry is used. For example, if the client's response evaluates to a single C1 category, the Computer-

Examiner reads the right hand cell values to be A = 6, and S = 1. This means that the Computer-Examiner is to terminate further questioning, go on to the next question, decrement the consecutive question failure count to zero, and award the client one point.

(b) DECISION TABLE 2

When the client's response has been evaluated to be a multiple response, and the last classifiable response is greater than or equal to, in terms of the response classification hierarchy, than any other response in the multiple response, Decision Table 2 is entered on the bottom of the table at the level of classification of the last response. The second table entry is made on the left hand side of the table with the highest level of classification of response available in the multiple response, disregarding the last response. The intersection of these two entries is the cell which the Computer-Examiner reads. For example, assume that a client's response has been evaluated and found to contain three classifiable responses in the following order:

C1, Z, C2.

The highest classifiable response is C2 and it is the last response. This value is used for the bottom table entry. The next highest response between the remaining C1 and C2 responses is the C1 response. This is used for the left hand side table entry. The intersecting cell reads A = 6 and S = 2. This means that the Computer-Examiner is to terminate further questioning, go on to the next question, decrement the consecutive question failure count to zero, and award two points.

(c) DECISION TABLE 3

When the client's response has been evaluated to be a multiple response, and the last classifiable response is less than, in terms of the response classification hierarchy, any other response in the multiple response, Decision Table 3 is entered at the bottom of the table at the level of classification of the last response. The second table entry is made on the left hand side with the highest level of classification of response available in the multiple response, disregarding the last response. The intersection of these two entries is the cell which the Computer-Examiner reads. For example, assume that a client's response has been evaluated and found to contain three classifiable responses in the following order:

C1 C2 DZ

The last response classification was the lowest, as DZ. This value is used at the bottom table entry. The highest response which can be evaluated among the remaining C1 and C2 responses is the C2 response. This is used for the left hand side table entry. The intersecting cell of these two entries reads A = 5 and S = 0. This means the Computer-Examiner is to terminate questioning, go to the next question, increment the consecutive question failure count by one, and award zero points.

(C) ANSWER ANALYSIS

Objective 57 (Appendix B, p. 208) states the human examiner is required to know the meanings of the words and concepts involved. The computer's ability to know can at best be only pseudo-knowledge because it is limited to functions of recognition and classification in

analysis of input. These functions can make the computer look as though it knows what has been said to it. This rationalization was accepted in development of the conceptual model. The previous two sections outlined conceptually the hierarchy and its seven possible categories. This section deals with conceptualization of how the Computer-Examiner is able to analyze the client's response and identify what parts belong to what category in the response hierarchy.

To each question in the Verbal Section of the WISC an expectation of what answer the client is most likely to give as a response could be established. That is to say, certain answers to a question have a much greater degree of probability of being elicited in the client's response than others. This concept stems from the statistical expectation that over a large number of clients, all asked the same question, a certain number of common answers will appear in the responses. The most frequent answer will have the highest probability of occurence, while the least frequent the lowest probability of occurence. In other words, when a question is directed towards a client in which the client is required to give a response (and it is assumed that the client will attempt to respond as honestly and sincerely as possible) it can be expected that he will likely give one of the more common answers to the question. Statistically, over a large number of persons this will be the expected case. In fact, the scoring schema used to derive the I.Q. scores is based upon the concept of statistical normality.

The concept of the common answer in the client's response is used in the WISC Manual and Massey Supplement in that typical (common)

answers have been cited as examples for each of the possible categories of answers for each question. However, no degrees of probability are assigned to these answers beyond the recognition that they are typical or common. It is also recognized that these answers are not exclusively the only possible ones.

It is important to summarize exactly what has been indicated to this point in order to clearly state what is being included in the conceptual design of the model. The concept of the probability of frequent occurrences of common answers in a client's response to a question can be considered in two parts:

(1) the client's response to a directed question from the examiner will likely contain an answer which has frequently occurred as an answer before, and

(2) it would be possible to statistically determine the frequency of occurrence of a common answer and to assign it a degree of probability of occurrence in reference to the question asked.

The second part of this concept cannot be considered in the development of the conceptual model for two reasons. First, the answers cited as examples in the manual and supplement do not have the quantitative aspect of the degree of probability of their evocation, and secondly, to attempt to assign such probabilities would be beyond the capabilities of the study. However, the first part forms the base for the conceptual design of the response analysis section of the Computer Examiner.

For the purpose of expediting development of the computer model, the manual and supplement were used as the only sources of answers for the response analysis section. These are by no means the only possible

answers, there are many more for each category of each question. The task of compiling a complete, or even partially complete, list of all possible answers is far beyond the scope of this study and is the primary reason for limiting the answer recognition capacity of the model to the two specified sources. In addition, this limitation clearly defines the boundaries of the model's capacity to recognize answers, thus permitting evaluation of the question How successful was the Computer-Examiner in recognizing the defined answer set? The question How common were the answers from the manual and supplement in relation to the answers received from the clients, can also be considered.

(1) EXAMPLE

Description of the response analysis section can be more readily understood through use of an example. A General Information question 23 has been selected for this purpose and will be referred to throughout the sections to follow. This question asks What does turpentine come from? The manual indicates that "Its source, pine or fir tree, must be indicated (p. 62)". The answer set of the supplement lists the following categories and answers:

C1: PINE tree
FIR tree
Sap of EVERGREEN tree

Q1: From SAP
From RESIN
From ROSIN
TREE
They BOIL it out of WOOD

Z: ANY TREE

FIGURE 2: LOGICAL FLOW CHART

ANSWER HIERARCHY SEQUENCE

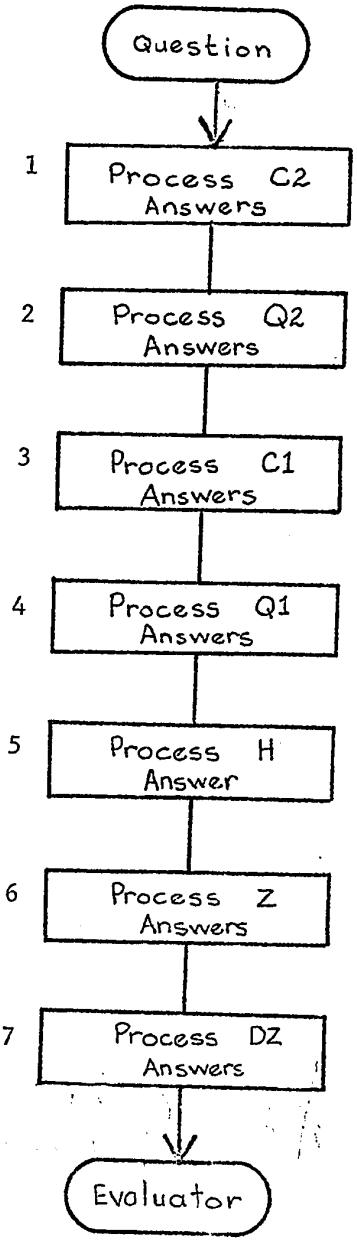
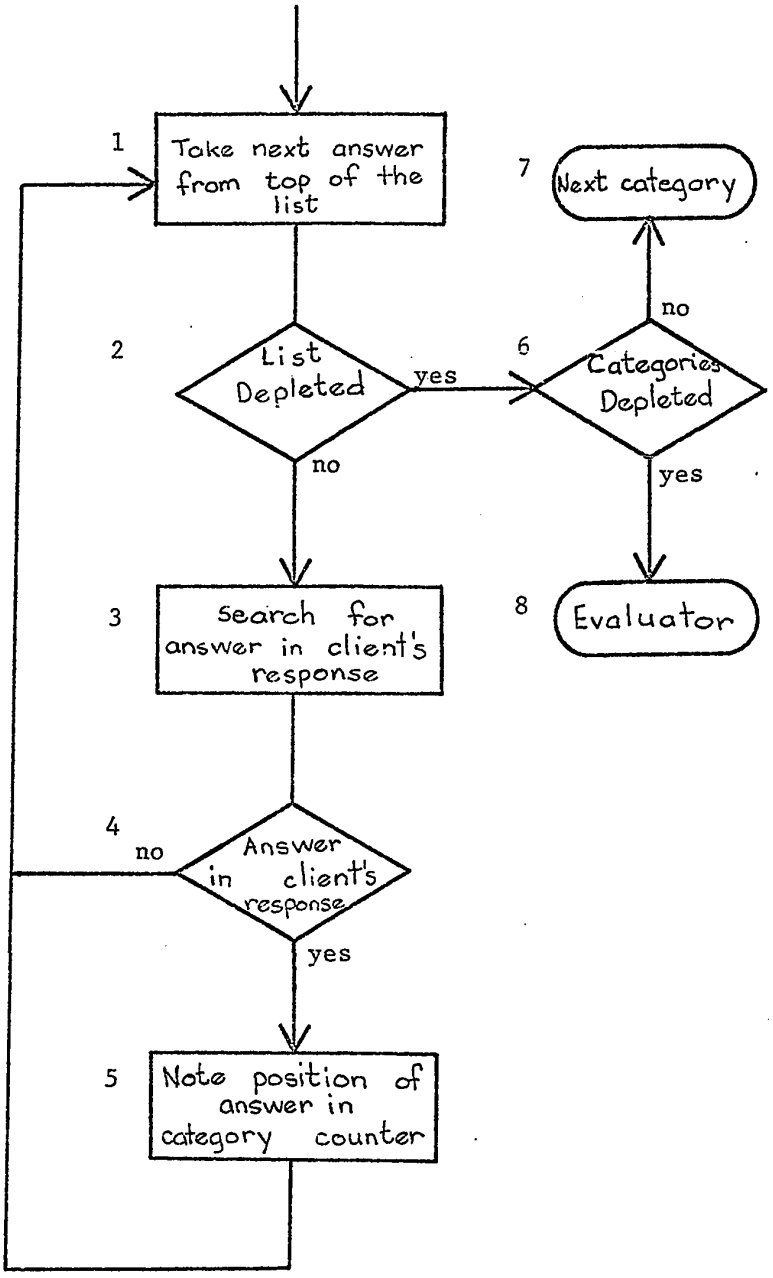


FIGURE 3: LOGICAL FLOW CHART

ANSWER ANALYSIS SEQUENCE



(2) PROCESSING CATEGORIES

It was previously suggested that to know for the computer is to recognize. This function of recognition was conceptually developed as follows. First, each question module (within module 8, Figure 11 (p. 103) and expanded in Figures 13A to 13F) would be provided with a source list of answers for that particular question as abstracted from the manual and supplement (see the example of General Information Question 23 (p. 76). A modified version of this list (the modification will be explained later) would be programmed into the question module. Upon administration of the question and subsequent receipt of the client's response, the computer would then start to search for the answers of the source list in the client's response, category-by-category, starting with the highest category present in the source list (for General Information Question 23 this would be category C1) and successively processing answers to the lowest category (for General Information Question 23 this would be category Z). Figure 2 shows the logical flow chart sequence for the processing of categories. The dotted lines outlining modules 1 to 7 serve to indicate the category may or may not be present for the question, depending on whether or not sources answers are available for the category from either the manual or the supplement. For General Information Question 23 this means that categories C1, Q1 and Z will receive processing in question module 23, while categores C2, Q2, H and DZ will not.

(3) PROCESSING SOURCE ANSWERS

Secondly, the processing of answers within each category would proceed as outlined by the logical flow chart of Figure 3. The computer

would select the first source answer in the category list (module 1) and attempt to locate it in the client's response (module 3). If it was not in the response the computer would iterate the process (module 4) and select the next source answer from the list (module 1). If the answer had been found in the client's response, the computer would locate its position. This is done by counting how many alphanumeric characters (punctutation, spaces, and special characters included) the first character of the answer was from the beginning character of the response, and then placing the numerical count of this position in the counter assigned to that category (module 5). At the conclusion of this action the computer would recycle to process the next source answer on the list (module 1).

Upon deletion of all source answers in the category list (module 2) the computer would proceed to the next lower category, provided that a category were available to be processed (modules 6 and 7). In lieu of this, it would proceed to the evaluator (module 8).

Using General Information Question 23 (p. 76), consider the following example of the process just described. Assume that the client had responded to the question with the response:

Client's Response	=	Turpentine comes from resin.
Character Position in Response Buffer	=	1 12 18

At initiation of the answer analysis sequence the first source answer in category C1 - pine tree - would be selected (module 1), then the computer would attempt to locate it in the client's response (module 3). Upon finding no match (module 4), the next source answer in the list - fir tree - would be selected (module 1). Processing would continue through the C1 category until all three source answers had been used, then

would continue to the next category Q1 via modules 2, 6 and 7. The same search for a match would continue until the source answer from resin (module 1) had been selected. With this answer the computer is able to make a match to the client's response (module 3). The match is recognized (module 4) and its position is noted to start at the 18th alphanumeric character (this included the spaces), at the letter f. The numerical value of this position, 18, is loaded into the counter used to note Q1 category matches (module 5). Processing then continues for the rest of the Q1 category, then follows the Z category. When the last source answer in this category has been processed the computer exits to the evaluator section via modules 2, 6 and 8.

(4) GENERATION OF KEYWORDS

The phrase source answer was used in the preceding description of the response analysis process to designate those answers used by the computer model as references in its search for answers in the client's responses. Source answers were not, however, used verbatim as the reference answers. They were modified to form keywords. In a preceding discussion on the selection of the source answers such modifications were indicated. They are modified in order to maximize the possibility of correctly categorizing answers in the response.

The source answers were modified in two ways. First, the number of words in each source answer was reduced to the absolute minimum needed in order to correctly identify input. Consider the source list for General Information Question 23 (p. 76) for example. The only words needed in each answer of this example list to indicate that the client had a classifiable answer would be those words which have

been deliberately capitalized in each answer. The remaining words are superfluous. According to the criteria of the manual, for a client to have his response categorized as C1, he needs only to indicate the source of turpentine, and obviously this would be done if one of the words pine, fir or evergreen were included in his answer. For the C1 category the word "tree" or the phrase "sap of" are not necessary, and in fact, limit the variability of the responses which can be correctly categorized as C1. Consider the following set of sample responses as an example of how decreasing the number of words in each source answer to a minimum increases the range of answers which can be accepted from the client by the computer model:

(1) I think it comes from pines.

(2) It comes from the fir tree.

(3) Evergreens are used to make it.

(4) Turpentine comes from pine resin.

(5) Pines!

If the source answer list were applied as the classification criteria the sample responses would be classified:

(1) no recognition

(2) C1

(3) no recognition

(4) Q1

(5) no recognition

If only the capitalized words in the example source list were applied as the classification criteria all of the sample responses would be classified as C1.

Consider, from the Q1 category, the following more complex

answer:

They boil it out of wood.

If this answer were required to be in the response verbatim, of the following responses only the first would be classified as Q1; the remainder would not be recognized by the computer model at all.

(1) They boil it out of wood.

(2) Turpentine is boiled out of wood.

(3) It is boiled out of wood.

(4) I think you boil wood.

(5) You boil wood.

If only the words boil and wood were used as the classification criteria each of the responses would be classified as Q1.

The procedure for reducing all source answers to the minimum number of words needed to make identification of answers does not contravene the evaluation criteria of the manual or supplement. In fact, objective 92 implies that to do so would be correct. It states: "The general rule for scoring is that any recognized meaning of the word is acceptable, elegance of language and precision being disregarded (Appendix B, p. 213)". The actual task of reducing each source answer was done by the modeler, using objectives 80 to 84, 87 to 89, and 91 to 94 as guidelines (pages 211 to 215). Objective 92 is the most important and explicit of these objectives.

(5) GENERATION OF SKELETON ANSWERS

Having the client type out responses on the Instructional Keyboard is the most major departure made in the computer administration of the test. In addition there is no correspondence to the possibility

of spelling errors in the WISC test because in the normal way of administering the test the client gives responses verbally. In order to increase the probability of accepting the answers in light of errors in spelling the keywords selected from the source answers were modified. The following two criteria were used to accomplish this task:

(1) the character length of the word was reduced to between three (the minimum) and six (the maximum) characters, and

(2) in light of the first criteria, as many vowels were eliminated from the word as possible.

Consider once again General Information Question 23 (p. 76) for an example of how keywords were modified. In the following list column one represents the source answers, column two the keywords, and column three the final skeleton answers ready for programming into the computer model.

	Source Answer	Keywords	Skeleton
C1:	pine tree	pine	pin
	fir tree	fir	fir
	sap of evergreen tree	evergreen	vrgrn
Q1:	from sap	sap	sap
	from resin	resin	rsn
	from rosin	rosin	rsn
	tree	tree	tre
	they boil it out of wood	boil wood	bol wod
Z:	any tree	any tree	any tre

The rationale for applying the character reduction criteria to the keywords was to attempt to increase the Computer-Examiner's ability to identify the keywords in a response in light of spelling or typing errors. Vowels were eliminated mainly because they are the most frequently misspelled parts of words. Consonants, in general, provide more information about the identity of words. In order not to completely obscure the identity of a keyword, the number of characters in a keyword was not allowed to fall below three, a criteria of purely arbitrary choice at this stage of development. It is obvious that the identity of a keyword would become more obscure as characters are taken out of it. In this case, it becomes more possible to mistake other words in the client's response for the keyword. The entire procedure of modifying keywords was very crude and was largely performed in a highly subjective and heuristic manner.

(5) COMMUNICATION BETWEEN CLIENT AND COMPUTER-EXAMINER

Two separate and different paths of communication exist between the client and the Computer-Examiner. First, all communication from the Computer-Examiner to the client was via prerecorded audio messages played through earphones from an audio cartridge loaded on the 1505 Audio Tape Unit, or via graphic messages displayed on the 1510 Instructional Display. Second, the client communicated his or her responses to the Computer-Examiner's questions by typing them on the 1510 Instructional Display Keyboard or by using the light pen.

(A) COMMUNICATION FROM COMPUTER-EXAMINER TO CLIENT

All of the audio messages to be given to the client were grouped

into the following categories:

(1) those messages used to introduce the Computer-Examiner to the client and demonstrate how the keyboard is to be operated,

(2) those messages used to show the client how to type responses to the Computer-Examiner's questions,

(3) those messages which explain to the client the use of the following control words:

(A) REPEAT.

(B) REST.

(C) STOP.

(D) I DON'T KNOW. and

(E) TEACHER.

(4) those messages associated with obtaining the AGE and SEX from the client,

(5) those messages used with the control word subroutines:

(A) REST.

(B) STOP. AND

(C) TEACHER.

(6) those messages associated with the TIME subroutine, which attempts to find out why a client is taking longer than two minutes to respond to a question,

(7) those messages required to instruct the client how questions in a subtest are to be answered, and

(8) those messages associated with each individual question.

(1) MESSAGES FOR QUESTIONING

Efforts to standardize the WISC have led to the need for very strict control over information the examiner is able to give to the client during testing. Rigid formats are specified for the various messages which can be given to the client. These are expressed in objectives 3, 4, 5, 8 through to 18, 26, 28, 31, 33, 34 and 41 (Appendix B, pp. 202 to 205) which suggest that there are five basic formats which can be used during administration of the test These were adopted as part of the conceptual model. They are as follows:

(1) THE QUESTION - each question is to be given as stated in the manual with no variation, except for Arithmetic questions 1, 2 and 3 which have to be presented differently.

(2) THE EXPLANATION INQUIRY - variations used were:

(A) Would you explain your answer more fully please.

(B) Tell me more about your answer.

(C) Please explain your answer.

(3) THE HOMONYM INQUIRY - only variation was:

(A) Would you give the other meaning please.

(4) THE DIFFERENCE INQUIRY - only variation was:

(A) Which answer is it?

(5) THE ENCOURAGEMENT RESPONSE - variations used were:

(A) Would you do the next question please.

(B) Let's go on to the next one.

(C) I wouldn't worry about missing a few of these questions. Some are a bit hard. Shall we try the next one?

(D) Alright! Let us go on to the next question.

(E) Okay! Let's go to the next question.

(F) That question was a little hard, wasn't it? Try the next one.

(G) That's okay! Try the next question.

(H) That question was a bit tricky, let's try the next one.

(I) Don't worry about missing a few of these questions. Some of them are really meant for older boys and girls.

For each of the 118 questions of the Verbal Section (there are actually 116, but two additional questions are required for the Comprehension questions 2 and 4, in order to handle the question of sex) a message group was constructed by taking one of the possible variations out of each of the five formats. The variations in each form were alternately selected so that each variation received equal use.

As an example the message group for Information question 1 was:

(1) How many ears do you have?
(2) Would you give me THE OTHER meaning please.
(3) Please explain your answer.
(4) Which answer is it?
(5) Would you do the next question please.

The structure for Information question 15 was:

(1) Why does oil float on water?
(2) Would you give me THE OTHER meaning please.
(3) Please tell me more about your answer.
(4) Which answer is it?
(5) That question was a little hard, wasn't it? Try the next one.

Note, the message which will be given to the client from a message group will depend upon how the client's response is categorized. For example, if a client's response was categorized as a homonym for Information question 1, the message would be Would you give me THE OTHER meaning please? If the response was recognized as I DON'T KNOW, for Information question 15, then the message would be That question was a little hard, wasn't it? Try the next one.

FIGURE 4 GRAPHIC MESSAGE FORMAT NO. 1

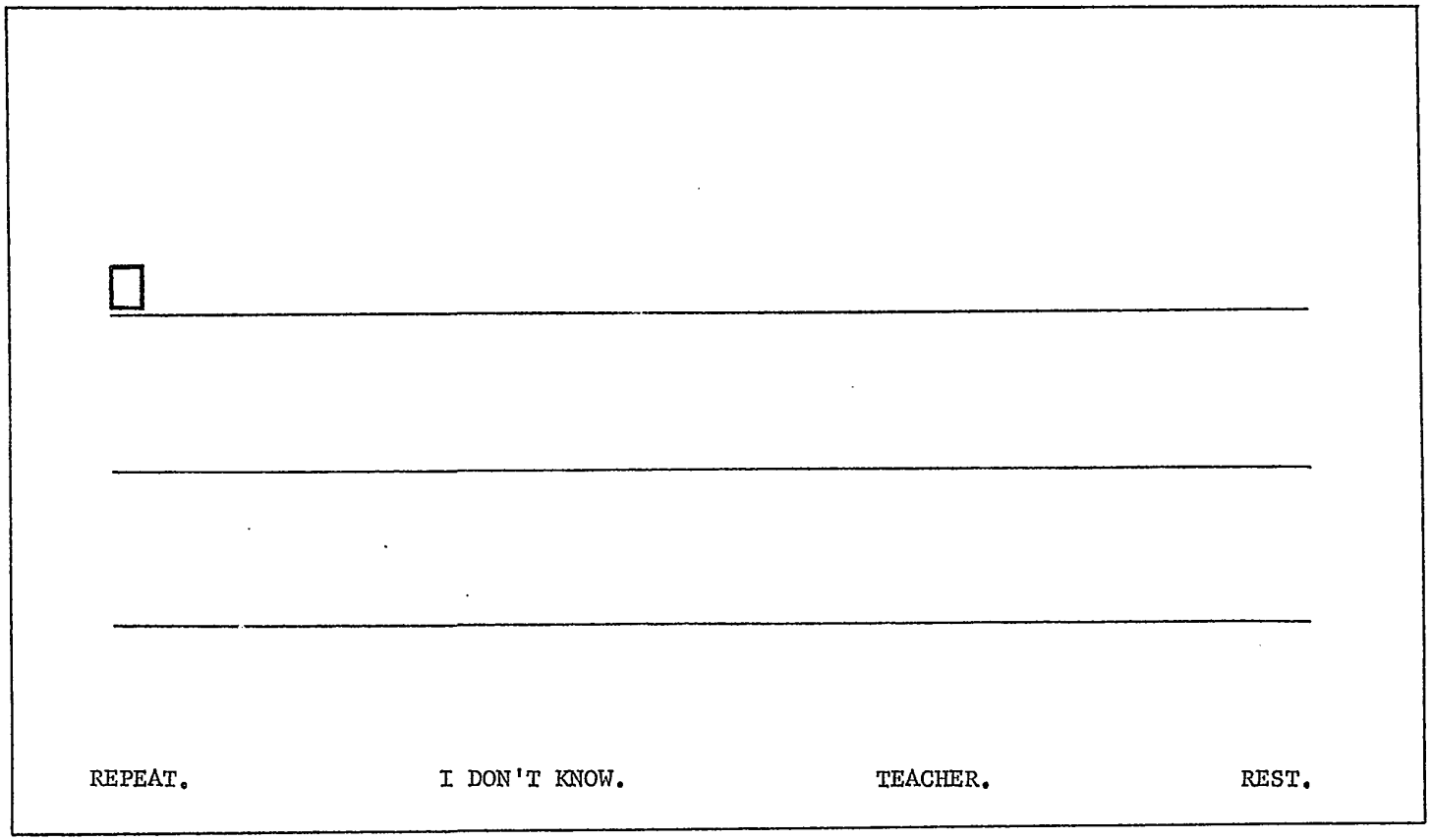


FIGURE 5 GRAPHIC MESSAGE FORMAT NO. 2

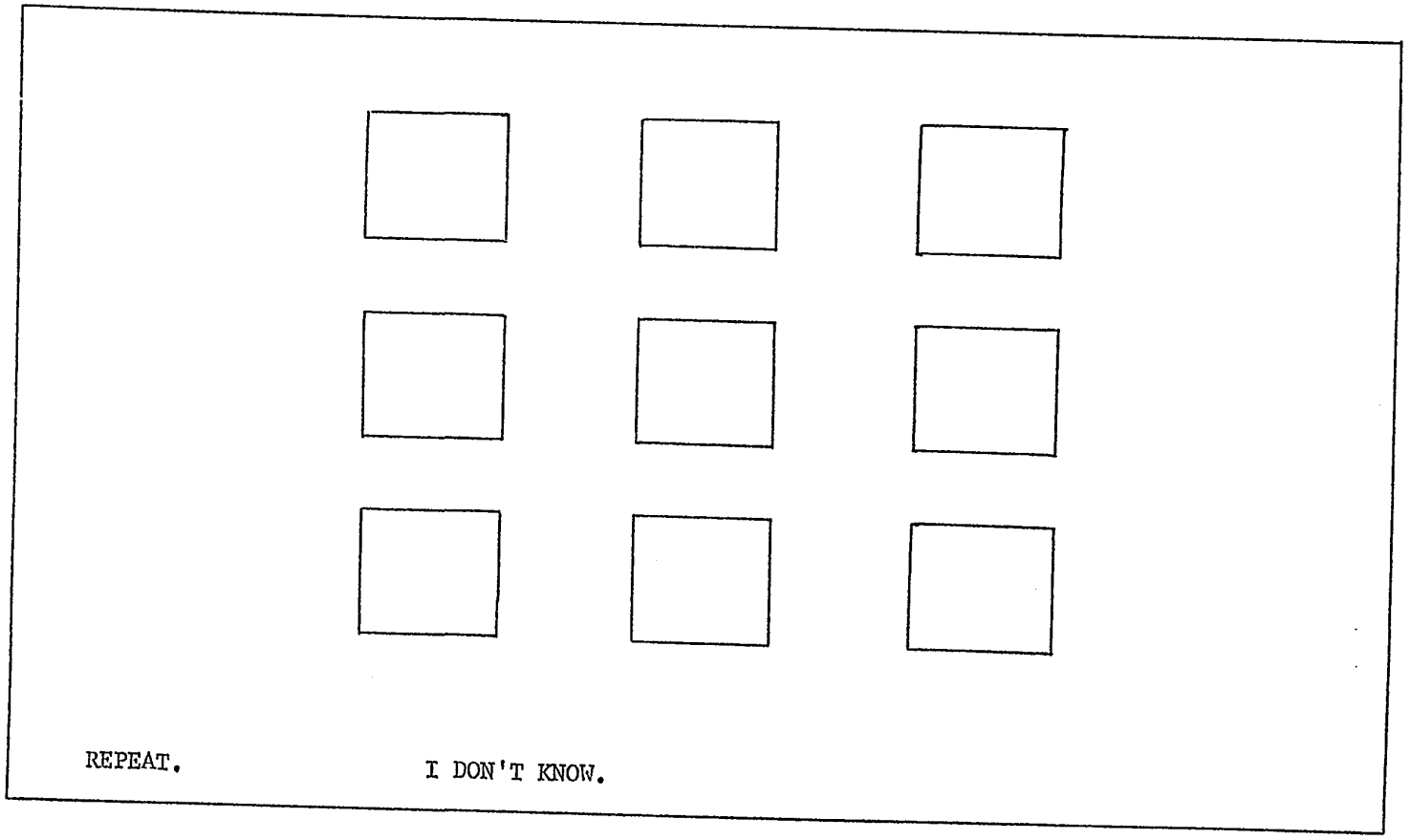


FIGURE 6 GRAPHIC MESSAGE FORMAT NO. 3

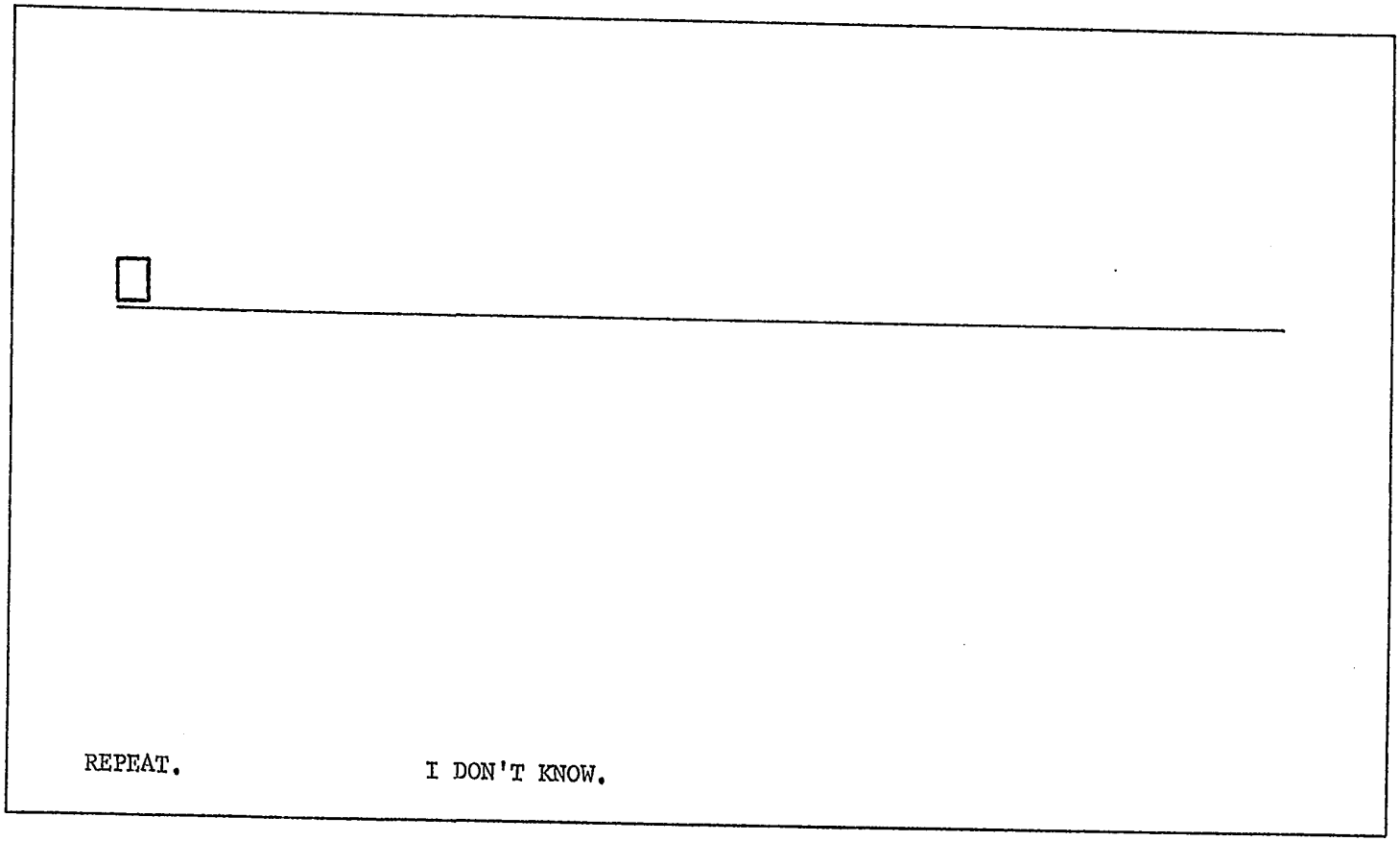
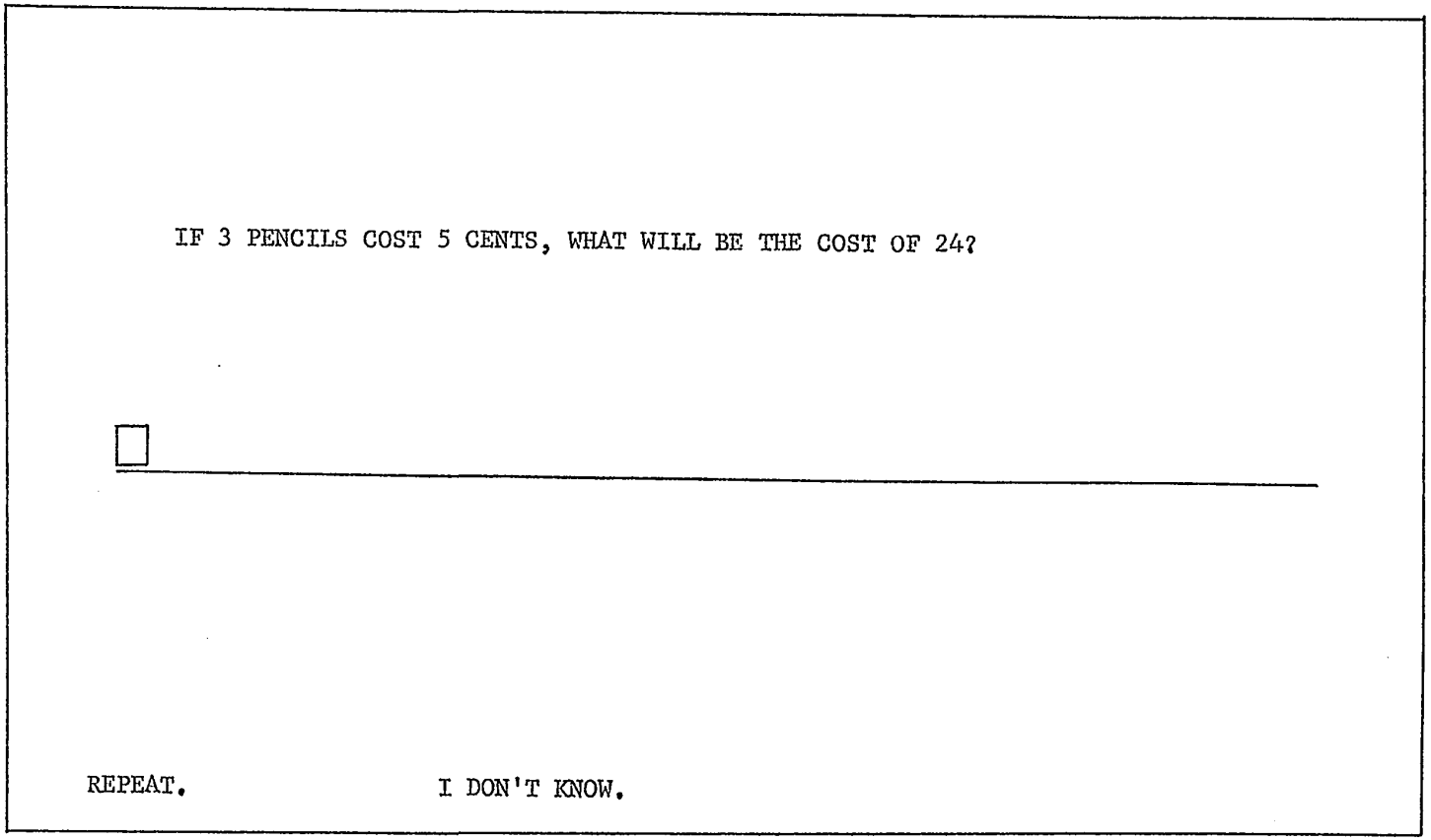


FIGURE 7 GRAPHIC MESSAGE FORMAT NO. 4



(2) GRAPHIC MESSAGES

A number of graphic messages were used in the conceptual model. The first (Figure 4, p. 89) was used for all the questions of the General Information, General Comprehension, Similarities, and Vocabulary subtests. The second format (Figure 5, p. 90) was used for Arithmetic questions 1, 2 and 3 while the third format (Figure 6, p. 91) was used for Arithmetic questions 4 through 13. The fourth format (Figure 7, p. 92) was used for Arithmetic questions 14, 15 and 16. The remainder of the graphic messages were used to introduce the Computer-Examiner to the client, familiarize him with the keyboard, the keywords and obtain age and sex information. The small square symbol in each of these figures is called a cursor and it indicates the position of the next character to be typed. When a client types a character, the typed character replaces the cursor in its position and the cursor moves one character position to the right, or to the left most character position of the next line.

(B) COMMUNICATION FROM CLIENT TO COMPUTER-EXAMINER

The client's responses are input to the Computer-Examiner by either the keyboard or the light pen of the 1510 Instructional Display. Except for those sections of the conceptual model which ask the client for information on sex and age (Figure 12G, p. 111) and answer the Arithmetic questions 1, 2 and 3 all responses are input through the keyboard.

(1) RESPONSE ACCEPTANCE VIA KEYBOARD

The maximum character input of any enter and process instruction is 100 characters. This is the length of the student response buffer (item 27, list 6, Appendix A, p. 200) the client would have to type,

FIGURE 8: LOGICAL FLOW CHART

KEYBOARD RESPONSE ACCEPTANCE ROUTINE

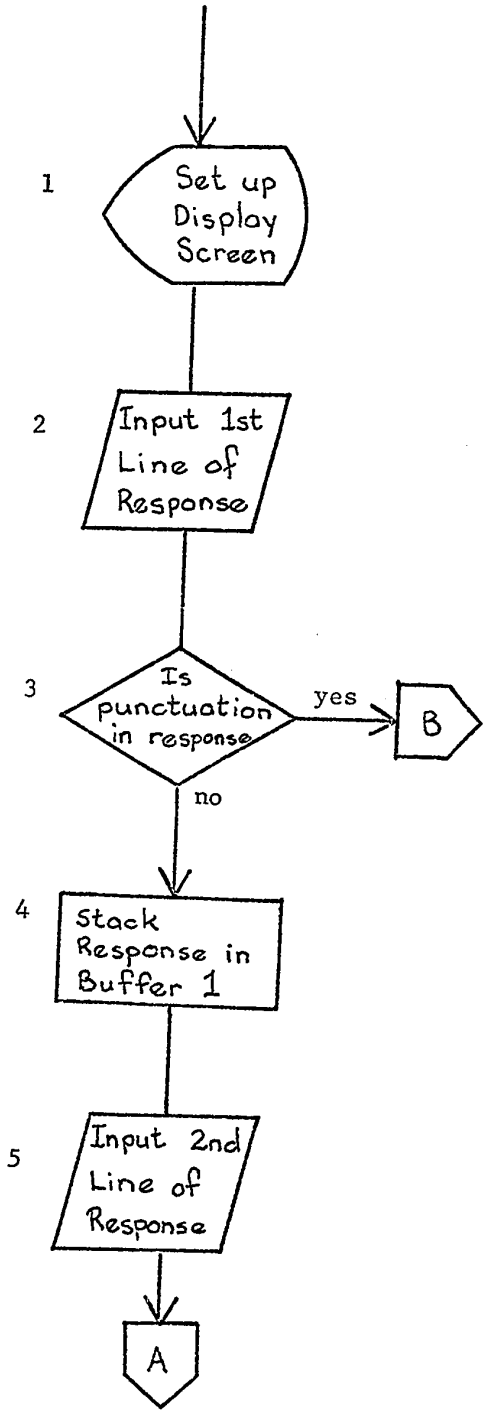
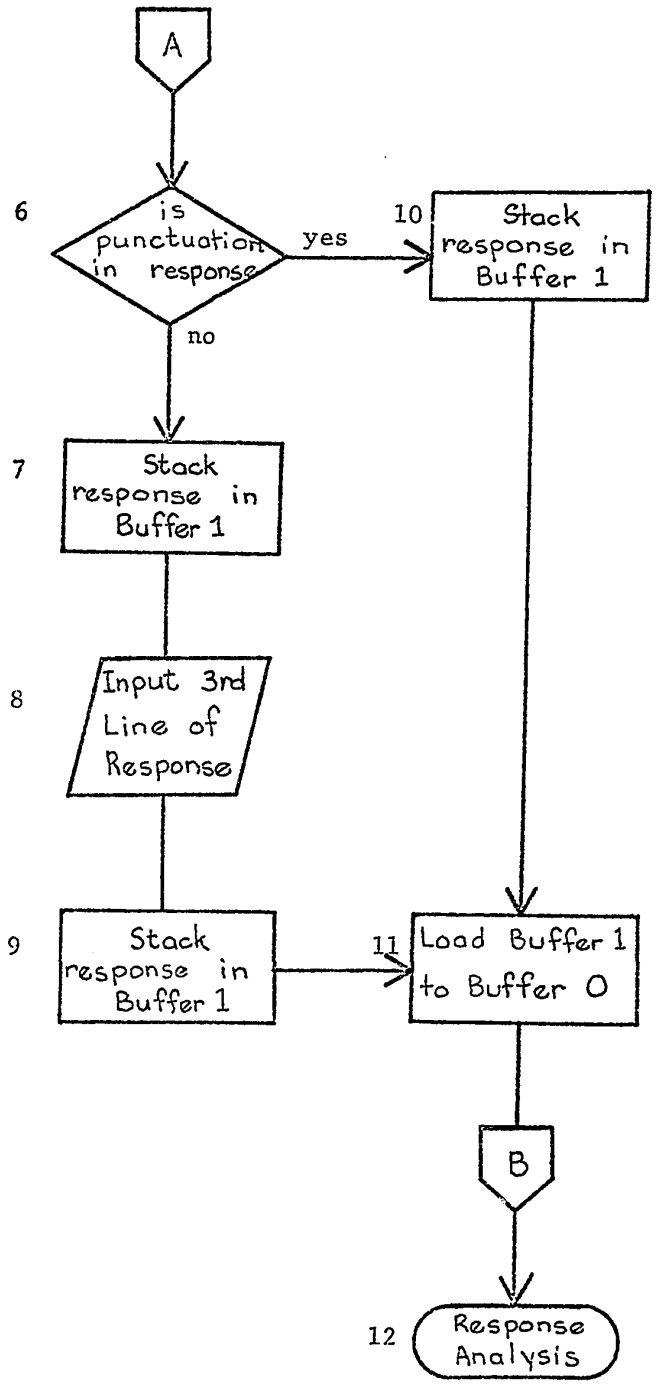


FIGURE 9: LOGICAL FLOW CHART

KEYBOARD RESPONSE ACCEPTANCE ROUTINE (CONTINUED)



at least twice, the combination of one return key and one index key in order to return the cursor to the right hand side of the display screen and advance the cursor to the next line. These keys would be typed at the end of the fortieth and eightieth characters. If this was not done the cursor would disappear from the display screen and subsequent typing by the client would not be seen.

Learning the function of and use of the return and index keys together within the administration period of the test would be a difficult task for many clients. For this reason it was decided to eliminate the use of these keys altogether and to adopt the following procedure as an alternate.

The character input for each line of response on the display screen would be limited to 33 characters and a total of three lines of response would be accepted. The format for the graphic display would be as shown in Figure 4 (p. 89). Thus the maximum character input would be limited to 99 characters or less.

In the process of accepting a response the display screen would be set up as per Figure 4 (module 1, Figure 8, p. 94). The cursor and system would then be set up to input the client's response (module 2). Upon condition of the client typing either 33 characters or typing the enter key the Computer-Examiner would proceed to analyze the response for either a period, question mark, or exclamation point in the last character position (module 3). If a punctuation mark were found, the Computer-Examiner would assume the client had finished typing the response and would proceed to the response analysis section (module 12).

If no punctuation mark was present in the response, the Computer-Examiner would stack the client's response in the first 33 characters of

buffer 1 (module 4) and proceed to set up the cursor and system for the second line of the client's response (module 5). Upon the condition of the client either entering the second line of response or typing 33 more characters, the Computer-Examiner would then proceed to analyze the response for a punctuation mark in the last character position (module 6). If a punctuation mark was present, the Computer-Examiner would stack the response in buffer 1 starting at position 34 (module 10), then load buffer 1 into the response buffer (module 11), and proceed to the response analysis section (module 12).

If no punctuation mark was present in the response, the Computer-Examiner would stack the client's response in buffer 1 starting at position 34 (module 7) and proceed to set up the cursor and system for the third line of the client's response (module 8). Upon the condition of the client entering the third line of response or typing 33 more characters, the Computer-Examiner would proceed to stack the response in buffer 1 starting at position 67 (module 9), then load buffer 1 into the response buffer (module 11), and proceed to the response analysis section (module 12).

(2) CLIENT TYPING PROCEDURE

In light of the response acceptance procedure adopted for the conceptual model the client was required to follow these steps in typing a response:

(1) type the response

(2) close it off with a punctuation mark, and

(3) depress the enter key.

To teach the client the steps, part of the client preparation (module 2, Figure 10, p. 100) was used for explanation and practice (module 6,

Figure 11B, p. 102). The detail of this step is specified by modules 38 and 39 (Figure 12F, p. 110). Modules 40 to 43 (Figure 12F) were used to check and reinforce this instruction.

6. SCOREKEEPING

The basic calculation which must be performed on subtest raw scores in order to derive a Verbal I.Q. score are outlined in objectives 73 to 77 (Appendix B, p. 211). In terms of the conceptual model, these objectives can be performed as a group once the Computer-Examiner has administered the Verbal Section. At this time the client can be released from the test situation. The procedure for calculating the I.Q. score is indicated by module 4, in Figure 10 (p. 100), and in greater detail as modules 19, 20, 21 and 22 in Figure 11D (p. 104).

It must be noted at this point that the Coursewriter II language does not have the normal table indexing facilities that are, for example, available to the Fortran IV language. The use of such a facility would extremely simplify the programming of the large number of Scaled Score Equivalents for Raw Score Tables and the Conversion of Verbal Scale Score into I.Q. Table of the WISC manual. Because of this, the problem of table look up and retrieval was approached by using the binary search technique.

7. RESPONSE RECORDING

Objective 72 (Appendix B, p. 211) specifies that the examiner must record each of the client's responses in full, while objective 6 specified that this must be done without hesitation. The ability to do these tasks is an inherent characteristic of the 1500 Instructional

System.

8. MULTIPLE TESTING PERIODS

Objective 7 (p. 202) specifies the need for more than one testing period if the human supervisor detects that the client needs it. This objective has been included in the conceptual model by creating a sub-routine which was accessed by typing the code word STOP. The human supervisor was responsible for initiating this action. At the beginning of the next testing session the Computer-Examiner would review the typing instructions, then proceed with the question where questioning was stopped. This latter task was accomplished mainly through use of the 1500 System's student restart facility.

9. MISCELLANEOUS OBJECTIVES

In order to account for all the objectives of Appendix B, the following objectives must be considered in relation to development of the conceptual model. Objective 71 (p. 211), which deals with the marking of the letter Q beside questionable responses, was not necessary in the model. Similarly objectives 79, 81, 86, 90 and 95 (pp. 211 to 215), which specify the maximum raw score counts for each subtest, are of no importance to the development of the conceptual model, mainly because of objectives 78, 80, 85, 87 and 88 (pp. 211 to 213) which provide a more precise description of the raw score points which are awarded to questions in each subtest.

Objective 32 (p. 205) which specifies that it is good practice to repeat the question if no response is obtained after ten or fifteen seconds, but no alteration or abbreviation is permitted was considered part of

FIGURE 10: LOGICAL FLOW CHART

COMPUTER-EXAMINER

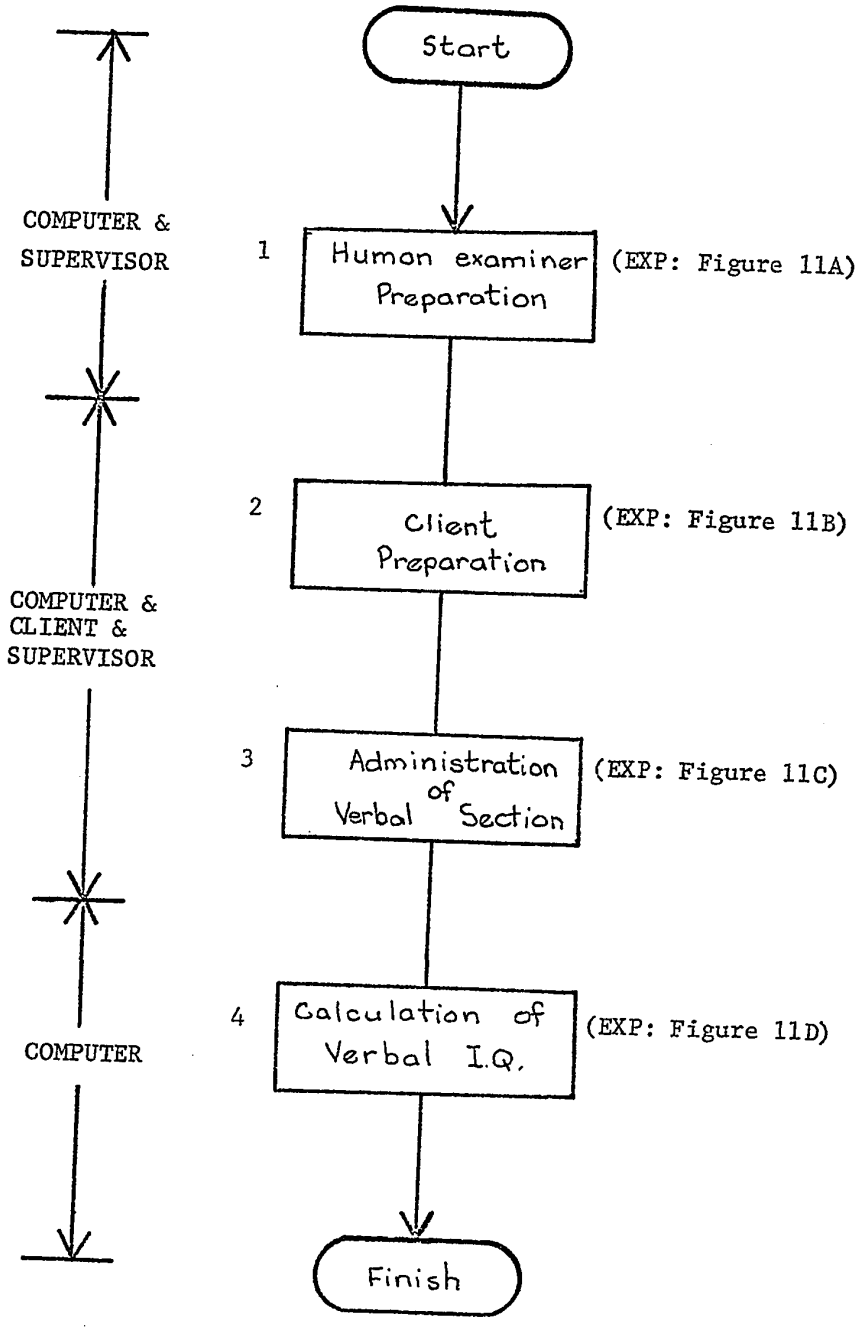


FIGURE 11A: LOGICAL FLOW CHART

HUMAN EXAMINER PREPARATION

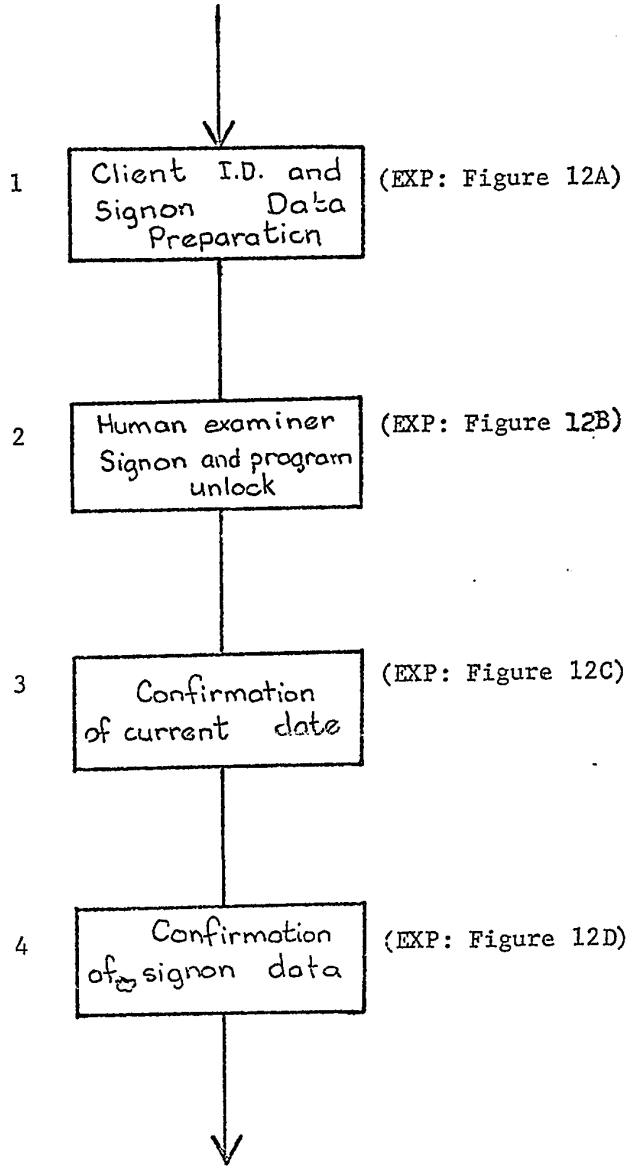


FIGURE 11B: LOGICAL FLOW CHART

CLIENT PREPARATION

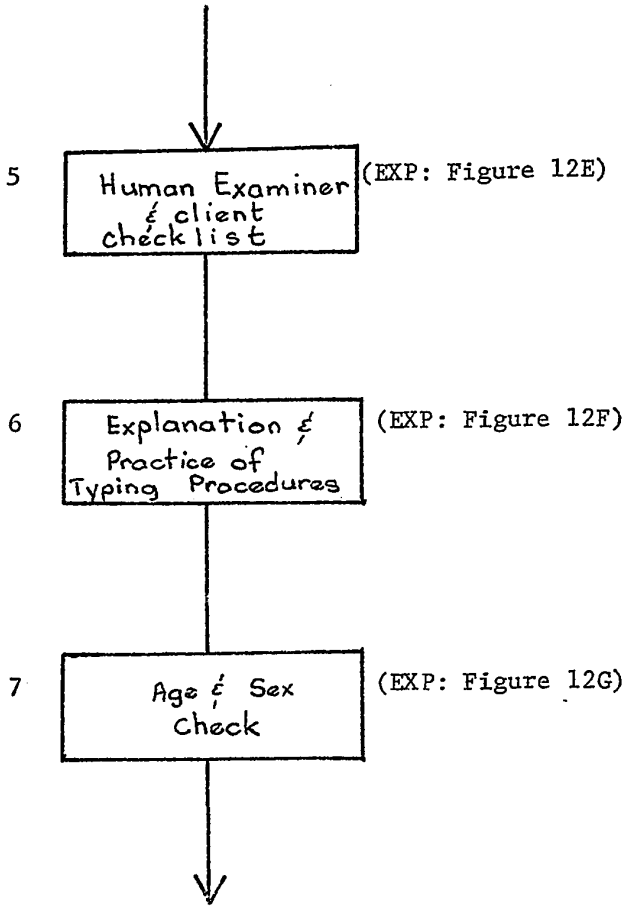


FIGURE 11C: LOGICAL FLOW CHART

ADMINISTRATION OF THE VERBAL SECTION

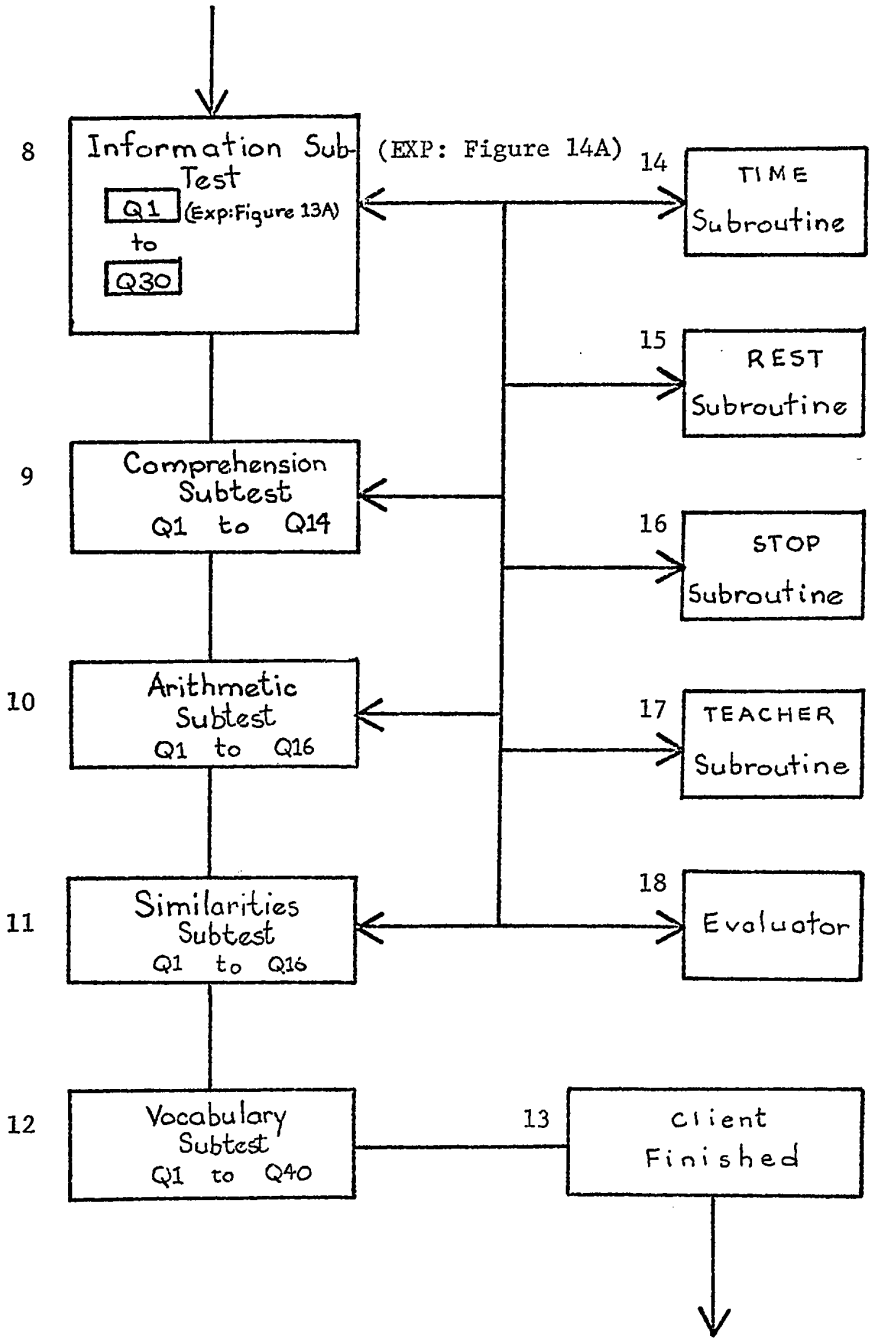


FIGURE 11D: LOGICAL FLOW CHART

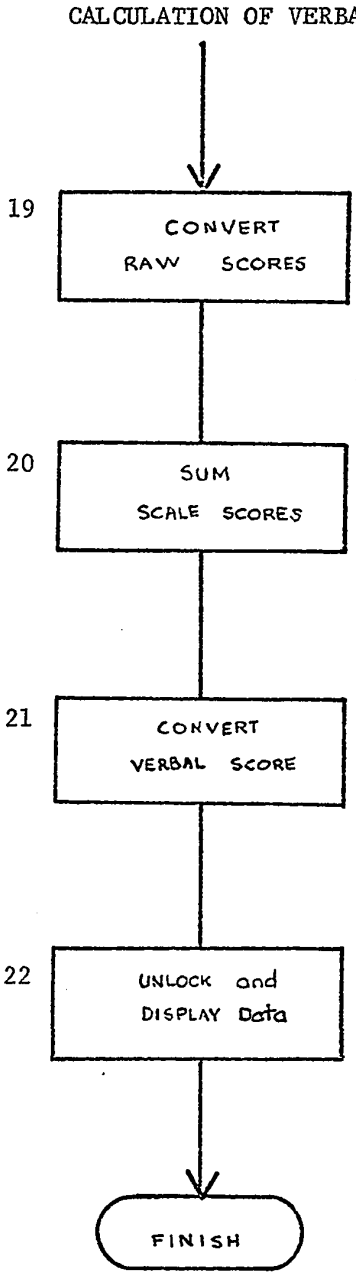


FIGURE 12A: LOGICAL FLOW CHART

CLIENT I. D. & SIGNON DATA PRESENTATION

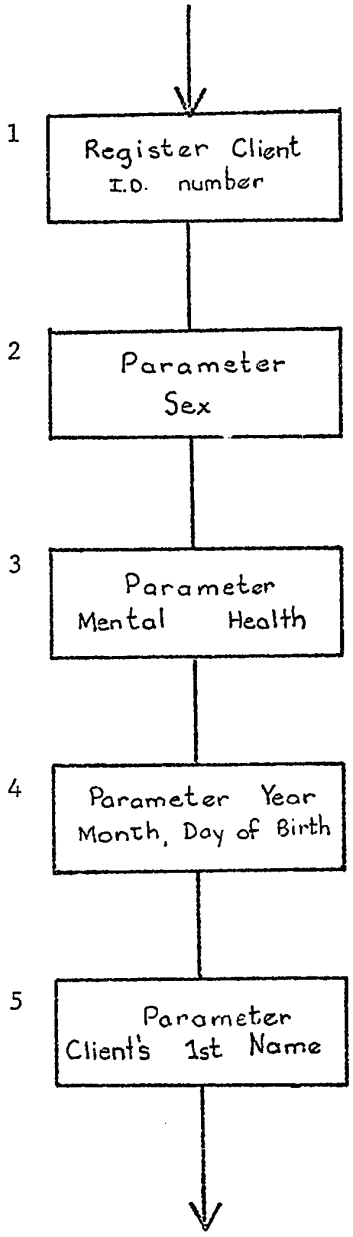


FIGURE 12B LOGICAL FLOW CHART

HUMAN EXAMINER SIGNON AND UNLOCK

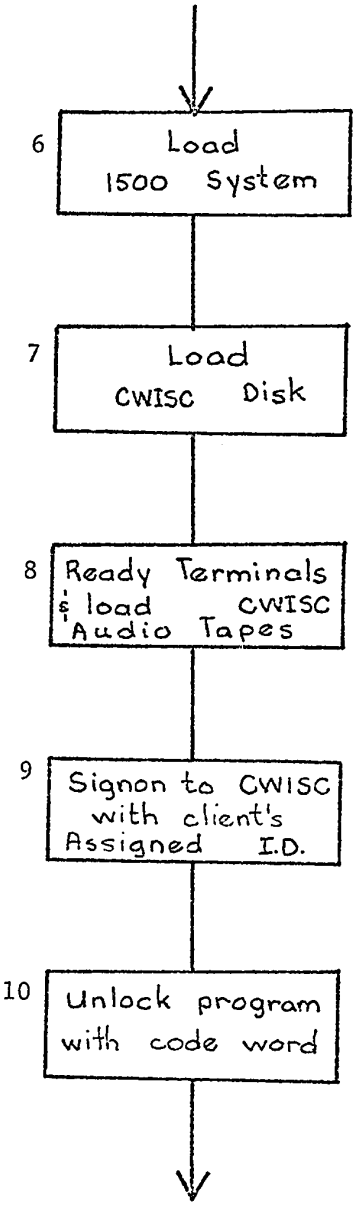


FIGURE 12C: LOGICAL FLOW CHART

CONFIRMATION OF CURRENT SYSTEM DATE

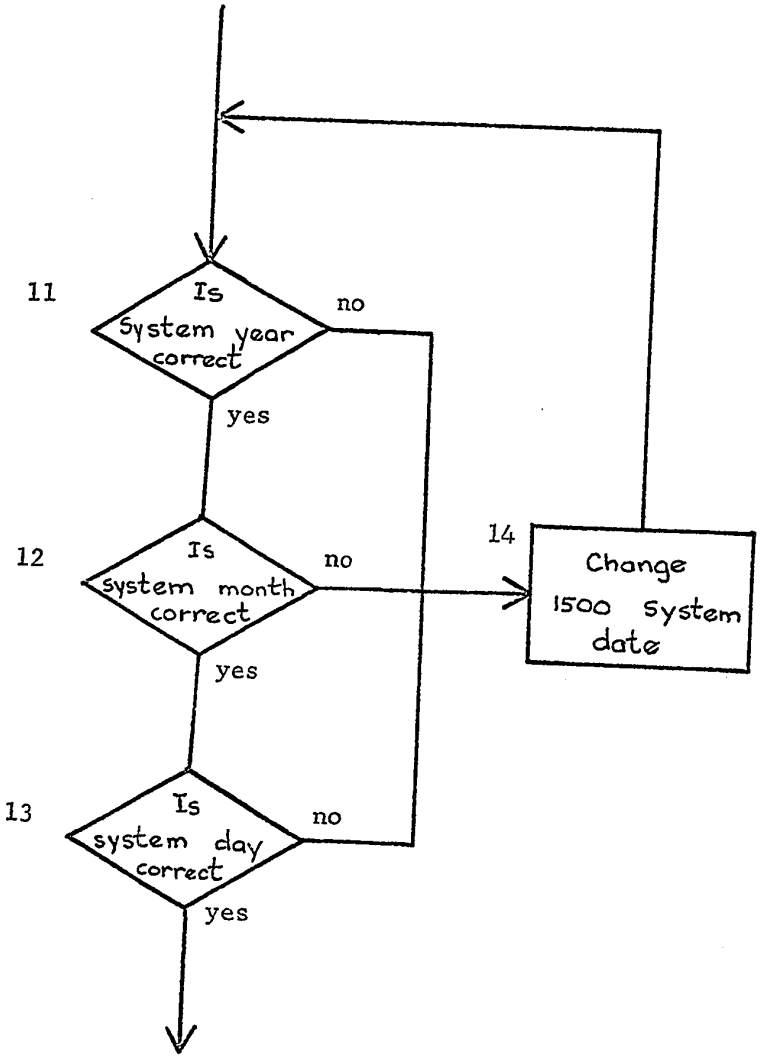


FIGURE 12D: LOGICAL FLOW CHART

CONFIRMATION OF SIGNON DATA

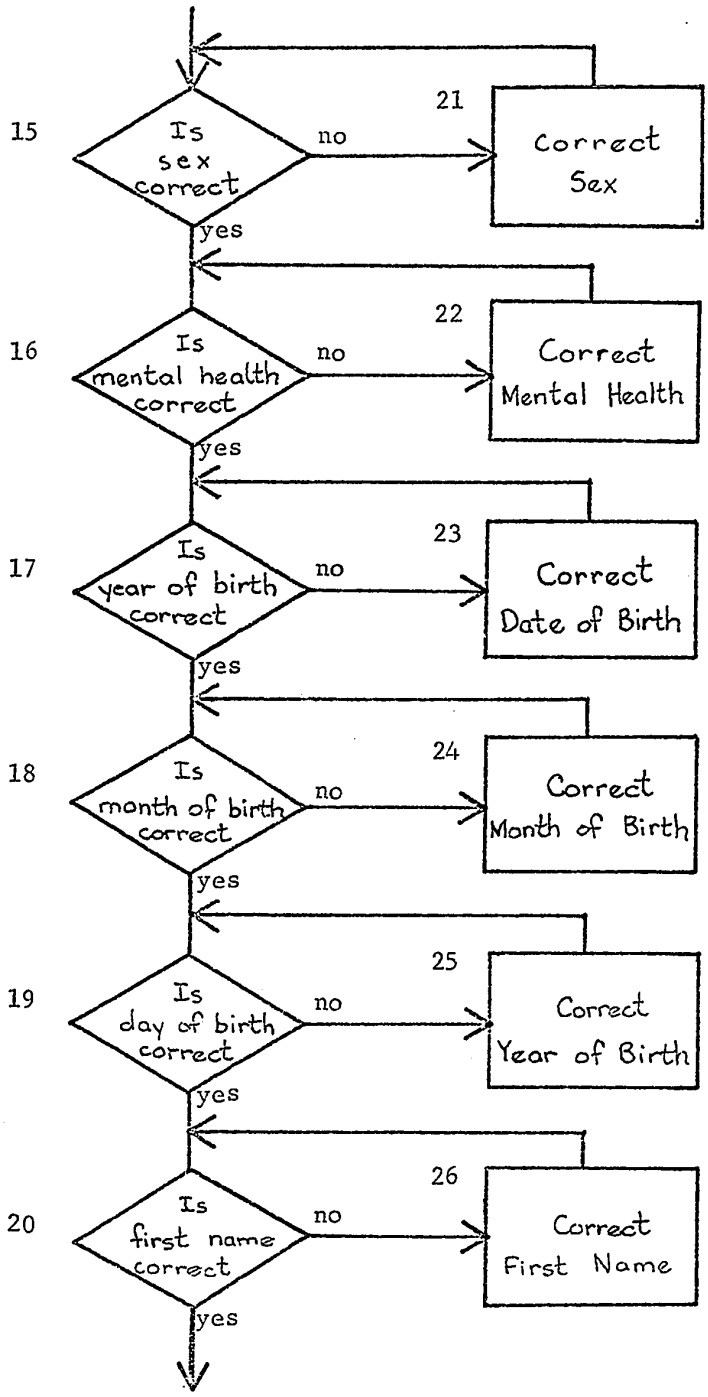


FIGURE 12E: LOGICAL FLOW CHART

HUMAN EXAMINER & CLIENT CHECKLIST

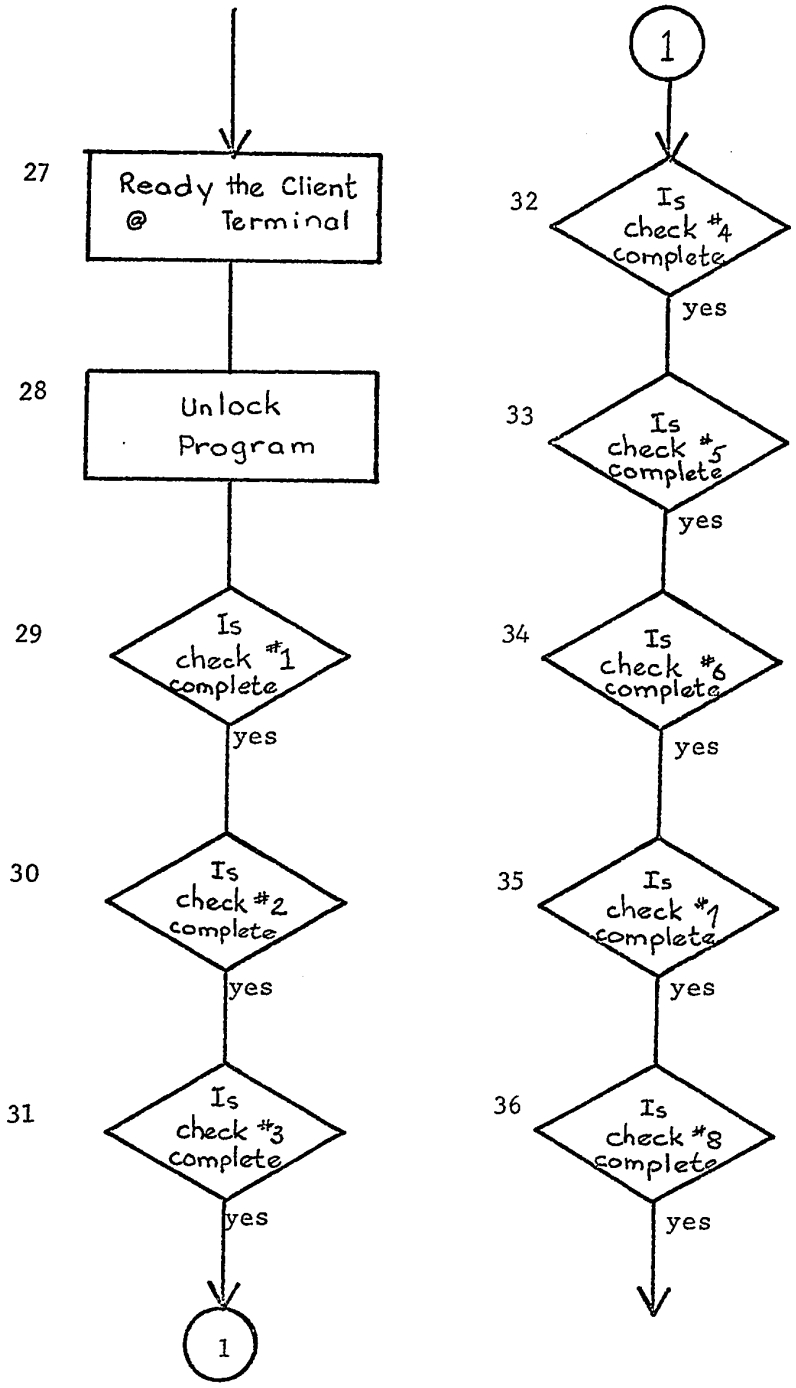


FIGURE 12F: LOGICAL FLOW CHART

EXPLANATION & PRACTICE OF TYPING PROCEDURES

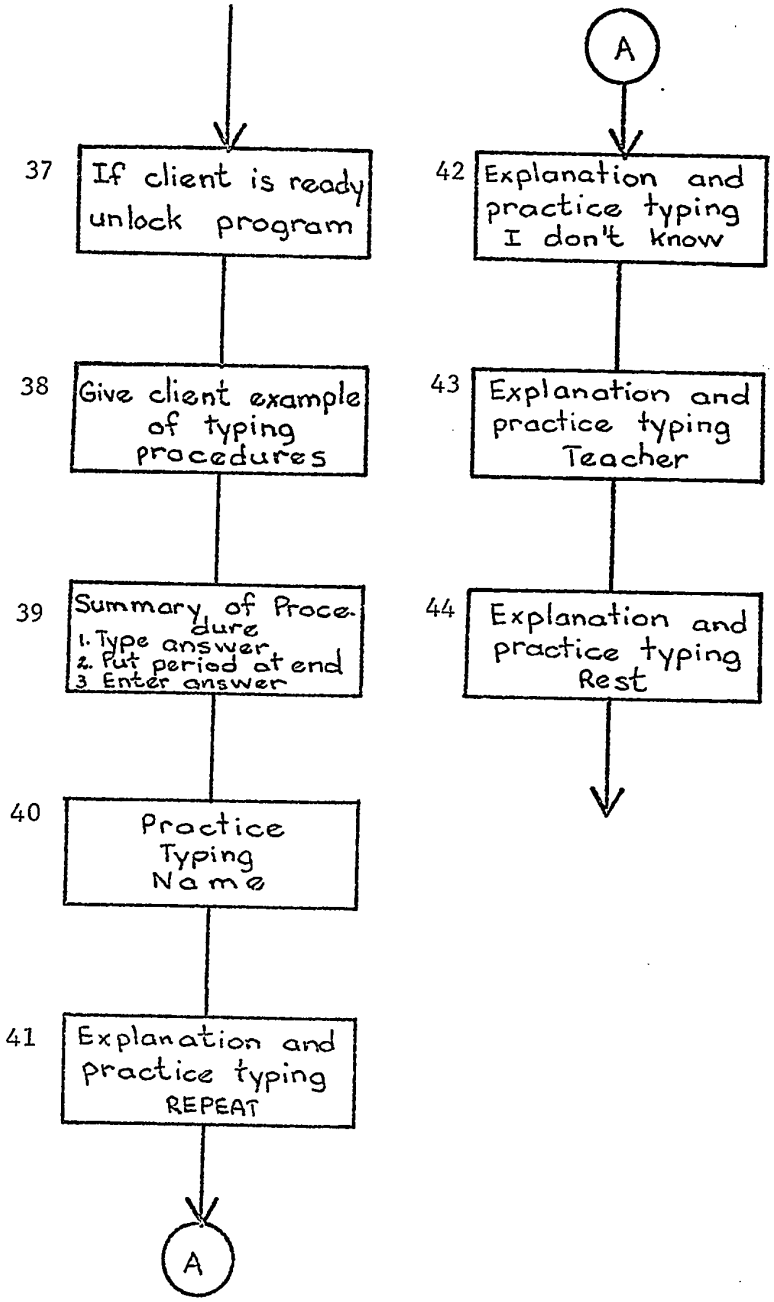


FIGURE 12G: LOGICAL FLOW CHART

AGE & SEX CHECK

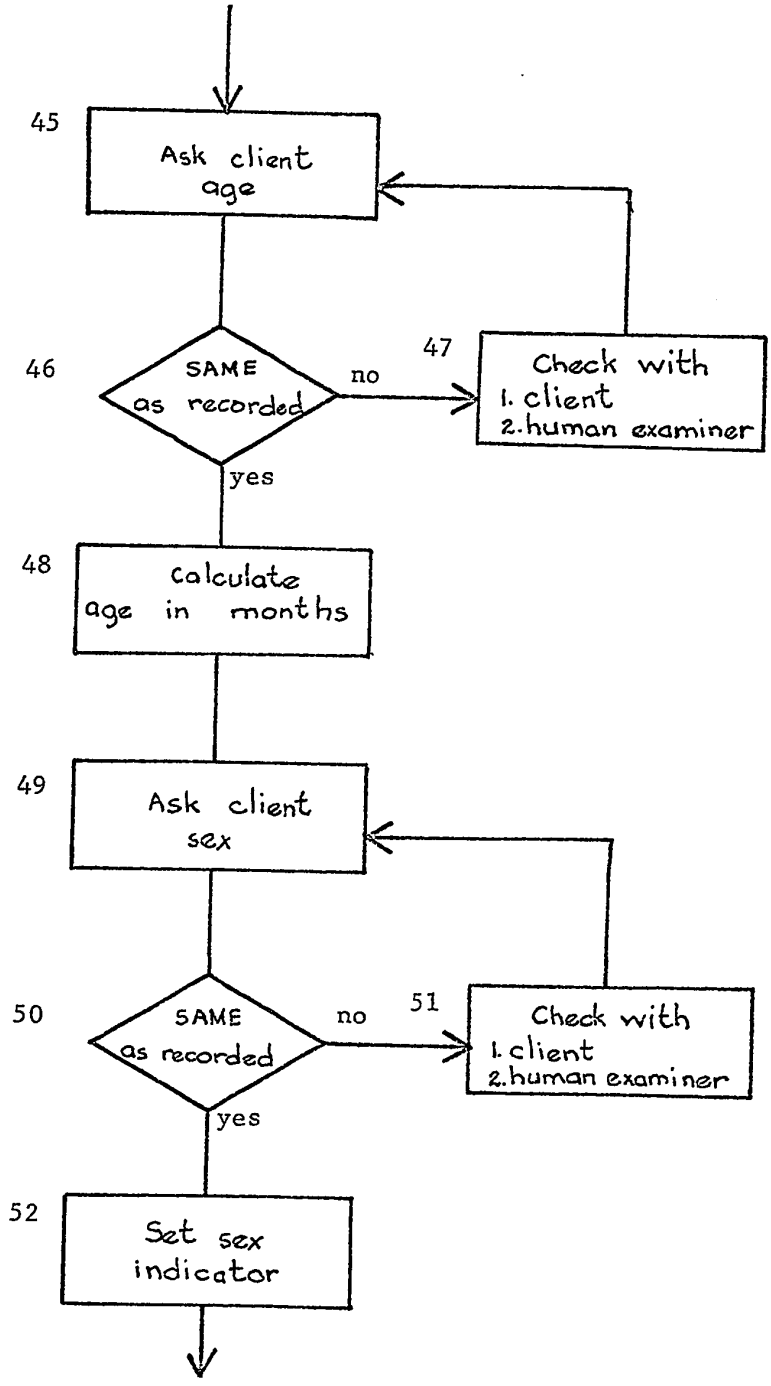


FIGURE 13A: LOGICAL FLOW CHART

QUESTION MODULE

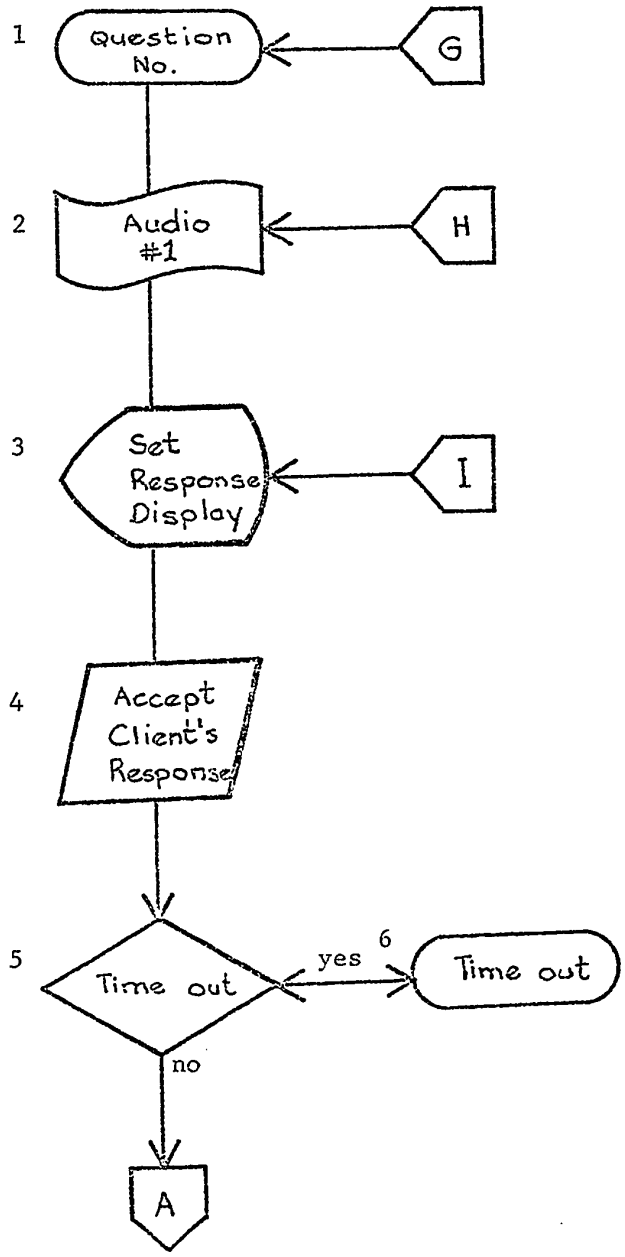


FIGURE 13B: LOGICAL FLOW CHART

QUESTION MODULE (CONTINUED)

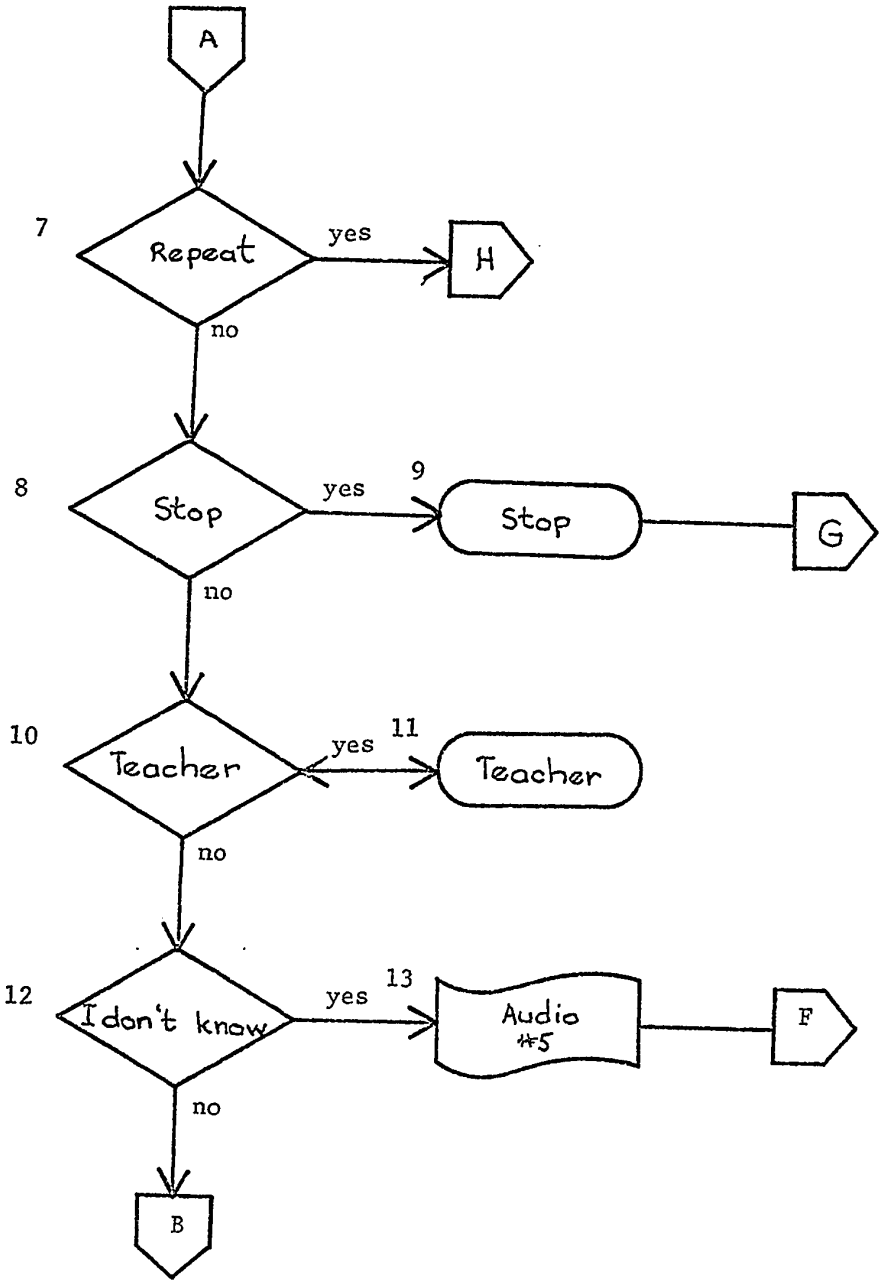


FIGURE 13C: LOGICAL FLOW CHART

QUESTION MODULE (CONTINUED)

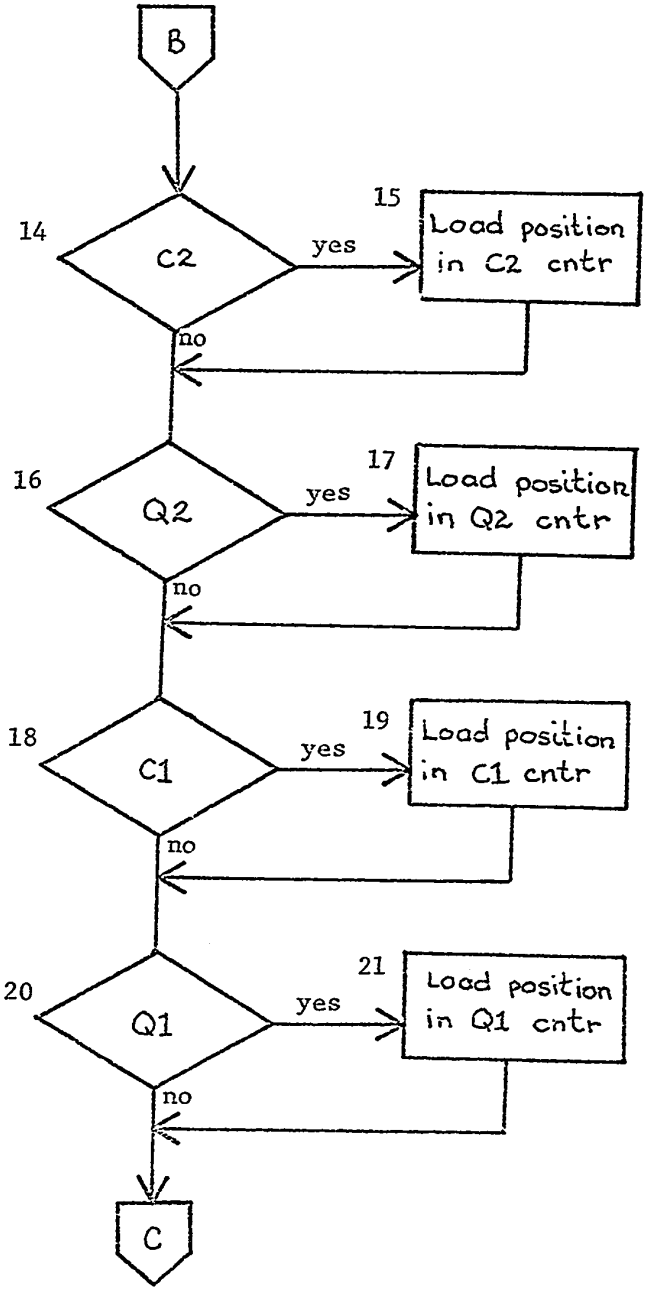


FIGURE 13D: LOGICAL FLOW CHART

QUESTION MODULE (CONTINUED)

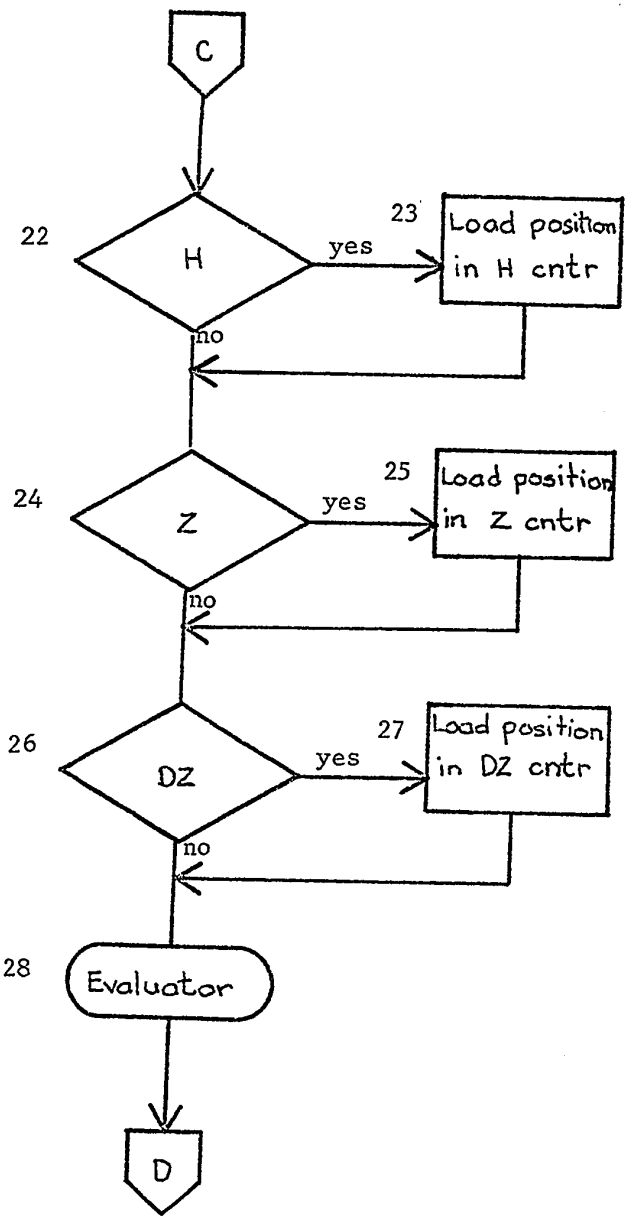


FIGURE 13E: LOGICAL FLOW CHART

QUESTION MODULE (CONTINUED)

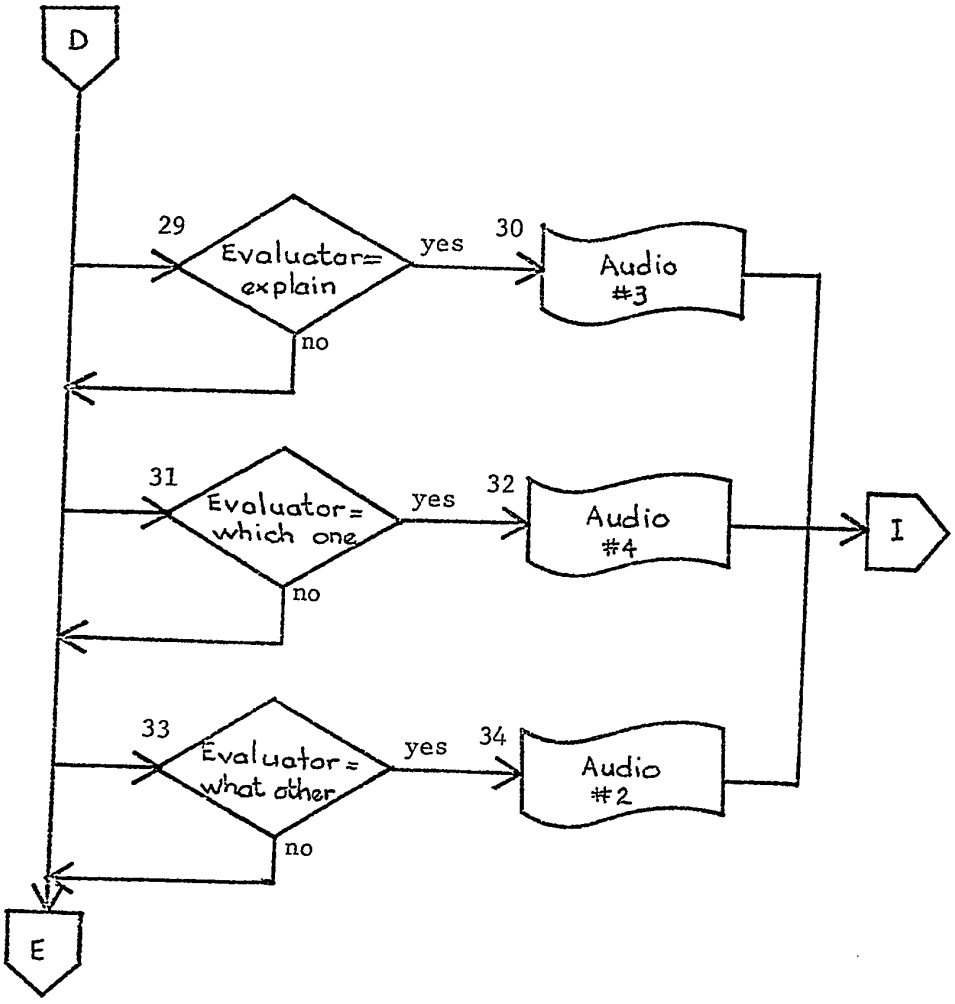
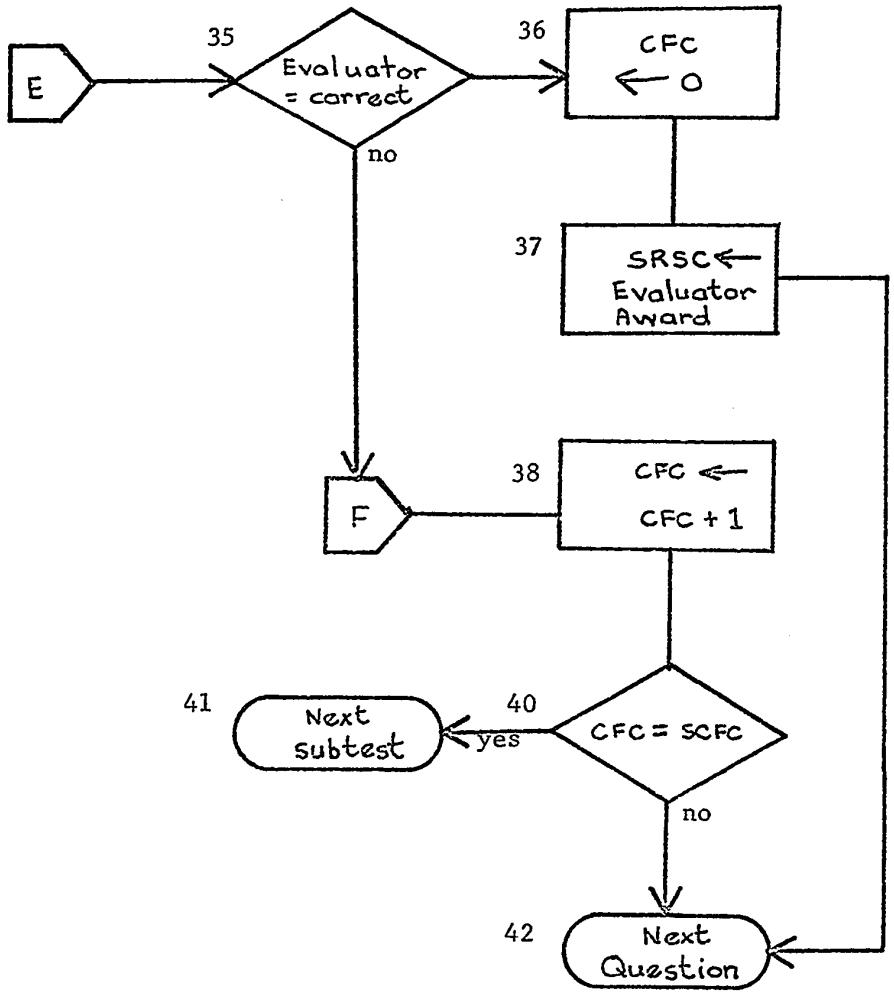


FIGURE 13F: LOGICAL FLOW CHART

QUESTION MODULE (CONTINUED)



the TIME subroutine which attempts, after two minutes, to find out why the client had not responded to the question. This was done by asking the client a series of questions, and requiring a yes or no response which attempts to pinpoint the client's reason for not responding.

4. TRANSFORMATION TO COMPUTER MODEL

There were four stages in the process of transforming the conceptual model into the computer model. These were:

(1) LOGICAL FLOW CHARTING,

(2) PROGRAM FLOW CHARTING,

(3) COMPUTER PROGRAMMING, and

(4) PROGRAM DEBUGGING.

In description of these stages only cursory attention will be paid to documentation (module 33, Figure 1D;, p. 45; module 39, Figure 1E, p. 46; module 45, Figure 1F, p. 47; and module 39, Figure 1G, p. 48). The volume of material generated is simply too great to be reasonably presented. As each stage was completed, the detail and specification for the next multiplied and consequently so did the volume of documentation. The absence of this material does not in any way influence or affect the merits of the validity establishing characteristics of the modeling process. The documentation for each stage was completed and is available for inspection.

In addition, the documentation of each stage played a progressively more important role as the conceptual model was transformed into the computer model by providing a consistent logical framework into which the detail of the computer model was structured. This helped to ensure that each specification of the conceptual model was not ignored or lost in the

increasing complexity of the computer model.

LOGICAL FLOW CHARTING (module 28, Figure 1D, p. 45)

The first activity was to prepare logical flow charts for the computer model based upon the concepts and ideas which constitute the conceptual model. This activity is represented in the developmental process as module 28 (Figure 1D), while associated modules 29, 30, 31 and 32 diagram the validity establishing process and module 33 documentation.

A number of logical flow charts were previously referenced in order to facilitate description of the conceptual model. These and a number of other flow charts will be used to describe the general logical organization of the Computer-Examiner.

Four sets of logical flow charts are presented to diagram the logical organization, each progressively more detailed than the next. These are: Figure 10 (p. 100), Figures 11A to 11D (pp. 101 to 104), Figures 12A to 12G (pp. 105 to 111), and Figures 13A to 13F (pp. 112 to 117). Figure 10 provides the most general description, while Figures 11A to 11D are a more detailed expansion of Figure 10. Figures 12A to 12G expand the detail of Figures 11A and 11B, while Figures 13A to 13F expand the question module indicated in module 8 (Figure 11C).

There are four sections to the Computer-Examiner. These are:

(1) HUMAN EXAMINER PREPARATION (module 1, Figure 10)

(2) CLIENT PREPARATION (module 2)

(3) ADMINISTRATION OF THE VERBAL SECTION (module 3), and

(4) CALCULATION OF THE VERBAL I.Q. (module 4).

Each of these are carried out in the sequence indicated starting from the point at which the client is scheduled to take the test until the

Computer-Examiner calculates and displays the Verbal I.Q. Score.

1. HUMAN EXAMINER PREPARATION (module 1, Figure 10, p. 100)

First, preparations must be made by the human examiner prior to seating the client at the Instructional Station (module 1). These are grouped in common sequences of activity as represented by modules 1, 2, 3 and 4 (Figure 11A, p. 101) and are individually listed in Figures 12A to 12D (pp. 105 to 108). In summary they are:

(1) module 1 - in order to identify the client on the Instructional System a separate identification number is assigned to each client.

(2) modules 2 to 5 - the sex (male or female), the mental health (normal or mentally deficient), the birth date (year, month and day) and the first name of the client are coded as parameters.

(3) modules 6 to 10 - are concerned with readying the System, Instructional Stations, assignment of the clients to the Stations, and signing the Stations on with the client's identification number.

(4) modules 11 to 14 - are concerned with confirming the System date (year, month and day) is current.

(5) modules 15 to 26 - are concerned with confirming that the client's parameters (modules 2 to 5) were coded correctly.

2. CLIENT PREPARATION (module 2, Figure 10, p. 100)

Second, the client is prepared by the human examiner at the Instructional Station (module 2). These preparations are grouped in common sequences of activity as diagrammed in modules 1, 2, 3 and 4 (Figure 11A, p. 101) and are listed in Figures 12E to 12G (pp. 109 to 111). In summary they are:

(1) module 27 - the client is readied at the Instructional Station.

(2) module 28 - when the client has been readied the human examiner starts the Computer-Examiner program by typing a code word.

(3) modules 29 to 36 - to ensure the client has been readied properly, the Computer-Examiner asks the human examiner to confirm that the following preparations have been carried out:

(A) Are you and the clients the only persons in the room (module 29)?

(B) Is the Terminal Room door locked and the testing sign hung up (module 30)?

(C) Is the client settled (module 31)?

(D) Is the client's seat adjusted for his or her height and position (module 32)?

(E) Is the keyboard properly positioned for the client's reach (module 33)?

(F) Show the client the keyboard, space bar and light pen (module 34)?

(G) Show the earphones and volume control to the client, place on the phones, adjust the bands and position until comfortable to the client (module 35).

(H) Examiner, is the client ready to begin (module 36)?

(4) module 37 - at the end of the checklist the Computer-Examiner stops. When the human examiner is ready to leave the client alone at the Station, he types a code word to restart the program.

(5) module 38 - the client is shown a simulated example of how responses are to be typed.

(6) module 39 - the client is presented with a summary of the steps needed to type a response. These are:

(A) type the answer,

(B) close the answer with a period, and

(C) enter the answer.

(7) module 40 - a check is made to ensure the client knows the typing procedure. This is done by having the client type his or her first name.

(8) modules 41 to 44 - the client is taught the function of the control words REPEAT. I DON'T KNOW. TEACHER. and REST. and requested to type each one as further reinforcement for the typing procedure. Each request is comprehensive enough to pick up errors in spelling, punctuation and use of the enter key.

(9) modules 45 to 47 - the client is asked for his or her age for confirmation of the System's data, to give practice in use of the light pen, and to make the Computer-Examiner look as though it knows if the client's response is wrong.

(10) module 48 - the client's age in years, months and days is converted to total age in months for use in administration of the test questions and for consulting scale score tables.

(11) modules 49 to 51 - the client is asked for his or her sex in order to confirm the System's data, give further practice with the light pen and make the Computer-Examiner look as though it knows if the client's response is wrong.

(12) module 52 - the sex switch is set. This is used to determine the administration of Comprehension questions 2 and 4.

3. ADMINISTRATION OF THE VERBAL SECTION OF THE WISC (module 3, Figure 10, p. 100)

Third, the Verbal Section of the WISC is administered to the client (module 3). This activity is diagrammed in the logical flow chart of Figure 11C (p. 103). In summary the modules were:

(1) module 8 to 12 - illustrates the administration of the subtests. Each module represents the administrative superstructure of the subtest whose function is to govern the administration of questions within. The administrative superstructure for Information subtest (module 8) is expanded in program flow chart form in Figure 14A to 14G (Appendix D, p. 222 to 237).

The question modules for each question are contained within the administrative superstructures of each subtest (see module 9). The logical flow chart of the question module is diagrammed in Figures 13A to 13F (pp. 112 to 117). It will be explained in the next section.

(2) module 13 - the client is told the test is finished, thanked, asked to quietly leave the Terminal Room, and the human examiner is notified via the System typewriter.

(3) module 14 - the TIME subroutine is called when the client fails to respond to a question within a period of two minutes. The client is asked to respond yes or no to a series of questions. If the answer is yes the Computer-Examiner takes corrective action. If the answer is no the next question in the series is asked. These questions are:

(A) I bet that you're really finished answering the question I asked but forgot to finish your answer with a period or forgot to press the enter key to make the box disappear from the screen. Am I correct?

If yes, the Computer-Examiner sequences to process the client's response to the question.

(B) Would you like some more time to think about your answer?

If yes, the Computer-Examiner sequences back to the question and waits another two minutes.

(C) Would you like the question repeated for you?

If yes, the Computer-Examiner sequences back to repeat the question.

(D) Do you know the answer to the question?

If no, the Computer-Examiner sequences back to the question and gives the client an Encouragement Response, then proceeds to the next question.

If yes, the Computer-Examiner calls the human examiner for assistance via the System typewriter.

(3) module 15 - the REST subroutine is called when the client types REST. questioning is stopped, the client is told to take a break, go to the bathroom, or take a drink and is instructed to return in five minutes. The subroutine monitors the time of the break and when exceeded informs the human examiner via the System typewriter. The client restarts the questioning by typing BEGIN.

(4) module 16 - the STOP subroutine is called by the human examiner in order to terminate questioning for the test period and sign off the Instructional Station. When ready to proceed with questioning the human examiner signs on to the program, goes through the client check list, the typing procedures are reviewed for the client, and then questioning is resumed.

(5) module 17 - the TEACHER subroutine is called when the client

types TEACHER. This subroutine informs the human examiner, via the System typewriter, that the client wants to see him, then asks the client to wait patiently until the human examiner comes to the Instructional Station.

(6) module 18 - the EVALUATOR subroutine is called by the question modules for the purpose of assigning score points to the categorized answers and to indicate what subsequent actions must be taken by the question modules. The EVALUATOR subroutine is the embodiment of Decision Tables 1, 2 and 3 as outlined in the desciption of the Evaluation of the Client's Response Section of the conceptual model (p. 63).

4. CALCULATION OF THE VERBAL I.Q. (module 4, Figure 10, p. 100)

Last, the client's Verbal I.Q. Score is calculated (module 4). This activity is diagrammed in Figure 11D (p. 104.) In summary the modules are:

(1) module 19 - each subtest raw score is converted to a scaled score by use of the Scale Score Tables of the manual and the client's age in months. This module basically represents the embodiment of these tables.

(2) module 20 - the subtest scaled scores are summed.

(3) module 21 - the scaled score total is converted to a Verbal I.Q. score by consulting the Verbal Score in the I.Q. Table of the manual which is embodied in this module.

(4) module 22 - when ready, the human examiner can see the results on the Instructional Display by typing a code word to unlock the program.

5. QUESTION MODULE (module 8, Figure 11C, p. 103)

The question module is the central and most essential part of the Computer-Examiner. There is one question module for each question in the test. It is responsible for the administration details of the question. In addition, the question module contains that section of the Computer-Examiner which has the function of analyzing and categorizing answers from the client's response, as outlined in the Evaluation of the Client's Response section of the conceptual model. The logical flow chart is diagrammed in Figures 13A to 13F (pp. 112 to 117). In summary the modules are:

(1) module 1 - the label number of the question

(2) modules 2, 34, 30, 32 and 13 - respectively the Question, the Explanation Inquiry, the Homonym Inquiry, the Difference Inquiry and the Encouragement Response audio messages as outlined in the Messages for Questioning Section of the conceptual model.

(3) module 3 - could be any one of the Instructional Display graphic messages shown in Figures 4 to 7 (pp. 89 to 92), depending upon the question asked, as outlined in the Graphic Messages section of the conceptual model.

(4) module 4 - the Instructional System is set up to receive the client's response.

(5) modules 5 and 6 - if the client does not respond to the question within two minutes (module 5), the TIME subroutine is called (module 6).

(6) modules 7 to 13 - the client's response is analyzed for the control words REPEAT. STOP. TEACHER. and I DON'T KNOW. (modules 7, 8, 10 and 12). If the question is to be repeated, a branch

is made back to Audio #1 (module 2). If the client wants to stop, the STOP subroutine is called (module 9). If the human examiner is wanted the TEACHER subroutine is called (module 11). If the client types I DON'T KNOW. the Encouragement audio message is given (module 13) and the question is counted as being wrong (module 38).

(7) modules 14 to 27 - the client's response is searched for the answers provided in each category (that is if answers have been supplied from the manual and supplement for a category; if not, the category is not included in the search) (modules 14, 16, 18, 20, 22, 24, and 26) and if answers are found, their position number in the response buffer is loaded into the category counters (modules 15, 17, 19, 21, 23, 25 and 27).

(8) module 28 - at the end of answer categorization, the EVALUATOR subroutine is called.

(9) modules 29 to 34 - the EVALUATOR provides a number in the action counter to tell the question module what action is to be taken next. This number is used in the question module to branch to the correct action (modules 29, 31, 33, 35 and 39). If the action counter is equal to two, the Explanation audio message (Please tell me more about your answer.) is given (module 30) and the question module recycles to accept another response (module 3). If the count is three, the Difference audio message (Which answer is it?) is given (module 32) and the question module recycles to accept another response (module 3). If the count is four, the Homonym audio message (Would you give me the other meaning please.) is given (module 34) and the question module recycles to accept another response (module 3).

If the action count is equal to five, the response is wrong and the

consecutive failure count (CFC) is incremented by one (module 38). This count is compared against the subtest consecutive failure count (SCFC) (module 40), which is the maximum value of consecutive failures that can be permitted for the subtest. If the CFC is less than the SCFC the next question is sequenced (module 42). If the CFC is equal to the SCFC then the next subtest is sequenced (module 41).

When the action count is six the response is correct and the consecutive failure count is set to zero (module 36), the subtest raw score counter (SRSC) is incremented by the evaluator's raw score point award for the question (module 37), and the next question is sequenced (module 42). In the case that the question module is the last question of the subtest, the administrative superstructure sequences to the next subtest.

PROGRAM FLOW CHARTING (module 34, Figure 1E, p. 46)

The second activity in the process transforming the conceptual model into the computer model was to prepare program flow charts of the computer model. This activity is illustrated in the developmental process as module 34 (Figure 1E). Modules 35 to 38 diagram the validity establishing steps and module 39 documentation.

As an example, the program flow chart for the administrative superstructure of the General Information subtest is diagrammed in Figures 14A to 14G (Appendix D, pp. 222 to 228). No explanation will be given of this flow chart other than definition of the symbols used int it.

COMPUTER PROGRAMMING (module 40, Figure 1F, p. 47)

The third activity was to transform the program flow charts in to the computer program (module 40). Modules 41 to 44 diagram the validity establishing steps and module 45 documentation.

As an example, the computer program of the question module for the General Information subtest question 5 is listed in Appendix E, p. 229 No explanation is given of this listing. It should be read by only those familiar with the Coursewriter II computer language.

PROGRAM DEBUGGING (module 46, Figure 1G, p. 48)

The last stage in transforming the conceptual model into the computer model was to debug the Computer-Examiner program (module 46). Modules 47 to 50 were used for establishing validity while module 51 for documentation.

Debugging was a three step operation. First, various sections of the Computer-Examiner's program were isolated (for example the EVALUATOR, the QUESTION MODULE, or the ADMINISTRATIVE SUPERSTRUCTURES). Simulation generators were then built around these sections to input generated data and monitor output. Unanticipated output was used as the source for investigating programming errors. When all output could be anticipated the section was considered debugged.

Secondly, the isolated sections were put together, one section at a time, and tested under simulated conditions in order to ensure that various sections work together. Last, the Computer-Examiner was then debugged as whole program by inputting simulated data and observing its response.

5. EVALUATION OF THE COMPUTER-EXAMINER

Three stages in development of the Computer-Examiner were used for the purpose of evaluation. These were:

(1) VOLUNTEER ADMINISTRATION,

(2) PROTOCOL ADMINISTRATION, and

(3) STUDENT ADMINISTRATION

VOLUNTEER ADMINISTRATION (module 52, Figure 1H, p. 49)

The Computer-Examiner was tested with simulated data in the debugging stage of development. To provide more realistic data, in order to facilitate further detection of modeling and programming errors, ten graduate students from the Division of Educational Research and Department of Educational Psychology volunteered to act the role of clients. Dummy identification numbers, name, sex, health and birthdate parameters were created for each of these volunteers and the test was administered to them (module 52). Several administration and evaluation modeling and programming errors were detected in the performance recordings of these administrations. Subsequently, steps were taken to correct these errors (modules 53 to 56) and update the documentation of the previous stages (module 57). This activity was really an extension of the debugging stage but because volunteers were used to obtain the performance recordings and because the following evaluation took place it was considered to be part of the evaluation process.

In order to obtain an accurate estimate of how capable the Computer-Examiner was in correctly identifying the source answer set (all the answers given in the manual and supplement) five of the

graduate students were instructed to take the test in the normal manner, but in place of their own answers to each question, they were to give only those answers provided in the manual and supplement. This was done by providing each volunteer with a xerox copy of the supplement, instructing them to select at random one answer from those provided for each question, and to type this answer as their response. In addition they were instructed to check with a pen the selected answer.

By comparing the raw score points awarded to each answer with the raw score points which should have been awarded for the answer according to the supplement, it was possible to evaluate what percentage of answers were correctly identified by the Computer-Examiner. The results of this comparison will be discussed in the next chapter.

PROTOCOL ADMINISTRATION (module 58, Figure 1I, p. 50)

The second stage of evaluation was to have the Computer-Examiner administer the test using six WISC protocols. This was done in the following manner. First, the protocols were selected from the files of the Clinical Services section of the Department of Educational Psychology by a school psychologist. The criteria for selecting these protocols was simply to pick five protocols in which the answers to each question in the Verbal Section were accurately and completely entered on the record form. The identification of the clients who took these tests was not taken from the file. Computer identity numbers and dummy first names were assigned to each protocol. The sex, mental health and age data from each protocol were recorded as input parameters. The mental health

parameter in each protocol was determined by examining the order of the administered questions in each test. The school psychologist then substituted for the client of each protocol as the test was administered. For each question asked by the Computer-Examiner the psychologist responded with the answer supplied in the protocol (module 58).

The purpose of the protocol administration was to move one logical step towards evaluation of the Computer-Examiner in a real situation. The data obtained from the performance recordings of these administrations will be evaluated in the next chapter.

In addition to the evaluation function of this stage, additional administration and evaluation errors in the model and computer program were looked for (modules 59 to 62). No administration errors were detected. A few errors were detected in the computer programming logic of the response analysis section. These were corrected and the documentation of the computer program stage was updated (module 63).

STUDENT ADMINISTRATION (module 66, Figure 1J, p. 51)

The last stage of evaluation was to have the Computer-Examiner administer the test to a number of clients. Three criteria were used to select the clients:

(1) that they would fall into the age range of the WISC,

(2) that they knew from previous experience how to make responses with the keyboard and light pen, and

(3) that they would be available for testing on the same day.

Eight children, ranging in age from 8 years, 8 months, to 15 years, 11 months, met the criteria and were willing to take the test (module 65).

The individual characteristics of the clients are listed in Appendix C.

The test date was scheduled to be Saturday, August 21, 1971 (module 64). On that date the test was administered at the same time to each of the eight clients (module 66). Evaluation of this administration (module 67) and of the resulting performance recordings will be described in the next chapter along with some observations, conclusions and a discussion of the results (module 68).

CHAPTER 5

EVALUATION, OBSERVATIONS, AND RECOMMENDATIONS

This study was initiated for the purpose of assessing the potential of the computer as a substitute examiner in place of the human examiner of the individualized intelligence test. It was oriented towards the development of a computer model, based upon the capabilities of the IBM 1500 Instructional System, which would duplicate the administrative and evaluative functions needed by the human examiner to give the Verbal Section of the WISC. The design, construction and implementation of this model and a description of the steps taken to generate evaluation data were discussed in Chapter 4.

Chapter 5 is concerned with tabulation, classification and interpretation of the evaluation data. In addition, this chapter records a number of observations that were made during development and implementation of the model and lists a number of recommendations for improving the design of future Computer-Examiners. It concludes with a summary of the study.

1. EVALUATION

The task of assessing the Computer-Examiner's performance was approached as a problem of providing answers to the following two questions:

(1) Did the Computer-Examiner make administrative mistakes while administering the test to clients?

(2) Did the Computer-Examiner score responses as accurately as

would a human examiner?

In conceptualizing the computer model it was clearly recognized that the mechanics of administering the questions were distinctly separate from those of evaluating the client's responses. The administrative function of the Computer-Examiner acts on the assumption that information received from the evaluative function is correct. In operation, the evaluative section of the model passes to the administrative section a numerical value which categorizes the client's response on the answer hierarchy as defined in the description of the conceptual model presented in Chapter 4. The administrative section assesses and weights this value in relation to a number of other factors, then takes a specified course of action.

In assessing the performance of the Computer-Examiner it was important to recognize that the evaluative section could incorrectly categorize the client's response and pass to the administrative section an incorrect value, which in turn, would act upon this information and take a prescribed course of action. Even though this course would be incorrect, in relation to the correctness of the client's response (as a human examiner would score the response and take action), the administrative section would not be in error. The evaluative section would be in error.

It is also possible to have the situation in which the evaluative section correctly categorizes the client's response but the administrative section takes the wrong course of action. Consideration of such possibilities led to the two question approach of evaluating the Computer-Examiner, that is a separate assessment of the administrative and

evaluative functions of the computer model.

ASSESSMENT OF THE ADMINISTRATIVE SECTION

Did the Computer-Examiner make administrative mistakes while administering the test to clients? Administrative errors were present in the performance recordings of the Volunteer Administration group. These were expected as it was the first time the Computer-Examiner had been used to administer the test to real, rather than simulated clients, and one purpose of having the Volunteer Administration group was to have the opportunity to obtain performance recordings for the purpose of ferreting out administrative errors. Subsequently, these were traced to faults in the logic of the model or instructions of the computer program. These faults were corrected and the model was retested to ensure their elimination.

No administrative errors were found by the author or assisting school psychologist in the performance recordings of either the Protocol or Student Administration groups. There were six clients represented by protocols in the Protocol Administration. A total of 655 responses were generated by this group. Eight clients made up the Student Administration group and they generated a total of 1,042 responses. The fact that no errors were detected over a total of 1,697 responses would strongly suggest that the administrative criteria for the Verbal Section of the WISC, as outlined in the list of Examiner's Administrative Procedures of Appendix A (pages 179 to 200) were successfully implemented into the computer model.

ASSESSMENT OF THE EVALUATION SECTION

Did the Computer-Examiner score responses as accurately as would a human examiner? A preliminary perusal of the performance recordings indicated clearly that it could not. Therefore, it was assumed that the objective to be met in assessing the Computer-Examiner's evaluative section was to measure how accurately it could. To make this assessment the human examiner's scoring was used as the referant, (assumed perfect) and the Computer-Examiner's scoring was measured relative to it.

1. RESPONSE ANALYSIS

Neither the Verbal I.Q. Score, the Scaled Scores, or Raw Scores are sufficiently accurate measures of the differences in the evaluation of responses between the computer and human examiners. A measure based on these scores would be difficult to interpret because the questions administered and their sequence in a subtest would not be the same between the computer and human administrations due to differences which would arise in categorizing responses. In addition, the interpretation of differences between Verbal I.Q. Scores or Scaled Scores would have been confounded by the nonlinear transformation of the scaling tables in the WISC Manual.

In the Volunteer Administration the client's responses were already scored in that the Volunteers were instructed to select their answers from the Massey Supplement in response to the Computer-Examiner's questions. Responses in the Protocol Administration were likewise previously scored. This had been done by the human examiners who

originally administered the tests.

For the Student Administration group each of the client's responses were abstracted from the performance recordings and given to the school psychologist for scoring in a list ordered by the questions. The computer score for each response was then added to the list and the computer and human scores for each response were compared. The results of this comparison are tabulated in Tables 4 through 18 of Appendix F (pages 234 to 253). Tables 4 to 8 tabulate the responses, by subtest, of the Volunteer Administration group. Tables 9 to 13 tabulate responses of the Protocol Administration group, while Tables 14 to 18 the Student Administration group.

The columns of each Table are the same and the purpose of each is as follows:

(1) QUESTION - refers to the number of the question in the subtest.

(2) NUMBER OF RESPONSES - tabulates the total number of responses made to a question by all of the clients in the administration group.

(3) 2 POINTS - tabulates responses which the Computer-Examiner and the human examiner agree were worth a full 2 points.

(4) 1 POINT - tabulates responses which the Computer-Examiner and the human examiner agreed were worth 1 point.

(5) 0 POINTS - tabulates responses which the Computer-Examiner and the human examiner scored 0 points.

(6) QUESTION - tabulates responses which the Computer-Examiner asked the client to explain further.

(7) PASSED - tabulates questions which the Computer-Examiner passed. A question was passed either because the client's age or

mental health required initial questions not to be asked, or because the client had failed a number of successive questions and further questioning was to be terminated.

(8) I DON'T KNOW - tabulates questions to which the client responded "I don't know".

(9) REPEAT - tabulates questions to which the client responded with the word "Repeat".

(10) TYPED INCORRECT - responses incorrectly typed or typographically incorrect were tabulated in this column.

(11) COMPUTER MISTAKES - All of those questions to which the Computer-Examiner and the Manual or human examiner were in disagreement were tabulated as Computer Mistakes. This included situations in which the Computer would award a single point while the human examiner or the WISC Manual awarded two points, or vice versa.

2. EVALUATION EFFICIENCY

In order to accurately compute the Computer-Examiner's evaluation capabilities it was necessary to delete the following columns from the computation formula:

(1) PASSED

(2) REPEAT

(3) TYPED INCORRECTLY

The PASSED tabulations were not included because they were the result of the Computer-Examiner's administrative, rather than evaluative, decisions. REPEAT responses were omitted because they did not influence the client's score in any way; the Computer-Examiner simply repeated

the question. Responses of the TYPED INCORRECTLY category occurred only in the Volunteer and Protocol Administration groups where responses were taken from specific sources and therefore possible to incorrectly type or make typographical errors while entering them. In the Volunteer Administration they were taken from the Massey Supplement while in the Protocol Administration they came from the protocols. These errors would confound the measurement of evaluation efficiency and were therefore removed. For example, if a two point response had been incorrectly typed it could be evaluated to be a zero point response and without the TYPED INCORRECTLY category would have been tabulated as a COMPUTER MISTAKE.

The columns of Tables 4 through to 18 used to compute the Computer-Examiner's response evaluation efficiency are:

(1) (A) - 2 POINT RESPONSES

(2) (B) - 1 POINT RESPONSES

(3) (C) - QUESTION

(4) (D) - ZERO

(5) (E) - I DON'T KNOW

(6) (F) - COMPUTER MISTAKE

Each letter is a substitute for the indicated column. The formula was:

$$\text{RESPONSE EVALUATION EFFICIENCY} = \frac{(A+B+C+D+E)}{(A+B+C+D+E) + F} \times 100\%$$

TABLE 19

COMPUTER-EXAMINER RESPONSE EVALUATION

EFFICIENCY OVER THREE ADMINISTRATIONS

ADMINISTRATION SUBTEST	VOLUNTEER	PROTOCOL	STUDENT
INFORMATION	96.0%	97.0%	89.7%
COMPREHENSION	71.9%	54.8%	42.7%
ARITHMETIC	100.0%	100.0%	100.0%
SIMILARITIES	97.1%	86.4%	68.0%
VOCABULARY	88.3%	74.2%	60.0%
AVERAGE/TOTAL RESPONSES IN ADMINISTRATION	91.8%	80.8%	73.4%

The formula for computing the Computer-Examiner's evaluation efficiency was applied to the totals of Tables 4 through 18. The results were tabulated in Table 19. Caution must be taken when attempting to interpret this Table. The percentages do not refer to a measure of the Computer-Examiner's ability to produce Raw or Scaled Scores comparable to those of the human examiner. They refer to the measure of the Computer-Examiner's ability to correctly identify and categorize the responses given by the client to each question.

Under the former definition the test would have been administered twice to each client, once by the Computer-Examiner and once by the human examiner. The two resulting Raw or Scaled Scores could then be compared and evaluation efficiency would have been defined as the ratio of either Raw or Scaled Scores. This procedure was not followed because it was assumed that such an estimate was premature. Evaluation efficiency was defined to measure how well the Computer-Examiner evaluated responses. This measure was assumed to be more important and a priori to the comparison of Raw or Scaled Scores. In the Student Administration the Computer-Examiner administered, evaluated and scored responses for each of the eight clients. The human Examiner was required to evaluate and score the responses of the computer administration. Evaluation efficiency was then computed from the results of a comparison of the two sets of evaluations. The percentages of Table 19 therefore estimate how well the Computer-Examiner evaluated responses.

It should be noted that the computation of the Raw and Scaled Scores was assumed to be an administrative function and as previously indicated during the Student Administration no administrative errors were made by the Computer-Examiner. Each of the computed Raw and

Scaled Scores for each of the eight administrations were traced from the tabulation of score points to the derivation of the Verbal I.Q. Score. No mistakes had been made by the Computer-Examiner.

(1) It was expected that the Computer-Examiner's response evaluation efficiency would progressively decrease as it was applied across the three administrations from the most simulated administration (the Volunteer Administration) to a less simulated administration (the Protocol Administration) and finally the non-simulated administration (the Student Administration). The average evaluation efficiency for all the responses in each of the three administrations showed this trend, as well as for each subtest when individually compared across the three administrations. For the Volunteer Administration this average was 91.8%, whereas it was 80.8% for the Protocol Administration and 73.4% for the Student Administration.

This result is in line with the observation that for the Volunteer Administration the range of acceptable responses which could be entered as responses to the Computer-Examiner's questions was the smallest; therefore the easiest to anticipate and to program into the computer model. The Student Administration created a situation in which the range of acceptable responses which could be entered as responses to the Computer-Examiner's questions was the greatest and therefore the most difficult to anticipate and to program into the computer model. The relationship between the Computer-Examiner's response evaluation efficiency and the range of possible responses to the test administration is inverse; that is, evaluation efficiency decreases as the range of acceptable responses increases.

(2) The trend observed in point one extends in another direction in the percentages of Table 19. An inverse relationship appears to exist between the Computer-Examiner's response evaluation efficiency and a combination of two things - the range of responses which could answer the questions and the grammatical structure of these responses. As this range and the grammatical structure become more complex, the Computer-Examiner's response evaluation efficiency decreases. Of the five subtests, the Comprehension subtest requires answers of the greatest grammatical complexity and could be answered by the largest number of acceptable responses. It had the lowest average evaluation efficiency. Of the five subtests, the Arithmetic subtest required answers of the least grammatical complexity and could be answered by the fewest number of acceptable responses. It had the highest average evaluation efficiency. The Information, Similarities, and Vocabulary subtests ranged between the Arithmetic and Comprehension subtests and it is not difficult to visualize, knowing the kinds of questions asked in these subtests, that the order they appeared in was to be expected.

(3) The observations of points one and two appear to apply to the ordering of the questions within the Information, Similarities, and Vocabulary subtests. It was observed that the Computer-Examiner's response evaluation efficiency decreased as the range of responses increased and the grammatical structure became more complex. In general, questions of this kind are ordered towards the end of each subtest. It was observed in the data of Tables 4 to 18 that computer mistakes tended to occur more frequently in the higher ordered questions.

For the Arithmetic subtests there were no computer mistakes and therefore the same observation could not be made. In the

Comprehension subtest, computer mistakes tended to be distributed evenly over all of the questions. It shows that Comprehension questions were consistently the most difficult to develop an adequate response repertoire in the computer model.

(4) The percentages for the column labelled Volunteer Administration in Table 19 suggest the task of incorporating the scored responses of the WISC Manual and Massey Supplement into the response repertoire of the computer model was successful. The Supplement lists all of the responses in the Manual. Each client was given a copy of this Supplement and instructed to select a response at random from the list available for the question asked and to enter this response as their own. The average evaluation efficiency for all responses evaluated in the Volunteer Administration was 91.8%.

The fact that four of the subtests did not achieve an evaluation efficiency of 100% attests to the degree of difficulty and the complexity encountered in the task of developing a response repertoire for the computer model; especially when it must be considered that the list of responses available from the Manual and Supplement are but a small part of the total spectrum of acceptable responses. The order into which subtests fell further attests to the individual difficulty and complexity of developing a response repertoire for each subtest; the Arithmetic subtest was the easiest, while the Comprehension subtest was the most difficult.

It is significant that the evaluation efficiency for each subtest under the Volunteer Administration was relatively high for a first generation model. It suggests that the procedure used for developing

the computer model's response repertoire might prove to be an extremely successful one, provided that a well defined list of acceptable responses was made available. As explained in Chapter 4, the key concept in this procedure was to select from each acceptable response one or more key-words which would be used as the criteria of identifying a correct client response.

(5) Even with its relatively limited response repertoire the Computer-Examiner identified a significant number of the client's responses under the Protocol and Student Administrations. The response evaluation efficiencies in Table 19 for these two groups were well above expectation. They suggest that it would be possible, maybe not feasible at the present time but definitely possible, to develop a computer model, which would be capable of administering and evaluating the Verbal Section of the WISC. Such a model could have response evaluation efficiencies in all subtests approaching 95% and 99% or even better. Undoubtedly the task would be extremely difficult with subtests such as Comprehension and Vocabulary, but as this study demonstrates, not impossible.

In making this prediction it should be kept in mind the Computer-Examiner was developed only to the end of a first generation stage and its response analysis capability was deliberately restricted. Under this condition it was still possible to correctly identify 42.7% of the Comprehension responses in the Student Administration - the Computer-Examiner's poorest performance. For demonstrating the feasibility of computer involvement in individualized intelligence testing, this low evaluation efficiency still indicates good future

TABLE 20

COMPUTER MISTAKES

VOLUNTEER ADMINISTRATION

ERRORS/SUBTEST	INFORMATION	COMPREHENSION	ARITHMETIC	SIMILARITIES	VOCABULARY	TOTAL
CONTROL WORD ERROR	-	2	-	-	1	3
SPELLING ERRORS	-	-	-	-	-	-
TYPING ERRORS	-	-	-	1	-	1
HUMAN ERRORS	-	-	-	-	-	-
WRONG CATEGORY	-	-	-	-	-	-
PROGRAMMING ERROR	5	16	-	-	6	27
NOT PROGRAMMED	-	-	-	-	-	-
TOTALS	5	18	-	1	7	31

TABLE 21

COMPUTER MISTAKES

PROTOCOL ADMINISTRATION

ERRORS/SUBTEST	INFORMATION	COMPREHENSION	ARITHMETIC	SIMILARITIES	VOCABULARY	TOTAL
CONTROL WORD ERROR	-	-	-	-	-	-
SPELLING ERRORS	-	-	-	-	-	-
TYPING ERRORS	1	-	-	-	-	1
HUMAN ERRORS	-	-	-	2	3	5
WRONG CATEGORY	-	-	-	2	-	2
PROGRAMMING ERROR	-	11	-	-	3	14
NOT PROGRAMMED	-	34	-	5	22	61
TOTALS	1	45	-	9	28	83

TABLE 22

COMPUTER MISTAKES

STUDENT ADMINISTRATION

ERRORS/SUBTEST	INFORMATION	COMPREHENSION	ARITHMETIC	SIMILARITIES	VOCABULARY	TOTAL
CONTROL WORD ERROR	3	7	-	-	4	14
SPELLING ERRORS	3	2	-	1	8	14
TYPING ERRORS	3	1	-	-	-	4
HUMAN ERRORS	-	-	-	-	3	3
WRONG CATEGORY	2	7	-	3	10	22
PROGRAMMING ERROR	-	2	-	4	4	10
NOT PROGRAMMED	5	27	-	15	39	86
TOTALS	16	46	-	23	68	153

possibilities.

3. ANALYSIS OF COMPUTER MISTAKES

Computer mistakes were defined to be those situations in which scores awarded for a response by the computer and human examiners did not match. These were identified and tabulated in Tables 4 through 18 (pages 234 to 253). Each of these mistakes were further reclassified into one of seven categories and tabulated in Tables 20, 21 and 22 for the Volunteer, Protocol and Student Administration groups. These categories were:

(1) Control Word Error - Several control words were active during administrations of the test. If typed, these words would cause the computer model to take specific courses of action as outlined in Chapter 4, page 58. They were: REPEAT, REST, STOP, I DON'T KNOW, and TEACHER. An error was made in the programming logic which searched for control words in the client's responses. The logic was set up to accept any occurrence of the letter sequence of the control words within the letters of other words in the response. The programming logic should have been designed to recognize the control words only if they appeared in the response exclusive of other words or letters. As a result the following kinds of errors occurred:

(A) Information question 9 - an acceptable response was Christopher Columbus. The underlined letters caused the model to interpret the response as STOP and attempted to sign the client off the test.

(B) Comprehension question 1 - an acceptable response included the phrase stop bleeding but the underlined word caused the same result as in (A).

(C) Comprehension question 6 - an acceptable response included the phrase it is more resistent to bad weather, but the underlined letters caused the model to interpret the response as a request for a REST and stopped its questioning.

(D) Vocabulary question 10 - an acceptable response contained the phrase a diamond is the hardest substance but the underlined letters caused the same result as in (C).

(2) Spelling Errors - Several mistakes were made because of words that were misspelled in the client's responses in spite of the efforts which were made to recognize them. Had these words been recognized, the responses would have been scored correctly. Examples were:

(A) Information question 6 - saway was meant to be safeway.

(B) Similarities question 2 - hass was meant to be hands.

(C) Vocabulary question 3 - hound was meant to be round.

(D) Vocabulary question 9 - it mes vat you rae bad was meant to be it means that you are bad.

(3) Typing Errors - The clients made typing errors and as a result responses were not recognized by the Computer-Examiner. Examples were:

(A) Spacing errors such as 2 000 for 2000.

(B) Run on words such as usedany for used any.

(C) Letter reversals such as watre for water.

(D) Striking the wrong key - such as grihds food for grinds food.

(4) Human Errors - It was found in a few client responses that

the computer's evaluation was correct, and the human's was not; correct in that the response was listed in the Massey Supplement.

(A) Similarities question 9 - both made into fuel was correctly categorized as a 2 point response by the Computer-Examiner but awarded 1 point by the human examiner.

(B) Similarities question 11 - metals was correctly categorized as a 2 point response by the Computer-Examiner while the human examiner awarded 1 point.

(C) Vocabulary question 9 - to join together was correctly categorized as a 2 point response by the Computer-Examiner. The human examiner questioned the response.

(5) Wrong Category Errors - The Computer-Examiner incorrectly categorized a number of client responses. Some examples were:

(A) Comprehension question 2 - buy him a new one (or find his lost ball and give her another doll were evaluated as being worth 1 point by the Computer-Examiner whereas the human examiner awarded 2 points.

(B) Vocabulary question 2 - a cushion is a bag filled with compressible material used by people for comfort was evaluated to be worth 1 point by the computer model while the human examiner awarded 2 points.

(C) Vocabulary question 7 - a nail is a straight piece used for fastening pieces of wood together (sharp at one end and blunt at the other) was evaluated as worth 1 point by the computer model and 2 points by the human examiner.

(6) Programming Errors - A number of client responses matched, word for w d, the key words of responses listed in the Massey Supplement, yet the computer model failed to recognize them because either the modeler failed to include them in the response repertoire or made mistakes in the programming logic. For example:

(A) Similarities question 10 - they are both used for measuring, the pound and the yard are used for measuring, they are both ways to measure something that is dry, and they are a unit of measure and both are not in the metric system, were not recognized by the Computer-Examiner. All should have been awarded 2 points as the key word in each was measure.

(B) Vocabulary question 8 - it is a little smaller than a horse and has smaller ears should have been recognized as the response like a horse only smaller which was included in the response repertoire.

(C) Vocabulary question 9 - the hair of an animal should have been recognized and evaluated as 2 points as hair was the key word in the response repertoire.

(7) Not Programmed - Those computer mistakes in which the response was not recognized and the keywords of the response were not included or only partially included in the response repertoire were classified as not programmed. Examples were:

(A) Information question 6 - in a dry goods store was not listed in the Massey Supplement.

(B) Information question 14 - it is where food begins to be digested and where different enzymes are used to break it down was only partially included in the response repertoire.

(C) Comprehension question 2 - try to keep him from ruining

himself should have been questioned but had not been included in the response repertoire.

4. INTERPRETATION OF COMPUTER MISTAKES

(1) In the Volunteer Administration 31 computer mistakes (Table 20) were generated over administration of the test to five clients; for the Protocol Administration 83 computer mistakes (Table 21) were generated with six clients, while in the Student Administration 153 computer mistakes (Table 22) were generated with eight clients. The ratio of computer mistakes per client for each administration was:

Volunteer Administration	6.20
Protocol Administration	13.83
Student Administration	19.12

The order of these ratios restate in a different way the previously obtained result in which the Volunteer Administration was found to have the highest response evaluation efficiency, followed by the Protocol Administration with the second highest and then the Student Administration with the lowest response evaluation efficiency.

(2) The subtest column totals of Tables 20 to 22 were abstracted from the corresponding subtest Computer Mistakes columns of Tables 4 through 18 (Pages 234 to 253). These totals cannot be interpreted in Tables 20 to 22 as they must be considered in relation to the other column totals of Tables 4 through 18. In this respect, they were used in the computation on subtest response evaluation efficiencies as tabulated in Table 19 and interpreted in point 2 of the discussion on

response evaluation efficiency (page 144).

(3) The row totals of Tables 20 to 22 show how Computer Mistakes were distributed over the seven error categories. The row totals of Tables 20 and 21 cannot be interpreted for the following reasons:

(A) There were not a sufficient number of responses over the seven categories to show a clear distribution of errors.

(B) The clients in the Volunteer and Protocol Administrations knew how to spell and type as they were graduate students and therefore biased the distribution of errors.

(C) These clients were also aware of the significance of typing control words inadvertently in their responses. This knowledge further biased the distribution.

(D) Human errors in the Volunteer Administration did not exist because the clients' responses were taken from the Massey Supplement.

Only the row totals of Table 22 gave an unbiased estimate of the distribution of computer mistakes over the seven error categories. These totals were each used as the numerator of a ratio to the sum total of computer mistakes for the Student Administration. The ratios are listed in the first column of Table 23. This column shows that 56.2% of the computer mistakes were due to the Computer-Examiner's inability to recognize client responses. Mechanical errors caused 30.1% of the computer mistakes. Of these, 14.3% were due to wrong categorization of responses, 9.3% due to the improper recognition of control words and 6.5% due to incorrect programming. Another 9.3% of the computer mistakes were attributable to the clients' inability to spell and 2.5% to their inability to type responses. Incorrect judgements by the human examiner accounted

TABLE 23

PERCENTAGE EFFECT OF COMPUTER
MISTAKE ERROR CATEGORIES ON THE STUDENT ADMINISTRATION
AVERAGE RESPONSE EVALUATION EFFICIENCY

ERROR CATEGORIES	% of COMPUTER MISTAKES	% CHANGE IN RESPONSE EVALUATION EFFICIENCY
NOT PROGRAMMED	56.2%	14.9%
WRONG CATEGORY	14.3%	3.8%
CONTROL WORDS	9.3%	2.5%
SPELLING	9.3%	2.5%
PROGRAM	6.5%	1.7%
TYPING	2.5%	.7%
HUMAN	1.9%	.5%
	100.0%	26.6%

for the remaining 1.9% of the mistakes.

The average response evaluation efficiency of the Student Administration was previously determined to be 74.3%. If it were possible to eliminate each of the error categories the corresponding increases in evaluation efficiency in the second column of Table 23 would take place. For example it is suggested in this column that if the response repertoire of the Computer-Examiner was expanded to include all of the responses received during the Student Administration, response evaluation efficiency would have increased by 14.9% from 74.3% to 89.2%. It is not unrealistic to suggest that response evaluation efficiency could be increased considerably by expanding the response repertoire beyond the limits of the WISC Manual and the Massey Supplement. It was observed that many of the responses of the performance recordings appeared to be quite common yet were not listed in either manual.

To what size would the response repertoire have to be expanded in order to increase evaluation efficiency by 14.9%; what would be the cost of planning, the number of man hours and the fiscal cost to achieve this goal? Such questions were not answered by this study. What has been suggested is that the goal is possible, provided that effective grammatical identification procedures can be developed.

In a second generation model it is reasonable to expect that errors due to incorrect categorization of responses, improper recognition of control words, or programming errors could almost be eliminated because they can be defined, traced to specific faults and corrected. Approximately 8.0% of the average response evaluation efficiency was accountable to mechanical errors.

Spelling and typographical errors lowered the average response evaluation efficiency by 3.2% and relative to other kinds of errors appear not to be significant. Yet, if the desire in constructing computer administered intelligence tests is to create models of practical significance, they must be considered relevant. Their elimination could be accomplished in two ways. First, they could be reduced by raising the lower age limit of the client. This is based on the assumption that older clients would have developed more skill in spelling and typing and therefore would make fewer mistakes. In the Student Administration, clients ranged in age from 8 years to 15. It was observed that spelling and typing errors tended to occur more with the younger clients. What would be the lowest age limit to which a practical computer administered intelligence test would be practical? This is a question which has been raised. Its answer would involve knowledge of a complicated interrelationship of many factors; for example, the kinds of questions being asked in the test (Information, Arithmetic, Comprehension, etc.), the types of terminal hardware being used to administer the test, or the required standards of test reliability. Before an effective lower age limit could be set, the relationships between such factors must be known.

A second way in which spelling and typographical mistakes could be reduced would be to develop algorithms which recognize and evaluate such errors. This approach would be more desireable than the first, because it would help to expand the age range of the clients to which a Computer-Examiner could administer the test. No attempt was made in this study to recognize spelling errors other than the simple

procedure of avoiding the use of vowels (as described in page 83). Several examples were observed in which this procedure increased the recognition of incorrectly spelled responses. For example in Information question 2 the answer was thumb, the skeleton answer was thm and thumb, thom, and thum were all recognized as correct. What has been suggested in this study is that spelling and typographical errors are identifiable and classifiable, and therefore it is reasonable to expect that procedures and algorithms could be developed and implemented to reduce them.

4. OBSERVATIONS AND RECOMMENDATIONS

During development, implementation and evaluation of the Computer-Examiner, the modeler made a number of observations. These are listed as follows with a number of recommendations and comments.

(1) The Computer-Examiner's response evaluation logic was developed on the basis of just five operators; the AND, OR and NOT operators, the GROUPING operator (parentheses sets), and an operator which instructs the computing system to recognize the indicated keyword or groups of keywords in the client's response only when they appear in the response exactly as they appeared in the reference list. These operators were developed for the 1500 Instruction System by Dr. Morris Peuchot (Roberge, 1971). It was observed, during the process of transforming the responses listed in the Massey Supplement into the response repertoire, that the five indicated operators were not sufficient. Several other operators and functions might have been used. In fact, the transformation would have been smoother and the resulting response

repertoire would have been more sophisticated had the APL (A Programming Language) set of operators and functions been available. This language has a vector capability which would have been extremely useful for referencing vectors of keywords and concatenation operator that could have been used to join keywords of vectors of keywords. The NAND, NOR or the EXCLUSIVE AND and OR operators could have been well used in the evaluation logic. The ARITHMETIC operators and the APL system's arithmetic computing power would have been useful for analyzing responses to questions of the Arithmetic Subtest. In addition, the APL system allows the programmer the option of creating his own functions. This facility would have been invaluable in the development of higher order functions of evaluation logic.

(2) During the Student Administration the modeler and consulting psychologist made a number of observations about the overt behavior of the clients. These were:

(A) The oldest clients tended to see the Administration as a challenge. The idea was to feed the Computer-Examiner an unusual, but correct response. For example, Vocabulary question 13 asks for a definition of the word sword and this kind of response was received: a weapon consisting of a sharp long blade with a handle designed to impale people and is rarely used now.

(B) The younger clients tended to think the Administration was a game, and except for the long introduction and some delays caused by the improper recognition of control words, they were very enthusiastic about the test. Most of them thought that it was a lot of fun.

(C) The younger clients had a great deal of difficulty in

catching on to the response entering procedure. They tended to leave out the punctuation marks, and of course, these were necessary for recognizing the end of a response.

(D) All of the clients, especially the older ones, were bored with the introduction on the typing of responses.

(E) Throughout the administration clients engaged each other in occasional play kidding and conversation, but this was very short because new questions were constantly asked, to which the client's attention would be drawn.

(3) The system used in the computer model to accept the client's responses proved to be generally cumbersome and frustrating to use. A number of times the older clients attempted to construct responses of more than 100 characters but were automatically cut off at the hundredth character. There were also several occurrences where the clients did not follow the strict routine for using two or three lines to enter their responses and as a result had to wait until the response acceptance program timed out or had their responses prematurely entered.

The response entering system was also cumbersome because clients were required to work their way through an introductory program designed to explain, demonstrate and exercise the procedures required to enter responses. It took clients from 5 to 10 minutes to complete this section depending on their skill to comprehend the instructions. Even after this introduction the younger children still made mistakes.

The problems encountered in entering responses were mainly due to hardware and software limitations of the 1500 Instructional System. The

System was unable to accept responses of more than 100 characters at a maximum of 40 characters per line. In addition, it was unable to sense the end of a line on the graphic display and automatically reset the cursor to the left hand side of the display on the next line. Future computer models should be implemented on a computing system which has the hardware and software capabilities to permit the client to type a response as though he were writing it on paper.

(4) Developing the Computer-Examiner's response repertoire was the most complex, longest and arduous task encountered while constructing the model. A great deal of effort was also expended in testing it out. Still programming errors and wrong categorizations accounted for a 5.5% drop in the average response evaluation efficiency of the Student Administration. Considering this in relation to the fact the repertoire was deliberately limited in size and sophistication, it suggests that future human involvement in the preparation of larger and more sophisticated response repertoires will be limited to a supervisory role. Because of the obvious complexity of the task, computers would be used. The basic steps of a procedure to develop such a repertoire would be:

(A) A computing system would be set up to administer questions of the proposed test to a group of clients of a statistically significant size. During administration of the test all responses to questions would be recorded but not evaluated.

(B) Each response in the performance recordings would then be examined, categorized and scored by a panel of human judges.

(C) The responses, and their specified categories and scores, would then be input as data to a computer program whose function would be to search out the commonalities in word patterns between responses within categories. This could be patterned upon the keyword approach used in this study or other types of pattern recognition algorithms like syntax or linguistic analysis, list processing, or some form of statistical averaging. The program would also list the recurring patterns in a descending order from the most probable pattern to the least probable to occur in a client's response.

(D) Data from the pattern recognition program would then be input into an encoding program whose function would be to translate the pattern recognition output into a series of logical statments (the response repertoire).

(E) The evaluation logic would be set up to check for the most probable patterns first.

(F) The pattern recognition program would be designed, not only to recognize patterns within response categories, but to recognize patterns that differentiate between categories in order to maximize separation. This would help to decrease the number of categorization errors.

(G) The process of categorizing responses, searching for keyword patterns, and generating evaluation logic could also be extended to the recognition of spelling and typographical errors. The human judges (Step B) could spot and mark such errors, the pattern recognition program (Step C) would include algorithms to categorize these errors, and the encoding program (Step D) would set up the evaluation logic to

to search for them.

(5) Not one response of either the Protocol or Student Administrations was categorized in the H or DZ categories (pp. 64 to 65). The H category was reserved for those responses evaluated to be homonyms of correct responses. The DZ category was reserved for those responses evaluated to be negations of correct responses. Set up of the DZ category was a complex task. This category was the sole purpose for the existence of Decision Table 3. It is recommended that these categories be dropped from future models. Questions that might possibly elicit homonyms as responses could be avoided. It might be possible to develop a generalized algorithm to recognize negations in responses. When a negation is detected, the client could be asked to restate the response without the negation.

(6) The control word functions were used well by the eight clients of the Student Administration, and are summarized as follows:

(A) The REST function was used six times by clients who left the testing room to take a break.

(B) The REPEAT function was called 43 times throughout the Administration (Tables 14 to 18, pp. 247 to 252); 14 of which were for Vocabulary questions 9 (Fur), 10 (Diamond), 11 (Join), and 31 (Ballast). Information question 30 (What is a lien?) was asked to be repeated 4 times. It is quite reasonable to expect that clients might wish to have these questions repeated.

(C) The I DON'T KNOW function was used 48 times (Tables 14 to 18, pp. 247 to 252). It was expected that this control function would tend to occur more frequently towards the higher numbered questions in each subtest but this was not the case. They tended to be scattered

throughout the questions. It appears the clients only used this function as the last resort to provide an answer, otherwise they tried to supply their own answer to the question.

(D) The TEACHER function was not used once during the Student Administration. Instead, the clients spoke to the supervisor directly.

(E) The STOP function was not made available to the client but several clients triggered the routine with the word STOP in their response.

The control word functions REST, REPEAT, and I DON'T KNOW should be retained in the next generation model. The TEACHER function should be eliminated. The STOP function should be set up so that only the supervisor can gain access to it.

(7) With little doubt the most frustrating and annoying experience encountered by most of the clients was due to improper recognition of control words. Each time these words were improperly recognized in a client's response a two or three minute delay was encountered and it became necessary for the supervisor to assist the client in restarting the questioning. One client had the misfortune to trigger the REST sub-routine four times in succession on the same question. If possible, control words such as STOP, REST, or I DON'T KNOW should be implemented as keyboard functions. If this is not possible, then the best alternative would be to set up the evaluatuion logic to recognize control words only when the client types them in absence of other words or characters.

(8) One advantage to having humans administer tests like the WISC is that humans are able to observe, qualitatively, a large range of behavior be it abnormal or routine, and thereby are able to capture

information about the client which might provide more insight and understanding of his or her intelligence. With the advent of computer examiners this capability might not be lost. It may change form; that is from observation of the client's overt behavior to a more detailed analysis of his responses through the performance recordings collected by the computer. From such an information base the human examiner might be able to continue making insightful interpretations or possibly gain an even greater understanding about the client's intelligence.

Consider as an example the following responses from the performance recordings of one eight year old client in the Student Administration (interpretation in parenthesis):

(A) I don't know - (I don't know)

(B) I wud go and am hip - (I would go and some help)

(C) hass - (hands)

(D) thea aer - (they are)

(E) theay aer anr in botl - (they are in bottle)

(F) saway - (safeway)

(G) cwo - (cow)

Setting aside the argument that the following suggestions could introduce a cultural or age bias into the measure of intelligence, these responses might tell something about the client. They could suggest that the client has specific spelling difficulties as well as a problem of reversing and also tends to read backwards. By correlating such a pattern of responses with a particular behavioral problem it might be possible for human examiners to make more accurate inferences about

behavior and intelligence and the client's ability to learn.

(9) Not only is there the problem of developing a response repertoire based on those clients who construct their responses with a correct grammatical style and spelling, but there is also the problem of developing a response repertoire, or series of repertoires, based on particular groups of clients who construct responses with a common style and form of spelling all their own. Before such response repertoires could be constructed, there would be a real need to collect data on clients who might form such groups.

(10) While examining the performance recordings of the Student Administration an interesting question developed. How would the client's typed response compare with his verbal response? This question might be answered in a project that asks the client to first give his response verbally and then to type the response. The verbal and typed responses would be recorded by the computing system for later analysis.

(11) Part of the introduction to the Computer-Examiner's administration was a section in which the supervisor was required to answer questions concerning the client's physical readiness to begin the test. When the eight clients of the Student Administration started the test, the supervisor had to hurry about the testing room and work with each client, one-by-one, through this section. The effect was to stagger the starting of each client. The eighth client was started approximately 10 minutes after the test began. Delays like this are obviously undesireable. The problem could be eliminated by having a bypass option built into the computer model that would permit the supervisor to checkout all the clients at once via group instruction, or the introductory section could be made more automatic.

(12) One other section of the Computer-Examiner's introductory section was concerned with confirming the client's name, age and sex. The section was very interactive in that it was designed to catch almost every conceivable mistake the client could have made in entering this information. It created a very good impression on those who made mistakes in stating their name, age or sex by asking them to type in the correct information. The program was also designed to be corrected by the supervisor if the information given by the client was correct while the computer was wrong. There was one such incident of this type. Information about a client's age was wrong, and had this mistake been allowed to continue the Verbal I.Q. score would have been calculated incorrectly. Introductions of this type should be continued in future models because it gives clients the impression or feeling that the machine knows who he is.

5. SUMMARY

The computer's presence is being increasingly felt in all areas of our society; one of these is education. Within education the computer's use is being explored on many frontiers. One such frontier is psychological testing. The purpose of this study has been to push back that frontier. It was an attempt to explore the feasibility of using the computer as a substitute in place of the human to administer and evaluate an individualized intelligence test. This was done through the construction of a computer model of the Verbal Section of the Wechsler Intelligence Scale for Children (WISC) which was developed, implemented and evaluated on an IBM 1500 Instructional System at the

University of Alberta.

The model's construction was made possible through use of well defined computer modeling process which has been successfully tried and tested in commercial, industrial and scientific areas. The major premise in this process is that models are abstractions of reality, and as such, can be carefully and progressively moved one stage at a time from the state of the real phenomena to the state of the abstracted model. This careful progression ensures that as much detail about the real phenomena as possible is retained in the model. Eight stages of progression were taken to develop the Computer-Examiner: inventory and classification of the phenomena's characteristics, selection of the characteristics to be modeled, conceptualization of the model, logical flow charting, program flow charting, writing the computer program, debugging and implementation of the program, and evaluation. Within each of these stages four basic steps of activity were executed: performance of the stage's function, validation of the function, modifications, and documentation.

Several characteristics of the WISC testing situation deliberately excluded from the computer model either because the necessary hardware and software needed to implement the characteristics did not exist on the 1500 Instructional System, or there was not enough time and resources to develop them in the model, or they could not be stated in operational terms. Development of the model was kept strictly to the basic characteristics of administering the test and evaluating responses.

The Computer-Examiner was evaluated with the aid of data obtained from the performance recordings of three administrations: the Volunteer Administration in which six graduate students acted as clients and took

the test under specified conditions; the Protocol Administration in which a school psychologist took the test six times and responded with responses from six protocols; and the Student Administration in which eight legitimate clients ranging in age from eight to fifteen years took the test.

During the Protocol and Student Administrations the Computer-Examiner did not make a single administrative mistake. It functioned perfectly in sequencing questions, tabulating Raw Scores and calculating Verbal I. Q. Scores. The Computer-Examiner's evaluation section performed exceptionally well in evaluating responses. It achieved an average response evaluation efficiency of 91.8% in the Volunteer Administration, which is a measure of how accurately the modeler programmed the response list of the WISC Manual and Massey Scoring Supplement into the response repertoire of the model. In the Student Administration the average response evaluation efficiency was 72.4%. This measure was high considering that the model's response repertoire was small and unsophisticated. It suggests there is a definite feasibility to the use of computers in individualized intelligence testing.

In Chapter 2 it was suggested that a number of reasons existed for justifying an investigation into the feasibility of computer examiners. It is appropriate to address these reasons based on the data and results generated from this study. First, it was suggested the concept of computer-human interaction was contemporary and hotly disputed by two groups. Some maintain that such interaction is feasible and would be acceptable to humans, while the others not only contend the computer is not capable of simulating such sophisticated human

skills as those needed to administer and evaluate individualized intelligence tests, but such activities could not successfully be carried out without the human element; the experience would be too dehumanizing. The author pointed out that he knew of no evidence to support either view and that studies such as this would begin to provide cumulative empirical evidence as to the degree of success or failure of the computer as an examiner. The results suggest it is possible to develop computer models which can duplicate the administrative and evaluative skills of the human examiner. No attempt was made to show that either the human element can be adequately simulated in the computer model, or that it would not be required to meet the needs of the client during administration of the test.

It was suggested that with commonplace use of computers as examiners it would be possible for human examiners to devote savings in administration time to tasks more demanding of their professional competence and skills. In the Student Administration the researcher supervised eight clients, and except for problems encountered in introducing the client to the test, helping clients who failed to follow the response input procedure, and restarting the administration after improper recognition of control words, he was hardly required at all. With the technical problems solved, a simpler hardware and software response acceptance system, and a comprehensive but simple and self-explanatory introduction, there would be no need to involve the professional psychologist in supervision of the administration.

Third, it was suggested the possibility of replacing a human examiner with a computer would become more apparent once the character-

istics of administering and evaluating an individualized intelligence test were more clearly known. The success of the computer model in this study adds a great deal of credibility to this possibility. It has been demonstrated that in spite of the WISC's apparent complexity a large proportion of the test consisted of very routine and highly standardized administrative and evaluative procedures. Good examples of this demonstration were the development of the Answer Hierarchy and Decision Tables. With these conceptualizations it was possible to evaluate responses, award correct score points and correctly sequence the order of questions. To the author's knowledge no similar operational description of the WISC's evaluative decision structure exists.

Additional benefits could be derived from conceptualizations such as the Answer Hierarchy and Decision Tables. They might be used to provide a clearer description of the WISC's evaluative process and could therefore be used as training aids to instruct potential testers in the correct evaluation of responses. In Chapter 1 it was indicated that development of computer models often lead to a clearer understanding of the original real phenomena. The author suggests that parts of the development of the conceptual model have done the same.

The last reason given justifying the need for studies of this kind was that construction of computer models would result in characteristics of the phenomena being modeled being translated into operational statements. The magnitude of the task undertaken to develop the Computer-Examiner was, to say the least, great and because of this it was only possible to create a conceptual model out of those characteristics that were readily expressible in near operational form. Modeling objectives

1 through 95 in Appendix B are those which are in a near operational form. Non-objective 1 through 17 in Appendix B represent those characteristics which could not be operationally stated. In spite of the roughness of the objectives list and the existence of the non-objectives list, it was still possible to create a computer model that closely approximated the administrative and evaluative characteristics of the test. Without a doubt with more time and resources subsequent development of the Computer-Examiner would result in sharper operational statements, a clearer conceptual model and therefore a much closer approximation to the real administrative and evaluative phenomena of the WISC.

The WISC's Verbal Section was selected as the test to be modeled because of its reputation in the field of intelligence measurement. It was speculated that if this test could be successfully modeled a great deal of credibility would be transfered to the concept of computer administered individualized intelligence tests and therefore this study would have achieved its purpose. In this respect the author is convinced this purpose has been achieved. To give further credibility to the concept of computer involvement, more studies of this type might be carried out with other tests such as the Stanford Binet. The Performance Section of the WISC would be an excellent choice for another study.

From the experience gained in this study the author is convinced that development of a practical computer examiner should not be geared towards development of a model which emulates completely any existing intelligence test. Too much time and effort would be geared towards the task of duplicating a test's particular idiosyncrasies. Rather

a practical computer model should be developed to take advantage of those qualities of the computer which would make a computer examiner superior to the human. In this respect, development should also be undertaken to explore what new sources of information the facilities of a computing system might be brought to bear upon the task of developing a more accurate measure of intelligence.

In conclusion, the Computer-Examiner represents a first step in what will hopefully be the first in a whole series of steps that will culminate in the practical use of computer administered individualized intelligence tests. This first step has shown the process by which such a goal can be achieved.

REFERENCES

Blocher, Donald H., "Counseling as a Technology for Facilitating and Guiding Change in Human Systems," Educational Technology. Vol. IX, No. 3, March 1969, pp. 15 - 19.

Borko, H., (ed.) Computer Applications in the Behavioral Sciences. Englewood Cliffs, N. J: Prentice-Hall, 1962.

Cogswell, J. F., and Estman, D. P., "Explorations in Computer-Assisted Counseling," Technical Memorandum Series, TM - 2582/000/00. Santa Monica, California: System Development Corporation, 1965.

French, Joseph L., and Tardibuono, John, "Numerical and Verbal Aptitude Tests Administered at the CAI Student Station," Computer-Assisted Instruction Laboratory, Semi Annual Progress Report, No. R - 9. December 31, 1967.

French, Joseph L., and Tardibuono, John, "Numerical and Verbal Aptitude Tests Administered at the CAI Student Station," Computer-Assisted Instruction Laboratory, Semi Annual Progress Report, No. R - 11. June 30, 1968.

Gathercole, C. E., Assessment In Clinical Psychology. Harmondsworth, Middlesex, England: Penguin Book Ltd., 1968.

Hickey, Albert E., Computer-Assisted Instruction: A Survey of the Literature. Newbury Port, Massachussetts: Entelek Incorporated, 3rd Ed., 1968.

International Business Machines, 1500 Instructional System, Introduction to Computer-Assisted Instruction and System Summary. San Jose, California: IBM Systems Development Division, 1967.

International Business Machines, 1500 Instructional System, Coursewriter II Author's Guide Part 1: Course Planning. San Jose, California: IBM Systems Development Division, 1967.

International Business Machines, 1500 Instructional System, Audio Programming Author's Guide. San Jose, California: IBM Systems Development Division, 1968.

International Business Machines, 1500 Instructional System, Coursewriter II, Author's Guide. San Jose, California: Special Systems Programming Documentation Department, 1968.

Lekan, Helen A., Index to Computer-Assisted Instruction. Boston, Massachusetts: Sterling Institute, 2nd Ed., 1970.

Littell, William M., "The Wechsler Intelligence Scale for Children: Review of a Decade of Research," Psychological Bulletin. Vol. 57, No. 2, 1960.

Loughary, John W., and Tandow, Murray, "Computers as Substitute Counsellors," Educational Technology. Vol. IX, No. 3, March 1969, pp. 33 - 36.

Martin, Francis T., Computer Modeling and Simulation. New York: John Wiley & Sons, Inc., 1968.

Massey, James D., WISC Scoring Criteria: Scoring Supplement for the Wechsler Intelligence Scale for Children. Palo Alto, California: Consulting Psychologists' Press, 4th Ed., 1969.

Naylor, Thomas H., The Design of Computer Simulation Experiments. Durham, N.C.: Duke University Press, 1969.

Pritsker, Alan B., and Kiviat, Philip J., Simulation with GASP. Englewood Cliffs, N.J.: Prentice-Hall, 1969.

Rath, G.J., Anderson, H.S., and Brainerd, R.C., "The IBM Research Center Teaching Machine Project," in Galanter, E.H. (ed.) Automatic Teaching: The State of the Art. New York: John Wiley and Sons Inc., 1959, pp. 117 - 130.

Roberge, Luc Paquin, Programmes de Support: Laboratoire de Pedagogie en Fomatique, Quebec: Ministere d'Education du Quebec, S.I.M.E.O, March 1971.

Romaniuk, Eugene William, Computer-Assisted Counselling. Unpublished Master's Thesis, The University of Alberta, Edmonton: 1968.

Treg, Ernest W., and Clark, Willis W., "URLTOO and URLTO1," The Harvard University Computer-Aided Laboratory Program Abstracts. Boston, Massachusetts: Harvard University, June 1969.

Wechsler, David, WISC Manual. New York: The Psychological Corporation, 1949.

APPENDIX A*

*The statements in lists 1 to 4 of this Appendix are direct quotations taken from the indicated pages in the WISC Manual.

1. LIST OF EXAMINER'S ADMINISTRATION BEHAVIOR

GENERAL TO WHOLE WISC

(1) The WISC should be administered carefully and according to the directions of the manual (p. 17).

(2) The absolute necessity that the examiner follow the directions in the manual cannot be overstressed (p. 17).

(3) It is necessary for the examiner to adhere to certain prescribed procedures which have been selected as superior to others on the basis of research (p. 17).

(4) The directions and the special conditions pertaining to certain tests should be so familiar to the examiner that he can follow the manual and record the responses without hesitation. However it is better to read from the manual directly than to change the wording of the instructions (p. 17).

(5) The administration of a standardized test requires a competent examiner, ...a friendly relationship between examiner and the subject (p. 17).

(6) Calm movements, and easy conversation between examiner can do much to relieve any uncertainty and tension which the child may feel on being subjected to a test (p. 17).

(7) In no case should the examiner indicate dissatisfaction with a response as given nor buildup an expectancy of approval in the subject so that giving no comment would be interpreted by him as disapproval (p. 19).

(8) He (the examiner) should not be discouraged if a young child is unable to finish the test at that time; more than one test period may be required (p. 20).

(9) Testing of very young subjects requires special consideration by the examiner ....he (the young subject) has a flow of alternating interest and indifference which must be felt sympathetically and manipulated by the examiner (p. 20).

(10) The examiner has to fit into the concepts the child has acquired through his experiences of satisfying adult-child relationships (p. 20).

(11) Some subjects are unable to participate in the test relationship if they cannot touch the examiner or are removed from him by the

width of the table; others express their satisfaction with the examiner by sharing with him personal experiences relating home and family. While sharing of experiences, spontaneously, is not credited on the test and does not enter into the final score, it provides the examiner with important information about the particular subject's level and flow of language, his ability to express his own ideas, his areas of interests, his concern with himself, and often provides insight into his special needs (p. 20).

(12) It is the responsibility of the examiner to know the routine to which the very young subjects are accustomed in their classroom and to fit his testing program into it (p. 21).

(13) The examiner of very young children must be prepared simultaneously to express affectionate, kindly concern for them, to manipulate test materials with ease, to meet the needs of a scientific attitude towards his work (p. 21).

(14) As regards the first situation it is important for the examiner not to assume any special attitudes but to treat the client (suspected abnormal) in as nearly as possible the same way as he would any normal client. Friendliness and warmth should, of course, characterize the examiner's approach, but gushiness and over solicitousness must be avoided (p. 21).

(15) One must, however, distinguish between negativism and shyness in a new situation (p. 21).

2. LIST OF EXAMINER'S ADMINISTRATION PROCEDURES

GENERAL TO THE WHOLE WISC

(1) If an adjustable table and chair are not available, the examiner should make certain that the feet of the young subject are resting on a box, and that his arms are in a normal position for handling the manipulative materials on the table (p. 17).

(2) The administration of a standardized test requires a competent examiner, properly organized materials, a quiet room... (p. 17).

(3) The performance materials which are to be handled by the subject should be kept out of his range of vision until the examiner is ready to present them (p. 17).

(4) The subject should be seated comfortably (p. 17).

(5) If necessary any directions may be repeated, but they may not be explained or paraphrased, except that some tests special rules about

about restating the questions are given (p. 18).

(6) The responses of the subject for the verbal tests should be recorded exactly as they are given (p. 18).

(7) If the subject does not do well on a test, it is helpful in maintaining rapport for the examiner to say quietly, "that one was a little hard; you will be able to do it when you are older," or some other supportive expression, and proceed to another test on which the subject is likely to do well (p. 18).

(8) Most examiners will find it convenient to administer the tests in the order in which they appear in the record form. Variations are in order and are permitted. For most subjects, practical experience has shown that the information test is a good starting point (p. 19).

(9) For various statistical, clinical, and administrative reasons the digit span test of the verbal series and mazes of the performance series are considered as supplementary tests to be given for special reasons:

(A) as an alternate if a regular test has been spoiled,

(B) if adequate time is available, and

(C) if the examiner recognizes that a specific handicap will prevent administration of a test (p. 19).

(10) The general rule, then, is to give the five verbal and five performance tests which are regularly scheduled unless one has clinically valid reasons for substituting an alternate (p. 19).

(11) The examiner working in his milieu (the public school environment) is required to adjust his schedule to the divisions of the school day. He must become familiar with the school program in which he is testing so that he may contribute to the excellence of school management. The school psychologist should be part of normal school living (p. 20).

(12) It is important that the examiner be known and accepted by the child to be tested and by the other children in his group. This may require that the examiner spend some time in the classroom, and for a friendly, unhurried relationship between the child's teacher and the examiner. The child should be made to feel that leaving the group for a test is a pleasant experience and that other children share his experience at other times (p. 20).

(13) Effort to establish rapport should be gradual; but if rapport cannot be established in a relatively short time, it is best to discontinue the test and arrange for another appointment (p. 21).

(14) Destruction of material, of course, cannot be permitted; but

the child should be allowed to roam about the room to get acquainted with his surroundings as well as with the examiner (p. 21).

(15) Children from undisciplined or undiscipling backgrounds sometimes try to take command of the situation. Here again leeway is permissible, but the examiner should maintain a sufficient degree of control to permit correct administration of tests (p. 21).

(16) In analyzing many test protocols and supervising test administrations, the author has observed two common administrative errors which affect scoring: "overquestioning" and attempts to elicit second responses, especially on the comprehension subtest. The statement "tell me more about it," implies that the answer is inaccurate or vague. Excessive questioning will usually channel additional responses which will artificially increase the score. The test, in part, is based on the ability of a child to express himself spontaneously without coaching or coaxing (Massey, p. 7).

(17) Some examiners typically alternate subtests from verbal to performance to avoid restlessness and fatigue. No standard pattern of sequence of subtest administration is mandatory (Massey, p. 7).

(18) Each subtest has a minimum number of consecutive errors before termination of the subtest. When testing children who display an inconsistent response pattern, it may be advisable to continue beyond the allowable number of errors. Additional points should be reported in parentheses with an explanation in the subsequent write-up. This method, in specific cases, gives a more meaningful appraisal of a child's intellectual potential (Massey, p. 7).

GENERAL INFORMATION

(1) Read each question as stated and in the given order.

(2) Subjects 8 and older, not suspected mental defectives. Subjects may begin with item 4 and be credited for items 1, 2, and 3 if he passes items 4, 5 and 6. If any of these is failed, administer items 1, 2 and 3 before proceeding further.

(3) If the response is not clear it is permissible to say "explain more fully," or "tell me more about it," but not to ask leading questions nor to spell the words (p. 61).

(4) Discontinue after 5 consecutive failures (p. 61).

GENERAL COMPREHENSION

(1) Read each question to the subject (p. 63).

(2) Sometimes subjects find it difficult to remember the entire question. It is, therefore, always permissible to repeat the question (p. 63).

(3) It is good practice to repeat the question if no response is obtained after ten or fifteen seconds, but no alteration or abbreviation is permitted (p. 63).

(4) It may be necessary to encourage the subject by such remarks as "yes" or "go ahead" (p. 63).

(5) If the response is not clear "please explain further" or "tell me more about it," may be added (p. 63).

(6) Discontinue after 3 consecutive failures (responses scored 0) (p. 63).

ARITHMETIC

(1) Problems 1 through 13 are read to the subject (p. 64).

(2) Problems 14, 15 and 16 are presented on separate cards for the subject to read (p. 64).

(3) There is a time limit for each problem (p. 64).

(4) Begin timing immediately after stating the problem (p. 64).

(5) For problems 1, 2 and 3 use the blocks for the block design test (p. 64).

(6) If the subject requests, or if it is apparent that he failed to understand, the problem may be repeated. However, the timing always begins at the end of the "first" statement of the problem (p. 64).

(7) For subjects 8 and older, not suspected mental defectives. Begin with problem 4 and give credit for problems 1, 2 and 3. If the subject fails to pass both problems 4 and 5, administer problems 1, 2 and 3. If the subject then passes problem 3, he is entitled to try problem 6, because he has not yet had three consecutive failures (p. 64).

(8) Discontinue after 3 consecutive failures (p. 64).

SIMILARITIES

(1) For subjects under 8, and older suspected mental defectives do the analogies section (p. 60).

(2) Analogies 0 before reading each item say "finish what I want to say." All four items are given to the subject (p. 60).

(3) For subjects 8 and older, not suspected mental defectives do similarities (p. 66).

(4) Similarities - say "in what way are a plum and a peach alike?" If the subject says that they are not alike, fails to answer, or gives an inferior reply, say "oh yes, they are both fruit, you eat both, and both have skins and seeds. Now tell me in what way are a cat and a mouse alike?" (p. 66).

(5) Similarities - if the subject fails on cat-mouse, explain again and go on to item 7, but give no further help (p. 66).

(6) Subjects who have not been given analogies and who score less than 3 points on similarities must be given analogies according to the instructions for younger children. They are given the total points earned on both parts of the test (p. 66).

(7) Discontinue after 3 consecutive failures (responses scored 0) on the similarities items (p. 66).

VOCABULARY

(1) Say "I want to see how many words you know. Listen carefully and tell me what these words mean. Bicycle ...what is a bicycle?" (p. 66).

(2) Proceed with the words in the order list, repeating at each presentation, "What is a --------? or What does ------- mean?" (p. 68).

(3) The examiner must be certain that he is using the standard pronunciations of the words (p. 68).

(4) Subjects 8 and older, not suspected mental defectives. Subject may be started with word 10. However, if the subject fails to give a 2 point definition for any one of the words 10 through 14, the examiner must return to word 9 and go backwards, word by word, until the requirement of five consecutive 2 point successes is met. Then he should proceed forward again from the word beyond the one which causes this digression (p. 68).

(5) Occasionally, it is difficult for the examiner to decide

whether the subject does or does not know the meaning of a word. In such instances it is permissible to say "Please explain a little more," or make some similar neutral inquiry. The following principles should govern the supplementary inquiry:

(A) if the examiner feels the response is vague (cannot readily be scored) questioning is permissible.

(B) the criterion examples of 0 responses (Appendix C) are given to illustrate marginal 0 responses, ones which reveal some understanding of the word but are nevertheless too poor to be scored 1 point. Responses scorable as 0 should be questioned if the examiner believes that the response actually given indicates that a better response can be evoked by neutral inquiry. The examiner, however, should avoid further questioning of a clearcut 0 response.

(C) Spoilage of response after inquiry. After questioning, a response may be scored lower if additional matter reveals serious misunderstanding, or if it reveals the first response as being mere verbalism.

(D) Under all circumstances the inquiry should be limited to saying "Please explain a little more," or very similar neutral (non-leading) statements.

(E) The following words sometimes evoke a response which is based on the homonym; fur (fir); sword (soared); gamble (gambol); recede (re-seed). If this occurs, the examiner pronounces the word again and asks "What else does ------- mean?" He does not spell the word for the subject. The scoring is based on the response to the intended word regardless of the response to the homonym (p. 68).

(6) Discontinue after 5 consecutive failures (responses scored 0) (p. 68).

3. LIST OF EXAMINER'S EVALUATION BEHAVIOR

GENERAL TO THE WHOLE WISC

(1) The examiner should be thoroughly familar with the criteria for scoring the tests before he undertakes to administer them (p. 18).

(2) This imposes on the examiner the responsibility for questioning the subject skillfully (p. 18).

(3) This places on the examiner a tremendous responsibility

(A) to be thoroughly acquainted with the principles of scoring

the test

(B) to know the meanings of the words and concepts involved and

(C) continually to be alert to the need for careful work in scoring each item (p. 22).

(4) All examiners will do well to check themselves from time to time against the following standards:

(A) Record points for each item correctly and legibly as the item is scored.

(B) If a test is begun beyond the first item (such as with item 10 on vocabulary with older children) enter the assumed credits properly.

(C) When bonuses are allowed, use the table carefully and record correctly.

(D) In determining the total number of points for a test, add them at least twice.

(E) Transfer the score for each test to the proper space on the front of the record form.

(F) Use special care in converting raw scores to scaled scores to IQS. If there is prorating to be done, check arithmetic.

(G) Compute age carefully after first verifying the birth date and testing date.

(H) Check all copying and conversions (p. 22).

(5) A final word of caution may be in order. In checking scores of protocols, the author was startled to discover how frequently simple errors of arithmetic or careless reading of conversion tables resulted in large changes in scores. The addition, the computation for chronological age, the child's birthdate, and other clerical work should always be carefully rechecked (Massey, p. 7).

4. LIST OF EXAMINER'S EVALUATION PROCEDURES

GENERAL TO THE WHOLE WISC

(1) Since the tests should not be scored at the time that they are administered; abbreviated recording provides for an unreliable basis for scoring; a full recording of the subject's answer permits

a subsequent evaluation and a fuller consideration of them in comparison with responses obtained on other tests (p. 18).

(2) It is well for the examiner to record any unusual reactions of the subject to test items (p. 18).

(3) Whenever the examiner questions a response he should record the letter Q with the subject's answer following; often the response to the query determines the score for the item (p. 16).

(4) To increase scoring accuracy and for better qualitative understanding of the subject's responses, the examiner may find it necessary to query his subject on some items. This imposes on the examiner the responsibility for questioning the subject skillfully (p. 18).

(5) One must not suggest to him that his answer can be improved (p. 18).

(6) Query must be neutral; i.e. nonevaluative. One may ask the subject to "Please explain further," or to "Tell me more about it;" this will enable him to clarify his statement so that it can be scored properly (p. 18).

(7) Questioning must not become a procedure for pressing for the limits of his knowledge (p. 18).

(8) When a child gives multiple responses, either spontaneously or after querying by the examiner, it is fair to credit him with the best responses if

(A) all parts of his responses are acceptable but one part is better, or

(B) if his second response really clarifies his first, more vague response. However, the subject may sometimes add to his initial response in such a way as to spoil it. The examiner should observe the suggestions given for some tests with regard to such spoilage. Upon querying he should be credited if he really knows what his words implied and should be failed if his added comments do not show that he knew the first meaning given (p. 18).

(9) Compute the age of the child. Having determined the exact birth date of testing the subject, age is computed by subtraction (p. 22).

(10) Select the proper scaled score table. A subject's age in years and months determines which table of scaled scores is appropriate for him. These tables are printed on the pages following this chapter and are provided for four-month spans. The first age given is the lowest age for that age span: 5 - 0 through 5 - 3 includes children who are past their fifth birthday but have not yet completed their fourth month; 6 - 8 through 6 - 11 includes children who have at

least completed the eighth month beyond their sixth birthday but are not yet seven years old (p. 23).

(11) Determine scaled scores. Having located the proper table for the subject's age, the examiner enters the column for a given test and finds the raw score earned on that by the subject. Then he finds in the extreme left or right column of the table the scaled score which is equivalent to the obtained raw score. He does this for each of the tests. He records the scaled scores on the record form in the spaces just to the right of the raw scores. When he has done this, the record form shows a column of raw scores and an adjacent column of scaled scores. No more attention need be given to the raw scores since they have all been transformed into scaled scores which are more meaningful and can be treated arithmetically (p. 23).

(12) The verbal score is secured by adding the five scaled scores of the five tests which have been administered. (See notes below regarding prorating if four or six tests rather than five have been given) (p. 23).

(13) The performance score is the sum of the five performance scaled scores (p. 23).

(14) The full scale score is the sum of the verbal score and the performance score. It is based on ten tests (p. 23).

(15) Notes on prorating scores. As discussed, there will be times when more than five or fewer than five verbal or performance tests will be given. Since the tables of intelligence quotients are based on the sum of five tests for the verbal score and five for the performance score and ten tests for the full scale score, it becomes necessary in such situations to prorate the sums before looking up the I.Q.'s (p. 23).

(16) If six tests have been given, the sum of the six scaled scores must be multiplied by five-sixths (5/6) to reduce it to the equivalent of five tests (p. 24).

(17) If four tests have been given (or five given and one has been rejected as being spoiled or invalidated), the sum of the four scaled scores must be multiplied by five-fourths (5/4) to expand it to the equivalent of five tests (p. 24).

(18) One can, of course, also prorate for three tests, but it is expected that the WISC I.Q.'s will always be based on at least four verbal and four performance tests (p. 24).

(19) If the verbal and performance sums have been prorated, the symbol PR (meaning prorated) should be written in the margin near the prorated score which is to be used for computing I.Q.'s. Of course, one never prorates the full scale score since it is secured by adding the previously prorated verbal score and performance score (p. 24).

(20) By use of Tables XIA and XIB the examiner can now convert the three scores (verbal, performance, and full scale) into intelligence quotients. These values are entered on the appropriate spaces on the record form (p. 24).

(21) All responses listed under "question" and some listed under "zero" should be clarified by questioning. Where there is a possible two point score, some one point responses should be questioned. However, it is not the examiner's role to try to elicit two point responses to each question, but primarily to clarify what the child attempted to say. General scoring principles should include:

(A) When a child changes a correct response or denies it, the latter answer is scored. Example: Discovered - America. "Columbus - or was it Washington? No, Washington." Score this zero.

(B) If one correct and one incorrect response are given, the child should be asked, "Which one is it?" Example: Color - rubies. "Red or white." Ask the subject, "Which is it?" If he cannot answer, score zero.

When two responses of different value are given, give credit for the higher one. Example: Cut - finger. "Tell my mama and then I'd put a bandaid on it."

(D) If a child gives a creditable response followed by a contradiction, score zero. Example: Fight. "I'd walk away - the next time I saw him I'd beat him up." (An exception to this rule is found in vocabulary. If a child gives a positive response and then an incorrect response, score the positive response. Example: Spade. "It's a shovel thing and also a spade ship (space ship).

(E) If questioning lowers the score, the examiner should consider scoring the initial response. The rationale for this is that the subject's anxiety or defensiveness may have been aroused by the examiner's questioning.

GENERAL INFORMATION

(1) Each item is scored 1 or 0. Essentials of acceptable answers are noted below (p. 61).

(2) Maximum score: 30 points (p. 61).

GENERAL COMPREHENSION

(1) Each item is scored 2, 1 or 0. See Appendix for specific scoring criteria and sample answers (p. 63).

(2) Maximum score: 28 points (p. 63).

(3) The examiner should match the subject's responses against the criteria and sample answers given below (Massey). The examiner will undoubtedly find unusual responses which are not typified, as no attempt has been made to cover all possible replies. In these instances he will have to use his own judgment. Most of the 0 point examples given typify marginal responses: those which contain evidence of some understanding may be queried in the neutral manner prescribed (Massey, p. 15).

(4) Items 1 - 5. They are "What to do" items. Allow 2 points if the subject indicates that he understands what should be done and assumes personal responsibility for doing it. Allow 1 point if the subject indicates that he understands what should be done but places the responsibility on someone else or sets up qualifying conditions for himself (p. 15).

(5) Items 6 - 14. For each of these items several acceptable responses are noted. In general, 2 points are given for a response which includes at least two of the general reasons. A response which has only one of them mentioned is given 1 point. Typical responses are credited 2, 1 or 0 points. Note that for item 14 the degree of abstraction is particularly stressed. (Do not encourage a second response other than a short period of silence following the subject's first response (Massey p. 19).

ARITHMETIC

(1) Each problem is scored 1 or 0. Problems 2 and 3 are given 1/2 credit each if the subject makes an error but corrects it within the time limit (p. 65).

(2) Maximum score: 16 points. (Round half score upward) (p. 65).

SIMILARITIES

(1) 1 point for each item correctly answered. The only correct analogies - 1 point for each item correctly answered. The only correct responses for the items are, respectively: 1. sweet; 2. arms, hands; 3. women, ladies, mothers; 4. cut (p. 66).

(2) Similarities - 2, 1 or 0 points for each similarities item (5 - 16), depending on the degree and quality of the generalization. See Appendix B for scoring criteria and sample answers (p. 67).

(3) Similarities - If the subject gives multiple acceptable

responses, credit him on the basis of the best response. If the subject gives differences or wrong answers in addition to some correct responses, then ask, "Now which one is it?" and score according to this decision. Spontaneous improvement of responses is easily recognized and query is not necessary. Added remarks obviously not part of the subject's answer do not enter into the scoring. Example: cat-mouse - Both have eyes, tails, legs ...my cat caught a mouse (p. 67).

(4) Similarities - maximum score: 28 points. The sum of points earned on analogies and similarities (p. 67)

VOCABULARY

(1) Each word is scored 2, 1 or 0, except that words 1 - 5 are scored only 2 or 0. Specific instructions are given in Appendix C (p. 68).

(2) The general rule for scoring is that any recognized meaning of the word is acceptable, elegance of language and precision being disregarded. Most of the words in the list can be scored 2, 1 or 0 without difficulty. (A distinction must be made in the scoring responses for the first five words. Words 1 - 5 should be scored as 2 or 0. This arbitrary decision has been necessary because of the young child's limited ability to give verbal expression to his concepts. Therefore, any definition which indicates that the young subject knows what one of these five words mean is credited 2 points.) While elegance of expression is disregarded, poverty of content is penalized to some extent. If a subject seems to know only vaguely what a word means, his response is credited with 1 point. Thus, 1 point is given for a response such as (nonsense) "Somebody trying to be funny." The following general scoring principles should be helpful:

(A) 2 point responses

(a) a good synonym
(b) a major use
(c) one or more definitive features or primary features of objects
(d) general classification to which the word belongs
(e) a correct symbolic use of the word - example: "An umbrella of artillery for the advancing infantry."
(f) several less definitive but correct descriptive features which cumulatively indicate understanding of the word (p. 32).

(B) 1 point responses

In general, a response which is not incorrect but shows poverty of content.

(a) a vague or less pertinent synonym
(b) a minor use, not elaborated

(c) attributes which are correct but not definitive or not distinguishing features
(d) examples using the word itself, not elaborated

(C) 0 point responses

(a) obviously wrong answers
(b) verbalisms - example: "A brave man," when no real understanding is shown after inquiry
(c) not totally incorrect responses, but ones which, even after questioning, are very vague or trivial or show a great poverty of content.

(3) The formal definitions are taken from Webster's New International Dictionary, second edition, unabridged; to save space, some of the most esoteric meanings are omitted (Massey, p. 32).

(4) With the aid of these three guides - general principles, examples, and definitions - the examiner should be able to evaluate each response without too much difficulty (Massey, p. 33).

(5) Maximum score: 80 points (p. 68).

5. LIST OF 1500 INSTRUCTIONAL SYSTEM PHYSICAL CHARACTERISTICS

(1) The 1500 System is divided into seven major physical units:

(A) The 1131 Central Processing Unit,

(B) The 1502 Station Control Unit,

(C) The 2310 Disk Storage Unit,

(D) The 1442 Card Reader Punch Unit,

(E) The 1132 Line Printer Unit,

(F) The 2415 Magnetic Tape Unit, and

(G) 16 Instructional Stations.

The organization of the 1500 Instructional System is diagrammed in Figure 1.

(2) The Central Processing Unit (CPU) is the computer for the 1500 Instructional System. It executes the stored program instructions that direct the operations of all other units and devices in the system. It contains the controlling circuitry needed to perform the physical operations of transferring data from one unit to another. The 1131 CPU has the following features:

(A) A core storage capacity of 32K (32,000) 4 BYTE (16 BITS) words,

(B) A core storage access cycle time of 3.6 micro-seconds,

(C) One on-line high speed single disk storage unit with a removable disk cartridge of a storage capacity of 512,000 words and a data transfer rate of 36,000 words per second, and

(D) Communication capabilities with external devices.

(3) The 1502 Station Control Unit connects the CPU to all of the system's instructional stations. It controls the movement of course materials and instructions to and from each of the subunits of the instructional stations. It houses the video buffer disk upon which the course material is stored. This video buffer regenerates the information to be displayed on the face of instructional display.

(4) The 2310 Disk Storage Unit is used in the 1500 System to store course material, performance recordings, and other administrative information. It has these features:

(A) Four disk storage drives which provide the system with a random access storage of up to a maximum of 2,048,000 words of storage.

(B) A data transfer rate of 36,000 words per second, and

(C) An access time of 22.5 milliseconds.

(5) The 1442 Card Reader Punch is used to read course material into the 1500 System on the format of standard IBM 80 column cards, and to punch out course material on the same format. It has a capacity of processing 300 cards per minute reading and 60 cards per minute punching.

(6) The 1132 Line Printer Unit is used in the 1500 System to print out student performance recordings, program listings, and control information for the general operation of the system. It prints a maximum of 80 lines per minute.

(7) The 2415 Magnetic Tape Unit is used in the 1500 System to store inactive material and to provide on-line storage for student performance recordings. It has two standard nine track tape drives, a read-write capability of 15,000 BYTES per second, and it is important to note that this unit is compatible with the IBM System 360. This will permit performance recordings to be processed directly from tape on the 360 System.

(8) There are 14 instructional stations in the 1500 System, each of which is equipped with:

(A) a 1510 Instructional Display with keyboard,

(B) a light pen,

(C) a 1512 Image Projector, and

(D) a 1505 Audio Tape Unit.

There are also two instructional stations in the system which are equipped with one 1518 typewriter. These stations are used by the proctor to monitor error messages from anyone of the 14 instructional stations.

(9) The 1510 Instructional Display with keyboard unit allows for two-way communication between the operating system and the student through a typewriter keyboard and cathode ray tube display (CRT). It has these features:

(A) The display screen is divided into 40 vertical columns and 32 horizontal rows which can display 40 characters on each of 16 lines,

(B) Characters are displayed as a configuration of white dots on a black background, each occupying an area of 8 dots wide by 12 dots high,

(C) A total of 640 characters may be displayed at one time,

(D) A graphic of any combination of dots may be generated that ranges in size from a 8 x 12 dot area to the entire screen, which has a dot area of 192 x 320,

(E) The student can enter responses or communicate with the system via the keyboard,

(F) The keyboard keys are analogous to the standard typewriter key set except for the alternate coding key, and the reverse index key. The alternate coding key is used to signal special control functions to the system, while the reverse index key moves the cursor one row up the display screen.

(G) A cursor is a special character in the dictionary that is displayed on the CRT to indicate where the next keyed-in character will appear on the screen.

(10) A special pen at the right of the keyboard is available for student response. The student points to a specific location on the screen, thereby indicating his response to a question asked. The program can be written to interpret responses from the light pen. The only requirement is that there is a lighted area on the display screen when the student pushes the pen against the screen over the lighted area.

(11) The 1512 Image Projector is a display screen which can show

color or black-and-white 16 millimeter still images. Interchangeable cartridges containing the film are inserted into the projector. These cartridges can hold as many as 1,000 frames and can be accessed by the author at random. The moving speed of the film during positioning is 40 frames per second. Frames may be prepositioned before they are called.

(12) The 1505 Audio Tape Unit plays and records audio messages on interchangeable magnetic tape cartridges. Audio messages can be played through a set of earphones to the student and recordings can be made of the student's voice via a microphone attached to the earphone set. Each tape unit has a volume control which can be adjusted for a comfortable listening level.

Each tape cartridge can hold up to two hours of messages. These messages can be recorded on any one of three tracks. Therefore, for any given tape address, any one of the three messages would be available to the author for a presentation to the subject. The tape can be moved forwards and backwards as required at a speed of 15 inches per second. Messages can be of variable length, and each message can be placed for optimal retrieval efficiency.

(13) The terminal room in which the instructional stations are housed is a large room with dimensions of 15 x 54 feet. The room is quiet for a room of that size. The quietness is due to the carpeted tile floor, acoustical tile ceiling, and draperies over the windows. The south wall is a complete bay of windows each covered with a different colorful pastel drapery. During normal operation these drapes are drawn and the room is lighted by a combination of fluorescent lights and a number of adjustable incandescent lamps. The north and east wall contain a number of one-way windows from which observations can be made of the activities taking place in the room. The west wall contains a blackboard.

In general, if the students are working at all of the terminals, the only noises that can be heard are those of the equipment cooling fans and the sounds of keys being struck on the instructional keyboards. There appears to be very little distraction in the room except for the movement of the proctor and others making observations. A proctor is usually available and moves about the room freely assisting and supervising the students.

6. LIST OF 1500 INSTRUCTIONAL SYSTEM OPERATING CHARACTERISTICS

(1) The 1500 Operating System (A Supervisory Program System) functions under the direction of a programming language called Coursewriter II (CW). Authors used the instructions available in this language to instruct the operating system as to what operations it is to perform, when they are to be performed, under what conditions to do

them, which of the instructional stations are to be used, where the data for the operation is located and where the data from the results of the operation are to be stored.

(2) There are two major parts to the operating system:

(A) The Coursewriter Assembler, and

(B) The Coursewriter Interpreter.

(3) The assembler converts the author's course material into a format that can be more readily used by the interpreter at the time of executing the program.

(4) The interpreter executes the operations that have been coded into the Coursewriter instructions that make up a program. This is the interpreter that presents course material to the student.

(5) The Coursewriter II language is an instruction set with special control facilities. The instructions are used to present course material to the students, control their path through the course, ask for responses, evaluate these responses, and to make recordings of their performance.

(6) The control facilities manipulate course flow and record the activities of students during instruction. These facilities are:

(A) Answer analysis,

(B) Performance recording,

(C) Timing control,

(D) Explicit branching,

(E) Implicit branching,

(F) Buffers,

(H) Switches,

(I) Labels,

(J) Return registers,

(K) Restart points,

(L) Macros, and

(M) Functions.

(7) Coursewriter provides for answer analysis through an automatic matching facility. The author defines the expected responses and

them, which of the instructional stations are to be used, where the data for the operation is located and where the data from the results of the operation are to be stored.

(2) There are two major parts to the operating system:

(A) The Coursewriter Assembler, and

(B) The Coursewriter Interpreter.

(3) The assembler converts the author's course material into a format that can be more readily used by the interpreter at the time of executing the program.

(4) The interpreter executes the operations that have been coded into the Coursewriter instructions that make up a program. This is the interpreter that presents course material to the student.

(5) The Coursewriter II language is an instruction set with special control facilities. The instructions are used to present course material to the students, control their path through the course, ask for responses, evaluate these responses, and to make recordings of their performance.

(6) The control facilities manipulate course flow and record the activities of students during instruction. These facilities are:

(A) Answer analysis,

(B) Performance recording,

(C) Timing control,

(D) Explicit branching,

(E) Implicit branching,

(F) Buffers,

(H) Switches,

(I) Labels,

(J) Return registers,

(K) Restart points,

(L) Macros, and

(M) Functions.

(7) Coursewriter provides for answer analysis through an automatic matching facility. The author defines the expected responses and

classifies them as a correct answer, wrong answer, synonymous correct answer, synonymous incorrect answer, additional acceptable answer, or an unrecognizable answer. When a student enters a response the interpreter automatically attempts to match it with one of the author's anticipated responses. The direction of subsequent material is determined by the instruction which follows the match of a response and this is author controlled.

(8) The performance recording is a record of the actual path taken by each student, the questions he encounters, and his responses to these questions. The interpreter automatically records the following types of information in the performance recording:

(A) The student's name and the name of the course,

(B) They time required to complete the response (response latency),

(C) They type of response (incorrect, correct, etc.),

(D) The actual response accompanied by a response identifier of up to 10 alphanumeric characters,

(E) The contents of any counters, switches, and buffers, and

(F) The restart points past.

(9) The interpreter has access to a system clock and thus the author can specify that certain operations are timed. This clock allows the author to specify time intervals to within a tenth of a second. For example the author can determine if the student responded to a question in 5 seconds or 4 minutes, or he can specify that the student is to respond to the question within 30 seconds. If the student responds "in time" the course will proceed in one direction. If he does "time out," he may be taken in another direction. The time control can be used to specify pauses between instructions in order to give the student a chance to read the course material, or to work on the problem.

(10) The interpreter normally executes Coursewriter instuctions sequentially (instruction 1 is followed by instruction 2, then 3, etc.). The facility exists for explicit branching which is simply directing the interpreter to any other instruction other than the next one in the sequence. Explicit branches are made as a direct result of a special class of branch instructions which the author deliberately codes into the Coursewriter program to change course flow if certain conditions occur. If the conditions do not exist then course flow takes place in a sequential fashion as normal.

(11) The interpreter has a capability for implicit branching. These are branches that are made automatically as a result of a decision which the interpreter makes. For example, the interpreter will analyze the next member of an answer set if the current member

being examined does not provide for a match to the student's response. If a match takes place, the interpreter will automatically decide not to examine the next member of the answer set, but will go on to the next problem. Implicit branching is used to save the author from having to code certain obvious operation.

(12) Coursewriter makes provision for six 100 character buffers for each instructional station. All response data from the student is recorded in buffer 0. Certain instructions allow the author to put information into a buffer, to transfer information between buffers, and to move information from the buffers to other parts of the system. Information in a buffer can also be displayed on the instructional display.

(13) Coursewriter makes provision for 30 counters for each instructional station. The counters can contain integers between -32,768 and +32,767. Basic arithmetic operations of add, subtract, multiply, and divide can be performed under control of the author's instructions in the coursewriter program. The contents of counters are automatically written into performance recordings. Counter 0 is reserved for the system to record response latency.

(14) Coursewriter makes provision for 32 switches that are binary recording devices used to keep track of yes or no conditions that can occur during execution of the program. When a switch is on, it will contain a 1; when it is off, it will contain a 0. Switches may be repeatedly used in a course and are initiated or set by a number of instructions. If the author has need of more switches, he can use any-one of thirty counters as switches. Each counter is equivalent to sixteen switches. The contents of switches are automatically written into performance records.

(15) Coursewriter provides for a label facility. Labels consist of up to six alphanumeric characters in any combination the author wishes. Each coursewriter instruction is associated with a particular address or area in the system where it will be located. If the author wishes to refer to a particular instruction he must do it through the use of labels which enable him to symbolically address the instruction he wishes to execute next.

(16) Coursewriter has six return registers reserved for each instructional station. The author uses these registers to store the location of a student's return point (location of the next instruction to be executed), or for a future transfer point.

(17) In the event that a student signs off a course before it is completed, or the operating system encounters an error that causes it to cease functioning, a restart facility is provided. This facility allows the author to define convenient restart points in the program. They may locate anywhere the author so chooses to put them.

(18) Certain sets of Coursewriter instructions are used repeatedly. To simplify the problem of coding these sets over and over, Course-

writer provides a macro facility. This allows the author to build a framework, or skeleton, of a set of instructions called macro routine. Once this routine has been written, it may be called by one instruction and assembled into the program. The macro processing facility thus relieves the author of a lot of repetitive coding operations.

(19) Coursewriter provides for the facility of using functions which are special sets of instructions coded in 1130 assembler language, and are available for the author to help perform special processing not provided for by the specific instruction which initiates the course flow decision table for functions, executes the function, then returns control back to the interpreter. The following list is an example of some of the routines that are available as functions that operate on student responses:

(A) Upshift or downshift all alphabetic characters,

(B) Replace any group of synonyms with a fixed word or group of words,

(C) Replace or delete specified characters or punctuation,

(D) Extract the integer portion of a response and place it into a buffer,

(E) Generate random numbers,

(F) Search for key letters in a response,

(G) etc.

(20) The 1500 Instructional System is capable of presenting course material through either the:

(A) Instructional display,

(B) Image projector, or

(C) Instructional display keyboard, or

(D) Light pen.

Only those responses entered via the last three devices can be processed by the system. It is important to note that the system cannot process verbally recorded responses. They can be stored for either replay to the student, or instructor after the student has finished the lesson.

(22) If a response is expected from the student in the form of a verbal message a "beep" start tone announces the beginning of the recording and its termination. The student is expected to record his response between these two tones.

(23) If a response is to be entered into the system via the light pen, a "P" is displayed in the lower right hand corner of the instructional display. The student then chooses the lighted area he feels is the correct response to the question and presses the light pen directly over this area.

(24) If a response is to be entered into the system via the keyboard, a "K" is displayed in the lower right hand of the instuctional display and a cursor appears in the position of the display where the first character of the response is expected. As each character is typed, the cursor moves to the right and is replaced by the typed characters.

(25) The time of accepting a response can be variably set from zero to approximately 16 minutes in increments of 1/10th of a second.

(26) Two types of responses are acceptable to the operating system:

(A) Single entry, and

(B) Multiple entry responses.

An entry is defined as either a single pressing of the light pen to the instructional display or a series of keyed characters terminated by the "enter" function. The enter function signals the system that the student has completed a response from the keyboard. This is also duplicated by the use of the return key. When a student presses the return key he has indicated that he is finished entering his response.

(27) All responses are initially placed in buffer 0 which has the capacity of holding 100 characters. Multiple entries are placed consecutively in this buffer. If the student attempts to key more than 100 characters into buffer 0, the remaining characters are ignored. Responses can be moved to other buffers from buffer 0 and thus the response capacity of the system at any one given time can be as great as 600 characters (one hundred in each buffer).

(28) Each response to a question is individually identified by they system and placed into the performance recording of the student along with the identifier which is a 1 to 10 alphanumeric character chain.

(29) A performance recording is made each time the student makes a response, or can be made by the author when he requires it.

APPENDIX B

1. LIST OF COMPUTER MODEL OBJECTIVES

EXAMINER'S ADMINISTRATION BEHAVIOR

1. GENERAL TO WHOLE WISC

(1) To provide a calm and easy, but skillful manipulation of the test material so as not to create any uncertainty or tension in the client.

(2) To have the client participate in a test relationship in which he is able to touch the examiner.

(3) Not to assume any special attitudes towards the subject suspected of being abnormal, but to treat him, as nearly as possible, in the same way as he would any normal subject.

(4) Not to indicate dissatisfaction with a response as given nor build up an expectancy of approval in the client so that giving no comment could be interpreted by him as disapproval.

(5) To administer the WISC carefully and according to the directions of the manual.

(6) To record all of the responses of the client without hesitation.

(7) To be able to allow the client to finish the test in more than one test period.

EXAMINER'S ADMINISTRATION PROCEDURES

1. GENERAL TO WHOLE WISC

A. QUESTIONING

(8) To question the client skillfully in order that his score is more accurate.

(9) It is better to read from the manual directly than to change the wording of the instructions.

B. REQUESTIONING

(10) If necessary, directions may be repeated but they may not be explained or paraphrased, except that in some tests special rules about restating the questions are to be followed.

(11) Questioning must be neutral; i.e. nonevaluative. The examiner may ask the client to "please explain further," or to "tell me more about it." This will permit the client to clarify his statement so that it can be properly scored. Other similar neutral (non-leading) statements may be made.

(12) When the client gives differences or wrong answers in addition to some correct answers ask "Now which one is it?"

(13) When a response is given based on a homonym, the examiner pronounces the word again and asks "What else doe _____ mean?" He does not spell the word regardless of the quality of the response to the homonym.

(14) The client must not be overquestioned.

(15) A second response must not be elicited by the examiner.

(16) Questioning is not to become a procedure for pressing the limits of the client's knowledge.

(17) Suggestions must not be given to the client that would indicate the answer could be improved.

(18) Do not ask leading questions or spell the words.

C. ORDER OF TEST ADMINISTRATION

(19) Variations in the order of administration of the subtests are permitted. Most examiners administer the subtests in the order in which they appear in the record form. Information has been considered as a good starting point.

D. CLIENT MANAGEMENT

(20) The Performance materials which are to be handled by the client should be kept out of his range of vision.

(21) Destruction of materials by the client cannot be permitted.

(22) The client should be allowed to roam about the room to get acquainted with his surroundings as well as with the examiner.

E. ENVIRONMENTAL CONDITIONS FOR TESTING

(23) The client must be seated comfortably.

(24) The client must be able to manipulate the test materials with ease.

(25) The testing room must be quiet.

2. GENERAL INFORMATION

(26) Read each question as stated in the given order.

(27) Clients 8 or older, not suspect mental defectives. Clients may begin with item 4 and be credited for items 1, 2 and 3 if he passes items 4, 5 and 6. If any of these is failed, administer items 1, 2 and 3 before proceeding further.

(28) If the response is not clear it is permissible to say "explain

more fully," or "tell me more about it," but not to ask leading questions nor to spell the words.

(29) Discontinue after 5 consecutive failures.

3. GENERAL COMPREHENSION

(30) Read each question to the client.

(31) Sometimes clients find it difficult to remember the entire question. It is, therefore, always permissible to repeat the question.

(32) It is good practice to repeat the question if no response is obtained after ten or fifteen seconds, but no alteration or abbreviation is permitted.

(33) It may be necessary to encourage the client by such remarks as "yes" or "go ahead."

(34) If the response is not clear "please explain further," or "tell me more about it," may be added.

(35) Discontinue after 3 consecutive failures (responses scored 0).

4. ARITHMETIC

(36) Problems 1 through 13 are read to the client.

(37) Problems 14, 15 and 16 are presented on separate cards for the client to read.

(38) There is a time limit for each problem.

(39) Begin timing immediately after stating the problem.

(40) For problems 1, 2 and 3, use the blocks for the block design test.

(41) If the client requests, or if it is apparent that he failed

to understand, the problem may be repeated. However, the timing always begins at the end of the "first" statement of the problem.

(42) For clients 8 and older, not suspected mental defectives begin with problem 4 and give credit for problems 1, 2 and 3. If the client fails to pass both problems 4 and 5, administer problems 1, 2 and 3. If the client then passes problem 3, he is entitled to try problem 6, because he has not yet had three consecutive failures.

(43) Discontinue after 3 consecutive failures.

5. SIMILARITIES

(44) For clients under 8, and older suspected mental defectives do the analogies section.

(45) Analogies - before reading each item say "finish what I want to say." All four items are given to the client.

(46) For clients 8 and older, not suspected mental defectives do similarities.

(47) Similarities - say "in what way are a plum and a peach alike?" If the client says that they are not alike, fails to answer, or gives an inferior reply, say "oh yes, they are both fruit, you eat both, and both have skins and seeds. Now tell me in what way are a cat and a mouse alike?"

(48) Similarities - if the client fails on cat-mouse, explain again and go on to item 7, but give no further help.

(49) Clients who have not been given analogies and who score less than 3 points on similarities must be given analogies according to the instructions for younger children. They are given the total points

earned on both parts of the test.

(50) Discontinue after 3 consecutive failures (responses scored 0) on the similarities items.

6. VOCABULARY

(51) Say "I want to see how many words you know. Listen carefully and tell me what these words mean. Bicycle ...what is a bicycle?"

(52) Proceed with the words in the order list, repeating at each presentation, "What is a ----------? or What does ---------- mean?"

(53) The examiner must be certain that he is using the standard pronunciations of the words.

(54) Clients 8 and older, not suspected mental defectives. Client may be started with word 10. However, if the client fails to give a 2 point definition for any one of the words 10 through 14, the examiner must return to word 9 and go backwards, word by word, until the requirement of five consecutive 2 point successes is met. Then he should proceed forward again from the word beyond the one which causes this digression.

(55) Occasionally, it is difficult for the examiner to decide whether the client does or does not know the meaning of a word. In such instances it is permissible to say "Please explain a little more," or make some similar neutral inquiry. The following principles should govern the supplementary inquiry:

(A) if the examiner feels the response is vague (cannot readily be scored) questioning is permissible.

(B) the criterion examples of 0 responses (Appendix C of the Manual) are given to illustrate marginal 0 responses, ones which reveal

some understanding of the word but are nevertheless too poor to be scored 1 point. Responses scorable as 0 should be questioned if the examiner believes that the response actually given indicates that a better response can be evoked by neutral inquiry. The examiner, however, should avoid further questioning of a clearcut 0 response.

(C) Spoilage of response after inquiry. After questioning, a response may be scored lower if additional matter reveals serious misunderstanding, or if it reveals the first response as being mere verbalism.

(D) Under all circumstances the inquiry should be limited to saying "Please explain a little more," or very similar neutral (non-leading) statements.

(E) The following words sometimes evoke a response which is based on the homonym; fur (fir); sword (soared); gamble (gambol); recede (re-seed). If this occurs, the examiner pronounces the word again and asks "What else does ------- mean?" He does not spell the word for the client. The scoring is based on the response to the intended word regardless of the response to the homonym.

(56) Discontinue after 5 consecutive failures (responses scored 0).

EXAMINER'S EVALUATION BEHAVIOR

(57) To know the meanings of the words and concepts involved.

(58) To be thoroughly acquainted with the principles of scoring the test, and to be continually alert to the need for careful work in scoring each item.

EXAMINER'S EVALUATION PROCEDURES

1. GENERAL TO WHOLE WISC

A. INITIAL SINGLE RESPONSE EVALUATION

(59) Supplementary inquiry should take place if the response is vague (i.e. cannot be readily scored).

(60) Where there is a clear 2 point response to a question whose maximum score is 2, or a clear 1 point response to a question whose maximum score is 1, award the maximum score.

(61) All responses listed as questionable should be clarified by question.

(62) The following words sometimes evoke a response which is based on the homonym: fur (fir); sword (soared); gamble (gambol); recede (re-seed). The examiner must requestion.

(63) Where there is a possible 2 point score, some 1 point responses should be questioned. It is not the examiner's role to elicit 2 point responses to each question, but primarily to clarify what the client has attempted to say.

(64) Responses scorable as 0 should be questioned if the examiner believes that the response actually given indicates a better response can be evoked by neutral inquiry. Most 0 point responses typify marginal responses; those which contain evidence of some understanding may be queried in the neutral manner prescribed.

B. MULTIPLE SUBJECT RESPONSES

(65) When a client gives multiple responses either spontaneously

or after query, it is fair to credit him for the "best" response if:

(A) all parts of the response are acceptable, but one part is better,

(B) the second response clarifies his first more vague response, and, (C) if the client really knows what his words implied.

(66) A multiple response should be scored zero if the client

(A) changes a correct response, or

(B) follows a creditable response with a contradicition, or

(C) denies the correct response.

An exception to Rule B is in vocabulary, where a positive response, when followed by a negative response, is scored with the value of the positive response.

(67) A multiple response should be scored appropriately 1 or 2, if a correct response follows an incorrect zero response..

(68) If the client gives differences or wrong answers "in addition" to correct responses, requestion and score according to this decision. Spontaneous improvement of responses is easily recognized and query is not necessary.

C. QUESTIONING

(69) If questioning lowers a client's score, the examiner should consider scoring the initial response. The rationale for this is that the client's anxiety or defensiveness may have been aroused by the examiner's questioning.

(70) Added remarks obviously not part of the client's response do not enter into the scoring.

(71) When a response is questioned, the letter "Q" should be recorded with the client's response. This will often determine the score for this item.

(72) To provide for a full recording of the client's answer which will permit subsequent evaluation and fuller consideration of these answers in light of other data.

D. SCOREKEEPING

(73) To record the client's birthday.

(74) To compute the client's age.

(75) To select proper scale score tables and to compute and record scaled scores.

(76) To compute the Verbal Score.

(77) By use of Tables XIA and XIB (WISC manual) the examiner can now convert the three scores (Verbal, Performance, and Full Scale) into intelligence quotients. These values are entered in the appropriate spaces in the record form.

GENERAL INFORMATION

(78) Each item is scored 1 or 0.

(79) Maximum score: 30 points.

GENERAL COMPREHENSION

(80) Each item is scored 2, 1 or 0. See Appendix (WISC manual) for specific scoring criteria and sample answers.

(81) Maximum score: 28 points.

(82) The examiner should match the client's responses against the criteria and sample answers given below. The examiner will undoubtedly find unusual responses which are not typified, as no attempt has been made to cover all possible replies. In these instances he will have to use his own judgment. Most of the 0 point examples given typify marginal responses; those which contain evidence of some understanding may be queried in the neutral manner presribed.

(83) Items 1 - 5. They are "What to do" items. Allow 2 points if the client indicates that he understands what should be done and assumes personal responsibility for doing it. Allow 1 point if the client indicates that he understands what should be done but places the responsibility on someone else or sets up qualifying conditions for himself.

(84) Items 6 - 14. For each of these items several acceptable responses are noted. In general, 2 points are given for a response which includes at least two of the general reasons. A response which has only one of them mentioned is given 1 point. Typical responses are credited 2, 1 or 0 points. Note that for item 14 the degree of abstraction is particularly stressed. (Do not encourage a second response other than a short period of silence following the client's first response.)

ARITHMETIC

(85) Each problem is scored 1 or 0. Problems 2 and 3 are given $\frac{1}{2}$ credit each if the client makes an error but corrects it within the time limit.

(86) Maximum score: 16 points. (Round half score upward.)

SIMILARITIES

(87) 1 point for each item correctly answered. The only correct analogies - 1 point for each item correctly answered. The only correct responses for the items are, respectively : 1. sweet; 2. arms, hands; 3. women, ladies, mothers; 4. cut.

(88) Similarities - 2, 1 or 0 points for each similarities item (5 - 16), depending on the degree and quality of the generalization. See Appendix B (WISC manual) for scoring criteria and sample answers.

(89) Similarities - If the subject gives multiple acceptable responses, credit him on the basis of the best response. If the client gives differences or wrong answers in addition to some correct responses, then ask, "Now which one is it?" and score according to this decision. Spontaneous improvement of responses is easily recognized and query is not necessary. Added remarks obviously not part of the client's answer do not enter into the scoring. Example: cat-mouse - Both have eyes, tails, legs ...my cat caught a mouse.

(90) Similarities - maximum score: 28 points. The sum of points earned on analogies and similarities.

VOCABULARY

(91) Each word is scored 2, 1 or 0 except that words 1 - 5 are scored only 2 or 0.

(92) The general rule for scoring is that any recognized meaning

of the word is acceptable, elegance of language and precision being disregarded. Most of the words in the list can be scored 2, 1 or 0 without difficulty. (A distinction must be made in scoring responses for the first five words.) Words 1 - 5 should be scored as 2 or 0. This arbitrary decision has been necessary because of the young child's limited ability to give verbal expression to his concepts. Therefore, any definition which indicates that the young client knows what one of these five words mean is credited 2 points. While elegance of expression is disregarded, poverty of content is penalized to some extent. If a client seems to know only vaguely what a word means, his response is credited with 1 point. Thus, 1 point is given for a response such as (nonsense) "Somebody trying to be funny." The following general scoring principles should be helpful:

(A) 2 Point Responses

(a) A good synonym.

(b) A major use.

(c) One or more definitive features or primary features of objects.

(d) General classification to which the word belongs.

(e) A correct symbolic use of the word. Example: "An umbrella of artillery for the advancing infantry."

(f) Several less definitive but correct descriptive features which cumulatively indicate understanding of the word.

(B) 1 Point Responses

(a) A vague or less pertinent synonym.

(b) A minor use, not elaborated.

(c) Attributes which are correct but not definitive or not distinguishing features.

(d) Example using the word itself, not elaborated.

(C) 0 Point Responses

(a) Obviously wrong answers.

(b) Verbalisms. Example: "A brave man," when no real understanding is shown after inquiry.

(c) Not totally incorrect responses, but ones which, even after questioning, are very vague or trivial or show a great poverty of content.

(93) The formal definitions are taken from Webster's New International Dictionary, second edition, unabridged; to save space, some of the most esoteric meanings are omitted.

(94) With the aid of these three guides - general principles, examples, and definitions - the examiner should be able to evaluate each response without too much difficulty.

(95) Maximum score: 80 points.

2. LIST OF COMPUTER MODEL NONOBJECTIVES

EXAMINER'S ADMINISTRATION BEHAVIOR

(1) To establish a friendly relationship which is affectionate and shows kindly concern for the subject, but is not gushy and oversolicitous.

(2) To be sensitive to the young client's alternating interests

and indifferences to the test.

(3) To adapt to the client's concepts of a satisfactory adult-child relationship which has been acquired through his experiences.

(4) To have the client share his personal experiences relating to home and family with the examiner.

(5) To gather important information about the particular client's level and flow of language, his ability to express his own ideas, his areas of interest, his concern with himself and to analyze this data to determine the client's special needs.

(6) To know the routines to which the clients are accustomed in their classroom and to adapt the testing program to this routine. Note: This does not appear to be directly related to the administration of the test, and therefore would most likely become the responsibility of the psychologist making the referral to gather this information.

(7) To be able to distinguish between a subject's negativism and shyness in a new situation.

EXAMINER'S ADMINISTRATION PROCEDURE

(8) It is important that the examiner is known and accepted by the subject. This may require the examiner to spend some time in the classroom for a friendly unhurried relationship to be maintained between the client's teacher and the examiner. The client should be made to feel that leaving the group for a test by the examiner is a pleasant experience and that other children share his experiences at other times.

(9) The general rule is to give the five Verbal subtests which are regularly scheduled, unless the following valid reasons for

substituting alternatives are used:

(A) if the regular test has been spoiled,

(B) adequate time is available, and

(C) the examiner recognizes that a specific handicap will prevent administration of a test.

EXAMINER'S EVALUATION BEHAVIOR

(10) A response may be scored lower if additional material reveals a serious misunderstanding, or that the first response was mere verbalism.

(11) To record any unusual reactions of the client to the test items.

(12) To compute Performance Score. Reasons: Not part of the Verbal Score.

(13) To compute Fullscale Score. Reasons: Not possible without the Performance Score.

(14) If six tests have been given, the sum of the six scaled scores must be multiplied by 5/6 to reduce it to the equivalentof 5 tests.

(15) If four tests have been given the sum of the four scaled scores must be multiplied by 5/4 to expand it to the equivalentof 5 tests.

(16) One can, of course, prorate on three tests, but it is expected that the WISC I.Q. will always be based on at least four Verbal and four Performance tests.

(17) The symbol (PR) should be written in the margin near the prorated score which is to be used for computing I.Q.

APPENDIX C

TABLE 24

CHARACTERISTICS OF CLIENTS

IN STUDENT ADMINISTRATION

STUDENT	AGE	SEX	MENTAL HEALTH	REMARKS
1	14	Male	Normal	
2	15	Male	Normal	
3	10	Female	Normal	
4	10	Female	Normal	
5	9	Male	Normal	
6	14	Female	Normal	
7	15	Male	Normal	
8	8	Female	Normal	

APPENDIX D

The symbols used in the program flow chart are defined as follows:

(1) INF00 to INF30 - are program labels.

(2) C1 - subtest consecutive failure count (maximum value)

C2 - consecutive failure counter

C3 - Evaluator's score point award counter

C4 - Evaluator's action counter

C14 - the client's age count in months (96 = 8 years)

C15 - information raw score counter

(3) S1 - mental health switch (if equal to 1 client is suspected mentally defective)

S14 to S18 - respectively the Q4 to Q8 bypass switches (bypass if equal to 1)

S19 - no bonus point switch (if equal to 1)

(4) Q1 to Q 30 - respectively the question modules for information questions 1 to 30.

FIGURE 14A: PROGRAM FLOW CHART

INFORMATION SUBTEST

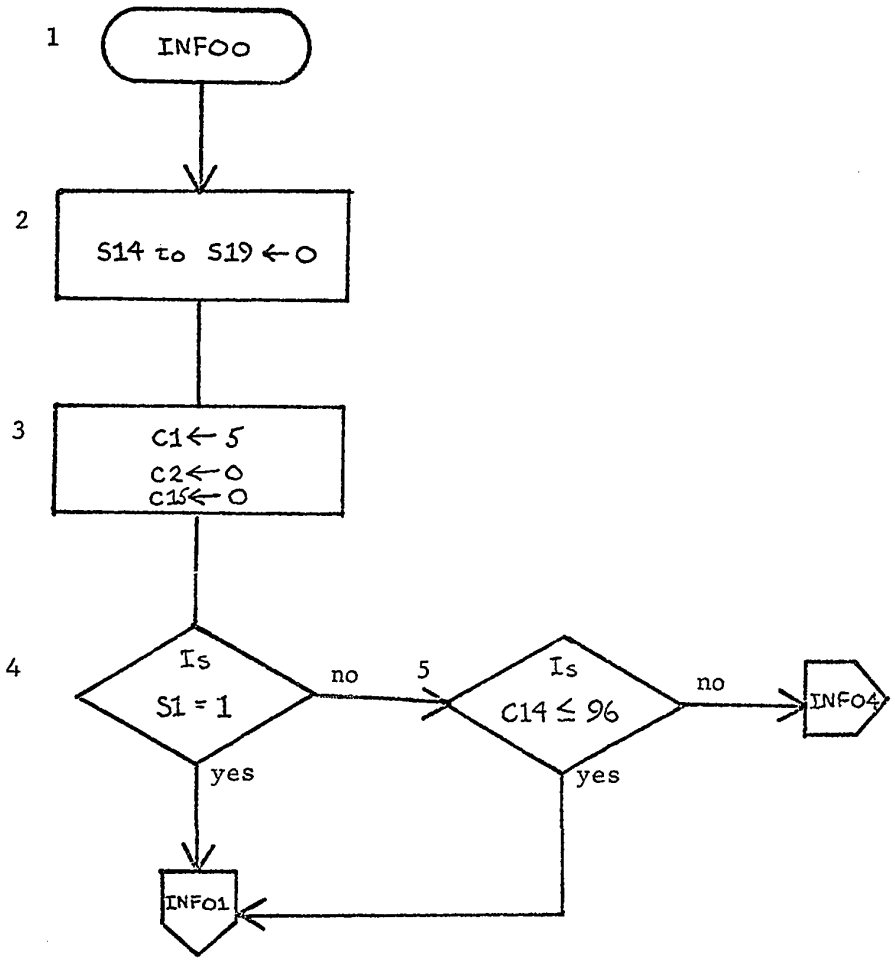


FIGURE 14B: PROGRAM FLOW CHART

INFORMATION SUBTEST (CONTINUED)

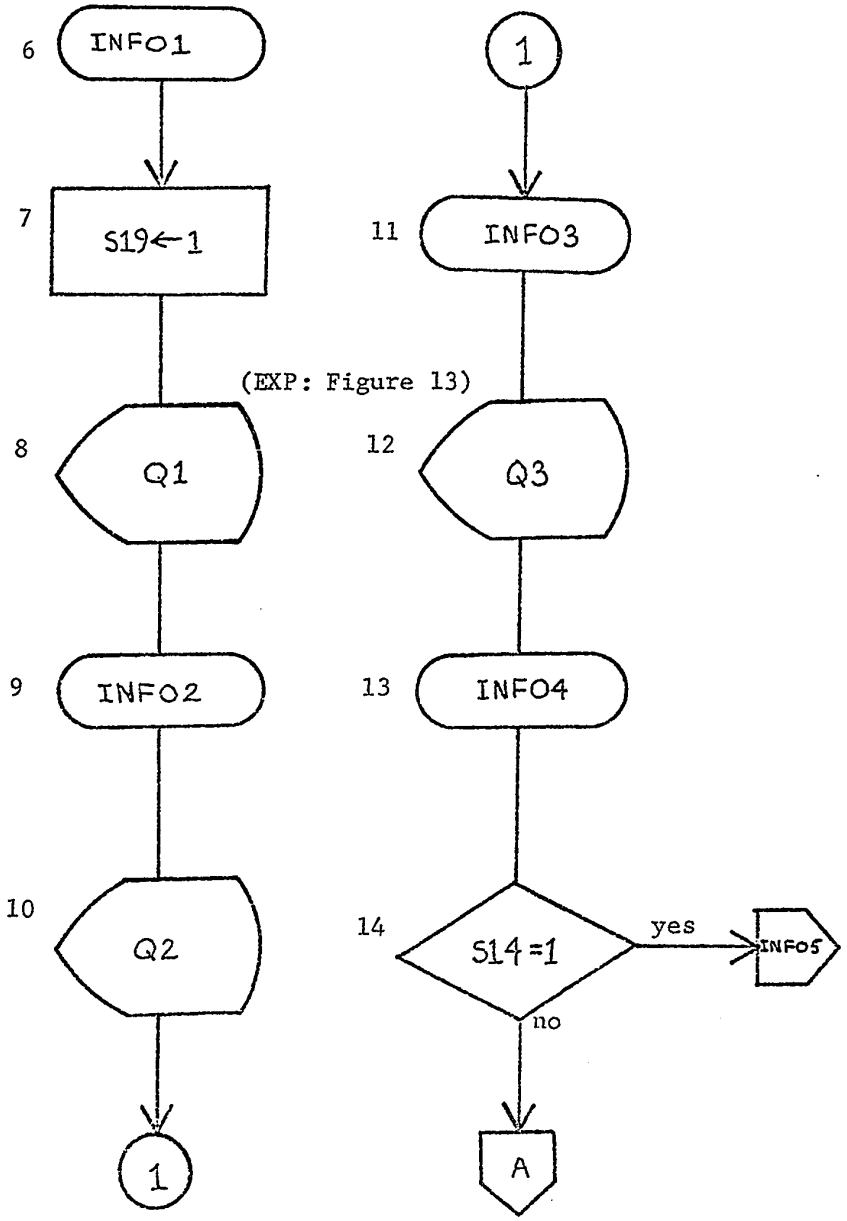


FIGURE 14C: PROGRAM FLOW CHART

INFORMATION SUBTEST (CONTINUED)

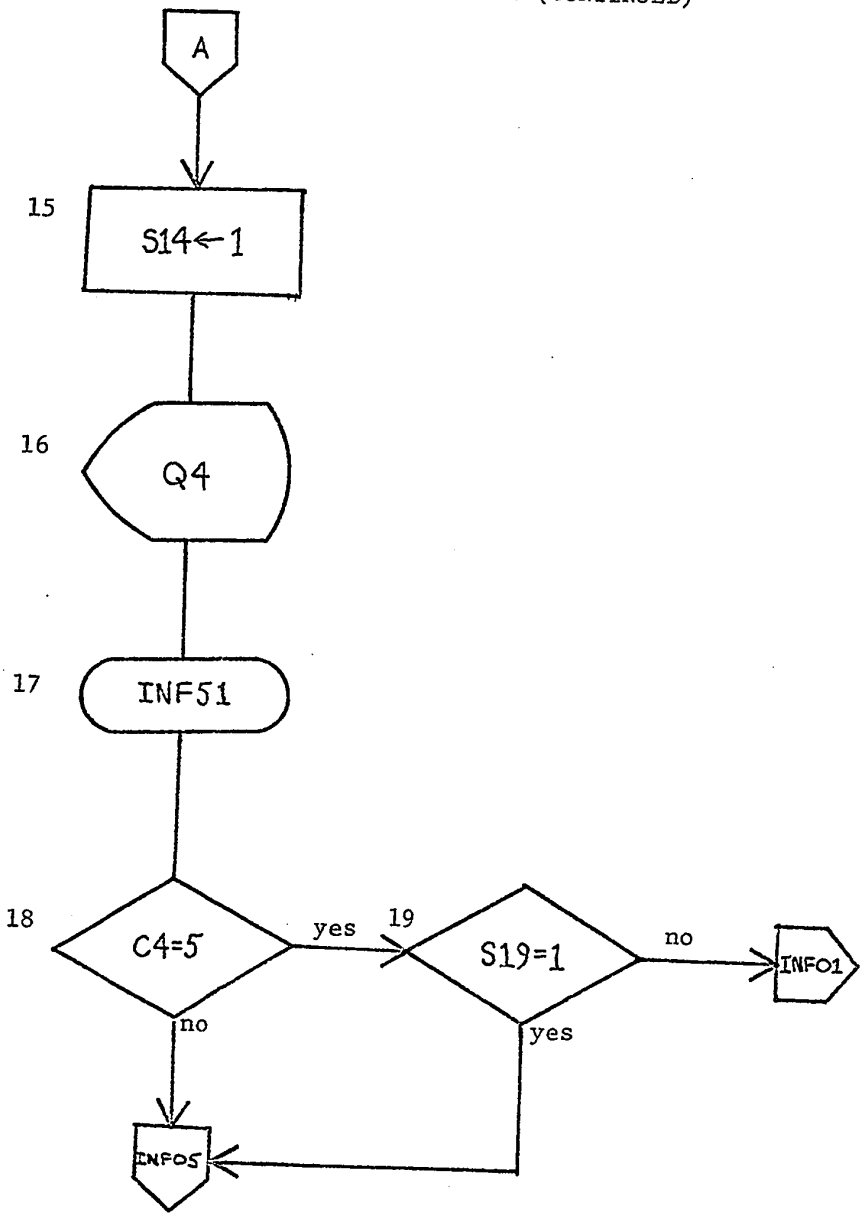


FIGURE 14.D: PROGRAM FLOW CHART

INFORMATION SUBTEST (CONTINUED)

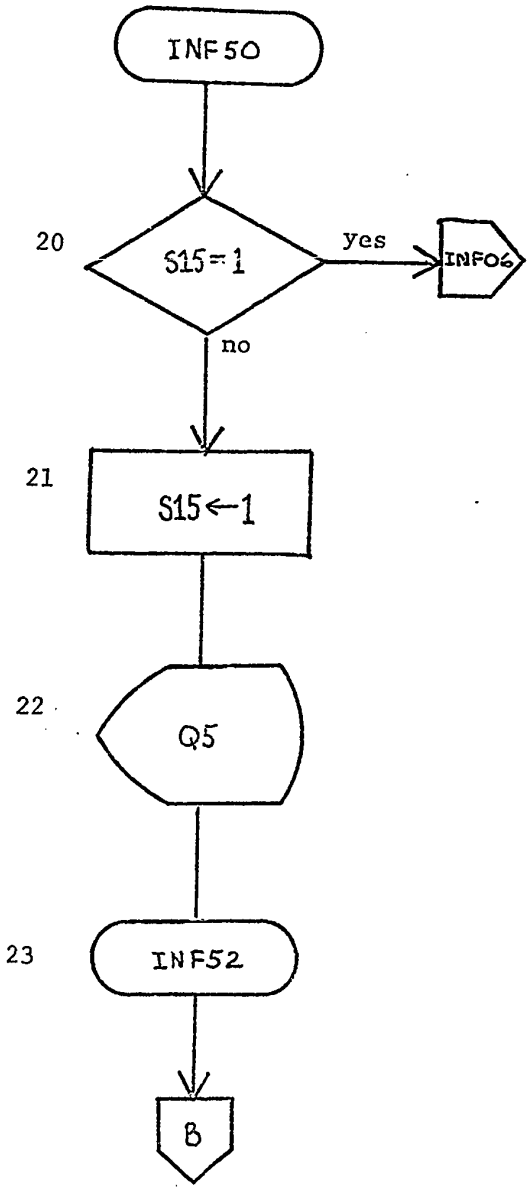


FIGURE 14E: PROGRAM FLOW CHART

INFORMATION SUBTEST (CONTINUED)

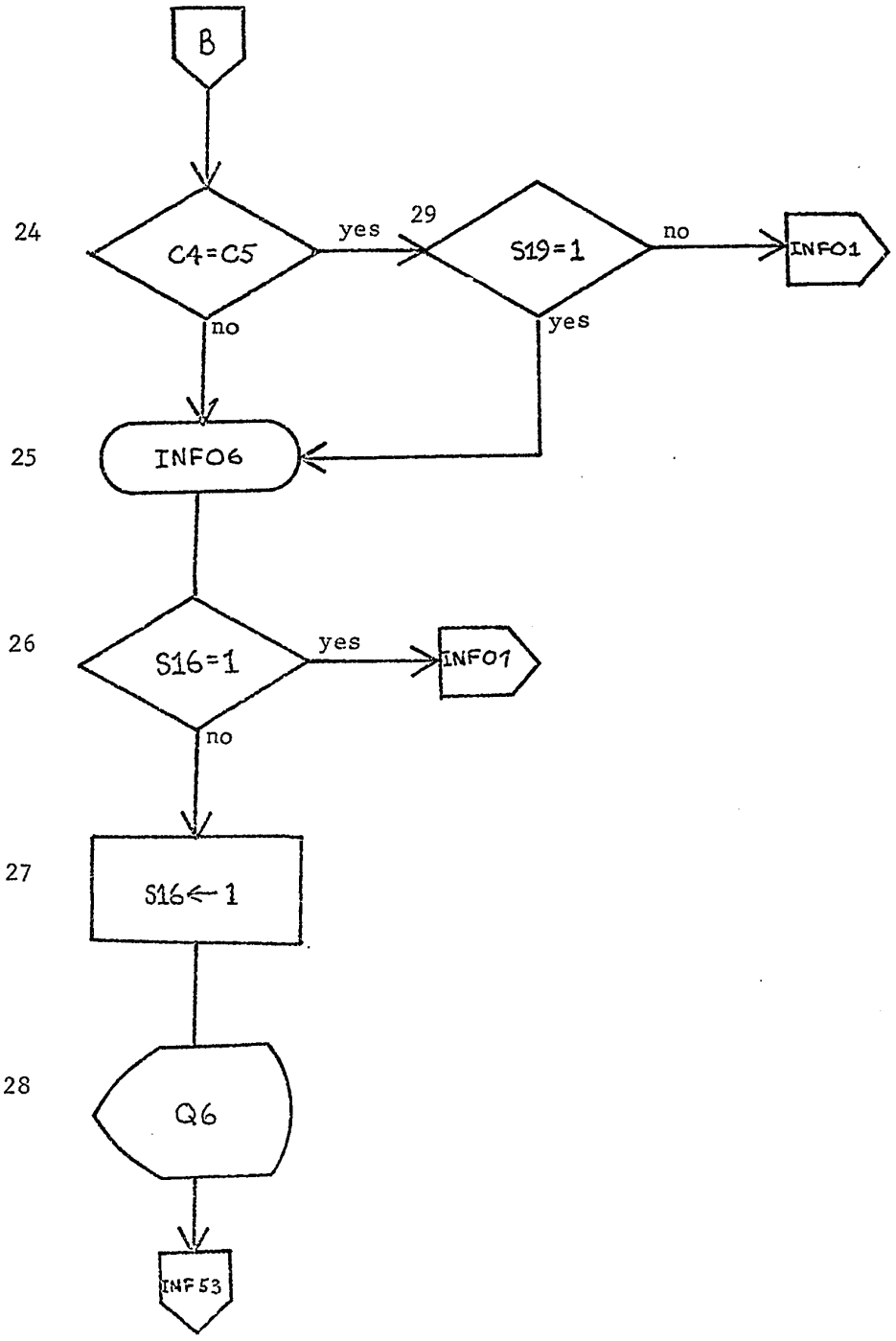


FIGURE 14F: PROGAM FLOW CHART

INFORMATION SUBTEST (CONTINUED)

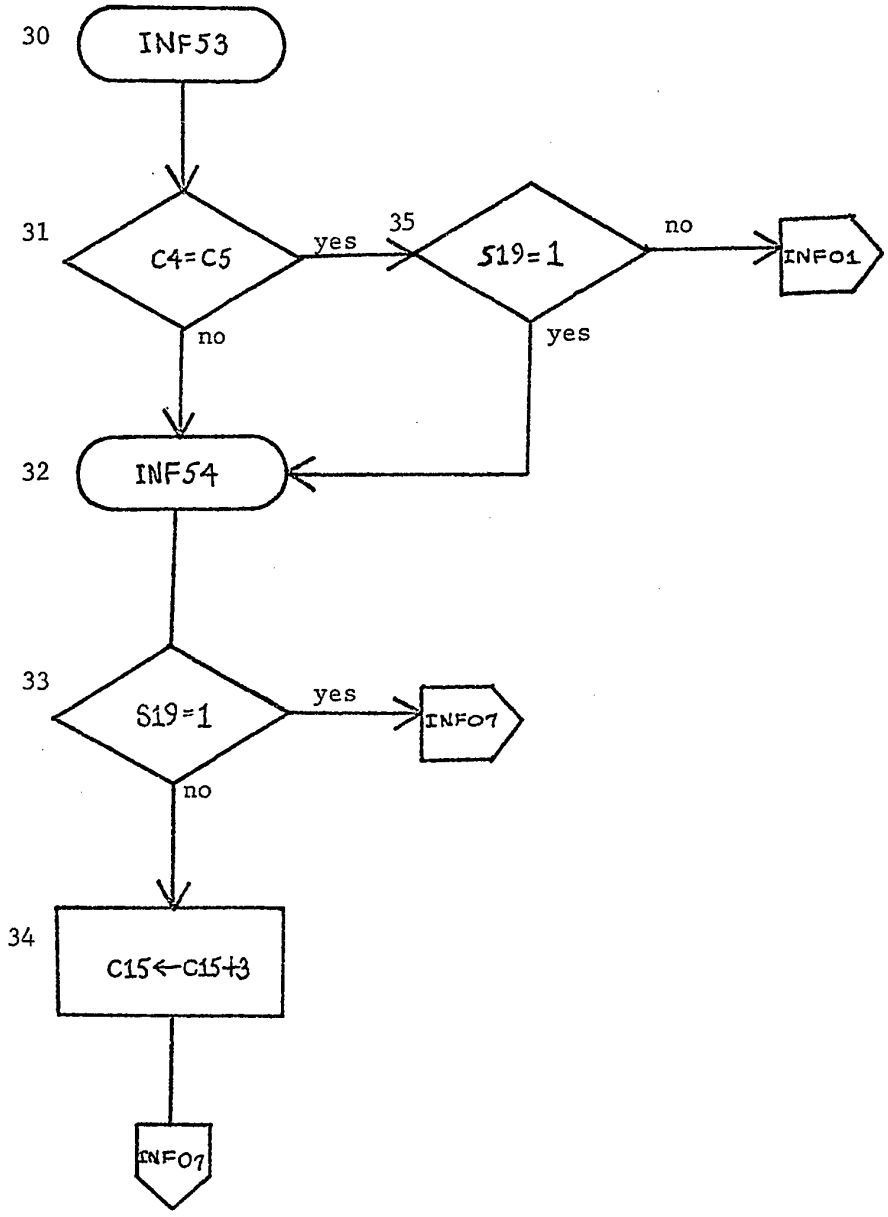
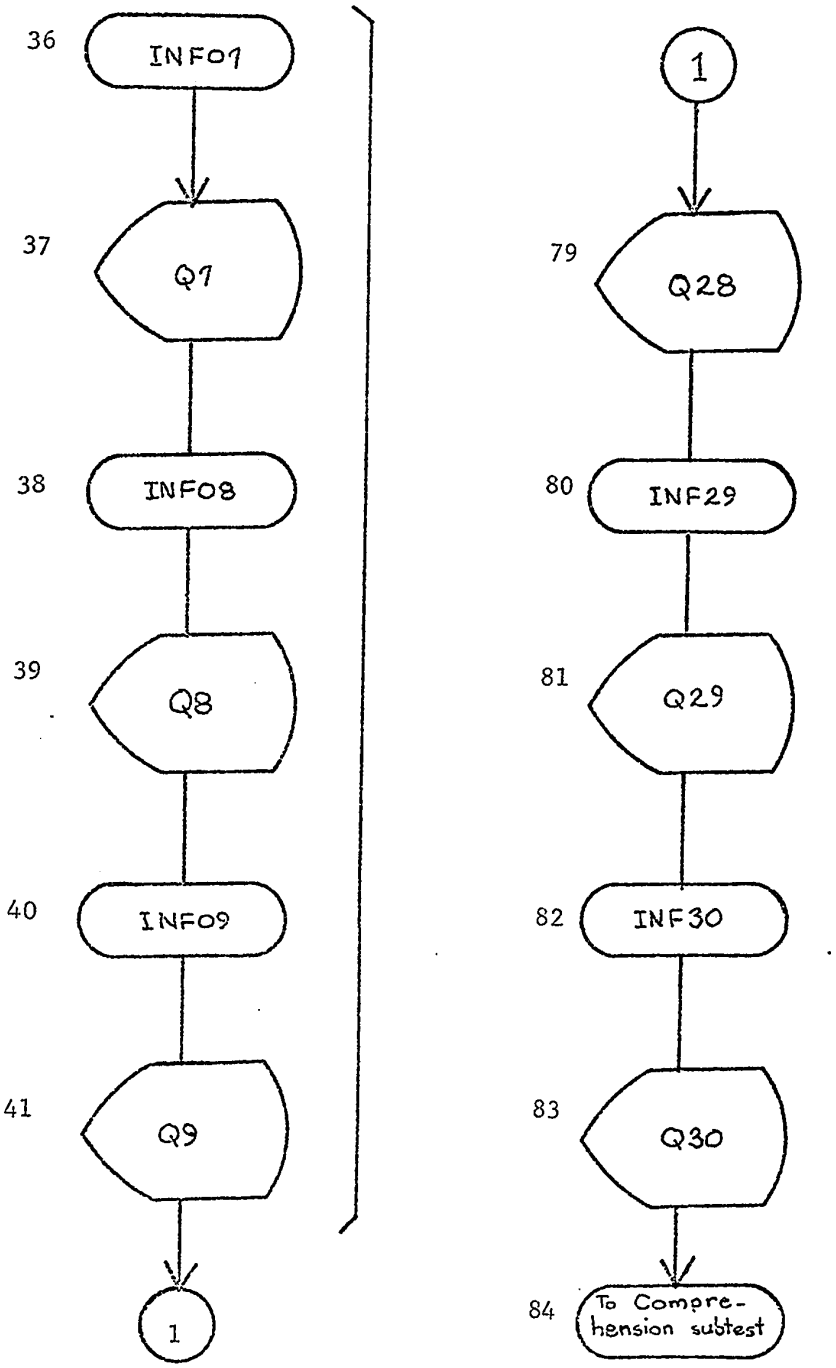


FIGURE 14G: PROGRAM FLOW CHART

INFORMATION SUBTEST (CONTINUED)



APPENDIX E

PROGRAM LISTING INFORMATION QUESTION 5

```
#A1##1
       PRR
       DE  0-/32
       AUP 6165,0-/22
       BR  #EP##5
#A3##5
       AUP 6188,0-/20
       AD  1-/C28
       EPP 0-/EXPLANIN05
       PR
       BR  #EP##5
#A4##5
       AUP 6188,1-/16
       AD  1-/C29
       EPP 0-/WCHONEIN05
       PR
       BR  #EP##5
#A5##5
       AUP 6165,1-/16
       AD  1-/C30
       EPP 0-/WOTHERIN05
       PR
       BR  #EP##5
#A2##5
       AUP 6165,0-/22
       AD  1-/C26
       EPP 0-/REPEATIN05
#EP##5
       PR
       DL  0-/32
       BR  PR1-/S24B-/0
       DT  2,13-/2,2-/40,0-/QUESTION INO5
       PR
       DL  13,3-/33
       DL  17,3-/33
       DL  21,3-/33
       DT  30,0-/2,30-/40-/REPEAT. I DON'T KNOW. TEACHER. REST.
       LD  . ? -/34
       EP  11,3-/2,11-/33,3-/1800-/33-/ZZ
       NX
       LD  B0-/B1
       LR  #TO##5-/RR5
       BR  TIME
       FN2 KEYL-/1-/C-/-/R1
       LD  B0-/B1
       BR  #RA##5
       PR
       LD  B0-/B1
       EP  15,3-/2,15-/33,3-/1800-/33-/ZZ
       FN  MV-/B0,,E-/B1,-/A-/
```

```
        NX
        FN  MV-/BO,,E-/B1,-/A-/
        LR  #TO##5-/RR5
        BR  TIME
        FN2 KEYL-/1-/C-/-/R2
        BR  #RA##1
        PR
        EP  19,3-/2,19-/33,3-/1800-/33-/ZZ
        NX
        FN  MV-/BO,,E-/B1-/A-/
        LR  #TO##5-/RR5
        BR  TIME
#TO##5
        BR  #EP##5-/C4-/E-/0
        BR  #A2##5-/C4-/E-/1
        BR  #12##5-/C4-/E-/6
        PR
        FN  MV-/BO,,E-/B1,-/A-/
#RA##5
        EP  30,39-/2,30-/1,39-/1-/1-/QUESTIN05
        PR
        LD  B1-/BO
        LD  REPT AGAN STOP REST TECHR DON KOW-/
        FN2 COD-/1,1
        FN2 SQ-/CR-/(5)
        LR  #TO##5-/RR5
        BR  TEACH
        FN2 SQ-/CR-/(1+2)
        BR  #A2##5
        FN2 SQ-/CS-/(3)
        LR  #TO##5-/RR5
        BR  STOP
        FN2 SQ-/CS-/(4)
        LR  #A2##5-/RR5
        BR  REST
        FN2 SQ-/CK-/(6.7)
        AD  1-/C27
        AUP 6188,2-/24
        LD  0-/C3
        LD  5-/C4
        BR  #09##5
#AN##5
        NO  WB-/STV COK FIR OVN PLG HOT STM LTRC BOL WTR TRN HET
        PR
        LD  STV COK FIR OVN PLG HOT STM LTRC BOL WTR TRN HET-/B1
        FN2 COD-/1,1
        NO  EM
        NO  WH-/(1+2+3+4+12)
#C1127
        PR
        FN2 SQ-/C5-/(1+2+3+4+12)
        NO  EM
        NO  WI-/(5+(6.(10)')+7+8)
#Q1128
```

```
        PR
        FN2 SQ¬/C4¬/(5+(6.(10)')+7+8)
        FN  POS¬/1,C10
        NO  EM
        NO  WK¬/(9+(11.6.10))
#ZZ129
        PR
        FN2 SQ¬/ C2¬/(9+(11.6.10))
        FN  POS¬/1,C12
        NO  EM
#AC##5
        PR
        LD  0¬/S6
        PR
        BR  EVA000¬/S6¬/0
        BR  #09##5¬/C4¬/G¬/4
        BR  #10##5¬/C4¬/G¬/2
        BR  #A3##5¬/C4¬/E¬/2
        BR  #A2##5
#09##5
        BR  #11##5¬/C4¬/E¬/6
        AD  1¬/C2
        AD  C3¬/C15
        BR  #13##5¬/C2¬/GE¬/C1
        BR  #12##5
#10##5
        BR  #A5##5¬/C4¬/E¬/4
        BR  #A4##5
#11##5
        LD  0¬/C2
        AD  C3¬/C15
#12##5
        EPP 0¬/ZZ
        BR  INF52
#13##5
        PM  ,0¬/2,¬/40,0¬/PM700: NOTE COM00
        EPP 0¬/NXTSTIN05
        PR
        BR  COM00
        NO  EM
INF52
```

APPENDIX F

TABLE 4

VOLUNTEER ADMINISTRATION

INFORMATION SUBTEST

INFORMATION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED WRONG	COMPUTER MISTAKE
1	5		3			2				
2	5		3			2				
3	6		3	1		2				
4	6		5	1						
5	7		5	2						
6	6		4	1	1					
7	5		4					1		
8	5		5							
9	6		4	1					1	
10	5		5							
11	6		5	1						
12	5		4		1					
13	7		5	2						
14	6		4	1					1	
15	5		5							
16	5		5							
17	5		3							2
18	7		3	2				1	1	
19	5		1						3	1
20	5		5							
21	5		3						2	
22	6		5	1						
23	6		5	1						
24	5		4							1
25	5		4						1	

TABLE 4 (Cont'd.)

VOLUNTEER ADMINISTRATION

INFORMATION SUBTEST

INFORMATION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED WRONG	COMPUTER MISTAKE
26	5		4		1					
27	6		5	1						
28	5		3						2	
29	6		4	1						1
30	5		4						1	
TOTALS	166	0	122	16	3	6	0	2	12	5

TABLE 5

VOLUNTEER ADMINISTRATION

COMPREHENSION SUBTEST

COMPREHENSION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECTLY	COMPUTER MISTAKE
1	5	4							1
2	5	4							1
3	5	5							
4	7	2		1				2	2
5	7	5		1					1
6	6	2	1	1				2	
7	7		3	2				1	1
8	6	2		1				2	1
9	5		1					1	3
10	6		3	1		1			1
11	5					1		1	3
12	5	3	1			1			
13	6		2	1		1			2
14	5					1		2	2
TOTALS	80	27	11	8	0	5	0	11	18

TABLE 6

VOLUNTEER ADMINISTRATION

ARITHMETIC SUBTEST

ARITHMETIC QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	5		3			2			
2	5		3			2			
3	5		3			2			
4	5		3					2	
5	5		4					1	
6	5		3					2	
7	5		3			2			
8	5		3			2			
9	5		3			2			
10	5		3			2			
11	5		3			2			
12	5		3			2			
13	5		3			2			
14	5		3			2			
15	5		3			2			
16	5		3			2			
TOTALS	80		49	0	0	26	0	5	0

TABLE 7

VOLUNTEER ADMINISTRATION

SIMILARITIES SUBTEST

SIMILARITIES QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	5		3			2			
2	5		3			2			
3	5					5			
4	5					5			
5	5	5							
6	5	5							
7	5	5							
8	6	4	1	1					
9	5	4	1						
10	7	4		1				1	1
11	7	4	1	2					
12	6	1	4	1					
13	5	4						1	
14	5	5							
15	5	5							
16	5	3						2	
TOTAL	86	49	13	5	0	14	0	4	1

TABLE 8

VOLUNTEER ADMINISTRATION

VOCABULARY SUBTEST

VOCABULARY QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	7	5		2					
2	5	5							
3	5	4							1
4	6	5		1					
5	6	5							1
6	5	2	1	1					1
7	5	4						1	
8	5	4						1	
9	7	1	1	2				1	2
10	5	2		1					2
11	3	2						1	
12	2	1						1	
13	1	1							
14	1	1							
15	2			1	1				
TOTAL	65	42	2	8	1	0	0	5	7

TABLE 9

PROTOCOL ADMINISTRATION

INFORMATION SUBTEST

INFORMATION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	6					6			
2	6					6			
3	6					6			
4	6					6			
5	6		6						
6	6		6						
7	6		6						
8	6		6						
9	6		5					1	
10	6		6						
11	6		6						
12	6		4		2				
13	6		5		1				
14	7		5	1	1				
15	6		2		2		2		
16	6						6		
17	6		2		1		3		
18	6		1		1		4		
19	6		2		3				1
20	6		2		3	1			
21	6		3		2	1			
22	6		1		1	1	3		
23	6				1	2	2		1
24	6		1		3	2			
25	6				2	3			1

TABLE 9 (Cont'd.)

PROTOCOL ADMINISTRATION

INFORMATION SUBTEST

INFORMATION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
26	6				2	3	1		
27	6		2		1	3			
28	6				1	3	2		
29	6				1	3	2		
30	6				1	3	2		
TOTAL	181		71	1	29	49	27	1	3

TABLE 10

PROTOCOL ADMINISTRATION

COMPREHENSION SUBTEST

COMPREHENSION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	6	3							3
2	6								6
3	6	2	1						3
4	6								6
5	6	5							1
6	6	2	1						3
7	6	2	1						3
8	6	2	2						2
9	6								6
10	6				1				5
11	6		1	2					3
12	6	1			4				1
13	6				4				2
14	6				4	1			1
TOTAL	84	17	6	2	13	1	0	0	45

TABLE 11

PROTOCOL ADMINISTRATION

ARITHMETIC SUBTEST

ARITHMETIC QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	6					6			
2	6					6			
3	6					6			
4	6		6						
5	6		6						
6	6		6						
7	6		6						
8	6		6						
9	6		5		1				
10	6		5				1		
11	6		2		4				
12	6		4		2				
13	6				4	2			
14	6		1		3	2			
15	6				4	2			
16	6				1	5			
TOTAL	96		47	0	19	29	1	0	0

TABLE 12

PROTOCOL ADMINISTRATION

SIMILARITIES SUBTEST

SIMILARITIES QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECTLY	COMPUTER MISTAKE
1	6		1			5			
2	6		1			5			
3	6		1			5			
4	6		1			5			
5	6	3	3						
6	6	5							1
7	6	2	1	1					2
8	6	3	3						
9	6		4		1				1
10	6	6							
11	6				4				2
12	6				5				1
13	6	1			3		1		1
14	6				2	3			1
15	6	1			1	3	1		
16	6				1	4	1		
TOTAL	96	21	15	1	17	30	3	0	9

TABLE 13

PROTOCOL ADMINISTRATION

VOCABULARY SUBTEST

VOCABULARY QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	6					6			
2	6					6			
3	6					6			
4	6					6			
5	6					6			
6	6					6			
7	6					6			
8	6					6			
9	6					6			
10	6	4							2
11	6	3	1						2
12	6	6							
13	6	2							4
14	6	5							1
15	6	4							2
16	6	4	2						
17	6	2			1				3
18	6	2	1						3
19	6	2			1		2		1
20	6	3							3
21	6	5							1
22	6				1		2		3
23	6	1			3		2		
24	6				1		4		1
25	6				1		3		2

TABLE 13 (Cont'd.)

PROTOCOL ADMINISTRATION

VOCABULARY SUBTEST

VOCABULARY QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
26	6								
27	6				2		4		
28	6				3		3		
29	6				1	2	3		
30	6				2	3	1		
31	6				1	3	2		
32	6					5	1		
33	6					5	1		
						6			
TOTAL	198	43	4	0	17	78	28	0	28

TABLE 14

STUDENT ADMINISTRATION

INFORMATION SUBTEST

INFORMATION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED WRONG	COMPUTER MISTAKE
1	11		3			7	1			
2	12		3			7	2			
3	10		3			7				
4	8		8							
5	8		8							
6	11		5	1			1	2		2
7	8		8							
8	8		8							
9	12		4				1	3		4
10	9		6		1		1	1		
11	8		5		1			2		
12	8		5		1			2		
13	10		3	1	5			1		
14	8		3			1		2		2
15	9		3		2	1	1	1		1
16	8		4			2		2		
17	9		5			2	1	1		
18	9		2		2	2	1	2		
19	8		3		2	3				
20	8		4		1	3				
21	8		3			3		1		1
22	8		3		2	3				
23	10		4	1		3		2		
24	8				4	3		1		
25	8		2		2	3		1		

TABLE 14 (Cont'd.)

STUDENT ADMINISTRATION

INFORMATION SUBTEST

INFORMATION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED WRONG	COMPUTER MISTAKE
26	8		2		2	4				
27	10		2	1		5				2
28	10		1		1	5	1			2
29	8		1			5				2
30	12				1	5	4	2		
TOTAL	272		111	4	27	74	14	26	0	16

TABLE 15

STUDENT ADMINISTRATION

COMPREHENSION SUBTEST

COMPREHENSION QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED WRONG	COMPUTER MISTAKE
1	10	6			1					
2	8	2								3
3	8	2	1							6
4	16	1	1	3						5
5	9	5		1			1			11
6	10	1	1		2			1		
7	8	1				3				5
8	8	1			1	3				3
9	9					3		1		3
10	9				1	3	1			4
11	9		1			5	1			3
12	8					6	1			1
13	8					7				1
14			1			7				1
						7				
TOTAL	128	19	5	4	5	44	4	2	0	46

TABLE 16

STUDENT ADMINISTRATION

ARITHMETIC SUBTEST

ARITHMETIC QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	8		1			7				
2	8				1	7				
3	8				1	7				
4	8		8							
5	8		8							
6	8		8							
7	8		6		2					
8	8		5		3					
9	8		4		3	1				
10	8		3		3	2				
11	8		3		2	3				
12	8		4			3		1		
13	8		4			4				
14	8		4			4				
15	8		3		1	4				
16	8		1		3	4				
TOTAL	128		62	0	19	46	0	1	0	0

TABLE 17

STUDENT ADMINISTRATION

SIMILARITIES SUBTEST

SIMILARITIES QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED WRONG	COMPUTER MISTAKE
1	10		3			7				
2	10		2			7				1
3	8		1			7				
4	8					8				
5	8	4	1		1					2
6	8	4			2			2		
7	11	2		1	3		1			4
8	8	3	2			2		1		
9	9	1				3	1	1		3
10	14	3		3		3		1		4
11	9	4				3	1	1		
12	8	2				4				2
13	8				1	4				3
14	8	1			1	4				2
15	8					6				2
16	8	1				7				
TOTAL	143	25	9	4	8	65	3	6	0	23

TABLE 18

STUDENT ADMINISTRATION

VOCABULARY SUBTEST

VOCABULARY QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
1	13	3	1	1		7		1		
2	12	4				7				1
3	12	1				7				4
4	14	4			1	7				2
5	14	5				7		1		1
6	19	4			2	7	2			4
7	17	3		1	1	7	1	2		2
8	17	5		1		7				4
9	27	1	2	4	1	7	4	3		5
10	15	3	1				3			8
11	12	4				2	4			2
12	10	5		1	2	2				
13	8	5				2				1
14	8	1				3				4
15	12	2		1	1	4	2			2
16	8				1	4				3
17	8	2				4		1		1
18	8					4		1		3
19	8	2			1	4		1		
20	8				2	4				2
21	8		1			5				2
22	10				1	5	1			3
23	8				1	5		1		1
24	8				1	5				2
25	8					5		1		2

TABLE 18 (Cont'd.)

STUDENT ADMINISTRATION

VOCABULARY SUBTEST

VOCABULARY QUESTION	NUMBER OF RESPONSES	2 POINT	1 POINT	QUESTION	ZERO	PASSED	REPEAT	I DON'T KNOW	TYPED INCORRECT	COMPUTER MISTAKE
26	9	1	1		1	5	1			
27	9	1			1	6	1			
28	8	1				6				1
29	8					6				2
30	8					6				2
31	13		1			6	3	1		2
32	8					7				1
33	8					7				1
34	8					8				
TOTAL	371	57	6	10	17	178	22	13	0	68

END OF REEL

ARC-30