NATIONAL LIBRARY
OTTAWA

BIBLIOTHÈQUE NATIONALE
OTTAWA

CANADA

NAME OF AUTHOR....DAVID...ARTHUR.....PEARSON.

TITLE OF THESIS.....THE...INFLUENCE...OF...THE..
REFLECTIVE/IMPULSIVE...DIMENSION.
ON PROBLEM-SOLVING SKILLS..IN......
ELEMENTARY SCHOOL SCIENCE

UNIVERSITY.....UNIVERSITY....OF.ALBERTA.........

DEGREE FOR WHICH THESIS WAS PRESENTED.......PhD.........

YEAR THIS DEGREE GRANTED..........1976.................

(Signed).....D.A. Pearson.....

PERMANENT ADDRESS:
..9830-105 St. #706
..EDMONTON, ALBERTA
....

DATED...30th....July.....1976

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

THE INFLUENCE OF THE REFLECTIVE/IMPULSIVE DIMENSION ON PROBLEM-SOLVING SKILLS IN ELEMENTARY SCHOOL SCIENCE

by

DAVID ARTHUR PEARSON

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELEMENTARY EDUCATION

EDMONTON, ALBERTA

FALL, 1976

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "The Influence of the Reflective/Impulsive Dimension on Problem-Solving Skills in Elementary School Science" submitted by David A. Pearson in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

K. G. Jacknicke
Supervisor

D. A. Wilson

Cathcart

Neil M. Purcis

M. A. Nay

James R. Wailes
External Examiner

Date July 2 1976

ABSTRACT

The main purpose of this investigation was to compare performances of grade six students, classified as reflective and impulsive, on response uncertainty problem-solving tasks demanding adoption of (a) a guided-discovery approach and (b) an open-ended approach for completion.

Kagan's Matching Familiar Figures Test was used to classify students participating in the investigation as being reflective or impulsive in disposition, or of neither a marked reflective or impulsive disposition.

Guided-discovery tasks required students to select observations and hypotheses pertaining to problem situations depicted on a film-loop, and to subsequently evaluate the selections made in terms of their effect upon the depicted problem situations. Open-ended tasks required students to generate multiple observations and hypotheses pertaining to the same depicted problem situations, and to evaluate the generations made in terms of their effect upon those problem situations.

Reviews of Buros (1965, 1972) indicated that published guided-discovery and open-ended tasks appropriate to the requirements of the present investigation at the grade six level, did not exist. Accordingly, all guided-discovery and open-ended tasks were designed by the investigator. Efforts were taken to ensure that reliabilities and validities of the tasks were brought to as high a level as possible.

In addition to guided-discovery and open-ended tasks, three other measures were employed in this investigation to assess students'

abilities in observing and in hypothesizing. Content areas and modes of content presentation incorporated in these three tasks, differed from those incorporated in the guided-discovery and open-ended tasks. These measures were used to provide information concerning the extent to which observation and hypothesizing abilities can be considered task-specific.

Another purpose of the investigation was to compare the types of question asked by Reflectives and Impulsives during involvement in those problem episodes used as foci for the guided-discovery and open-ended tasks. Questions asked were subsequently assigned to specific category types according to criteria and format prescribed by Suchman (1962).

One hundred and eighty-four students participated in the study. Of these, sixty-eight were subsequently classified as being reflective in disposition and seventy as being impulsive in disposition.

Analysis of data resulting from administration of instruments employed in this investigation indicated the following:

1. Reflectives perform about as well as Impulsives on tasks measuring abilities in selecting and in generating observations.

2. Reflectives perform about as well as Impulsives in selecting and in generating hypotheses on tasks utilizing the particular content area and presentation mode associated with the film-loop.

3. Reflectives' performance is superior to that of Impulsives in selecting and in generating hypotheses on tasks utilizing content areas and presentation modes different to those associated with the film-loop. In the case of ability to generate hypotheses, Reflectives' performance is significantly superior to that of Impulsives.

4. Reflectives perform about as well as Impulsives on all tasks measuring abilities in evaluating selected and generated observations and hypotheses.

5. Reflectives ask similar types of Suchman questions and similar quantities of questions as Impulsives during involvement in problem-solving sequences.

6. Gagné's (AAAS, 1965) claim that observing and hypothesizing abilities manifest high inter-task generality was only partially supported in this investigation. In instances where tasks did not differ greatly in terms of content area, presentation modality and response requirement, moderately high correlations were obtained between task scores. Where these factors did differ greatly, essentially zero correlations were obtained between associated task scores.

ACKNOWLEDGEMENT

I should like to express my appreciation to those who have given assistance and cooperation at various stages of this study. In particular I am indebted to:

Dr. K. G. Jacknicke, my supervisor, for his constant support and constructive commentary.

Dr. M. A. Nay who remained a source of encouragement throughout the investigation. Completion of this thesis owes much to the perceptive and scholarly assistance he readily gave.

Dr. W. G. Cathcart, Dr. D. W. R. Wilson, Professor N. M. Purvis and Dr. J. Wailes for their interest and helpful advice.

TABLE OF CONTENTS

CHAPTER PAGE

CHAPTER PAGE

CHAPTER		PAGE

LIST OF TABLES

TABLE | PAGE

TABLE PAGE

TABLE		PAGE

CHAPTER I

THE PROBLEM

INTRODUCTION TO THE PROBLEM

Traditionally science in the schools has been viewed as a body of indisputable knowledge or 'rhetoric of conclusions' (Schwab, 1966, p. 24). The major responsibilities of the science educator have been to impart this knowledge and periodically to assess the extent to which students were retaining such information. Current major elementary science curricula such as Science a Process Approach (SAPA), Elementary Science Study (ESS) and Science Curriculum Improvement Study (SCIS), contend that teaching science merely through impartation of knowledge promotes an oversimplified and static conception of the nature of science. Accordingly these programs maintain that students can best develop awareness of the complex and multidimensional nature of science not only by learning about the products of the investigations of others, but additionally by participating in investigations giving rise to these products. These programs encourage and provide substantial opportunity for problem-solving activity on the part of students. Through direct participation in problem-solving sequences it is commonly held that students are able to more closely identify with the <u>modus operandi</u> of the practising scientist, and to achieve more mature insight into the complexity of science than would otherwise be possible where no such active involvement was entailed. Primarily for such reasons, problem-solving today occupies a central position within elementary science programs.

While SAPA, ESS and SCIS programs encourage active problem-solving activity, each of these programs adopts a distinctive approach towards such activity. This difference in approach is manifested in terms of the extent to which opportunities for teacher and student decision-making are provided. In general, SAPA advocates a 'guided-discovery' approach (Mosston, 1972) towards problem-solving, whereas ESS advocates an 'open-ended' approach (Keislar, 1969) towards problem-solving. In basic terms, within the guided-discovery approach decision-making concerning problem identification and solution evaluation, become the total responsibility of the teacher/program. The student is provided limited opportunity for decision-making only with respect to procedures adopted in achieving solution to the problem. Contrasting the guided-discovery approach, the open-ended approach permits considerably more opportunity for student decision-making. Although problems are still predominantly teacher identified, the student is provided abundant opportunity for making decisions concerning which procedures to utilize in solution attainment, and additionally provided opportunity to evaluate his own solution outcomes. In general terms, other current curricula, such as SCIS, adopt approaches to problem-solving which reflect a mix of the decision-making characteristics associated with the SAPA and ESS approaches. In initial stages of a SCIS problem-solving episode for example, students are given problems and provided considerable license to make decisions concerning procedures to adopt in solution attainment (exploration). After completion of these initial stages, however, few opportunities for further student decision-making exist (invention). In the final stages of a SCIS problem-solving episode, problem identification, procedures for solution attainment and solution evaluation are predominantly under the control of the teacher (discovery).

THE PROBLEM

With extensive implementation of newer elementary science programs currently taking place, an urgent need exists for assessing program effectiveness in teaching science to students. The fact that several programs have recently become widely and extensively adopted, and in many instances have totally replaced traditional modes of science teaching, undoubtedly indicates that these programs are widely considered to be generally effective in teaching science. However it has not been shown that all students were able or willing to adapt to the requirements of certain programs. This is hardly surprising since not only do large differences occur between program approaches as already ind ated, but marked differences also undoubtedly exist in terms of inte   ual and personological characteristics of problem-solvers involved (1970). Thus the common practice of wholesale implementation of programs of distinctive character for use by all students, may be delite us t some participants. In accordance with differential theory, Hunt (1970) maintains that the effectiveness of any approach should not be measured in terms of its general superiority (or inferiority) over another:

> Rather than ask whether one educational approach is generally better than another, one asks, "given this kind of a person

> which of these approaches is more effective for a given objective?" (Hunt, 1970, p. 68)

Assessment of the effectiveness of elementary science programs should therefore entail determination of the outcomes of meshing particular problem-solving approaches with specific individual characteristics known to influence problem-solving abilities. By so doing, any subsequent attempt to maximize particular students' science learning can be made through careful selection of approaches most appropriate to the individuals concerned and/or by the modification not only of those approaches, but also of those individual characteristics affecting problem-solving behaviour.

Individual differences in problem-solving have generally been attributed to differences in age, intellectual ability, concern for performance and to differences in overall knowledge of associated content (Kagan, 1965a). Recent research initiated by Jerome Kagan and associates indicates that a further attribute, the Reflective/Impulsive Dimension (R/I), has been shown to markedly affect problem-solving ability (Kagan and Kogan, 1970). This dimension is specifically concerned with the extent to which an individual reflects on the validity of responses during the solving of problems that entail response uncertainty (Kagan et al, 1964). In these situations an individual who characteristically requires lengthy periods of time to assess the validity of alternatives provided, is classified as 'Reflective.' In contrast to the behaviour of the Reflective, an 'Impulsive' makes selection rapidly (Kagan et al, 1964). Studies involving consideration of R/I set and problem-solving performance in several non-science subject areas demonstrate that problem-solving of the type noted is indeed markedly affected by R/I set (Kagan, 1965b, 1966a, 1966b; Kagan and Kogan, 1970). However no study has yet been cited investigating the impact of R/I upon problem-solving within an elementary science context. The intimate association between R/I and problem-solving, according to the Kagan definition (Kagan et al, 1964), suggests that investigation of the influence of R/I within elementary science problem-solving environments might profitably be undertaken. Furthermore, with noted elementary science programs adopting different approaches towards problem-solving the determination of the influence, if any, of reflective and impulsive sets upon each approach would serve to indicate favourable and unfavourable approach/set combinations with respect to problem-solving performance in an elementary science context.

THE PURPOSE

The principal purpose of this study is to investigate performance of students classified as reflective and impulsive on certain response uncertainty problem-solving tasks demanding adoption of (a) a guided-discovery approach and (b) an open-ended approach for their completion. Decision to utilize guided-discovery and open-ended approaches in this investigation is based on the fact that these particular approaches contrast markedly in terms of their teacher/student decision-making character. In addition (as earlier noted, p. 2) both approaches or combinations of these approaches are commonly utilized in several of the major elementary science curricula (e.g. in SAPA, ESS and SCIS).

In keeping with the character of a guided-discovery approach towards problem-solving, guided-discovery tasks will require the search for solution to those problem-situations that have been previously depicted and identified by non-student sources. Only limited opportunity will be provided for student decision-making during task completion. To this end alternative solutions will be provided, and the student will be required to make solution *selection* from those provided. In keeping with the character of an open-ended approach to problem-solving, open-ended tasks will require the search for solution to the same problem situations similarly depicted and identified by non-student sources. However, in contrast to the guided-discovery approach, substantial opportunity for student decision-making during task completion will be provided. Students will be required to *generate* multiple responses, rather than to make selections, from provided responses. With respect to the nature of the problem-solving tasks involved, all tasks accommodating guided-discovery and open-ended approaches will be constructed in a fashion enabling assessment of ability in utilizing the skills of visual observation, hypothesizing and evaluation. Decision to involve these particular skills was based upon the following rationale:

1. These skills predominantly feature within the Matching Familiar Figures Test (MFF), a test designed by Kagan (1965a) specifically for identifying and classifying reflective and impulsive behaviour (for details see p. 29 and p. 70). Measurement of Reflectives' and Impulsives' abilities to utilize each of these skills could provide information concerning the extent to which each

influences overall reflective and impulsive behaviour.

2. Gagné considers problem-solving skills, including observation and hypothesizing, to exhibit high inter-task generality (AAAS, 1965). If this proposition is correct, then the performance outcomes of Reflectives and Impulsives in each of these skills within the particular science content setting of this study, may serve to predict their performance in tasks utilizing similar skills in different content settings. Thus results of reflective and impulsive performance in observation, hypothesizing and evaluation from this study could have important implications for performance of Reflectives and Impulsives involved in science programs such as SAPA, ESS and SCIS that deliberately encourage the use, and even the teaching, of such problem-solving skills in a wide variety of content areas.

In summary, guided-discovery tasks will be used to determine ability of reflective and impulsive students to make selection of observations and hypotheses pertaining to depicted and identified problem situations. Open-ended tasks will be used to assess ability of reflective and impulsive students to generate observations and hypotheses pertaining to the same problem situations. Students will subsequently be required to evaluate all selections and generations made.

Another purpose of this study is to determine the types of questions asked by Reflectives and Impulsives during certain problem-solving activities. Studies by Ault (1973) and Denney (1973) indicate that Reflectives and Impulsives ask different types of questions during problem-solving. However, information concerning

Reflectives' and Impulsives' question-asking strategies from these studies is limited since questions asked were assigned to one of merely two defined categories. A more detailed question-type analysis would provide more information concerning question-asking strategies of reflective and impulsive types. The analysis format prescribed by Suchman (1962) appears to offer such a system. In this model the types of questions asked by problem-solvers may be allocated to six 'verification' and to six 'implication' categories (identification of these categories and associated definitions are given on p. 9). Verification questions are those attempting to identify and clarify objects, events and general conditions present in a problem sequence. Implication questions are those searching for relationships between associated variables. According to Suchman (1962), the relative frequency of question types employed by problem-solvers reveal their strategies and goals. Thus Suchman's model will be used to provide information that may assist in any subsequent attempt to modify ineffective question-asking strategies adopted by reflective and/or impulsive students during problem-solving episodes.

DEFINITION OF TERMS

For the purposes of this investigation the following definitions of terms will apply.

Reflective: Characterized by an individual who requires lengthy periods of time to assess the validity of responses during the solving of problems that entail response uncertainty (Kagan et al, 1964).

Impulsive: Characterized by an individual who requires short periods of time to assess validity of responses (Kagan et al, 1964).

Problem-solving: The activities involved in a search for the solution or solutions to a depicted and identified problem situation. This search can result in the selection of an acceptable response from among several alternatives provided (guided-discovery approach), or the generation of multiple acceptable responses (open-ended approach).

Observation ability: The ability to collect qualitative information concerning the physical attributes of objects and of events using the sense of sight.

Evaluation of observations ability: The ability to determine the effect, if any, that selected and generated observations (of objects and of events) have upon the depicted problem situation.

Hypothesis: A tentative explanation of the depicted problem situation.

Hypothesizing ability: The ability to recognize a possible cause or causes of the depicted problem-situation.

A-type Hypothesis: An hypothesis indicating one of the possible causes of the depicted problem.

B-type Hypothesis: An hypothesis indicating at least two of the possible causes of the depicted problem.

Evaluation of hypotheses ability: The ability to designate selected and generated hypotheses to an A-type or B-type category.

Suchman question types:

a. Categorical verification: nominal question. A question identifying the class to which the object observed belongs.

b. Categorical verification: normative question. A question designating the object observed according to the norm of the particular class.

c. Analytical verification: condition-descriptive question. A question specifying quantitative or qualitative conditions.

d. Analytical verification: condition-comparative question. A question identifying a relationship between two conditions.

e. Analytical verification: structural-component question. A question determining relationships between parts of a structural whole.

f. Analytical verification: properties check question. A question ascertaining the physical properties of an object.

g. Abstract-conceptual: diffuse question. A question in which the nature of the relationship between variables is not specified.

h. Abstract-conceptual: directed question. A question in which the nature of the relationship is specified.

i. Concrete-inferential: elimination question. A question eliminating an object to determine the effect upon the outcome of the experiment.

j. Concrete-inferential: substitution question. A question replacing an object by another in order to determine its effect upon the outcome.

k. Concrete-inferential: addition question. A question adding a new object to determine its effect upon the outcome.

l. Concrete-conceptual question. A question determining the necessity of an object or condition for occurrence of the phenomenon.

HYPOTHESES

In this investigation the following hypotheses will be tested:

H_1: There is no significant difference existing between Reflectives' and Impulsives' ability in selecting objects observed present in depicted problem episodes from objects not observed present.

H_2: There is no significant difference existing between Reflectives' and Impulsives' ability in selecting events observed occurring in depicted problem episodes from events not observed occurring.

H_3: There is no significant difference existing between Reflectives' and Impulsives' ability in selecting hypotheses that pertain to depicted problem episodes from hypotheses that do not pertain.

H_4: There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating the objects they selected.

H_5: There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating the events they selected.

H_6: There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating the hypotheses they selected.

H_7: There is no significant difference existing between Reflectives' and Impulsives' ability in generating objects observed present in depicted problem episodes.

H_8: There is no significant difference existing between Reflectives' and Impulsives' ability in generating events observed occurring in depicted problem episodes.

H_9: There is no significant difference existing between Reflectives' and Impulsives' ability in generating hypotheses that pertain to the depicted problem episodes.

H_{10}: There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating the objects they generated.

H_{11}: There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating the events they generated.

H_{12}: There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating the hypotheses they generated.

H_{13}: There is no significant difference in the total number of Suchman type questions asked between reflective and impulsive groups.

H_{14}: There is no significant difference in the distribution of stions of specific Suchman types asked between reflective and impulsive groups.

H_{15}: With respect to subjects asking one Suchman type question only, there is no significant difference in the number of questions of Verification and Implication type asked between reflective and impulsive groups.

H_{16}: With respect to subjects asking two Suchman type questions, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first and second questions asked, between reflective and impulsive groups.

H_{17}: With respect to subjects asking three Suchman type questions, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first, second and third questions asked, between reflective and impulsive groups.

H_{18}: With respect to subjects asking four Suchman type questions, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first, second, third and fourth questions asked, between reflective and impulsive groups.

ASSOCIATED QUESTIONS

In addition to the research hypotheses stated, the following Associated Questions will be examined in this investigation:

1. Are correlations among all Selection, Generation and Evaluation of Selection and Generation Scores, of similar magnitude for reflective and impulsive groups?
2. What common factors exist among the Selection, Generation and Evaluation of Selection and Generation variables?
3. Do significant differences exist between reflective and impulsive groups with respect to factor scores derived from common factors?

DELIMITATIONS OF THE STUDY

Although every effort will be taken to standardize procedures in administration and evaluation of tasks, experimenter bias may be present. To minimize bias in task administration the three central problems which provide the focal points for all guided-discovery and open-ended tasks will be depicted by means of a film-loop presentation, rather than by a demonstration on the part of the investigator. The film-loop will present problem situations in identical form and fashion to students participating from different classes. In efforts to minimize task evaluation bias, guided-discovery and open-ended tasks will be evaluated by the investigator, as well as by a trained science education doctoral student. Inter-rater reliability will be calculated to determine the degree of agreement among scoring procedures.

Short-term memory is demanded in any task requiring cognitive effort. To minimize influence of memory loss on performance during testing, tasks will be administered immediately following film-loop viewings.

All guided-discovery and open-ended tasks used in this study are of a paper-and-pencil format appropriate for administration on a group basis. To avoid problems of readability, task instructions and items will be read aloud by the investigator.

All open-ended tasks require written responses. It is assumed that the substance of these responses accurately reflects abilities in the various skills involved.

LIMITATIONS OF THE STUDY

Students from eight grade six classes will participate in the investigation. The eight classes will be randomly selected from schools located within middle-income areas as defined by the City of Edmonton Census (1971). Accordingly the generalizability of this study is limited to grade six students within such areas.

Studies indicate that the performance of Reflectives is significantly superior to that of Impulsives with respect to the Thurstone Primary Mental Abilities Test (Michenbaum and Goodman, 1969), reading ability and science achievement tests (Gupta, 1970). Accordingly covariate statistical techniques will be employed in the study to remove any bias due to I.Q., reading ability and general science knowledge on the performance of participating students.

Since most instruments employed in this study are investigator-designed, the problem of reliability and validity of instrumentation is an extremely important one. Several pilot studies will be conducted to assess the reliability and validity of these instruments, and to aid in their subsequent modification and refinement where necessary. By these procedures the reliabilities and validities of tasks used in the main study will be brought to as high a level as possible.

CHAPTER II

THEORETICAL FRAMEWORK AND RELATED LITERATURE

PROBLEM-SOLVING APPROACHES IN ELEMENTARY SCIENCE

According to Mosston (1972) differences in general teaching style may be identified according to the amount of decision-making manifested on the part of the teacher and/or student, and according to the location of impact of that behaviour. In more specific terms, the influence of decision-making may be felt at the following distinct places within the teaching process:

1. At the 'pre-impact' stage entailing content preparation,
2. At the 'impact' stage wherein content is implemented according to pre-impact decisions, and finally
3. At the 'post-impact' stage involving evaluation of that content.

Consideration of the impact of teacher decisions at each of these stages enabled Mosston to identify a spectrum of teaching styles ranging from the 'command style' in which teacher decision dominates all of the above stages, through approaches involving ever-decreasing degrees of teacher decision-making, to finally the 'individual program-students design' approach in which decisions at each stage become the total responsibility of the student. Utilization of Mosston's model within an elementary science context could effectively serve to differentiate between traditional and current programs in terms of the opportunities each program provides for active decision-making on the part of the students. More specifically, this model would indicate the extent and location of decision-making behaviour during problem-solving activities. In the ensuing discussion, comments concerning provision of opportunities for decision-making by different elementary science programs are based upon analyses of claims made by respective program designers and not upon classroom observations of opportunities provided. It is possible that provisions made at the classroom level for decision-making differ from those intended. Accordingly the following statements must be interpreted with this possible limitation in mind.

Traditionally science in the schools has been taught by heavy adoption of the command style, the teacher having had total authority over content, the implementation of content, and the subsequent evaluation of content and implementation. With decision-making in almost every instance under the control of the science teacher, it is not surprising that science in the schools has traditionally been viewed as merely a body of indisputable knowledge or scientific fact. Current elementary science programs provide considerable opportunity for active investigation on the part of students. Nevertheless opportunities providing for student decision-making vary from program to program. Thus in Mosston's terms, each program may be said to adopt a characteristic style or approach towards problem-solving. In contrast to the command-style traditionally adopted in elementary science, the SAPA program adopts a 'guided-discovery' approach (Mosston, 1972) towards problem-solving activity. In this program, decision-making on the part of the teacher/program is total at the pre- and post-impact stages, in the sense that objectives are succinctly specified and instruments are employed to assess the degree to which these specified behaviours are mastered. Student competence is thus measured entirely in terms of ability to meet these rigid requirements. At the impact stage, scientific knowledge is not imparted directly from teacher to student. Instead, students are expected to discover facts through active participation in series of sequentially organized activities. Each activity is deliberately contrived to elicit one 'correct' discovery response following each of the many activities. Exhibition of a response, whether correct or not, reflects the making of a decision on the part of the student with respect to the stimuli provided. Post-impact teacher evaluation of any response indicates whether or not participation in further activities, providing for further discovery-making, can be entertained. In contrast to the SAPA approach, the ESS program advocates a 'problem-solving' approach (Mosston, 1972), or as Keislar (1969) prefers an 'open-ended' approach towards scientific investigation. In this program, pre-impact subject matter decisions (including identification of problem-situations) remain the responsibility of the teacher. At the impact stage however, liberal opportunity now exists for student decision-making. The student is given license to make decisions concerning subject matter. In addition he is encouraged to produce multiple solutions. Decisions concerning these generations are predominantly his responsibility. In the post-impact stage, both teacher and student are involved in decision-making. As noted earlier (see p. 2) other major science programs (e.g. SCIS) adopt decision-making approaches that reflect various combinations of characteristics similar to the SAPA and ESS approaches.

PROBLEM-SOLVING SKILLS IN ELEMENTARY SCIENCE

In traditional science programs emphasis was commonly place upon the acquisition of the content or products of science to such a degree that active problem-solving was afforded at best scant attention. A call by the National Society for the Study of Education in 1947 for more consideration to be given in science education to problem-solving had little, if any, impact at the classroom level. School science continued to be taught largely as a body of indisputable and stable fact (Schwab, 1966). In instances where some effort was made to involve students in problem-solving activity, the Deweyian concept of the Scientific Method was commonly adopted as guide providing general direction in conducting such activity. This guide contained the following sequentially order stages or skills considered to be those which a problem-solver must utilize in order to attain solution to the problem at hand:

1. Definition of the problem,
2. Location and evaluation of associated data,
3. The formulation of an hypothesis,
4. The evaluation of the hypothesis, and
5. The conclusion. (Dewey, 1953)

In this sequencing of skills it is evident that a highly logical nature of science and sciencing was implied. The limited and generally ineffective usage made of this model was due largely to the demonstrated lack of existing guidelines that offered ways and means by which each of the noted skills might be identified, and their acquisition and application guaranteed some measure of success. Kneller (1966) maintains that the model was of limited benefit to students of elementary grades on account of its highly abstract nature, and even for older students it was limited in scope since these children would likely have advanced to a stage where more specific skills were needed. Thus the absence of clearly defined guidelines breaking down the skills into more specific ones capable of manipulation by students, ensured at best ineffective and infrequent application of the method.

With the launching of Sputnik in 1957, a large scale attempt was made to upgrade as quickly as possible the general quality of science education. In response to this demand several elementary science programs were designed and implemented, which, in contrast to traditional programs, provided substantial opportunity for active problem-solving participation by students. Furthermore these programs generally provided specific guidelines by which problem-solving sequences could ensue. Foremost among programs in this regard was the SAPA program. This program maintains that guided sequences incorporated within its framework provide involvement in activities approximating those of the practising scientist. SAPA regards the scientist's behaviours as complex sets of intellectual skills capable of analysis into simpler components. Advocates consider that these skills can in fact be learned, and that a sequence of instruction may be devised to facilitate their acquisition. Acquisition of these skills is based upon initial familiarity with each of the following basic skills or processes: observing, communicating, predicting, classifying, inferring, using numbers, measuring and using space/time relationships. Students are subsequently introduced to the following more complex 'integrated' skills: formulating hypotheses, controlling variables, interpreting data, defining operationally and experimenting. This subsequent incorporation of the complex processes purportedly takes place in accretionary form. By following this program it is maintained that participants become increasingly adept in making decisions and in problem-solving in general (AAAS, 1965). It can be seen that SAPA adopts a similar stand to that of the Deweyian approach in terms of the noted preference for the logical and sequential acquisition and employment of problem-solving skills. However, unlike the Deweyian approach, which fails to provide precise guidelines by which specific skills can be identified, used and evaluated, SAPA adopts a heavy behavioristic position. In this regard SAPA defines each skill in operational terms, and devises exercises to develop abilities in the particular skill involved. Objectives of these exercises are written in terms of observable and hence measurable aspects of student behaviour. Following completion of these exercises emphasis is placed upon the objective evaluation of performance in the particular skill highlighted.

It is interesting to note that designers of SAPA have recently started to modify their original program. In the new SAPA II curriculum, an increased emphasis given to the acquisition of scientific content has meant that the process skills do not completely dominate the program, as had previously been the case in the original SAPA program. While the sequentially oriented process hierarchy is still evident, it appears to be less rigidly defined and more flexible in operation. Complex integrated skills, for example, are given increased attention at earlier stages in the program. The acquisition and employment of these skills is apparently considered less dependent upon mastery of other less complex skills.

A departure from the basic SAPA rationale is manifested in the elementary science approach proposed by the Department of Education, Alberta (1969). While adopting a behavioristic stand, the Alberta Elementary Science Curriculum Guide (ESCG) contrastingly features each of the SAPA skills, both basic and integrated, at every elementary grade level. Thus while adopting SAPA processes wholesale, the ESCG appears to dispute the fundamental SAPA proposition that the acquisition of any problem-solving skill is dependent upon initial familiarity with the less complex skills. Nay (1971) provides further support for this proposition that problem-solving skills need not always follow one another in prescribed and logical sequence. Indeed Nay claims that certain skills fundamental to scientific investigation may be of a non-logical nature. Viewing each scientific investigation as a unique undertaking dependent for completion upon unique combinations of logical and non-logical processes, the author maintains that each skill is used only:

> when required and in a manner dictated by the problem being researched in the discipline. (Nay, 1971, p. 200)

Thus some processes may fail to function during a particular investigation, being effectively bypassed by others. In summary, teaching for the acquisition of problem-solving skills today enjoys widespread support, the behavioristic position generally finding favour. By emphasizing the acquisition and application of these skills it is believed that the general ability to solve problems will increase. Furthermore these problem-solving skills are believed to have high transfer potential, and claims have been made for their effective transfer to non-science areas (AAAS, 1965).

Visual Observation

Science curricula that deliberately accommodate and teach for the acquisition of problem-solving skills place heavy emphasis upon the development of visual observation skills. Two major curricula in this regard, SAPA and SCIS consider observation to be the most fundamental of skills within the scientific endeavour. SAPA furthermore claims that facility in observation skills is a fundamental prerequisite for the acquisition and demonstration of abilities in all other problem-solving skills. In accordance with this viewpoint, SAPA stresses development of observation skills in the first year of the program. Contrasting this policy, the ESCG, while recognizing the essential place occupied by observation in any problem-solving oriented program, places no emphasis upon the early acquisition of the skill. Instead exercises featuring observation skills are entertained at each of the six grade levels involved.

Definition of the term observation (visual). An analysis of SAPA, SCIS and ESCG programs indicates common acceptance of the following definition appropriate for use at the elementary school level. Observation (visual) entails the collection of qualitative data using the sense of sight. In more specific terms, this skill involves the identification of objects and object properties and the identification of changes in physical systems.

Hypothesis Formulation

According to the Deweyian model (Dewey, 1953) and to many other similar problem-solving skill compilations (reported in Williams, 1960) the formulation of hypotheses represents one of the most major skills basic to successful problem-solving. Several major science curricula, including SAPA, SCIS, ESS and the ESCG, and researchers such as Nay (1971), Quinn (1971), and Quinn and George (1975) consider teaching for the acquisition of ability in formulating hypotheses an essential part of the science program.

Definition of the term hypothesis. Although differences abound in definition of the term, most philosophers and science educators including Atkin (1956), Moore (1967), Quinn (1971), and Quinn and George (1975) consider the element 'explanation' to be a crucial constituent of an hypothesis. For the purposes of this study an hypothesis will be defined as a tentative explanation of a depicted problem.

Quality and Quantity of Hypothesis Production

The writings of Moore (1967), Quinn (1971) and Woodburn and Obourn (1965) indicate that the quality of an hypothesis depends upon the following factors:

1. The extent to which the presently known facts are accounted for,

2. The extent to which the hypothesis is capable of directing further action,

3. The extent to which the hypothesis lends itself to testing, and

4. The precision with which the hypothesis is stated.

The need for formulating hypotheses in copious supply has been emphasized by several researchers, including Chamberlain (1890), Moore (1967) and Osborn (1963). According to Chamberlain (1890), such measures reduce the likelihood of an hypothesizer developing a 'fondness' for one hypothesis, with the result that only supportive evidence for that one hypothesis tends to be sought. Osborn (1963) maintains that the greater the quantity of hypotheses formulated, the greater is the likelihood of good quality hypothesis formulation.

Mechanisms involved in hypothesis formulation. Opinions concerning the nature of hypothesis formulation differ widely. The following two opposing viewpoints are commonly held:

1. The claim that formulating hypotheses is a purely logical act, and

2. The claim that formulating hypotheses is dependent upon both logical and non-logical mechanisms.

Advocates of the SAPA strongly uphold the first position claiming that ability in formulating hypotheses is solely dependent upon demonstration of mastery of all specified prerequisite skills. Thus students are rarely exposed to hypothesis formulation before the third or fourth year of program implementation. In contrast, advocates of the second position (e.g. Martin, 1972; Osborn, 1963; Gordon, 1960) claim that non-logical as well as logical procedures are involved in hypothesis formulation, and accordingly must be allowed some sway. Thus logical procedures are often carefully devised to stimulate the assumed non-logical production of ideas. Common among advocates of these procedures is belief in the efficacy of incubation periods—periods which according to Freudian theo y precipitate in some unknown and non-logical way the production of hypotheses. The deliberate setting up of such periods is encouraged in several creative techniques that include Brainstorming (Osborn, 1963) and Synetics (Gordon, 1960). It is interesting to note that the ESS program considers the formulation of hypotheses to be a highly creative act, and accordingly devises classroom environments considered to stimulate and encourage such productions.

THE REFLECTIVE/IMPULSIVE DIMENSION

In recent years psychologists have identified several cognitive processes or 'cognitive styles' that have been shown to influence the form and lity of general cognitive performance. These cognitive styles are considered as information processing habits capable of modifying learning. To Messick (1970) they:

> represent a person's typical mode of perceiving, remembering, thinking and problem-solving . . . (p. 188)

and:

> are usually conceived as spontaneously applied, without conscious choice in a variety of situations. (p. 198)

Among other investigators, Kogan (1971), Messick (1970) and Witkin (1969) consider the influence of cognitive styles upon learning in general, sufficiently marked to warrant their attention in educational context. According to Messick (1970):

> . . . cognitive styles by embracing both perceptual and intellectual domains and by their frequent implication in personality and social functioning, promise to provide a more complete and effective characterization of the student than could be determined from intellectual tests alone. . . . these stylistic characteristics should have relevance, although direct research is admittedly scanty, not only for the course of individual learning in various subject areas, but also for the nature of teacher/pupil interaction and of social behaviour in the classroom, the family and the peer group. . . . cognitive styles, by virtue of their widespread operation, appear to be particularly important dimensions to assess in the evaluation of instruction. (pp. 195-196)

Of the several cognitive styles currently objects of theoretical and empirical examination, Kogan (1971) considers the Reflective/ Impulsive dimension (R/I) to have the most direct implications for education since this particular dimension is intimately involved with a person's evaluation of his own cognitive products during problem-solving.

Definitions of Reflective/Impulsive

1. Kagan's definition: The Reflective/Impulsive dimension to Kagan et al (1964) is concerned with the extent to which an individual reflects on the validity of responses during the solving of problems that entail response uncertainty. In these situations, an individual who characteristically requires lengthy periods of time to assess the validity of alternatives provided, is classified as a 'Reflective' (Kagan et al, 1964). In contrast to the behaviour of the Reflective, an 'Impulsive' makes selection rapidly (Kagan et

al, 1964). In more specific terms, the R/I dimension exerts an influence at stages (b) and (d) of the following generalized problem-solving sequence (Kagan, 1966b):

(a) Decoding or comprehension of the problem

(b) Selection of a likely hypothesis on which to act in order to arrive at solution

(c) Implementation of this hypothesis

(d) Evaluation of the validity of the solution arrived at in stage (c), and

(e) Reporting of solution to others.

Kagan et al (1964) warn against glib assignment of positive and negative values to reflective and impulsive behaviours (as defined by these authors) suggesting that each behaviour in turn may be desirable in certain instances:

> It seems reasonable to assume that effective learning and performance on varied intellectual tasks will sometimes be facilitated by a reflective or analytic approach, and sometimes by a more impulsive or less analytic orientation. (p. 35)

2. Freudian or Neo-Freudian definitions: In contrast to Kagan's conception of R/I which by definition is limited to the domain of problem-solving, Freudian and Neo-Freudian theory considers R/I as an all-pervasive characteristic affecting the total functioning of the individual. In common with such theory, impulsive behaviour has been associated with both psychopathic and anti-social activity, and has generally been assigned negative value by such theory. In the present investigation, Kagan's definition of R/I rather than that of Freudian-Neo-Freudian Theory will be adopted. Since as earlier noted, attention will focus upon reflective and impulsive

performance in problem-solving situations involving response uncertainty, Kagan's definition, though restricted, is particularly appropriate for this purpose.

Measures of Reflective and Impulsive Behaviour

Kagan's Matching Familiar Figures Test: This test is used to classify individuals as reflective or impulsive in disposition, as well as classifying those of neither marked disposition. The student is shown a picture (the standard) and six similar stimuli, one of which is identical to the standard. The subject is required to select the stimulus identical to the standard. This procedure is repeated using different groups of pictures. Two scores derived from the test provide standard measures of Reflective/Impulsive behaviour. A score of 'response latency' indicates the time elapsed before any response is made. The other score indicates the number of incorrect responses made prior to correct response. The inclusion of this latter score was recommended by Kagan (1965c) as a result of substantial evidence (see Kagan, 1965a, 1965b, 1965c) indicating negative correlations ranging from -.40 to -.65 between response latency and number of errors. These studies indicate that the Impulsive, though faster than the Reflective in responding, is typically less accurate in selection than the Reflective. Procedures adopted in classifying Reflectives and Impulsives on the basis of the two scores obtained are described in Chapter III (see p. 70).

Several other instruments exist capable of assessing reflective/impulsive behaviour. These include the Barrett Impulsiveness Scale (Barrett, 1965), the Guilford-Zimmerman Temperament

Survey 16 PF Scale (Cattell and Sheier, 1963) and the Thurstone Temperament Schedule (Thurstone, 1953). A decision to assess reflective/impulsive behaviour according to the referents of Kagan's instrument was based upon the following considerations.

1. Correlations in the order of .56 to .66 exist between the Barrett, Guilford-Zimmerman and Thurstone measures (Gupta, 1970). A significant relationship however, has been found lacking between the Barrett and Kagan tests, Kagan's instrument apparently identifying behaviours essentially dissimilar to those identified by the other measures (Gupta, 1970). If Gupta (1970) is correct in maintaining that impulsivity, as measured by the Barrett Scale, is more related to motor latency than to cognitive latency, then the employment of this Scale (and possibly those of Guilford-Zimmerman and Thurstone) in assessing performance in perceptual/cognitive domains involved in this study would appear to be inappropriate.

2. The adoption of an instrument measuring R/I specifically in terms of the definition of R/I accepted in this study, would appear consistent and justified.

3. The MFF test has been widely used in studies involving R/I measurement. The test is suitable for use at the elementary school level, and its administration is economic in terms of effort expended and time involved.

Determinants of Reflective/Impulsive Behaviour

Kagan (1965a, 1966a) postulates that both constitutional and psychological factors are responsible for differences in reflective/impulsive behaviour. The extent to which these factors operate singly or in combination to influence behaviour is thought to vary from individual to individual.

<u>Constitutional factors</u>. Kagan (1966a) maintains that constitutional factors are more powerful determinants of extreme impulsive behaviour than are psychological factors. Defects in the central nervous system (CNS) may be responsible, at least in part, for this form of behaviour. Kagan (1966a) believes that excessive motor activity in young children may be traced to deficits arising from brain damage effected during pre- and post-natal periods. It is possible that biological factors unrelated to CNS deficits predispose some to impulsivity. Schaefer and Bayley (1963) found that extremely active one year olds were minimally attentive to intellectual pursuits. Kagan and Moss (1962) found that hyperkinesis among three year olds correlated negatively with intellectual performance of these children at adolescence.

<u>Psychological factors</u>. Except in the case of extreme impulsivity mentioned, Kagan and Kogan (1971) maintain that standing on the R/I continuum is due more to psychological than to constitutional factors. According to Kagan (1966a) competence criteria may determine in part the extent of reflective and impulsive behaviour. On the one hand the Impulsive equates speed of response with competency and accordingly tends to respond quickly, performance accuracy being seemingly of lesser importance. The Reflective on the other hand, equates performance accuracy with competence and so takes time to respond in order to maximize performance accuracy. Developing these

ideas further, Kagan and Kogan (1971) suggest that reflective/impulsive behaviour may be substantially determined by anxiety over performance outcome. Anxious to avoid errors, the Reflective requires time to respond accurately. Minimal anxiety in making errors on the other hand, propels the Impulsive to respond quickly. A study by Ward (1969) involving kindergarten children supports this notion. Whenever errors were made by students engaged in an MFF test, they were informed, and a call for performance improvement made. Significant increases ($p<.0001$) in response latency were manifested by the Impulsives following the call. Furthermore Impulsives showed a significant tendency ($p<.0001$) to make fewer errors subsequent to notification of errors. Essentially similar results were obtained in a study by Messer (1970) involving grade three children.

Reliability of R/I

Temporal stability. A tendency exists for both Impulsives and Reflectives to become increasingly reflective with increase in age (Kagan, 1965a). Nevertheless each disposition is well established at an early age, and manifests both short- and long-term stability. A study by Yando and Kagan (1968) involving second grade children, yielded correlations of .70 between response latencies on MFF tests administered prior to, and following, a period of ten weeks. A study by Kagan (1965a) involving grade three and four children yielded correlations of .62 between response latencies of MFF tests given before, and after, a one year period. Finally a study by Messer (1970) indicated no significant change in response latency over a two and a half year span.

Cross-task generality. A tendency to be reflective or impulsive appears to exist across diverse tasks. Correlations between response latency on the MFF task and a haptic/visual task for groups of grade one, two, and three children, were within the .61-.87 range (Kagan et al, 1964). Further evidence for the generality of response latency comes from a study by Kagan (1965a) in which grade four and five boys and girls answered questions entailing some degree of response uncertainty. Correlations obtained between average temporal delay between termination of question and start of student's reply, and MFF response latency, were of the order of .30 ($p<.05$) for boys, and .38 ($p<.05$) for girls.

Correlates of R/I

R/I and intelligence score. A study by Michenbaum and Goodman (1969) indicated that reflective kindergarten students scored significantly higher than did impulsive peers on all subtests of the K-1 Thurstone Primary Mental Abilities Test. Contrasting these findings, a study by West (1973) revealed that no significant differences were found in scores obtained on the Wechsler Intelligence Scale for Children (WISC) between fourth grade reflective and impulsive groups. Furthermore the study indicated that moderate latency, rather than extreme fast or slow latency, is optimal to performance on the WISC. Children classified as moderately reflective achieved superior scores to those classified as highly impulsive. In addition, moderately reflective girls achieved scores superior to those of highly reflective girls.

R/I and school achievement. A study by Gupta (1970) indicated that scores of grade nine Reflectives on achievement tests in verbal ability, language, reading, mathematics and science were significantly superior to those of Impulsives. The superior performance of Reflectives on the verbal ability test calls into question Kagan's proposition (1965a, 1966a) that R/I is relatively independent of verbal ability. Superior, though not significant, performance of Reflectives on a social studies achievement test noted in Gupta's (1970) study, also questions Kagan's proposition (1966a) that performance in social studies may be hampered by a reflective set. Gupta's finding that Reflectives perform significantly better than Impulsives on mathematics tests is further supported in a study by Cathcart and Liedtke (1969).

R/I and memory for words. In a study by Kagan (1966b) grade three children participated in a serial-learning task requiring recall of familiar words read to them. While Reflectives made fewer errors than did Impulsives, the differences in performances were not significant.

R/I and reading prose. Grade one students (Kagan, 1965a) were required to identify the one word among five others that matched a word previously spoken aloud by the investigator. The task was repeated using different words. After partialling out verbal ability, Reflectives made significantly fewer reading errors than did Impulsives ($p<.05$).

R/I and reasoning. In a study by Kagan et al (1966) grade one students were given tasks involving inductive reasoning. In most task Impulsives exhibited significantly more errors than did Reflectives, even when verbal ability was controlled.

R/I and sex difference. Where studies have identified sex differences, these differences have generally been of low order and non-significant. Studies by Kagan (1966a) indicate that the relationship between number of errors and verbal ability is less pronounced for boys than for girls. In terms of response latency, girls manifest more stability than do boys (Kagan, 1966a). Meichenbaum and Goodman (1969) found that kindergarten girls are more impulsive than boys on response latency scores. Ward (1969) however, found no sex difference with respect to response latency scores manifested by kindergarten children.

R/I and economic status. A study by Hess and Shipman (1965) showed that mothers of low economic status and their children are more impulsive than their middle income counterparts. Information concerning the significance of these findings was unavailable. A study by Souch (1970) indicated that children of low income groups tend to exhibit shorter response latencies than do middle income peers. In contrast, a study by Yando and Kagan (1968) found that economic status bore no relationship to response latency.

R/I and eye-tracking patterns. Studies ascertaining behavioral operations have been conducted in efforts to comprehend the bases for the typically rapid and frequently incorrect responses of the Impulsive, and the slow, often correct responses of the Reflective. Siegelman (1969) used a scanning device to record visual fixations of fourth graders performing on an MFF task. Behaviours in terms of frequency and duration of looks focused on the standard, and on each of the alternatives were recorded. Significantly higher scores ($p<.001$) on measures of frequency and duration of looking behaviour were obtained by Reflectives over Impulsives. In terms of relative deployment of attention, Reflectives spent less time looking at the standard, looking at the most observed alternative stimulus, and looking at the alternative subsequently chosen, than did Impulsives. In addition, Reflectives made fewer numbers of looks at these particular stimuli than did the Impulsives. Impulsives ignored two and a half times as many alternatives as did Reflectives, suggesting both a biased and more peaked distribution of attention. Reflectives appear to first compare alternatives for differences, consulting the standard for confirmation or rejection. Impulsives on the other hand seem to compare the standard globally with one alternative at a time. Failing to perceive any difference between an alternative and the standard, the Impulsive tends to choose that alternative, paying little attention to the others.

R/I and cognitive maturity. In a study by Mosher and Hornsby (1966) marked differences in the question-asking strategies employed during problem-solving episodes were noted between groups of six year old and ten year old children. To Mosher and Hornsby (1966) these results are indicative of differences in level of cognitive maturity existing between the age groups. Shown different pictures, these groups of children were required to ask questions ascertaining the one picture that the investigator had in mind. Question types were assigned to one of two categories. 'Constraint-seeking' questions were those that indicated evidence of attempts to interrelate properties of different pictures during solution of the problem. 'Hypothesis-seeking' questions were those indicating consideration of pictures as wholes, rather than as combinations of parts. Questions of this type showed no evidence of attempts to interrelate properties of different pictures. Six year olds asked primarily hypothesis-seeking questions and rarely asked constraint-seeking ones. Ten year olds in contrast, primarily asked constraint-seeking questions and rarely asked hypothesis-seeking questions. Employing this format, studies by Ault (1973) and by Denney (1973) found that elementary grade Impulsives predominantly asked hypothesis-seeking questions, whereas Reflectives predominantly asked constraint-seeking forms. These findings suggest that the cognitive behaviour of Impulsives may resemble that of young children. This idea had originally been hypothesized by Siegelman (1969). She maintained that the Impulsive's acceptance of the immediately perceived global stimulus resembled the centration described by Piaget in young children (Siegelman, 1969).

Modification of Impulsive Behaviour

The accumulation of evidence indicating the general desirability of a reflective set over an impulsive one has stimulated attempts to modify impulsive behaviour. First attempts in this direction studied the effects of increasing response latency. Kagan et al (1966) reported that impulsive first graders engaged in inductive reasoning tasks could be trained to inhibit rapid decision-making. After three hours of direct tutoring, significantly lower response latencies ($p<.05$) were obtained; accuracy however, remained unaffected. A study by Yando and Kagan (1968) showed that grade one Impulsives, at the end of an eight month period of instruction under a reflective teacher, exhibited a significantly greater increase in response latency ($p<.01$) than did Impulsives who had had an impulsive teacher. Again error scores were unaffected. A study by Rowe (1973) involving elementary school children of different ages manifested increases in accuracy of performance following manipulation of response latency.

Limited success in modifying impulsive behaviour through manipulation of response latency led other investigators to focus attention upon the strategies employed during problem-solving. In a study by Debus (1970), impulsive third graders observed one of the following types of sixth grader actively involved in solving an MFF task: (1) an Impulsive, (2) a Reflective, (3) a model who initially exhibited an impulsive strategy, and then shifted to a reflective one, (4) two successive models, an Impulsive followed by a Reflective. In each case the model gave a verbal description of his strategy and the experimenter indicated the correctness of the model's choice. The reflective model had been previously instructed to succeed at the task, and the Impulsive to fail. Results indicated that all models served to increase significantly ($p<.01$) response latencies on MFF tasks administered, but failed to better the accuracy of the impulsive children. It is difficult to identify the factors responsible for the outcomes of this study. In an effort to isolate these factors, Ridberg, Parke and Hetherington (1971) utilized models who also verbalized response strategies. In addition they demonstrated scanning strategies by means of finger tracings. Impulsives were exposed to a film, in which a reflective model successfully solved a MFF task. Results of pre- and post-MMF test administration indicated that Impulsives showed significant increases in response latencies over control Impulsives ($p<.01$) and furthermore made significantly fewer errors ($p<.01$). In a study conducted by Zelnicker et al (1972), eye movements of a group of nine year old Impulsives were recorded while working first on an MFF task, requiring the normal selection of one variant identical to the standard, second, on a differentiation task requiring selection of a variant different from the standard, and finally on an MFF task similar to the first. A control group of nine year olds was tested under similar conditions except that no participation took place on the differentiation task. Results indicated that the experimental Impulsives (i.e. those who participated on all tasks) made fewer errors on the final MFF task than on the first, whereas the control Impulsives made more errors on the final task than on the first. Improved performance of the experimental Impulsives was not accommodated by longer latencies, but rather by a change in scanning strategy effected by the differentiation task. To Zelnicker and Oppenheimer (1973) the performance improvement resulted from a change in the type of processing of the visual information, the perceptual process being modified according to the task requirements, in this case looking for differences rather than for similarities. In the matching task Impulsives made selections by overlooking features distinguishing the alternative from the standard. The differentiation task, in contrast, did not allow selection to be made without detection of the features distinguishing the alternative from the standard. A study by Egeland (1974) indicated that training Impulsives working on MFF tasks to attend to all alternatives provided, to break down alternatives into component parts, and to look for similarities and differences across the alternatives, is as efficient as training for response delay in the immediate modification of impulsive behaviour. Both types of training manifested significant increases in response latency, and decreases in number of errors made, while the search strategy group maintained a low level of error response.

COMMENTS PERTAINING TO THE LITERATURE

According to Kagan (1966b) the R/I dimension is influential in both hypothesis selection and evaluation. Although Kagan intimates in his generalized problem-solving schema that hypothesis evaluation follows hypothesis selection it is doubtful whether the two operate separately, evaluation likely taking place before, during and after hypot[illegible]ection. The study by Siegelman (1969) cited page 35 ind[illegible] Reflectives make a more detailed comparison of similar eleme[illegible]s vari[illegible] alternative hypotheses than do Impulsives, who commonly fai[illegible]observe and hence compare certain variants. These findings sug[illegible] the Reflective places heavier emphasis upon evaluation of alternatives than does the Impulsive, by subjecting details to more efficient visual scrutiny. These differences may explain, at least in part, the lengthier responses and more accurate replies of the Reflective. In summary, when placed in problem-solving situations that involve selection from a group of alternatives, Reflectives, in contrast to Impulsives, utilize more efficient visual search skills and more efficient evaluation skills.

In spite of Kagan's insistence (see p. 27) that performance in certain intellectual tasks will sometimes be best facilitated by a reflective set and sometimes by an impulsive one, findings of studies reviewed above indicate a reflective disposition to be generally more favourable to performance in problem-solving than an impulsive one. Since the R/I dimension has also been shown to exhibit generality across task boundaries (see p. 33) a reflective disposition would appear in turn to be generally appropriate and conducive to performance on science-related tasks. Caution however should be exercised in the too ready acceptance of this generalization. In all studies reviewed, substantial external structuring was imposed upon the tasks, in the sense that in each case a problem was posed, and from the several alternatives provided, the subject was required to select the one response judged acceptable by external sources. In terms of Mosston's (1972) decision-making model, here the impact stage alone may be considered to provide for limited decision-making on the part of the student. Both pre- and post-impact decision-making remain the total responsibility of external sources. Thus the structure of these tasks provides for the tapping of essentially guided-discovery problem-solving activity. No opportunity for open-ended student inquiry was provided. Thus it can be argued that while a reflective disposition has indeed been shown more appropriate than an impulsive one in response uncertainty activities demanding a preponderance of guided-discovery, its superiority over impulsive behaviour in more open-ended problem-solving situations remains in doubt. Involvement in tasks with less externally imposed structure in which alternatives are not provided, such as in tasks which require generation of multiple alternatives, would enable the appropriateness of reflective and impulsive behaviours to be explored in more open-ended situations. Only then should the modification of disposition in elementary school science be contemplated.

CHAPTER III

DESIGN FOR THE STUDY

THE SAMPLE POPULATION

The sample utilized in this study was drawn from the population of grade six classrooms of schools within the Edmonton Public School System located in middle-income areas. Data from the most recent City of Edmonton Census (1971) indicated a mean income of household heads for each of the eighty census tracts identified, as well as a mean City of Edmonton income. For the purposes of this study a middle-income area is defined as one in which the mean income of household heads falls within a range plus or minus one thousand dollars of the mean City of Edmonton income. Decision to experimentally control for economic status as indicated by income, rested upon evidence marshalled by Hess and Shipman (1965) and Souch (1970) indicating substantial correlation between this variable and R/I. Thirty-three elementary schools were found to be located within middle-income areas. Ten schools were randomly chosen, and from these eight agreed to have students participate in the study. From each of the eight schools involved, students from one complete grade six class participated in the investigation. Table 1 summarizes the composition of the sample employed with respect to school affiliation and sex.

TABLE I

DISTRIBUTION OF STUDENTS WITH RESPECT TO SEX AND SCHOOL AFFILIATION

School Identification	Males	Females	Students per School
1	12	13	25
2	10	15	25
3	10	13	23
4	11	15	26
5	8	10	18
6	15	7	22
7	13	17	30
8	6	9	15
TOTALS	85	99	184

PROBLEM EPISODE PRESENTATION

Problem episodes forming foci for all guided-discovery and open-ended tasks were presented by means of a film-loop. This means of presentation appeared preferable to one involving investigator demonstrations for the following reasons. (1) Employment of a film-loop ensured that the episodes were presented in identical form and fashion to each of the eight classes involved in the study. (2) The nature of the study demanded several showings of the problem situations, a requirement easily facilitated by means of the loop mode of presentation, but considerably less easily by investigator-demonstration means. (3) Finally, by careful arrangement beforehand of film equipment and students' desks, the film-loop presentation provided participating students clear and unhampered viewing access.

For the purposes of this study, the Suchman (1962) film-loop 'The Knife' was used. In this film the following sequences are depicted: A straight bimetallic strip (the knife) is heated which subsequently bends downward. Then the strip is placed in a liquid and straightens out. After being turned over the strip is again heated, this time gradually bending upwards. Placed again in the liquid the bent strip once more straightens out. Three distinct problem situations can thus be identified in the loop, serving as foci for subsequent student activity. First, the strip bent downwards, second the strip bent upwards and third, the strip straightened out on two occasions. Decision to employ this particular film-loop among the several others available was based upon the following considerations. First, this loop has been effectively used by

Suchman (1962) with grade six children. Secondly, according to Piagetian theory, conceptual development is best facilitated by provision of environments that confront the learner with partially familiar occurrences promoting some degree of cognitive disequilibrium (Inhelder and Piaget, 1958). In efforts to assimilate the new information, thereby removing the disequilibrium, the individual is forced to modify and refine existing conceptual frameworks. In accordance with this position, problems depicted in 'The Knife' are neither completely familiar nor completely foreign to the experiences of grade six children. Yet the problem situations portrayed are novel enough to cause some disequilibrium on the part of students who cannot account for all events in terms of concepts normally acquired at this age. According to Suchman (1962) such a problem environment is highly effective in stimulating active investigation on the part of students. Thirdly, in the film-loop the number of significant variables involved in the production of the problem situations are few. Use of a film-loop in which complex causal relationships are absent appears appropriate at the grade six level.

INSTRUMENTATION

Introduction

A review of Buros (1965, 1972) indicated that very few published instruments exist capable of assessing visual observation and hypothesizing skills at the elementary level. This lack of suitable instrumentation is explained in large part by the fact that problem-solving skills in general have only recently been afforded considerable attention by certain elementary science curricula (e.g. SAPA, SCIS, ESS and ESCG). In the following two sections brief reviews of instruments currently available for evaluating observation and hypothesizing skills will be given. Although some of the instruments reviewed were designed for use at the junior high school level only, their inclusion here appears justified in view of the fact that certain investigator-designed tasks utilized in the present investigation were based in part upon criteria associated with those instruments.

Evaluation of Observational Skills

Tannenbaum (1971) developed a multiple choice instrument assessing abilities of junior high school students in visual observation skills. The following abilities were measured using this test: The demonstration of an operational knowledge of physical properties of objects, the identification and description of objects and object systems, the discrimination between spatial relationships of objects within a system, and finally the noting of visually observed characteristics of phenomena. Using grade seven, eight and nine students (N = 3673) KR_{20} reliabilities of .41 and .42 were reported for this test. Evidence establishing content and curricula validities was also provided. Hungerford and Miles (1969) devised an observation skill test assessing ability of seventh grade students in the identification and communication of morphological characteristics pertaining to deciduous twig specimens provided. Observation ability was measured in terms of three derived scores: (1) the proportional accuracies with which drawings of the twigs were made, (2) the accuracy of communication of morphological features and (3) the accuracy of labelling of major features. Inter-judge scoring reliabilities of .90 and .94 were reported. Using parallel-form procedures, reliabilities of .25 and .49 were reported (Hungerford and Miles, 1969).

Evaluation of Hypothesis Formulation

Works noted earlier (see p. 24) indicate that both quality and quantity factors should be accounted for in any effective and objective evaluation of ability in formulating hypotheses. Clearly the evaluation of quality of production is far more complex an undertaking than that of the quantity of production. Unfortunately, little research as yet exists pertaining to this particular area. In one effort to objectively evaluate the quality of hypothesis formulation, Quinn (1971) and Quinn and George (1975) designed an instrument identifying a spectrum of qualities against which hypotheses can be evaluated. The following criteria upon which these qualities are based, essentially reflect the commonly accepted hypothesis quality factors earlier noted on page 25.

Quality Hierarchy	Criteria
0	No explanation, such as a nonsense statement, a question, an observation, an inference.
1	Non-scientific explanation.
2	Partial scientific explanation . . . incomplete reference to variables . . . an analogy.
3	Scientific explanation relating at least two variables in general or non-specific terms.
4	Precise scientific explanation, a qualification and/or quantification of the variables.

5 Explicit statement of a test of an hypothesis. (Quinn and George, 1975, p. 290)

An inter-judge scoring reliability of .94 was reported for this scale (Quinn and George, 1975). The evaluation of hypothesis formulation procedures [illegible] by Quinn and George did not include assessment of quantity o[illegible]es generated. A study by Pearson (1975) involved the [illegible]us assessment of both quality and quantity of hypothesis production. Employing visual/mechanical and visual task modes, this study assessed the quantity of 'high quality' hypotheses generated by students in elementary grades. Hypotheses qualities were defined in terms of functional capabilities within the specific experimental situation entailed. Using parallel-form procedures a reliability of .91 was obtained on the visual/mechanical task and a reliability of .86 was obtained on the visual task.

Investigator-designed Instruments

The majority of tasks employed in this investigation were designed by the investigator. Such construction was deemed necessary since, as noted above, reviews of Buros (1965, 1972) indicated that published tasks appropriate to the specific requirements of this particular investigation at the grade six level, were not in existence. For this reason all guided-discovery tasks, guided-discovery evaluation tasks, open-ended tasks and open-ended evaluation tasks, were constructed by the investigator. As noted earlier (see p. 7), guided-discovery tasks require students to select observations and hypotheses pertaining to the problems depicted on the loop by indicating choices made in response categories provided. The general format of these tasks thus facilitates task-item analysis. Accordingly, guided-discovery tasks were subjected to item-analysis at several junctures in their development in order to determine the following information:

1. The Kuder-Richardson 20 reliability coefficient of the task. This value indicates the degree to which the task is internally consistent.

2. Item difficulty index (p) values. A p value indicates the proportion of the total group answering the test item correctly. The acceptability range for p values of .2 to .8 proposed by Stanley and Hopkins (1972) will be adopted throughout this study.

3. Item discrimination index (D) values. A D value indicates how well a particular item distinguishes between students who comprehend the test content, and those who do not. The acceptability range for D values of .2 and above, proposed by Ebel (1965), will be adhered to in this investigation.

As noted earlier (see p. 6) open-ended tasks require students to generate multiple observations and hypotheses pertaining to the depicted problems. The open-ended nature of these tasks did not facilitate task-item analysis.

In the following sections a general task description, and the procedures adopted in refining and scoring investigator-designed tasks, will be given.

G·ided-discovery Tasks

Object Selection Task. Refining procedures: A task consisting of sixteen items was constructed, of which eight items consisted of objects present on the loop and eight consisted of objects not present on the loop. Students were required to select objects present from those not present. The task in initial form was piloted on one complete grade six class (N = 26) that did not subsequently participate in the main study. Results obtained were scored and subjected to item analysis. Item-analysis revealed a very low KR_{20} reliability of .28. Only two of the sixteen items manifested acceptable p and D values. Unacceptable p and D values obtained on all other items indicated the need for rejection or extensive modification of these items. Using the item-analysis data a second edition of the task was constructed. The previous task content comprising merely a list of individual objects was abandoned and replaced by content in which the objects were described in some detail. By these means assessment of ability to observe objects solely as wholes now gave way to assessment of the students' ability to observe specific attributes of objects (e.g. shape, size, colour, etc.). Following administration to a second class of grade six children (N = 28), which was not involved in the major study, an item-analysis indicated that seven of the fifteen items received acceptable p and D values. Based on these acceptable items a KR_{20} reliability of .42 was obtained. This value compares favourably with values of .41 and .42 obtained by Tannenbaum (1971) on Observation sections of a Test of Science Processes. However since Jacknicke (1974) obtained a higher KR_{20} reliability of .61 on a Process Skill

Test, which contained several items assessing ability in visual observation, the noted KR_{20} reliability of .42 was not deemed acceptable. A final edition of the task was then prepared, consisting of eight items (see Appendix A for a copy of the instrument). Item p and D values are reported in Appendix G. The items were examined for face and content validity by a panel of judges consisting of three science educators, one mathematics educator and one science education doctoral student at the University of Alberta. Following administration to the major sample employed in this study (N = 138), a KR_{20} reliability of .54 was obtained. Since there are no published tests of a similar nature other than Tannenbaum's (1971) reported in Buros (1972) with which to compare KR_{20} reliabilities, the reliability of .54 obtained was judged to be at an acceptable level for purposes of this investigation. In order to determine if the task manifested stability over time, a test-retest was conducted on one class of students (N = 26) employed in the main sample. Analysis revealed a test-retest reliability based over a two week period of time of .78 (Winer, 1967, pp. 105-132). On the basis of this reliability, the test was judged to have stability over time.

Scoring procedures: An Object Selection score was derived by summing the total number of objects correctly selected on the task. One point was given for each correct selection made. Independent scoring by a trained science education doctoral student revealed a 100% investigator-doctoral student agreement.

Event Selection Task. Refining procedures: A task consisting of sixteen items was constructed. Eight items of the task consisted of events occurring on the film-loop and the remaining items consisted of events not observed to occur on the loop. Students were required to select those events which occurred on the loop. Results from first pilot administration of this task (N = 26) were subjected to item-analysis and indicated a KR_{20} reliability of .30. Only four of the items had acceptable p and D values. The rejection or modification of all other items appeared necessary. A second edition of the task was subsequently constructed, consisting of thi[illegible]n items which were described in more detail than were those of the first edition. Results obtained from the administration of the second edition indicated that seven items achieved acceptable p and D values. Based on these items a KR_{20} reliability of .46 was obtained. This value compares favourable to those of .41 and .42 obtained by Tannenbaum (1971) on Observation sections of the Test of Science Processes. A final edition of the task was prepared, consisting of eight items (see Appendix B for a copy of the instrument). Item p and D values are given in Appendix G. All items were examined by the same panel of judges (see p. 52) to determine face and content validity. Using the final version with the main sample (N = 138), a KR_{20} reliability of .61 was realized. In view of the lack of similar tasks with which to compare KR_{20} reliabilities, this value of .61 was judged adequate for purposes of this investigation. A test-retest reliability of .80 (N = 26) was obtained over a two week period of time on the final edition. Stability of the task over time was thus established.

Scoring procedures: An Event Selection score was derived from the total number of events correctly selected. One point was given for each correct solution. An investigator-doctoral student scoring agreement of 100% was obtained.

Hypothesis Selection Task. Refining procedures: Three distinct problem situations were identified within the film-loop 'The Knife' (see p. 45). Consistent with this fact, the Hypothesis Selection Task incorporates three subtasks, each subtask being exclusively associated with one of the three different problem situations depicted. Each subtask required students to select between hypotheses offering explanations of the particular problem identified and hypotheses offering explanations of a problem other than the one identified. In their initial forms, subtask 1 consisted of eight items, subtask 2 eight items and subtask 3 six items. Results obtained from first pilot administration of task (N = 26) indicated that six subtask 1 items, five subtask 2 items and two subtask 3 items had acceptable p and D values. Subsequent rejection or modification of other items appeared necessary. Results of a second pilot administration of a revised edition of the task (N = 28), indicated that two items associated with subtask 1, two items with subtask 2 and one associated with subtask 3 failed to reach acceptable p and D values. Based on acceptable subtask items, KR_{20} reliabilities of .32, .32 and .34 respectively were obtained on each subtask. There are no published tasks of a similar nature reviewed in Buros (1972) with which to compare reliabilities. However since Jacknicke (1974) obtained a KR_{20} reliability of .61 on his Test of Process Skills, a test containing several items measuring ability in hypothesizing, the noted KR_{20} reliabilities were deemed unacceptable. Following subsequent mir modification to certain items reliabilities associated with each subtask of .40, .38 and .39 respectively were obtained (N = 28). (See Appendix C for a copy of the instruments. Item p and D values are given in Appendix G.) All items were examined by a panel of judges (see p. 52) to assess face and content validity. Employing a final version with the major sample (N = 138), KR_{20} reliabilities of .48, .45 and .44 were obtained on each of the subtasks repectively. Based on the complete sixteen item Hypothesis Selection Task, an overall KR_{20} reliability of .52 was obtained. In view of the noted lack of similar tasks with which to compare reliabilities, reliabilities obtained were judged adequate for purposes of this investigation. Based on a two week interval of time, test-retest reliabilities of .78, .82 and .75 were obtained on each of the final edition subtasks respectively. Stabilities of the subtasks over time were thus established.

Scoring procedures: Hypothesis Selection subscores pertaining to each of the three subtasks were derived from the number of hypotheses offering explanations of the particular problem depicted which were correctly selected. One point was awarded each correct solution. A total Hypothesis Selection score was obtained by adding the three subscores. Investigator-doctoral student scoring agreements of 100% were obtained on each subtask.

Guided-discovery Evaluation Tasks

Evaluation of Selected Objects Task. Refining procedures: This task is closely associated with the Object Selection Task (see

p. 51). Students were required to evaluate the extent to which objects, which were previously selected as present in the Object Selection Task, had an effect upon occurrence of the depicted problem. In the first administration of the task (N = 28), 50% of the students were provided dichotomous response keys in which objects previously selected by students were evaluated in terms of having 'no effect' or 'some effect' on the depicted problem. Other students were provided trichotomous response keys in which objects were evaluated in terms of having 'no effect,' 'some effect' or 'large effect' upon the problem. These tactics were employed to ascertain whether or not students in general were able to differentiate between three response categories. Very low agreements between students and investigator in terms of evaluating responses according to trichotomous categories resulted, indicating widespread difficulties in making often fine distinctions between the three category system. Considerably higher investigator-student agreements were obtained on tasks utilizing the dichotomous keys. Thus for the specific purposes involved, dichotomous evaluation response categories appeared more appropriate than trichotomous ones. Following the second refinement of the associated Object Selection Task, item analysis indicated that all items on the Evaluation of Selected Objects Task manifested acceptable p and D values. Based on these items, a KR_{20} reliability of .45 was obtained. All items were examined at this point by the panel of judges (see p. 52) to assess face and content validity. A final version of the Task utilized with the main sample (N = 138) yielded a KR_{20} reliabilit of .50. With the absence of any similar tasks published in Buros

(1972), this reliability was deemed acceptable for the purposes of this investigation. (See Appendix A for a copy of the instrument. Item p and D values are given in Appendix G.) A test-retest reliability of .81 (N = 26) was obtained over a two week period on the final edition. The stability of the task over time was thus established.

Scoring procedure: An Evaluation of Selected Objects score was derived from the total number of objects previously correctly selected as present, and subsequently correctly classified as having no effect or some effect upon the depicted problem. One point was awarded each correct evaluation. An investigator-doctoral student scoring agreement of 100% was obtained.

A major weakness exists in the scoring format associated with this Evaluation of Selected Objects Task (and with scoring formats of all other Evaluation Tasks). Scores obtained are in part determined (limited) by scores previously obtained on the associated Object Selection Task. For example a student who previously failed to select any objects correctly on the Object Selection Task would be unable to score higher than zero on the Evaluation of Selected Objects Task since only objects correctly selected are required to be evaluated on this Task. Only in the case where maximum scores are obtained on the Object Selection Task, can scores on the Evaluation of Selected Objects Task be considered independent of Object Selection scores. Due to this weakness spuriously high correlations are likely to be manifested between Evaluation of Selected Objects scores and associated Object Selection scores. (Spuriously high correlations are also likely to exist between all other Evaluation Tasks and respective Selection or Generation Tasks.)

Evaluation of Selected Events Task. Refining procedures: This task is closely connected with the Event Selection Task (see p. 52). Students were required to evaluate the extent to which events previously selected in the Event Selection Task had an effect upon occurrence of the depicted problem. Results obtained from employing dichotomous and trichotomous keys identical to those utilized in the Evaluation of Selected Objects Task, indicated the superiority of the dichotomous system over the trichotomous one. Using trichotomous keys a very low investigator-student agreement was obtained, whereas a substantially higher agreement in evaluating responses was obtained using dichotomous keys. Acceptable p and D values on all items were obtained following the second refinement of the associated Event Selection Task. Based on these items, a KR_{20} reliability of .42 was obtained. At this point all items were scrutinized by judges (see p. 52) to assess face and content validity. The final version of the task used with the main sample (N = 138) manifested a KR_{20} reliability of .48. With no published task available with which to compare reliabilities, this reliability was deemed adequate for the purposes of this investigation. (See Appendix B for a copy of the instrument. Item p and D values are given in Appendix G.) A test-retest reliability of .85 was obtained over a two week interval (N = 26). The stability of the task over time was thus established.

Scoring procedure: An Evaluation of Selected Events score was derived from the number of events previously selected as present and subsequently correctly classified as having no effect or some effect upon the problem depicted. One point was given each correct evaluation. An investigator-doctoral student scoring agreement of 100% was achieved.

Evaluation of Selected Hypotheses Task. Refining procedures: As noted earlier (see p. 48) few attempts have been made to evaluate objectively the quality of formulated hypotheses. However the Quality Scale devised by Quinn and George (1975) is the outcome of one such attempt. This instrument provides a six-point scale against which the quality of hypotheses may be assessed (for details see p. 48). Since it was considered unlikely that students in the present investigation would be able to differentiate between these six qualities, a decision to modify the Quality Scale to a three-point one was taken. Accordingly students using the first edition of the Evaluation of Selected Hypotheses Task were required to evaluate hypotheses previously selected, in terms of 'poor,' 'fair' and 'good' qualities. A poor hypothesis was considered one equivalent to Quinn's 0 and 1 hypothesis qualities (see p. 48). A fair hypothesis was considered to be one equivalent to a 2 quality, and finally a good hypothesis was seen as one equivalent to 3 and 4 qualities. [Quinn's 5 quality (see p. 48) was not included in the response key since this particular quality indicates evidence of ability to test a hypothesis, a requirement not demanded in the present investigation.] Definitions of poor, fair and good qualities, in terms of the specific problem situations depicted, were provided students. Results of the administration of this trichotomous system to a complete grade six class (N = 26) indicated that the system was too difficult for general usage at this level. A very low investigator-student agreement was obtained, indicating difficulty in discriminating between the three categories. In view of these findings it appeared likely that a dichotomous response system would be preferable to a trichotomous one. In this regard selected hypotheses were now evaluated in terms of A-type and B-type qualities. An A-type hypothesis was defined as one which indicated one possible cause of the depicted problem. A B-type hypothesis was defined as one indicating at least two possible causes of the depicted problem. Use of this dichotomous evaluation system manifested a considerably higher investigator-student agreement than that manifested by the trichotomous system. Thus the superiority of the dichotomous evaluation system over the trichotomous one was established. Use of the dichotomous key with the second edition of the Hypothesis Selection Task yielded a KR_{20} reliability of .47, with one item failing to manifest acceptable p and D values. All items were scrutinized by judges (see p. 52) to assess face and content validity. Employing the final version of the Evaluation of Selected Hypotheses Task with the major sample (N = 138) all items manifested acceptable p and D values, and a KR_{20} reliability of .53 was obtained. In view of the lack of similar tasks with which to compare reliabilities, this value was judged adequate for purposes of this investigation. (See Appendix C for a copy of the instrument. Item p and D values are given in Appendix G.) Based on the total of the three scores associated with the subtasks, a test-retest reliability of .87 (N = 26) was obtained over a two week interval of time. The stability of the task was thus considered established.

Scoring procedures: Consistent with the noted need for consideration being given to both quantity and quality in any hypothesis evaluation program (see p. 48), an Evaluation of Selected Hypotheses score was derived from the total number of hypotheses previously correctly selected as offering explanation of the depicted problem, and subsequently correctly evaluated as being of A-type or B-type quality. One point was awarded for each correct evaluation. An investigator-doctoral student scoring agreement of 100% was obtained.

Open-ended Tasks

Object Generation Task. Refining procedures: Students were required to generate a list of all of the objects observed on the film-loop. In an initial administration of this task (N = 26), a test-retest reliability of .83 was achieved over a one week lapse of time. The task was then submitted to scrutinization on the part of the judges (see p. 52) in order to assess face and content validity. Following minor revisions, a further administration (N = 28) yielded a test-retest reliability of .82 based over a two week period of time. Since there were no similar tasks reported in Buros (1972), this reliability was judged acceptable for the purposes of this study. (See Appendix D for a copy of the task.)

Scoring procedures: All responses that were not objects (i.e. events, inferences, questions, etc.) were eliminated. An Object Generation score was derived from the total number of objects correctly listed as present on the loop. One point was given each correct generation. An investigator-doctoral student scoring agreement of 96% was achieved.

Event Generation Task. Refining procedures: Students were required to generate a list of all the events observed on the loop. In an initial administration of the task (N = 26), test-retest reliability of .70 was achieved based on a one week interval of time. The task was then examined by the panel of judges (see p. 52) to assess face and content validity. Following minor revision, the task was administered to another group of students (N = 28). A test-retest reliability of .77 was obtained over a two week interval. With the absence of any similar published task reported in Buros (1972), this reliability was deemed acceptable for purposes of this investigation. (See Appendix E for a copy of the task.)

Scoring procedures: All responses that were not events (i.e. objects, inferences, questions, e were eliminated. An Event Generation score was derived from the total number of events correctly listed as present on the loop. One point was given each correct generation. An investigator-doctoral student scoring agreement of 100% was obtained.

Hypothesis Generation Task. Refining procedures: To be consistent with the noted need for quality and quantity hypothesis production (see p. 48), this task comprised of three hypothesis generation subtasks, requiring students to generate quantities of hypotheses offering possible explanation of each of the three identified problem areas depicted on the loop. In an initial administration of the task (N = 26), reliabilities of .80, .80 and .70 were obtained over a one week interval on the three subtasks respectively. Subtasks were scrutinized by judges (see p. 52) to assess face and content validity. Following minor revision, a further administration (N = 28) yielded test-retest reliabilities of .82, .75 and .80 on each of the subtasks respectively. No similar tasks were reported in Buros (1972) with which comparisons of reliabilities could be made. Accordingly these reliabilities were deemed acceptable. (See Appendix F for a copy of the task.)

Scoring procedures: All responses that were not hypotheses [i.e. statements that did not meet criteria associated with A-type and B-type hypotheses (see p. 60)] were rejected. In addition all responses that were hypotheses but offered explanation of a problem other than the one depicted were also rejected. Hypothesis Generation subscores pertaining to each of the three subtasks were derived from the number of A-type and B-type hypotheses generated, one point being awarded each A-type hypothesis and two points being awarded each B-type hypothesis. By summing the three separate subscores a total Hypothesis Generation score was obtained. Investigator-doctoral student scoring agreements of 100%, 100% and 92% were obtained on the three subtasks respectively.

Open-ended Evaluation Tasks

Evaluation of Generated Objects Task. Refining procedures: This task is closely associated with the Object Generation Task (see

p. 61). Students were required to evaluate the extent to which objects previously generated had an effect upon occurrence of the depicted problem. Consistent with the format of the Evaluation of Selected Objects Task (see p. 55) a dichotomous response key rather than a trichotomous one was employed. As in the case of the Evaluation of Selected Objects Task, students were required to evaluate the role played by each object in terms of no effect or some effect. Following its first administration (N = 26), a test-retest reliability of .80 was obtained based over a one week time span. Following minor revisions advised by judges (see p. 52), a second administration (N = 28) resulted in a test-retest reliability of .78 over a two week time period. Since no similar tasks were cited in Buros (1972) this reliability was judged acceptable. (See Appendix D for a copy of the task.)

Scoring procedure: An Evaluation of Generated Objects score was derived from the total number of objects previously correctly generated and subsequently correctly classified as having no effect or some effect upon the depicted problem. One point was awarded each correct evaluation. An investigator-doctoral student scoring agreement of 100% was obtained.

Evaluation of Generated Events Task. Refining procedures: This task is closely associated with the Event Generation Task (see 62). Students were required to evaluate the extent to which events previously generated had an effect upon occurrence of the problem depicted. First administration of this task yielded a test-retest reliability of .70 based on a one week period of time. Following minor revision, a second administration (N = 28) yielded a test-retest reliability of .72 based on a two week time span. The reliability was judged acceptable for the purposes of the study in view of the absence of any similar task published in Buros (1972). (See Appendix E for a copy of the instrument.)

Scoring procedures: An Evaluation of Generated Events score was derived from the total number of events previously correctly generated and subsequently correctly classified as having no effect or some effect upon the depicted problem. One point was given each correct evaluation. An investigator-doctoral student scoring agreement of 100% was realized.

Evaluation of Generated Hypotheses Task. Refining procedures: Students were required to evaluate hypotheses previously generated in terms of A-type and B-type qualities. In the first administration of the task (N = 26), a test-retest reliability of .79 based on a total of the three scores associated with the subtasks was achieved over a one week period of time. Second administration of the task (N = 28) after revision yielded a test-retest reliability of .80 based over a two week time span. With the absence of similar tasks reported in Buros (1972) this reliability was judged acceptable for the purposes of the investigation. (See Appendix F for a copy of the instrument.)

Scoring procedures: An Evaluation of Generated Hypotheses score was derived from the total number of hypotheses generated on the three subtasks offering explanation of the depicted problem and subsequently correctly evaluated as being of A-type or B-type quality. One point was awarded each correct evaluation. An investigator-doctoral student scoring agreement of 92% was obtained.

Other Investigator-designed Instruments

The two hypothesizing tasks designed by the investigator [i.e. the Hypothesis Selection Task (see p. 54) and the Hypothesis Generation Task (see p. 62)] depend for their effective operation upon presentation of the film-loop 'The Knife.' Because of the nature of the loop and the particular episodes depicted, the content area associated is highly specific and restricted in scope. The assessment of ability in hypothesizing in other content areas and the employment of other modes of problem depiction would help to determine the extent to which hypothesizing abilities manifest inter-task generality. Assessment of hypothesizing abilities in other domains could also aid in the validation of the two above-mentioned hypothesizing tasks. Since a survey of Buros (1965, 1972) failed to indica[illegible]xistence of any reliable and valid instrument assessing hypothesizing ability at appropriate grade levels, a need for construction of suitable instrumentation was apparent. Accordingly the following two Tests were constructed by the investigator.

Pearson Hypothesis Formulation Test. This test is a modification of the Hypothesis Machine Test (Pearson, 1975) claimed to assess ability in generating hypotheses. In this modified version students were shown a diagram of a series of channels on the classroom chalkboard. Each channel permitted a 'ball' to pass freely from top to bottom. At various 'gaps' located in the channels, certain triangles having specific characteristics could be inserted to deflect the ball either one or three channels to the right, or one or three channels to the left. Once the capabilities of the triangles had been demonstrated, students were provided the test in which a series of eighteen identical representations of the diagram shown on the board had been drawn. On each representation a ball had been positioned at the top of one particular channel and another ball had been placed at the bottom of a different channel. Students were required to draw in all the possible alternative rout[illegible]hich the ball could successfully pass from the identified top [illegible] to the identified bottom position. Any combination of triangles was permitted in order to facilitate this objective. (See Appendix H [illegible] a copy of the instrument.) A Hypothesis Formulation score was derived from the number of routes drawn in that were functionally possible within the defined limits of the experimental situation. One point was awarded each functional route. Results obtained from the administration of this test to a complete grade six class (N = 26) not subsequently utilized in the main study, yielded a test-retest reliability of .77 based over a one week interval of time. With the noted absence of tests of a similar nature with which comparisons could be made, this reliability was deemed acceptable for the purposes of this investigation.

<u>Pearson Hypothesis Selection Test</u>. In an effort to enable measurement of students' ability in selecting hypotheses in a content area similar to that of the above Pearson Hypothesis Formulation Test, students were presented eighteen identical representations of the above-mentioned board drawing. This time however, routes have been

drawn in for the students prior to test administration. Some routes were functionally possible while other routes were not. Students were required to distinguish between those that were functional and those that were not. (See Appendix J for a copy of the instrument.) A Hypothesis Selection score was derived from the number of correct identifications made, one point being awarded each correct identification. Using a test-retest procedure a reliability of .85 was obtained over a one week period (N = 28), indicating the stability of the test over time. Results of an item-analysis conducted on the main sample employed in this investigation (N = 138), indicated a KR_{20} reliability = .75. With the lack of any similar test with which to compare this reliability, the value obtained was judged adequate for purposes of the investigation.

Published Instruments

Test of Science Processes (Observation). The four observation tasks designed by the investigator [i.e. Object Selection Task (see p. 51), Event Selection Task (see p. 52), Object Generation Task (see p. 61), and the Event Generation Task (see p. 62)] depend for their effective operation upon the content of the film-loop 'The Knife.' The assessment of ability in visual observation in other content areas and the employment of other modes of problem presentation would help determine the extent to which visual observation abilities manifest inter-task generality. Asses nt of these abilities in other areas could also aid in the validation of the four observation tasks noted above. Of the few available tests that claim to measure visual observation ability (see p. 47) Tannenbaum's (1971) Observation Test appears most suitable for use, since this particular test involves several different content areas and has established reliability and validity. However since this test was recommended for grade seven, eight and nine usage only, the test was first administered to a group of grade six students (N = 26) to determine whether effective employment was possible at this grade level. Data obtained from this administration indicated a 48% correct student response, with scores ranging from 1 to 8 (out of a possible 9) (St. Devn. = 1.8). These results indicated that the test was appropriate for general use at the grade six level. Administration of the test to the main sample employed in the investigation (N = 138) yielded a KR_{20} reliability of .39 which compares favourably with the .41 and .42 reliabilities obtained by Tannenbaum (1971) with grade seven, eight and nine students.

Suchman Question-Asking Test. This test was used to classify the functional types of question asked by Reflectives and Impulsives in efforts to gather information pertaining to the problem depicted on the film-loop. Following the format prescribed by Suchman (1962), questions were deliberately written in a manner promoting Yes or No responses. Practice sessions had been conducted prior to administration of the main investigation in order to train students in constructing questions of this type. For purposes of the practice session, the Suchman film-loop 'Boiling Coffee' was utilized. (See Appendix K for a copy of the practice instrument.) The major Question-Asking Test utilized the film-loop 'The Knife.' (See

Appendix L f copy of this instrument. Questions asked were subsequently cl ified according to the Suchman question types identified on page 9. (See Appendix M for specific examples of question types as they pertain to the film-loop.) To establish the reliability of the Question-Asking Test, one complete class (N = 26) was employed not subsequently involved in the main study. This class was shown both the 'Boiling Coffee' loop and 'The Knife' loop and invited to ask questions related to each film. Using a parallel-form procedure, comparisons of the total number of questions asked with respect to both film-loops yielded a reliability of .71. With no published data available with which comparisons could be made, this reliability was deemed acceptable for purposes of the present investigation.

Scoring procedure: A Question-asking score was derived from the total number of questions asked, one point being allotted each question. Specific question-type scores (e.g. Concrete-conceptual Question scores, etc.) were derived from the total number of such questions asked, one point being awarded per question. As a check on the investigator's classification of question types, an independent categorization was made by a trained science education doctoral student. An investigator-doctoral student scoring agreement of 89.6% was achieved.

<u>Matching Familiar Figures Test</u>. This test was used to classify students as reflective or impulsive in disposition, as well as classifying those of neither marked disposition. A general description of the test and the rationale for its inclusion in this stud has been given on page 29. Two scores derived from the test provided

measures of R/I behaviour. A score of 'response latency' indicated the time elapsed before any response is made. The other score indicated the number of errors made prior to correct response. Students scoring above the median on response latency and below the median on number of errors were classified as reflectives. Students scoring below the median on response latency and above on number of errors were classified as impulsives. Of the 184 students utilized in the study, 68 were subsequently classified as being reflective in disposition, 70 as being of an impulsive disposition and the rest as being of neither a pronounced reflective or impulsive disposition. Table II provides a distribution of students according to disposition, sex and school affiliation.

<u>Science Achievement Test</u>. Findings from Gupta's (1970) study indicated that grade nine Reflectives manifest significantly superior performances over Impulsive peers on science achievement tests (see p. 34). Since tasks in this investigation assessed performance in problem-solving skills associated specifically with science-oriented situations, performance on these tasks may be affected by differences in knowledge of science content. Hence in subsequent data analysis, science knowledge may be statistically controlled. The Stanford Achievement Test: Science, intermediate II grades 5.5-6.9 (Form W, 1964 edition) was used to assess general science knowledge. This test places heavy emphasis upon knowledge of scientific facts. The areas of physical and earth sciences, and the biological sciences are given equal emphasis. Bryan (in Buros, 1965) and Mallinson (in Buros, 1972) maintain that this particular standardized test is the

TABLE II

MFF CLASSIFICATION OF STUDENTS

School Identification	Reflective		Impulsive		Not Reflective + Not Impulsive		School Total
	M*	F*	M	F	M	F	
1	3	[illegible]	4	4	5	6	25
2	2	9	3	1	5	5	25
3	3	4	6	6	1	3	23
4	5	3	5	7	1	5	26
5	2	5	4	3	2	2	18
6	6	4	4	2	5	1	22
7	1	6	9	11	3	0	30
8	4	8	1	0	1	1	15
Subtotals	26	42	36	34	23	23	
Totals	68		70		46		184

*M = Males
*F = Females

best available for assessment of science content at the grade five and six levels. Authors claim content validity was insured by examination of appropriate courses of study and textbooks as a basis for determining the types of knowledges to be measured. The authors also report split-half (odd-even) reliabilities of .90 (N = approximately 1,000) and KR_{20} reliabilities of .89 (N = approximately 1,000) for the test at the grade six level.

Intelligence test. I.Q. may be statistically controlled in the analysis of data since findings by Michenbaum and Goodman (1969), (see p. 33) indicate significant correlations between this variable and the R/I dimension. School Cumulative Records supplied I.Q. scores obtained from the Canadian Lorge Thorndike I.Q. Test, recently administered (February 1976). This test yields a verbal and non-verbal I.Q. score. Test authors report split-half (odd-even) reliabilities of .92 on both verbal and non-verbal batteries (N = approximately 3,000) and KR_{20} reliabilities of .93 and .92 on the verbal and non-verbal batteries respectively (N = 259) at the grade five level.

Reading ability test. Reading ability may be statistically controlled in the analysis of data since findings by Gupta (1970) (see p. 34) indicate significantly superior performance of Reflectives over Impulsives on Reading ability tests. School records supplied Reading ability scores obtained from the Stanford Achievement Test: Reading Tests, intermediate II, grades 5.5-6.9 (Form W, 1964 edition) administered June, 1975. These Reading tests provide a Word Meaning score and a Paragraph Meaning score. Test authors claim that validity was ensured by examination of courses of study and textbooks as a basis for determining the types of skills, knowledges and understandings to be assessed. Traxler (in Buros, 1972) however considers that little solid evidence exists concerning the validities of these tests since the exact sources of test words employed in the Word Meaning Test and sources giving rise to paragraph classification used in the Paragraph Meaning Test were not specified. In spite of these shortcomings, Traxler (in Buros, 1972) claims these tests are among the best available for annual assessment of reading achievement at the elementary school level. Authors report split-half (odd-even) reliabilities of .89 and .92 for the Word Meaning and Paragraph Meaning Tests respectively (N = approximately 1,000), and KR_{20} reliabilities of .89 and .92 for the two tests (N = approximately 1,000).

TESTING PROCEDURES

All subjects involved in the investigation (N = 184) participated in all tasks. To aid efficiency in task administration and to minimize fatigue, students participated in the following three testing sessions:

Session 1: During this session the MFF test was administered on an individual basis. Time required for completion was approximately ten minutes per pupil.

Session 2: During this session all guided-discovery and open-ended tasks were administered on a group basis. In addition practice and main Question-Asking Tests were given on a group basis. Time required for completion of this session was approximately sixty minutes. The following list indicates the sequence of administration of tasks. This sequence was rigidly adhered to in the eight classes investigated in the study.

1. Group direction given in constructing questions of a form promoting Yes and No responses.
2. Film-loop 'Boiling Coffee' shown. Students given opportunity to write questions on the form described, concerning any data presented on the loop that would help explain the depicted problem.
3. Students informed that a different loop ('The Knife') would be shown, and asked to watch film for objects and events, and to expect questions on what has been seen immediately following viewing.
4. Film-loop 'The Knife' shown.
5. Object Generation Task administration.
6. Event Generation Task administration.
7. Students informed will watch film again, after which they will be given opportunity to write questions about data presented on the loop that would help explain the problem depicted.
8. Film-loop reshown.
9. Question-Asking Test administration.
10. Film reshown.
11. Hypotheses Generation Subtasks administration.
12. Film-loop reshown.
13. Evaluation of Generated Objects Task administration.

14. Evaluation of Generated Events Task administration.

15. Evaluation of Generated Hypotheses Task administration.

16. Film-loop reshown.

17. Object Selection Task administration.

18. Event Selection Task administration.

19. Hypothesis Selection Task administration.

20. Evaluation of Selected Objects Task administration.

21. Evaluation of Selected Events Task administration.

22. Evaluation of Selected Hypotheses Task administration.

As indicated above, the six tasks requiring generation were administered before the six tasks requiring selection. Decision to utilize this particular order was based upon the following rationale. Since selection tasks provide lists of data (i.e. of objects, events, hypotheses), memory of these data from participation in such tasks would likely influence performance on subsequent generation tasks requiring application of similar data. In contrast to the selection tasks, generation tasks provide no such data; thus performance on these tasks would likely less influence subsequent performance on selection tasks.

Since all Selection and Generation tasks have certain common elements (e.g. involve similar objects, etc.) consideration of the content of one task may benefit performance on a subsequent task. To minimize this possible inter-task influence, each task when completed was placed out of reach of student access before administration of subsequent tasks took place.

Session 3: During this session the Pearson Hypothesis

Formulation Test, Pearson Hypothesis Selection Test, Test of Science Processes (Observation), and the Stanford Achievement Test (Science) were group administered in the indicated order. Time required for completion of this session was approximately sixty minutes.

To prevent problems of readability of tasks employed, efforts were made to keep written instructions simple and concise. In addition all task instructions were read aloud by the investigator.

THE MAIN TESTING PROGRAM

The main testing program was conducted in January and February, 1976. Administration of the tasks to students involved in each of the eight schools used was conducted by the investigator.

TYPES OF DATA ANALYSIS USED

To test Hypotheses One to Thirteen, task scores were subjected in turn to Stepwise Regression Analysis to determine which, if any, of the variables I.Q., Reading Ability, Science Knowledge and Age, contributed significantly to the variance of the measure under investigation. In instances where one or more of the variables made a significant contribution, subsequent covariate analysis was performed on reflective and impulsive group means using the variable(s) that contributed significantly as covariate(s). In instances where no variable made significant contribution to the variance, t-Tests were subsequently performed between reflective and impulsive group means.

To test Hypothesis Fourteen, task scores were subjected to the Kolmogorov-Smirnov Two Sample Test (for details of the Test see

p. 128.

To test Hypotheses Fifteen to Eighteen, task scores were subjected to a χ^2-Test of Independence. In instances where cell frequencies were very small (i.e. <5) values for χ^2 were calculated using Yate's Correction for Continuity (Ferguson, 1971, p. 188).

To answer the first Associated Question, inter-task correlations between reflective and impulsive groups were compared, and Fisher's Z transformations executed to determine whether differences between correlations were significant.

To answer the second Associated Question, task scores were subjected to Factor Analysis and the resultant correlation matrix rotated orthogonally according to Varimax criteria.

To answer the third Associated Question, Factor Scores were subjected to statistical procedures similar to those adopted in testing Hypotheses One to Thirteen.

CHAPTER IV

RESULTS AND DISCUSSION

INTRODUCTION

In this chapter the results obtained from the statistical analysis of data are presented and discussed. The chapter is organized into three sections. In Section A attention focuses upon those hypotheses that are related to performances of Reflectives and Impulsives on tasks demanding ability in observation, hypothesizing and evaluation. Section B deals with those hypotheses relating performance of Reflectives and Impulsives on the Suchman Question-asking Test. Section C deals with associated questions relating to several of the research hypotheses investigated.

Statistical calculations were executed on the University of Alberta IBM 360/67 computer employing documented programs of the Division of Educational Research. Throughout the investigation differences between reflective and impulsive group means are considered statistically significant at the 0.05 level or less.

SECTION A

1. Intercorrelations

Based on the total sample employed in this study (N = 184) a substantial correlation of -.54 was obtained between the two R/I variables, Number of Errors and Response Latency. This negative correlation accords with values commonly reported by other researchers, ranging from -.4 to -.65 (see Kagan, 1965a, 1965b and 1965c).

As noted in Chapters II and III, Age, Intellectual ability, Reading ability and Science Knowledge have been shown by some researchers to correlate significantly with the R/I variables. Correlations obtained in this investigation between Age, Non-Verbal I.Q., Verbal I.Q., Word Meaning ability, Paragraph Meaning ability, Science Knowledge, and Number of Errors and Response Latency are shown in Table XII. It can be seen that one or both of the R/I variables correlated significantly at the .05 level with all variables 3-8 except Age. The lack of a significant correlation between the R/I variables and Age is not surprising in this instance, since grade six students alone were employed in the study and most of the students were of similar age. High positive correlations in the range .42 to .80 were manifested between Non-Verbal I.Q., Verbal I.Q., Word Meaning, Paragraph Meaning, and Science Knowledge. All values were significant at the 0.01 level. The existence of substantial negative correlations obtained between Age and these variables at first appears surprising. However although most students were of similar age, some students were twelve months or more older than group peers, having been held back from grade advancement at one or more junctures during school careers. Low scores manifested by these students on tests measuring above-mentioned abilities help to explain the negative correlations obtained.

In view of the noted significant correlations between the two R/I variables and Age, Non-Verbal and Verbal Intellectual abilities, Word and Paragraph Meaning abilities and Science Knowledge, it would

TABLE III

CORRELATIONS BETWEEN RESPONSE LATENCY, NUMBER OF ERRORS AND SPECIFIED VARIABLES (REFLECTIVE AND IMPULSIVE GROUPS COMBINED)

Variable	1	2	3	4	5	6	7	8
1. Response Latency	1.00							
2. Number of Errors	-.70**	1.00						
3. Age	.05	.10	1.00					
4. Non-Verbal I.Q.	.21	-.27**	-.45**	1.00				
5. Verbal I.Q.	.14	-.20*	-.60**	.68**	1.00			
6. Word Meaning	.25**	-.21*	-.19*	.42**	.70**	1.00		
7. Paragraph Meaning	.16	-.20*	-.27**	.46**	.67**	.80**	1.00	
8. Science Knowledge	.17*	-.14	-.32**	.55**	.67**	.69**	.67**	1.00

* Denotes significance at 0.05 level.
** Denotes significance at 0.01 level.

appear expedient at this juncture to control for sampling bias due to these variables in subsequent data analysis pertaining to performance of reflective and impulsive groups on tasks involved in this investigation. Employing the variables as covariates in subsequent data analysis would serve to equate reflective and impulsive groups by removing bias due to these factors. The appropriateness of employing such covariates in subsequent analysis was further examined following the calculation of correlations existing between these variables and all other variables investigated in this study. Correlation matrices are presented in Table IV and Table V. The tables indicate that each of the control variables 3-8 manifested significant, though generally low, correlations with several of the criterion variables 9-24. In some instances all of the variables 3-8 correlated significantly with criterion variables (as for example in the case with criterion 23—Evaluation of Generated Events). In other instances some of the variables correlated significantly with criterion variables (as in the case with criterion 19—Object Generation). In still other instances none of the variables correlated significantly with criterion variables (as in the case with criterion 18). Further examination of the correlation matrices indicates that no one variable from among variables 3-8 manifested a greater number of significant correlations or higher correlation magnitudes with criterion variables, than did any other. These findings indicate that no one variable or combination of variables from among control variables 3-8 could serve effectively as a <u>general</u> covariate, or as <u>general</u> covariates during subsequent criterion analysis. Expressed in different terms, no

TABLE IV

CORRELATIONS BETWEEN VARIABLES 1-15
(REFLECTIVE AND IMPULSIVE GROUPS COMBINED)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Response Latency	1.00														
2. Number of Errors	-.70**	1.00													
3. Age	.05	.10	1.00												
4. Non-Verbal I.Q.	.21*	-.27**	-.45**	1.00											
5. Verbal I.Q.	.14	-.20*	-.60**	.68**	1.00										
6. Word Meaning	.25**	.21*	-.19*	.42**	.70**	1.00									
7. Paragraph Meaning	.16	-.20*	-.27**	.46**	.67**	.80**	1.00								
8. Science Knowledge	.17*	-.14	-.32**	.55**	.67**	.69**	.67**	1.00							
9. Question-Asking	.14	-.12	-.02	.15	.13	.14	.22*	.20*	1.00						
10. Tannenbaum Observation	.00	-.03	-.11	.43**	.40**	.41**	.35**	.41**	.09	1.00					
11. Pearson Hyp. Sel.	.17*	-.12	-.07	.38**	.17	.17	.26**	.25**	.19*	.12	1.00				
12. Pearson Hyp. Form.	.22*	-.21*	-.12	.46**	.34**	.33**	.44**	.38**	.04	.14	.55**	1.00			
13. Object Selection	.06	-.03	-.03	.08	.01	-.01	.02	-.04	.16	-.12	.12	.02	1.00		
14. Event Selection	-.04	-.01	-.02	.13	.14	.14	.13	.12	.02	.08	.21*	.20*	-.03	1.00	
15. Hyp. Selection	.10	-.11	-.06	.05	.08	[illegible]	.00	.08	.06	-.04	.06	.17	-.07	.24**	1.00

* Denotes significance at 0.05 level.
** Denotes significance at 0.01 level.

TABLE V

CORRELATIONS BETWEEN VARIABLES 16-24
(REFLECTIVE AND IMPULSIVE GROUPS COMBINED)

Variable	16	17	18	19	20	21	22	23	24
1. Response Latency	-.04	-.08	.05	-.05	.10	.16	.05	.14	.14
2. Number of Errors	-.06	.00	-.03	.02	-.15	-.12	-.07	-.17*	-.10
3. Age	-.23**	-.22**	.02	-.12	-.15	.05	-.18*	-.24**	-.05
4. Non-Verbal I.Q.	.06	.29**	-.01	.17	.33**	.08	.24**	.38**	.16
5. Verbal I.Q.	.26**	.28**	.06	.22*	.34**	.08	.26**	.36**	.13
6. Word Meaning	.23**	.18*	.08	.19*	.37**	.12	.20*	.37**	.13
7. Paragraph Meaning	.19*	.22*	.09	.17	.33**	.18*	.17	.26**	.20*
8. Science Knowledge	.14	.25**	.07	.23*	.43**	.11	.25**	.36**	.14
9. Question-Asking	.07	.07	.07	.25**	.35**	.34**	.27**	.25**	.32**
10. Tannenbaum Observation	.00	.11	-.04	.11	.34**	.17	.15	.31**	.14
11. Pearson Hyp. Sel.	.03	.18*	-.11	.11	.24**	.14	.15	.30**	.16
12. Pearson Hyp. Form.	.06	.15	.07	.02	.14	.02	.07	.17	.07
13. Object Selection	.46**	.01	.12	.08	.04	.15	.03	.00	.11
14. Event Selection	.15	.60**	.11	.13	.09	.21*	.12	.07	.18*
15. Hyp. Selection	.02	.17*	.48**	.05	.07	.11	.09	.10	.17*
16. Evaln. of Sel. Objects	1.00	.23**	.10	.17*	.12	.07	.18*	.12	.06
17. Evaln. of Sel. Events		1.00	.13	.12	.11	.09	.21*	.21*	.11
18. Evaln. of Sel. Hyps.			1.00	.08	.06	.16	-.02	.01	.11
19. Object Generation				1.00	.47**	.14	.81**	.42**	.14
20. Event Generation					1.00	.16	.48**	.32**	.18*
21. Hyp. Generation						1.00	.21*	.09	.86**
22. Evaln. of Gen. Objects							1.00	.59**	.23**
23. Evaln. of Gen. Events								1.00	.14
24. Evaln. of Gen. Hypos.									1.00

* Denotes significance at 0.05 level.
** Denotes significance at 0.01 level.

variable or group of variables would contribute significantly to the variance of all criterion variables under investigation. Instead it appears that different combinations of variables 3-8 would be needed to serve as covariates as different criterion variables are subjected in turn to analysis. In order to identify which variable, if any, or which group of variables made a significant contribution to variances of criterion measures, Stepwise Regression Analysis was performed, employing control variables 3-8, on each of the criterion measures in turn. In cases where Stepwise Regression Analysis identified one or more variables as contributing significantly to the variance of a particular measure, analysis of covariance using the identified variable(s) as covariate(s) was used whenever reflective and impulsive group means pertaining to that measure were calculated.

As noted earlier, Tables IV and V (p. 83 and 84) show the correlations obtained between criterion measures 10-24 assessing abilities in selecting, generating and evaluating observations and hypotheses. Examination of these tables reveals that correlations obtained between Selection Tasks (Object Selection, Event Selection and Hypothesis Selection) and respective Generation Tasks were low and non-significant. Thus the abilities to select and to generate objects, events and hypotheses (as measured on tasks employed in this investigation) appear essentially unrelated. Knowledge of performance on Selection Tasks, for example, provides no information meaningful in predicting performance on respective Generation Tasks. In contrast to these findings a moderately high correlation ($r = .55$, significant at the 0.01 level) existed between the Pearson Hypothesis Selection

Test and the Pearson Hypothesis Formulation (Generation) Test. Since the content areas and modes of selection and generation used with the two Pearson measures, and with the six Selection and Generation Tasks differ greatly, these results would appear to indicate that the degree of relationship existing between ability to select, and ability to generate hypotheses is determined, at least in part, by the particular content area and mode of investigation adopted.

Tables IV and V reveal that the correlations between the two Selection of Hypotheses tasks, the Hypothesis Selection (variable 15) and Pearson Hypothesis Selection measures (variable 11), and between the two Generation of Hypotheses tasks, the Hypothesis Generation (variable 21) and Pearson Hypothesis Formulation measures (variable 12) were essentially zero. Furthermore the correlation obtained between the two Selection of Objects tasks, the Object Selection (variable 13) and Tannenbaum Observation (variable 10) measures were also essentially zero. These results indicate that abilities to select and to generate hypotheses, and abilities to select objects appear to be highly task-specific. Performance on one task, for example, appears to be of no use in predicting performance on another.

Correlations between the three Selection Tasks (variables 13, 14, 15) were of low order. It is interesting to note that a significant correlation did not exist between ability to Select Objects and ability to Select Events (as measured by tasks used in this investigation). Apparently abilities in these areas are essentially unrelated.

Correlations between the three Generation Tasks (variables

19, 20, 21) were of slightly higher order. A moderately high correlation of .47 was obtained between the Object Generation and Event Generation Tasks indicating that a fair degree of commonality apparently exists between abilities in these particular areas.

Correlations obtained between the three Evaluation of Selection Tasks (variables 16, 17, 18) and respective Selection Tasks (variables 13, 14, 15) were moderately high in magnitude (ranging from r = .46 to .60) while correlations between the three Evaluation of Generation Tasks (variables 22, 23, 24) and respective Generation Tasks (variables 19, 20, 21) were high in magnitude (ranging from r = .81 to .86). Since as earlier noted in Chapter III (see p. 57) scores obtained on Evaluation of Selection and Generation Tasks are in part determined by scores obtained on respective Selection and Generation Tasks, relatively high correlations were to be expected. Because these correlations may be spuriously high caution should be exercised if generalizations based on these intercorrelations are contemplated in other areas assessing similar abilities.

The investigation indicated that students were more effective evaluators of observations and hypotheses that had previously been generated, than of observations and hypotheses that had previously been selected. It is possible that more cognitive effort was required in generating observations and hypotheses than in selecting observations and hypotheses. Because of this, students were possibly more familiar with the components of generated products, than with those of selected products. This in turn may explain, at least in part, the subsequent more effective evaluation of generated products than of selected products.

2. Unadjusted Means, Standard Deviations and t-Tests

To determine if significant differences existed between reflective and impulsive groups on variables available as covariates (variables 3-8), t-Tests were performed. In addition t-Tests were performed on all criterion variables involved in the investigation. These t-Tests were executed on unadjusted data to provide bases for comparisons of data obtained from subsequent analysis of covariance. Table VI provides results of all t-Tests performed. The table reveals that reflective group means associated with Non-verbal and Verbal intellectual abilities, Word Meaning and Paragraph Meaning abilities and Science Knowledge were higher than those associated with the impulsive group. In no instance however were differences significant at the 0.05 level.

Reflective group means associated with criterion variables 10-24 were, with three exceptions (Tannenbaum Observation Test, Evaluation of Selected Objects Task, and Evaluation of Selected Events Task), higher than those associated with the impulsive group. In two instances (Pearson Hypothesis Selection and Pearson Hypothesis Formulation Tests) means were significantly higher. In the three noted cases where reflective group means were lower than impulsive group means, the differences were not significant.

TABLE VI

REFLECTIVE AND IMPULSIVE GROUP UNADJUSTED MEANS, STANDARD DEVIATIONS, AND t-TESTS FOR ALL VARIABLES

Variable	Reflective Group Mean	Impulsive Group Mean	Reflective Std. Devn.	Impulsive Std. Devn.	Df	t	Probability (Two Tailed Test)
1. Response Latency	268.60	120.10	82.88	32.15	136	13.80	0.00
2. Number of Errors	2.79	8.59	1.49	3.44	136	-12.92	0.00
3. Age (Months)	164.34	164.46	5.30	6.16	136	-0.12	0.86
4. Non-Verbal I.Q.	104.90	97.63	26.77	29.23	136	1.52	0.13
5. Verbal I.Q.	98.22	94.47	25.39	27.43	136	0.83	0.41
6. Word Meaning	21.57	18.90	9.27	7.73	136	1.84	0.07
7. Paragraph Meaning	31.25	29.47	12.48	10.82	136	0.89	0.37
8. Science Knowledge	32.38	31.01	7.88	8.92	136	0.96	0.34
9. Question Asking	2.41	2.19	1.57	1.47	136	0.85	0.39
10. Tannenbaum Observn.	4.10	4.24	1.78	1.54	136	-0.49	0.62
11. Pearson Hyp. Sel.	14.25	13.00	3.35	3.48	136	2.15	0.03
12. Pearson Hyp. Formn.	11.47	9.33	4.69	4.43	136	2.76	0.01
13. Object Selection	4.65	4.36	1.81	1.98	136	0.90	0.37
14. Event Selection	5.69	5.72	1.81	1.64	136	-0.13	0.90
15. Hyp. Selection	7.09	6.91	2.88	2.51	136	0.38	0.71
16. Eval. of Sel. Obj.	2.93	2.94	1.70	1.75	136	-0.06	0.96
17. Eval. of Sel. Events	2.78	2.90	1.45	1.47	136	-0.49	0.63
18. Eval. of Sel. Hyps.	2.65	2.59	1.84	1.86	136	0.19	0.85
19. Object Generation	6.82	6.69	2.09	1.97	136	0.40	0.69
20. Event Generation	5.71	5.44	1.82	1.76	136	0.86	0.39
21. Hyp. Generation	5.29	4.99	1.20	0.99	136	1.65	0.10
22. Eval. of Gen. Obj.	4.31	3.93	2.12	2.00	136	1.07	0.29
23. Eval. of Gen. Events	3.81	3.40	1.88	1.88	136	1.28	0.20
24. Eval. of Gen. Hyps.	4.91	4.62	1.25	1.09	136	1.41	0.16

3. Research Hypotheses

Hypothesis One (H_1) (As originally stated):

> There is no significant difference existing between Reflectives' and Impulsives' ability in selecting objects.

Since scores obtained from both the Object Selection Task and the Tannenbaum Observation Test were utilized in testing this hypothesis, Hypothesis One was divided into two parts. Hypothesis 1.1 deals with analysis pertaining to the Object Selection Task and Hypothesis 1.2 deals with analysis pertaining to the Tannenbaum Observation Test.

Hypothesis 1.1. To determine which, if any, of the control variables Age, Science Knowledge, Non-Verbal I.Q., Verbal I.Q., Word Meaning and Paragraph Meaning ability made significant contribution to the variance of the Object Selection measure, Stepwise Regression analysis was performed on these variables using the Object Selection measure as criterion. Results of the analysis indicated that no variable contributed significantly to the variance of the measure. These results are not surprising since correlations obtained between Object Selection scores and scores associated with each of the noted variables were essentially zero (see Tables IV, p. 83 and V, p. 84). Of the six variables analysed, Non-Verbal I.Q. was the major contributor (variance accounted for, 0.62%, $F = 0.85$, $p = 0.36$). On the basis of these results indicating the lack of significant contribution by any of the noted variables, subsequent covariate analysis of Object Selection scores appeared inappropriate.

Hypothesis 1.1 (restated):

There is no significant difference between Reflective and Impulsive group means on the Object Selection Task.

To test Hypothesis 1.1 a t-Test was performed on Object Selection scores. The results of the analysis are presented in Table VII. An examination of the table reveals that the t-Test was not significant (p = 0.37). Therefore the investigation failed to reject Hypothesis 1.1. It appears that there is no significant difference in ability to select objects as measured by the Object Selection Task between reflective and impulsive groups.

TABLE VII

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE OBJECT SELECTION TASK

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
4.65	4.36	1.81	1.98	136	0.90	0.37

Hypothesis 1.2. The classification of the Tannenbaum Observation Test as a measure assessing ability in selecting objects (rather than as one assessing ability in selecting events) appears justified, since according to Tannenbaum (1971) eight of the nine questions included on the test deal with the observation of object properties. Only one question deals specifically with the observation of events. Stepwise Regression analysis was performed to determine which of the six noted variables, if any, contributed significantly to the variance of the Tannenbaum Observation measure. Major results of this analysis are presented in Table VIII. An examination of the table reveals that

two variables, Science Knowledge and Non-Verbal I.Q., made significant contribution to the variance of the measure, the four remaining variables failing to act as significant contributors. On the basis of these results subsequent covariate analysis of Tannenbaum Observation scores employing Science Knowledge and Non-Verbal I.Q. scores as covariates appeared appropriate.

TABLE VIII

SIGNIFICANT CONTRIBUTIONS OF VARIABLES WITH THE TANNENBAUM OBSERVATION MEASURE AS CRITERION

Contributors	Multiple Regression R	F_R	% Variance Accounted for	F	p
X_1 (Science Knowledge)	0.41	27.46	16.80	27.46	0.00
X_1, X_2 (Non-Verbal I.Q.)	0.46	18.06	21.11	7.38	0.01

Regression Equation:

$$Y = 0.07X_1 + 0.01X_2 + 0.63$$

<u>Hypothesis 1.2</u> (restated):

> There is no significant difference between reflective and impulsive group means on the Tannenbaum Observation Test when means are adjusted for performances on the Stanford Achievement Test (Science) and on the Lorge-Thorndike Non-Verbal I.Q. Test.

In order to test Hypothesis 1.2 covariate analysis was performed on Tannenbaum Observation scores, using scores associated with the two noted Tests as covariates. Results of the analysis are presented in Table IX. Examination of this table reveals that the F-Tests for the covariates were highly significant indicating that regression coefficients of the covariates were not zero and that

TABLE IX

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE TANNENBAUM OBSERVATION MEASURE

Source	SS	Df	MS	F	p
Treatment	3.82	1	3.82	1.74	0.19
Covariate 1 Non-Verbal I.Q.	18.06	1	18.06	8.22	0.00
Covariate 2 Science Know.	47.94	1	47.94	21.81	0.00
Error	294.54	134	2.20		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1,\ \chi^2 = 0.02,\ p = 0.89$$

Test for Homogeneity of Regression Coefficients:

Covariate 1: $F_{(1,132)} = 1.02,\ p = 0.32$

Covariate 2: $F_{(1,132)} = 2.30,\ p. = 0.13$

	Unadjusted Means	Adjusted Means
Reflective Group	4.10	4.00
Impulsive Group	4.24	4.34

adjustments were being made to group means. The table reveals that the reflective group mean was adjusted down 0.[illegible] points and the impulsive group mean up 0.10 points. The difference between adjusted means w found to be 0.34, which was significant at the 0.19 level. Therefore the investigation failed to reject Hypothesis 1.2. It appears that there is no significant difference in ability to select objects, as measured by the Tannenbaum Observation Test, between reflective and impulsive groups.

Hypothesis Two (H_2) (As originally stated):

> There is no significant difference existing between Reflectives' and Impulsives' ability in selecting events.

Scores obtained from the Event Selection Task were utilized in testing this hypothesis. A Stepwise Regression analysis was performed to determine which of the variables if any contributed significantly to the variance of the Event Selection measure. Results of the analysis indicated that no vari contributed significantly to the variance of the measure. This absen of significant contribution is not surprising since, as in the the Object Selection analysis discussed above, correlatic ained between Event Selection scores and scores associated with the six variables 3-8 are non-significant (see Table IV, p. 83 and Table V, p. 84). Of the variables analysed, Word Meaning ability was the major contributor (variance accounted for, 2.53%, $F = 3.53$, $p = 0.07$). With the lack of significant contribution by any of the variables, subsequent covariate analysis of Event Selection scores appeared inappropriate.

Hypothesis H_2 (Restated):

> There is no significant difference between reflective and impulsive group means on the Event Selection Task.

To test H_2 a t-Test was performed on Event Selection scores. The results of the analysis are provided in Table X. Examination of this table indicates that the t-Test was not significant ($p = 0.90$). Therefore the investigation failed to reject H_2. It appears that there is no significant difference in ability to select events, as measured by the Event Selection Task, between reflective and impulsive groups.

TABLE X

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE EVENT SELECTION TASK

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
5.69	5.72	1.81	1.64	136	-0.13	0.90

Hypothesis Three (H_3) (As originally stated):

> There is no significant difference existing between Reflectives' and Impulsives' ability in selecting hypotheses.

Since scores obtained from two measures, the Hypothesis Selection Task and the Pearson Hypothesis Selection Test, were employed in testing H_3, the hypothesis was divided into two parts. Hypothesis 3.1 deals with analysis pertaining to the Hypothesis Selection Task and Hypothesis 3.2 deals with analysis pertaining to the Pearson Hypothesis Selection Test.

Hypothesis 3.1. Stepwise Regression analysis was performed to determine which of the six control variables, if any, contributed significantly to the variance of the Hypothesis Selection measure. Results of the analysis indicated that no variable contributed significantly to the variance. As in the cases of the Object Selection and Event Selection Tasks, absence of significant contribution was expected since correlations between Hypothesis Selection scores and those associated with all six variables were essentially zero (see Table IV, p. 83 and Table V, p. 84). Among the variables, Science Knowledge was the major contributor (variance accounted for, 0.58%, $F = 0.79$, $p = 0.38$). In view of these results, subsequent covariate analysis appeared inappropriate.

Hypothesis 3.1 (Restated):

> There is no significant difference between reflective and impulsive group means on the Hypothesis Selection Task.

To test $H_{3.1}$ a t-Test was executed on Hypothesis Selection scores. The results of the analysis are presented in Table XI.

TABLE XI

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE HYPOTHESIS SELECTION TASK

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
7.09	6.91	2.88	2.51	136	0.38	0.71

The table reveals that the t-Test was not significant ($p = 0.71$). Therefore the investigation failed to reject Hypothesis 3.1. It seems that there is no significant difference in ability to select hypotheses as measured by the Hypothesis Selection Task between reflective and impulsive groups.

Hypothesis 3.2. Stepwise Regression analysis was performed to determine which of the control variables, if any, contributed significantly to the variance of the Pearson Hypothesis Selection Test. Major results of this analysis are presented in Table XII. An examination of this Table shows that three variables, Science Knowledge, Non-Verbal I.Q., and Verbal I.Q., made significant contribution to the variance of the measure, the other three variables failing to contribute significantly. On the basis of these findings, subsequent covariate analysis appeared appropriate. It is interesting to note that Verbal I.Q. (variable X_3) in contrast to Non-Verbal I.Q. (variable X_2) acted as a suppressor variable (see the Regression Equation, Table XII).

TABLE XII

SIGNIFICANT CONTRIBUTIONS OF VARIABLES WITH THE PEARSON HYPOTHESIS SELECTION MEASURE AS CRITERION

Contributors	Multiple Regression R	F_R	% Variance Accounted for	F	p
X_1 (Science Knowledge)	0.25	9.21	6.34	9.21	0.00
X_1, X_2 (Non-Verbal I.Q.)	0.32	7.44	9.93	5.37	0.02
X_1, X_2, X_3 (Verbal I.Q.)	0.40	8.73	16.35	10.29	0.00

Regression Equation:

$$Y = 0.10X_1 + 0.09X_2 - 0.08X_3 + 8.83$$

Hypothesis 3.2 (Restated):

> There is no significant difference between reflective and impulsive group means on the Pearson Hypothesis Selection Test when means are adjusted for performances on the Stanford Achievement Test (Science) and on the Lorge Thorndike Non-Verbal and Verbal I.Q.

To test $H_{3.2}$ covariate analysis was performed on Pearson Hypothesis Selection scores using scores associated with the noted tests as covariates. Results of the analysis are presented in Table XIII. This table reveals that F-Tests for covariates were highly significant indicating that regression coefficients of the covariates were not zero and that adjustments were being made to group means. The table reveals that the reflective group mean was adjusted down 0.24 points and the impulsive group mean up 0.24 points. The difference between adjusted means was found to be 0.77, which was significant at the 0.16 level. Therefore the investigation failed to reject $H_{3.2}$. It seems that there is no significant difference in ability to select hypotheses, as measured by the Pearson Hypothesis Selection Test, between reflective and impulsive groups.

It is interesting to note that a t-Test performed on the unadjusted data was significant at the 0.03 level (see Table VI, p. 89). Once selection bias was removed by employment of covariates the significant difference between reflective and impulsive group means failed to persist.

Hypothesis Four (H_4) (As originally stated):

> There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating selected objects.

Scores obtained on the Evaluation of Selected Objects Task

TABLE XIII

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE PEARSON HYPOTHESIS SELECTION TEST

Source	SS	Df	MS	F	p
Treatment	19.94	1	19.94	1.96	0.16
Covariate 1 Non-Verbal I.Q.	138.32	1	138.32	13.59	0.00
Covariate 2 Verbal I.Q.	94.23	1	94.23	9.26	0.00
Covariate 3 Science Knowledge	93.63	1	93.63	9.20	0.00
Error	1353.34	133	10.18		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1,\ \chi^2 = 0.069,\ p = 0.79$$

Test for Homogeneity of Regression Coefficients:

Covariate 1: $F_{(1,130)} = 0.33,\ p = 0.57$

Covariate 2: $F_{(1,130)} = 0.87,\ p = 0.35$

Covariate 3: $F_{(1,130)} = 1.69,\ p = 0.20$

	Unadjusted Means	Adjusted Means
Reflective Group	14.25	14.01
Impulsive Group	13.00	13.24

were utilized in testing this hypothesis. Stepwise Regressio.. analysis was conducted using the Evaluation of Selected Objects measure as criterion. Major results of the analysis are presented in Table XIV. An examination of the table reveals that the variable Age made significant contribution to the variance of the measure. Subsequent covariate analysis of Evaluation of Selected Objects scores, using Age as covariate, appeared desirable.

TABLE XIV

SIGNIFICANT CONTRIBUTION OF AGE WITH THE EVALUATION OF SELECTED OBJECTS MEASURE AS CRITERION

Contributor	Multiple Regression R	% Variance Accounted for	F	p
X_1 (Age)	0.23	5.30	7.6	0.01

Regression Equation:

$$Y = 0.07X_1 + 15.31$$

Hypothesis H_4 (Restated);

> There is no significan[illegible] [illegible]rence between reflective and impulsive group means on the [illegible]ation of Selected Objects Task, when means were ad[illegible] [illegible]ge.

To test H_4 covariate an[illegible] [illegible] performed o[illegible] Evaluation of Selected Objects scores using A[illegible] [illegible]ari[illegible]e. Res[illegible]lts of the analysis are provided in Table X[illegible] [illegible] [illegible]veals that the F-Test for the covariate was highly si[illegible] [illegible]ant indicating that the regression coefficient of the covariate was no[illegible] zero, and that an adjustment was being made to group means. The table re[illegible]eals that the reflective group mea[illegible]

TABLE XV

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE EVALUATION OF SELECTED OBJECTS TASK

Source	SS	Df	MS	F	p
Treatment	0.02	1	0.02	0.01	0.93
Covariate: Age	21.56	1	21.56	7.58	0.01
Error	383.97	135	2.84		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1,\ \chi^2 = 0.58,\ p = 0.45$$

Test for Homogeneity of Regression Coefficient:

$$F_{(1,134)} = 0.20,\ p = 0.66$$

	Unadjusted Means	Adjusted Means
Reflective Group	2.93	2.92
Impulsive Group	2.94	2.95

impulsive group mean up 0.01 point. The difference between adjusted group means was found to be 0.03, significant at the 0.93 level. Therefore the investigation failed to reject H_4. It appears that there is no significant difference in ability to evaluate selected objects, as measured by the Evaluation of Selected Objects Task, between reflective and impulsive groups.

Hypothesis Five (H_5) (As originally stated):

> There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating selected events.

Scores obtained on the Evaluation of Selected Events Task were used in testing this hypothesis. Stepwise Regression analysis was conducted using Evaluation of Selected Events measure as criterion. Major results of the analysis are presented in Table XVI. The table indicates that Science Knowledge contributed significantly to the variance of the measure, the other five variables failing to contribute significantly. Subsequent covariate analysis of Evaluation of Selected Events scores, using Science Knowledge as covariate, appeared appropriate.

TABLE XVI

SIGNIFICANT CONTRIBUTION OF THE VARIABLE SCIENCE KNOWLEDGE WITH THE EVALUATION OF SELECTED EVENTS MEASURE AS CRITERION

Contributor	Multiple Regression R	% Variance Accounted for	F	p
X_1 (Science Know.)	0.24	5.98	8.66	0.00

Regression Equation:

$$Y = 0.04X_1 + 2.50$$

Hypothesis H_5 (Restated):

> There is no significant difference between reflective and impulsive group means on the Evaluation of Selected Events Task, when means are adjusted for performance on the Stanford Achievement Test (Science).

To test H_5 covariate analysis was performed on Evaluation of Selected Events scores, using scores associated with the Science Knowledge test as covariate. Results of the analysis are provided in Table XVII. The table reveals that the F-Test for this covariate was highly significant, indicating that the regression coefficient of the covariate was not zero and an adjustment was being made to group means. The table indicates that the reflective group mean was adjusted down 0.03 points and the impulsive group mean up 0.03 points. The difference between adjusted group means was found to be 0.18, significant at the 0.45 level. Therefore the investigation failed to reject H_5. There is no significant difference in ability to evaluate selected events, as measured by the Evaluation of Selected Events Task, between reflective and impulsive groups.

Hypothesis Six (H_6) (As originally stated):

> There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating selected hypotheses.

Scores obtained from the Evaluation of Selected Hypotheses Task were used in testing this hypothesis. Results of Stepwise Regression analysis indicated that no variable contributed significantly to the variance of the Evaluation of Selected Hypotheses measure. As in the case of the Hypothesis Selection Task, this finding was not surprising since correlations between the Evaluation

TABLE XVII

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE EVALUATION OF SELECTED EVENTS TASK

Source	SS	Df	MS	F	p
Treatment	1.14	1	1.14	0.57	0.45
Covariate: Science Knowledge	18.94	1	18.94	9.44	0.00
Error	271.05	135	2.01		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1,\ \chi^2 = 0.01,\ p = 0.97$$

Test for Homogeneity of Regression Coefficient:

$$F_{(1,134)} = 0.02,\ p = 0.91$$

	Unadjusted Means	Adjusted Means
Reflective Group	2.78	2.75
Impulsive Group	2.90	2.93

of Selected Hypotheses scores and those associated with all six variables were essentially zero (see Table IV, p. 83 and Table V, p. 84). Of the six variables, Science Knowledge was the major contributor (variance accounted for, 0.52%, $F = 0.71$, $p = 0.40$). On the basis of these findings subsequent covariate analysis did not seem appropriate.

Hypothesis H_6 (Restated):

> There is no significant difference between reflective and impulsive group means on the Evaluation of Selected Hypotheses Task.

To test this hypothesis a t-Test was conducted on Evaluation of Selected Hypotheses scores. Results of the analysis are given in Table XVIII. The table reveals that the t-Test was not significant ($p = 0.85$). Therefore the investigation failed to reject H_6. It appears that there is no significant difference in ability to evaluate selected hypotheses, as measured by the Evaluation of Selected Hypotheses Task, between reflective and impulsive groups.

TABLE XVIII

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE EVALUATION OF SELECTED HYPOTHESES TASK

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
2.65	2.59	1.84	1.86	136	0.19	0.85

Comments Pertaining to Hypotheses One to Six

Hypotheses One, Two and Three dealt with the ability of Reflectives and Impulsives to select objects, events and hypotheses.

Although Reflectives performed superiorly to Impulsives on most of the tasks involved, in no instance was performance superiority statistically significant. It appears the r ective and impulsive students performed about as well on tasks measuring abilities in selecting objects, events and hypotheses used in this investigation. In cases where covariate analysis was performed (with the Tannenbaum Observation and Pearson Hypothesis Selection measures), both Science Knowledge and Non-Verbal intellectual ability made major significant contributions to the variance of respective measures. These findings indicate that ability to select objects and hypotheses (as measured by these tests) is in part a function of Science Knowledge and Non-Verbal intellectual ability. The absence of significant contribution to the variance of the Object Selection, Event Selection and Hypothesis Selection measures indicates that ability to select objects, events and hypotheses (as measured by these Tasks) is essentially independent of Science Knowledge and Non-Verbal Intellectual ability. It appears from these findings that the two tests measuring ability in selecting objects, and the two tests measuring ability in selecting hypotheses, demanded markedly different types of cognitive activity on the part of students involved. This suggestion is further supported by the fact that essentially zero correlations were found to exist between the Tannenbaum Observation Test scores and the Object Selection scores, as well as betweeen the Pearson Hypothesis Selection scores and Hypothesis Selection scores (see Table IV, p. 83 and Table V, p. 84). These differences in cognitive requirement may arise, at least in part, as a result of evident differences in content areas, and in modes of content presentation associated with the different tasks.

Hypotheses Four, Five and Six dealt with the ability of Reflectives and Impulsives to evaluate selected objects, events and hypotheses. In all instances reflective and impulsive students performed about as well on tasks measuring these abilities used in this investigation. Only in the case of the Evaluation of Selected Events Task was there evidence that Science Knowledge played a significant role in ability to evaluate selections made.

Hypothesis Seven (H_7) (As originally stated):

> There is no significant difference existing between Reflectives' and Impulsives' ability in generating objects.

Scores from the Object Generation Task were utilized to test this hypothesis. Stepwise Regression analysis was performed using the Object Generation measure as criterion. Major results of the analysis are presented in Table XIX. The table reveals that Science Knowledge was the only variable contributing significantly to the

TABLE XIX

SIGNIFICANT CONTRIBUTION OF THE VARIABLE SCIENCE KNOWLEDGE WITH THE OBJECT GENERATION MEASURE AS CRITERION

Contributor	Multiple Regression R	% Variance Accounted for	F	p
X_1 (Science Know.)	0.23	5.43	7.80	0.01

Regression Equation:

$$Y = 0.06X_1 + 4.98$$

variance of the measure. Hence subsequent covariate analysis, using Science Knowledge as covariate, appeared appropriate.

Hypothesis H_7 (Restated):

> There is no significant difference between reflective and impulsive group means on the Object Generation Task when means are adjusted for performance on the Stanford Achievement Test (Science).

To test H_7 covariate analysis was performed on Object Generation scores sing scores from the Science Knowledge Test as covariate. Results of the analysis are provided in Table XX. The table shows hat the F-Test for the covariate was highly significant, indicating at the regression coefficient of the covariate was not zero and that an adjustment was being made to group means. The table also shows that the reflective group mean was adjusted down 0.03 points and the impulsive group mean up 0.03 points. The difference between adjusted group means was found to be 0.07, significant at the 0.86

Therefore the investigation failed to reject H_7. There to be no significant difference in ability to generate cs, as sured by the Object Generation Task, between reflective and impulsive groups.

Hypothesis E ght (H_8) (As originally stated):

> There is no significant difference existing between Reflectives' and Impulsives' ability in generating events.

Scores obtained from the Event Generation Task were employed to test this hypothesis. Stepwise Regression analysis was performed using the Event Generation measure as criterion. Major results of the analysis are presented in Table XXI. The table reveals that

TABLE XX

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE OBJECT GENERATION TASK

Source	SS	Df	MS	F	p
Treatment	0.13	1	0.13	0.03	0.86
Covariate: Science Knowledge	28.68	1	28.68	7.30	0.01
Error	530.32	135	3.93		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1, \chi^2 = 0.49, p. 0.48$$

Test for Homogeneity of Regression Coefficient:

$$F_{(1,134)} = 0.18, p = 0.67$$

	Unadjusted Means	Adjusted Means
Reflective Group	6.82	6.79
Impulsive Group	6.69	6.72

again Science Knowledge was the only variable contributing significantly to the variance of the measure. Subsequent covariate analysis, using Science Knowledge as covariate appeared appropriate.

TABLE XXI

SIGNIFICANT CONTRIBUTION OF THE VARIABLE SCIENCE KNOWLEDGE WITH THE EVENT GENERATION MEASURE AS CRITERION

Contributor	Multiple Regression R	% Variance Accounted for	F	p
X_1 (Science Knowledge)	0.43	18.47	30.8	0.00

Regression Equation:

$$Y = 0.09X_1 + 2.68$$

Hypothesis H_8 (Restated):

> There is no significant difference between reflective and impulsive group means on the Event Generation Task, when means are adjusted for performance on the Stanford Achievement Test (Science).

To test H_8 covariate analysis was performed on Event Generation scores, using scores of the Science Knowledge Test as covariate. Results of the analysis are given in Table XXII. The table indicates that the F-Test was highly significant for this covariate, indicating that the regression coefficient of the covariate was not zero and an adjustment was indeed being made to group means. The Table also indicates that the reflective group mean was adjusted down 0.07 points and the impulsive group mean up 0.07 points. The difference between adjusted means was found to be equal to 0.13, significant at the 0.62 level. Therefore the investigation failed to reject H_8. There is no

TABLE XXII

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE EVENT GENERATION TASK

Source	SS	Df	MS	F	
Treatment	0.66	1	0.66	0.25	0.62
Covariate: Science Knowledge	78.86	1	78.86	28.9	0.00
Error	358.56	135	2.66		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1,\ \chi^2 = 0.32,\ p = 0.57$$

Test for Homogeneity of Regression Coefficient:

$$F_{(1,134)} = 0.01,\ p = 0.99$$

	Unadjusted Means	Adjusted Means
Reflective Group	5.71	5.64
Impulsive Group	5.44	5.51

significant difference in ability to generate events as measured by the Event Generation Task between reflective and impulsive groups.

Hypothesis Nine (H_9) (As originally stated):

> There is no significant difference between Reflectives' and Impulsives' ability in generating hypotheses.

Since scores obtained from two measures, the Hypothesis Generation Task and the Pearson Hypothesis Formulation Test, were used to test H_9, the hypothesis was divided into two parts. Hypothesis 9.1 deals with analysis pertaining to the Hypothesis Generation Task and Hypothesis 9.2 deals with analysis pertaining to the Pearson Hypothesis Formulation Test.

Hypothesis 9.1. Stepwise Regression analysis was performed using the Hypothesis Generation Task as criterion. Results of the analysis indicated that no variable contributed significantly to the variance. Again this finding is expected since correlations between the Hypothesis Generation scores, and those of the six control variables, were all of non-significant magnitude. Among the variables, Science Knowledge was the major contributor (variance accounted for, 1.24%, $F = 1.71$, $p = 0.19$). In view of these findings, subsequent covariate analysis appeared inappropriate.

Hypothesis 9.1 (Restated):

> There is no significant difference between reflective and impulsive group means on the Hypothesis Generation Task.

To test $H_{9.1}$ a t-Test was executed on Hypothesis Generation scores. The results of the analysis are given in Table XXIII. The table reveals that the t-Test was not significant ($p = 0.10$).

Therefore the investigation failed to reject $H_{9.1}$. It appears that there is no significant difference in ability to generate hypotheses, as measured by the Hypothesis Generation Task, between reflective and impulsive groups.

TABLE XXIII

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE HYPOTHESIS GENERATION TASK

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Ste. Devn. Imp. Group	Df	t	p
5.29	4.99	1.20	0.99	136	1.65	0.10

Hypothesis 9.2. Stepwise Regression analysis was performed to determine which of the six control variables, if any, contributed significantly to the variance of the Pearson Hypothesis Formulation measure. Major results of the analysis are presented in Table XXIV.

TABLE XXIV

SIGNIFICANT CONTRIBUTIONS OF VARIABLES WITH THE PEARSON HYPOTHESIS FORMULATION MEASURE AS CRITERION

Contributors	Multiple Regression R	F_R	% Variance Accounted for	F	p
X_1 (Science Knowledge)	0.38	23.46	14.71	23.46	0.00
X_1, X_2 (Non-Verbal I.Q.)	0.43	15.65	18.82	6.83	0.01
X_1, X_2, X_3 (Verbal I.Q.)	0.47	12.35	21.66	4.85	0.03
X_1, X_2, X_3, X_4 (Paragr. Mning.)	0.50	11.16	25.13	6.17	0.01

Regression Equation:

$$Y = 0.12X_1 + 0.11X_2 - 0.09X_3 + 0.10X_4 + 1.71$$

An examination of the table shows that four of the six control variables, Science Knowledge, Non-Verbal I.Q., Verbal I.Q., and Paragraph Meaning ability made significant contributions to the variance of the measure. On the basis of these findings subsequent covariate analysis of Pearson Hypothesis Formulation scores appeared appropriate. It is interesting to note that Verbal I.Q. (variable X_3) in contrast to Non-Verbal I.Q. (variable X_2) acted as a suppressor variable (see the Regression Equation, Table XXIV).

Hypothesis 9.2 (Restated):

> There is no significant difference between reflective and impulsive group means on the Pearson Hypothesis Formulation Test, when means are adjusted for performance on the Stanford Achievement Test (Science and Paragraph Meaning sections) and on the Lorge Thorndike Non-Verbal and Verbal I.Q. Batteries.

To test $H_{9.2}$ covariate analysis was performed on Pearson Hypothesis Formulation scores using scores associated with the four noted variables as covariates. Results of the analysis are presented in Table XXV. This table reveals that the F-Tests for each of the covariates were highly significant, indicating that regression coefficients of the covariates were not zero and that adjustments were being made to group means. The table also reveals that the reflective group mean was adjusted down 0.37 points and the impulsive group mean up 0.37 points. The difference between adjusted group means was found to be 1.42, significant at the 0.04 level. Therefore the investigation rejected $H_{9.2}$. It appears that there is a significant difference in ability to generate hypotheses, as measured by the Pearson Hypothesis Formulation Test, between reflective and impulsive groups.

TABLE XXV

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE PEARSON HYPOTHESIS FORMULATION TEST

Source	SS	Df	MS	F	p
Treatment	67.05	1	67.05	4.11	0.04
Covariate 1: Science Knowledge	77.20	1	77.20	4.74	0.03
Covariate 2: Non-Verbal I.Q.	169.39	1	169.39	10.39	0.00
Covariate 3: Verbal I.Q.	111.32	1	111.32	6.83	0.01
Covariate 4: Paragraph Meaning	105.39	1	105.39	6.47	0.01
Error	2151.02	132	16.30		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1, \chi^2 = 0.64, p = 0.42$$

Test for Homogeneity of Regression Coefficients:

Covariate 1: $F_{(1,128)} = 2.78, p = 0.10$

Covariate 2: $F_{(1,128)} = 0.20, p = 0.65$

Covariate 3: $F_{(1,128)} = 1.01, p = 0.32$

Covariate 4: $F_{(1,128)} = 0.94, p = 0.33$

	Unadjusted Mean	Adjusted Mean
Reflective Group	11.47	11.11
Impulsive Group	9.33	9.69

Hypothesis Ten (H_{10}) (As originally stated):

There is no significant difference existing between Reflectives' and Impulsives' ability in evaluating generated objects.

Scores obtained from the Evaluation of Generated Objects were utilized in testing this hypothesis. Stepwise Regression analysis was performed using the Evaluation of Generated Objects measure as criterion. Major results of the analysis are presented in Table XXVI. The table reveals that Science Knowledge was the only variable contributing significantly to the variance of the measure. Hence subsequent covariate analysis employing Science Knowledge as covariate appeared appropriate.

TABLE XXVI

SIGNIFICANT CONTRIBUTION OF THE VARIABLE SCIENCE KNOWLEDGE WITH THE EVALUATION OF GENERATED OBJECTS MEASURE AS CRITERION

Contributor	Multiple Regression R	% Variance Accounted for	F	p
X_1 (Science Know.)	0.25	6.13	8.88	0.00

Regression Equation:

$$Y = 0.06X_1 + 3.17$$

Hypothesis H_{10} (Restated):

There is no significant difference between reflective and impulsive group means on the Evaluation of Generated Objects Task, when means are adjusted for performance on the Stanford Achievement Test (Science).

To test H_{10} covariate analysis was performed on Evaluation of Generated Objects scores using scores from the Science knowledge test as covariate. Results of the analysis are provided in Table XXVII. The table reveals that the F-Test for the covariate was highly significant indicating that the regression coefficient for this covariate was not zero, and that an adjustment was being made to group means. The table further reveals that the reflective group mean was adjusted down 0.04 points and the impulsive group mean up 0.04 points. The difference between adjusted group means was found to be 0.30, significant at the 0.39 level. Therefore the investigation failed to reject H_{10}. It appears that there is no significant difference in ability to evaluate generated objects, as measured by the Evaluation of Generated Objects Task, between reflective and impulsive groups.

Hypothesis Eleven (H_{11}) (As originally stated):

> There is no significant difference in Reflectives' and Impulsives' ability in Evaluating Generated Events.

Scores from the Evaluation of Generated Events Task were used to measure this hypothesis. Stepwise Regression analysis was performed, using the Evaluation of Generated Events measure as criterion. Major results of the analysis are presented in Table XXVIII. The table reveals that Science Knowledge was the only variable which made a significant contribution to the variance. Hence subsequent covariate analysis, employing this variable as covariate, appeared appropriate.

Hypothesis H_{11} (Restated):

> There is no significant difference between reflective and impulsive group means on the Evaluation of Generated Events Task, when means are adjusted for performance on the Stanford

TABLE XXVII

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE EVALUATION OF GENERATED OBJECTS TASK

Source	SS	Df	MS	F	p
Treatment	3.04	1	3.04	0.74	0.39
Covariate: Science Knowledge	33.63	1	33.63	8.14	0.01
Error	557.53	135	4.13		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1,\ \chi^2 = 0.25,\ p = 0.62$$

Test for Homogeneity of Regression Coefficient:

$$F_{(1,134)} = 0.10,\ p = 0.75$$

	Unadjusted Means	Adjusted Means
Reflective Group	4.31	4.27
Impulsive Group	3.93	3.97

TABLE XXVIII

SIGNIFICANT CONTRIBUTION OF THE VARIABLE SCIENCE KNOWLEDGE WITH THE EVALUATION OF GENERATED EVENTS MEASURE AS CRITERION

Contributor	Multiple Regression R	% Variance Accounted for	F	p
X_1 (Science Know.)	0.36	13.08	20.45	0.00

Regression Equation:

$$Y = 0.08X_1 + 2.04$$

Achievement Test (Science).

To test H_{11} covariate analysis was performed on Evaluation of Generated Events scores, using scores from the Science Test as a covariate. Results of the analysis are provided in Table XXIX. The table reveals that the F-Test for the Science Knowledge covariate was highly significant, indicating that the regression coefficient for the covariate was not zero and that an adjustment was being made to group measures. The table indicates that the reflective group mean was adjusted down 0.06 points and the impulsive group mean up 0.06 points. The difference between adjusted group means was found to be 0.29, significant at the 0.32 level. Therefore the investigation failed to reject H_{11}. There is no significant difference in ability to evaluate generated events, as measured by the Evaluation of Generated Events Task, between reflective and impulsive groups.

<u>Hypothesis Twelve (H_{12})</u> (As originally stated):

> There is no significant difference in Reflectives' and Impulsives' ability in evaluating generated hypotheses.

Scores obtained from the Evaluation of Generated Hypotheses Task were employed to test this hypothesis. Stepwise Regression analysis was performed using the Evaluation of Generated Hypotheses measure as criterion. Results of the analysis indicated that no variable made significant contribution to the variance of the measure. Among the variables, Science Knowledge was the major contributor (variance accounted for, 1.95%, F = 2.70, p = 0.10). The fact that no variable made significant contribution to the variance of the measure is not surprising since correlations between scores associated

TABLE XXIX

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE EVALUATION OF GENERATED EVENTS TASK

Source	SS	Df	MS	F	p
Treatment	3.04	1	3.04	0.98	0.32
Covariant: Science Knowledge	60.91	1	60.91	19.56	0.00
Error	420.40	135	3.11		

Bartlett Test for Homogeneity of Group Variance:

$Df = 1,\ \chi^2 = 0.00,\ p = 0.99$

Test for Homogeneity of Regression Coefficient:

$F_{(1,134)} = 0.06,\ p = 0.80$

	Unadjusted Means	Adjusted Means
Reflective Group	3.81	3.75
Impulsive Group	3.40	3.46

with the measure and those of the six v   .les were very low in magnitude (see Table IV, p. 83 and Table V, p, 84). In view of these findings subsequent covariate analysis appeared inappropriate.

Hypothesis H_{12} (Restated):

> There is no significant difference between reflective and impulsive group means on the Evaluation of Generated Hypotheses Task.

To test H_{12} a t-Test was performed on the Evaluation of Generated Hypotheses scores. The results of the analysis are given in Table XXX. The table indicates that the t-Test was not significant ($p = 0.16$). Therefore the investigation failed to reject H_{12}. It appears that there is no significant difference in ability to evaluate generated hypotheses, as measured by the Evaluation of Generated Hypotheses measure, between reflective and impulsive groups.

TABLE XXX

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE EVALUATION OF GENERATED HYPOTHESES TASK

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
4.91	4.62	1.25	1.09	136	1.41	0.16

Comments Pertaining to Hypotheses Seven to Twelve

Hypotheses Seven, Eight and Nine dealt with the ability of Reflectives and Impulsives to generate objects, events and hypotheses. With respect to each task, Reflectives performed superiorly to Impulsives but only in the case of the Pearson Hypothesis Formulation Test was performance superiority at a statistically significant level. Thus while it appears that reflective students perform about as well as impulsive students on tasks measuring abilities in generating objects, events and hypotheses (as measured by the Object Generation, Event Generation and Hypothesis Generation Tasks respectively), the performance of Reflectives on the Pearson Hypothesis Formulation measure is significantly superior to that of Impulsives. In all tasks (except the Hypothesis Selection Task) Science Knowledge played a significant role in contribution to the variances. It appears that ability to generate objects, events and hypotheses is in part a function of Science Knowledge. As in the case of the Hypothesis Selection and Pearson Hypothesis Selection measures, the two measures assessing ability in generating hypotheses, the Hypothesis Generation Task and the Pearson Hypothesis Formulation Test, appeared to demand markedly different types of cognitive activity on the part of students involved. Evidence supporting this claim is found in the essentially zero correlations existing between these measures (see Table V, p. 84). Further evidence is found in the fact that Science Knowledge, Intellectual ability and Reading ability contributed significantly to the variance of the Pearson Hypothesis Formulation Test, but failed to do so in the case of the Hypothesis Generation Task. Again these differences in cognit[illegible] requirement appear to result at least in part from differences in content areas and modes of content presentation associated with respective measures.

Hypotheses Ten, Eleven and Twelve dealt with the ability of Reflectives and Impulsives to evaluate generated objects, events and hypotheses. In all tasks Reflectives performed superiorly to

Impulsives, but in no instance was performance superiority statistically significant. Thus it appears that reflective and impulsive students perform about as well on tasks demanding ability in evaluating generated objects, events and hypotheses, as measured by tasks used in this investigation. While ability in evaluating generated objects and generated events is dependent in part upon Science Knowledge, ability in evaluating generated hypotheses appears essentially independent of Science Knowledge.

SECTION B

This section deals with those hypotheses which relate to the performance of Reflectives and Impulsives on the Suchman Question-Asking Task.

Hypothesis Thirteen (H_{13}) (As originally stated):

> There is no significant difference in the total number of Suchman-type questions asked between reflective and impulsive groups.

Scores obtained on the Suchman Question-Asking Test were used in testing this hypothesis. Table XXXI indicates the total number of questions asked by reflective and impulsive groups, and the number of questions asked by these groups according to the two major Verification and Implication question-type categories. As in the case of Hypotheses H_1-H_{12}, Stepwise Regression analysis was conducted with control variables Science Knowledge, Non-Verbal and Verbal I.Qs., Age and Word and Paragraph Meaning abilities, using the Suchman Question-Asking measure as criterion. Major results of the analysis are presented in Table XXXII. The table reveals that

TABLE XXXI

TOTAL NUMBER OF VERIFICATION AND IMPLICATION QUESTION TYPES ASKED BY REFLECTIVE AND IMPULSIVE GROUPS

	Question Type					
	Verification		Implication		Combined	
	Number		Number		Number	
Group	Total	Per Student	Total	Per Student	Total	Per Student
Reflective	118	1.74	46	0.68	164	2.41
Impulsive	108	1.54	45	0.64	153	2.19
Combined	226	1.64	91	0.66	317	2.30

Science Knowledge was the only variable contributing significantly to the variance. Subsequent covariate analysis employing Science Knowledge as covariate appeared appropriate.

TABLE XXXII

SIGNIFICANT CONTRIBUTION OF THE VARIABLE SCIENCE KNOWLEDGE WITH THE QUESTION-ASKING TEST AS CRITERION

Contributor	Multiple Regression R	% Variance Accounted for	F	p
X_1 (Science Know.)	0.20	3.84	5.43	0.02

Regression Equation:

$$Y = 0.04X_1 + 2.19$$

<u>Hypothesis H_{13}</u> (Restated):

> There is no significant difference in the total number of Suchman-like questions asked between reflective and impulsive groups when group means are adjusted for performance on the Stanford Achievement Test (Science).

To test H_{13} covariate analysis was performed on Suchman Question-Asking scores using scores from the Science Knowledge Test as covariate. Results of the analysis are provided in Table XXXIII. The table reveals that the F-Test for the Science Knowledge covariate was highly significant, indicating that the regression coefficient for the covariate was not zero and that an adjustment was being made to group means. The table indicates in addition that the reflective group mean was adjusted down 0.02 points and the impulsive group mean up 0.02 points. The difference between adjusted group means

TABLE XXXIII

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON THE QUESTION-ASKING TEST

Source	SS	Df	MS	F	p
Treatment	1.10	1	1.10	0.49	0.48
Covariate: Science Knowledge	10.97	1	10.97	4.89	0.03
Error	302.86	135	2.24		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1,\ X^2 = 0.58,\ p = 0.45$$

Test for Homogeneity of Regression Coefficient:

$$F_{(1,134)} = 0.47,\ p = 0.49$$

	Unadjusted Means	Adjusted Means
Reflective Group	2.41	2.39
Impulsive Group	2.19	2.21

was found to be 0.18, significant at the 0.48 level. Therefore the investigation failed to reject H_{13}. It appears that there is no significant difference in the total number of Suchman-type questions asked, as measured by the Suchman Question-Asking Test, between reflective and impulsive groups.

Hypothesis Fourteen (H_{14}):

> There is no significant difference in the distribution of questions of specific Suchman types asked between reflective and impulsive groups.

To test this hypothesis the Kolmogorov-Smirnov Two Sample Test was employed. Among other researchers, Guilford (1956) and Siegel (1962) recommend use of this test for comparing distributions of two samples. Application of this test necessitates allocation of questions asked to specific Suchman categories and the subsequent calculation of the cumulative proportions of questions within each category. By these procedures a standard value is obtained that is unaffected by differences in sample size. Table XXXIV presents the number of questions asked by reflective and impulsive groups according to specific Suchman categories. Cumulative proportions and differences between cumulative proportions are shown. In testing for significant difference between reflective and impulsive distributions a two-tailed test was employed that utilized the differences between the cumulative proportions per question category. For the distribution to be significantly different at the 0.05 level, the difference between the cumulative proportions of any category must exceed a value equal to

$$1.36 \left[\frac{N_{Ref.} + N_{Imp.}}{N_{Ref.} \times N_{Imp.}} \right]^{\frac{1}{2}}$$

TABLE XXXIV

NUMBER, PROPORTION, CUMULATIVE PROPORTION AND DIFFERENCE BETWEEN CUMULATIVE PROPORTIONS OF SUCHMAN QUESTION TYPES

Suchman Question Type**	Reflective Group			Impulsive Group			Difference Between Cum. Propns.
	No.	Propn.	Cum. Propn.	No.	Propn.	Cum. Propn.	
1	1	.006	.006	5	.033	.033	.027
2	21	.128	.134	9	.058	.091	.043
3	86	.524	.658	88	.575	.666	.008
4	0	.000	.658	0	.000	.666	.008
5	0	.000	.658	0	.000	.666	.008
6	10	.061	.719	6	.039	.705	.014
7	10	.061	.780	2	.013	.718	.062
8	10	.061	.841	5	.033	.751	.090*
9	11	.067	.908	14	.092	.843	.065
10	8	.049	.957	8	.052	.895	.062
11	1	.006	.963	2	.013	.908	.055
12	6	.037	1.000	14	.092	1.000	.000
Total	164			153			

* Maximum difference observed.

** 1 = Categorical verif.: Nominal
2 = Categorical verif.: Normative
3 = Analytical verif.: Condn.-Desc.
4 = Analytical verif.: Condn.-Comp.
5 = Analytical verif.: Struc.-Comp.
6 = Analytical verif.: Properties Check
7 = Abstract-Conceptual: Diffuse
8 = Abstract-Conceptual: Directed
9 = Concrete-Inferential: Elimination
10 = Concrete-Inferential: Substitution
11 = Concrete-Inferential: Addition
12 = Concrete-Conceptual

where $N_{Ref.}$ and $N_{Imp.}$ refer to Reflective and Impulsive sample respectively. Substituting values of $N_{Ref.} = 164$ and $N_{Imp.} = 153$ into the above equation a critical value of 0.152 was obtained. Examination of Table XXXIV reveals that the distribution of questions between reflective and impulsive groups within the first major Verification category (encompassing question types 1-6) was remarkably uniform. Reflectives, for example, had 71.9% of the questions within Verification categories, while Impulsives had 70.5% of the questions within the categories. Within the second major category, the Implication category, the distribution of questions was less uniform between groups, but still this distribution failed to reach a significant level of difference. The maximum difference in question distribution was attained in the Directed subcategory of Implication question. However again this difference failed to attain the critical value of 0.152 required for significance. Since in no instance did the difference between cumulative proportions reach this critical value, the investigation failed to reject H_{14}. According to the Kolmogorov-Smirnov Test it appears that there is no significant difference in the distribution of questions of specific Suchman-type asked between reflective and impulsive groups.

Comments Pertaining to Hypotheses Thirteen and Fourteen

Hypothesis H_{13} revealed that reflective and impulsive groups ask a similar number of questions (regardless of type) when seeking information during problem-solving. Hypothesis H_{14} revealed that reflective and impulsive groups ask similar quantities of the twelve distinct Suchman-type questions. These findings indicate that there is no preference for students of one disposition over another to ask questions of either an analytic type (as manifested by questions asked within Verification categories) or of a more global nature (such as those contained within Implication categories). Thus this investigation provides no evidence to support the claims of Ault (1973) and Denney (1973) (see p. 37) that Reflectives and Impulsives ask different types of questions during problem-solving.

Hypotheses Fifteen to Eighteen

These four hypotheses provide further information pertaining to the general issue of question-type usage. These hypotheses were used to investigate whether significant differences exist between reflective and impulsive groups in terms of the frequency with which Verification and Implication question types are used, in instances where one, two, three and four questions are asked during the question-asking session. Table XXXV provides data common to Hypotheses H_{15}-H_{18}. In this table comparisons are made between frequency of usage of Verification and Implication question types pertaining to those reflective and impulsive students who ask one, two, three and four questions respectively. Examination of the table reveals that for both reflective and impulsive groups, as the total number of questions asked increased, the proportion of Verification type questions increased and the proportion of Implication type questions decreased.

TABLE XXXV

NUMBER OF QUESTIONS ASKED AND FREQUENCY OF USAGE OF VERIFICATION AND IMPLICATION QUESTION TYPES FOR REFLECTIVE AND IMPULSIVE GROUPS

Group	Number of Questions Asked	Number of Students	Propn. of Group	Verification Types		Implication Types	
				No.	Propn.	No.	Propn.
Ref.	0	4	.06				
	1	20	.29	9	.45	11	.55
	2	13	.19	16	.62	10	.38
	3	17	.25	36	.71	15	.29
	4	9	.13	35	.97	1	.03
	> 4	5	.08				
Imp.	0	7	.09				
	1	20	.29	13	.65	7	.35
	2	16	.24	21	.66	11	.34
	3	14	.20	30	.71	12	.29
	4	8	.11	18	.57	14	.43
	> 4	5	.07				

Hypothesis Fifteen (H_{15})

> With respect to subjects asking one Suchman-type question only, there is no significant difference in the number of questions of Verification and Implication type between reflective and impulsive groups.

Table XXXVI shows the number of questions allotted to the two major Verification and Implication categories. In addition the proportion of questions of specific type are shown. The table reveals that 20 Reflectives (i.e. 29.4% of the reflective population and 14.5% of the total population) and 20 Impulsives (i.e. 28.6% of the impulsive population and 14.5% of the total population) asked one question only during the question-asking session. Data was subjected to a χ^2-Test of independence (Ferguson, 1971, p. 184). A χ^2 of 1.62 was obtained, significant at the 0.20 level. Therefore the investigation failed to reject H_{15}. In instances where one question is asked there is no significant difference in the number of questions of Verification and Implication types between reflective and impulsive groups.

Hypothesis Sixteen (H_{16})

> With respect to subjects asking two Suchman-type questions there is no significant difference in the number of questions of Verification and Implication type pertaining to the first and second questions asked between reflective and impulsive groups.

Table XXXVII indicates the number of questions allotted to the Verification and Implication categories for the first and second questions and the proportion of questions of specific type. The table reveals that 13 Reflectives (i.e. 19.1% of the reflective population and 9.4% of the total population) and 16 Impulsives

TABLE XXXVI

QUESTION TYPE USAGE FREQUENCY—ONE QUESTION ASKED

		Question Type		
		Verification	Implication	
Reflective Gr.	No.	9	11	20
	Propn.	[illegible]5	.55	
Impulsive Gr.	No.	13	7	20
	Propn.	.65	.35	
		22	18	40

$\chi^2 = 1.62$, df = 1, p = 0.20

TABLE XXXVII

QUESTION TYPE USAGE FREQUENCY—TWO QUESTIONS ASKED

FIRST QUESTION:

		Question Type		
		Verif.	Implic.	
Reflective Gr.	No.	6	7	13
	Propn.	.46	.54	
Impulsive Gr.	No.	9	7	16
	Propn.	.56	.44	
		15	14	29

$\chi^2 = 0.29$, $df = 1$, $p = 0.60$

SECOND QUESTION:

		Question Type		
		Verif.	Implic.	
Reflective Gr.	No.	10	3	13
	Propn.	.77	.23	
Impulsive Gr.	No.	12	4	16
	Propn.	.75	.25	
		22	7	29

$\chi^2 = 0.01$, $df = 1$, $p = 0.85$

(i.e. 22 [illegible] of the impulsive population and 11.6% of the total population) asked two questions during the Test. Data pertaining to the first and second questions were subjected to a χ^2-Test of independence. With repect to the first question a χ^2 of 0.29 was obtained, significant at the 0.60 level. With respect to the second question a χ^2 of 0.01 was achieved, significant at the 0.90 level. Therefore for both the first and second questions the investigation failed to reject H_{16}. In instances where two questions are asked, there is no significant difference in the number of questions of Verification and Implication types pertaining to the first and second questions asked between reflective and impulsive groups.

Hypothesis Seventeen (H_{17})

> With respect to subjects asking three Suchman-type questions, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first, second and third questions asked between reflective and impulsive groups.

Table XXXVIII indicates the number of questions allotted to the Verification and Implication categories for the first, second and third questions and the proportions of questions of specific type. The table reveals that 17 Reflectives (i.e. 25.0% of the reflective population and 12.3% of the total population) and 14 Impulsives (i.e. 20.6% of the impulsive population and 10.1% of the total population) asked three questions during the question-asking session. Data pertaining to each question were subjected to a χ^2-Test of independence. Since, as indicated in Table XXXVIII some cell frequencies were of small magnitude (i.e. <5), values for χ^2 were calculated using Yate's Correction for Continuity. Ferguson (1971, p. 188) recommends

TABLE XXXVIII

QUESTION TYPE USAGE FREQUENCY—THREE QUESTIONS ASKED

FIRST QUESTION:

		Question Type		
		Verif.	Implic.	
Reflective Gr.	No. Propn.	11 .65	6 .35	17
Impulsive Gr.	No. Propn.	12 .86	2 .14	19
		23	8	31

χ^2 (Using Yate's Correction) = 0.84, df = 1, p = 0.40

SECOND QUESTION:

		Question Type		
		Verif.	Implic.	
Reflective Gr.	No. Propn.	11 .65	6 .35	17
Impulsive Gr.	No. Propn.	9 .64	5 .36	14
		20	11	31

χ^2 (Using Yate's Correction) = 0.12, df = 1, p = 0.70

THIRD QUESTION:

		Question Type		
		Verif.	Implic.	
Refelctive Gr.	No. Propn.	14 .82	3 .18	17
Impulsive Gr.	No. Propn.	9 .64	5 .36	14
		23	8	31

χ^2 (Using Yate's Correction) = 0.54, df = 1, p = 0.47

application of this Correction Factor whenever cell frequencies less than 5 are obtained. With respect to the first question, a χ^2 of 0.84 was obtained, significant at the 0.40 level. With respect to the second question a χ^2 of 0.12 was obtained, significant at the 0.70 level. Finally in the third question a χ^2 of 0.54 was achieved that was significant at the 0.47 level. Therefore for all questions the investigation failed to reject H_{17}. It appears that in instances where three questions are asked, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first, second and third questions asked between reflective and impulsive groups.

Hypothesis Eighteen (H_{18})

> With respect to subjects asking four Suchman-type questions, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first, second, third and fourth questions asked between reflective and impulsive groups.

Table XXXIX indicates the number of questions allotted to the two question categories for the first, second, third and fourth questions, and the proportion of questions of specific type. The table reveals that 9 Reflectives (i.e. 13.2% of the reflective population and 6.5% of the total population) and 8 Impulsives (i.e. 11.4% of the impulsive population and 5.8% of the total population) asked four questions during the question-asking session. As in the case of H_{17}, data pertaining to each question were subjected to χ^2 tests of independence using Yate's Correction for Continuity. With respect to the first question, a χ^2 of 1.92 was obtained, significant at the 0.15 level. With respect to the second question, a χ^2 of 2.51 was

TABLE XXXIX

QUESTION TYPE USAGE FREQUENCY— QUESTIONS ASKED

FIRST QUESTION:

		Question Type: Verif.	Question Type: Implic.	
Reflective Gr.	No. Propn.	9 1.00	0 0.00	9
Impulsive Gr.	No. Propn.	5 0.63	3 0.37	8
		14	3	17

χ^2 (Using Yate's Correction) = 1.92, df = 1, p = 0.15

SECOND QUESTION:

		Question Type: Verif.	Question Type: Implic.	
Reflective Gr.	No. Propn.	9 1.00	0 0.00	9
Impulsive Gr.	No. Propn.	4 0.50	4 0.50	9
		13	4	17

χ^2 (Using Yate's Correction) = 2.51, df = 1, p = 0.12

THIRD QUESTION:

		Question Type: Verif.	Question Type: Implic.	
Reflective Gr.	No. Propn.	9 1.00	0 0.00	9
Impulsive Gr.	No. Propn.	5 0.63	3 0.37	8
		14	3	17

χ^2 (Using Yate's Correction) = 1.92, df - 1, p 0.15

FOURTH QUESTION:

		Question Type: Verif.	Question Type: Implic.	
Reflective Gr.	No Propn.	8 0.89	1 0.11	9
Impulsive Gr.	No. Propn.	4 0.50	4 0.50	8
		12	5	17

χ^2 (Using Yate's Correction) = 1.50, df = 1, p = 0.22

obtained, significant at the 0.12 level. In the case of the third question, a χ^2 of 1.92 was obtained, significant at the 0.15 level. Finally in the case of the fourth question a χ^2 of 1.50 was obtained, significant at the 0.22 level. Based on these results this investigation failed to reject H_{18}. It appears that in instances where four questions are asked, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first, second, third and fourth questions asked between reflective and impulsive groups.

As noted earlier, Table XXXV, page 132, reveals that changes in the proportion of Verification and Implication question-types asked, occur with changes in the total number of questions asked during the Suchman Question-Asking Test. These changes however are manifested by both reflective and impulsive groups. Hypotheses H_{15}-H_{18} indicate that there is no significant difference in the number of questions of Verification and Implication types asked by reflective and impulsive students when these questions are the first, second, third and fourth to be asked.

SECTION C

Associated Question 1

Student criterion correlations for the combined reflective and impulsive groups have already been presented. Several significant correlations were found to exist between the Selection, Generation and Evaluation of Selected and Generated scores (see Table IV, p. 83 and Table V, p. 84). The investigator wished to determine whether correlations between these criterion variables were of similar magnitude for reflective and impulsive groups. To enable this to be done, correlations were found between variables for the reflective group and for the impulsive group alone. Table XL and Table XLI provide the correlation matrices for reflective and impulsive groups respectively. When values are compared, several differences are evident. For example, with respect to the impulsive group, correlations between the Tannenbaum Observation Test and both the Object Generation and Event Generation Tasks were significant at the 0.05 and 0.01 levels respectively (r = 0.28, r = 0.49). With the reflective group however, correlations between the same measures failed to reach significant levels (r = -0.02 and r = 0.49). In a similar fashion, with the impulsive group, Tannenbaum's Test correlated significantly at the 0.01 level with both the Evaluation of Generated Objects and the Evaluation of Generated Events (r = 0.32, r = 0.42 respectively), but with the reflective group, this Test failed to correlate significantly with the same tasks (r = 0.17, r = 0.19). Other correlations appearing different on the two groups were between:

1. The Pearson Hypothesis Selection Test and both the Evaluation of Selected Objects and Event Generation Tasks;

2. The Pearson Hypothesis Formulation Test and the Event Selection, Evaluation of Selected Events and Evaluation of Selected Hypotheses Tasks;

3. The Event Selection Task and both the Hypothesis Generation and Evaluation of Generated Hypotheses Tasks; and

4. The Hypothesis Selection Task and Evaluation of Generated

TABLE XL

INTERCORRELATION MATRIX FOR REFLECTIVES ON CRITERION MEASURES

Task	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
10. Tannenbaum Observation	1.00														
11. Pearson Hyp. Sel.	.06	1.00													
12. Pearson Hyp. Form.	.17	.58**	1.00												
13. Object Selection	-.22	.13	-.02	1.00											
14. Event Selection	.12	.28*	.31**	.04	1.00										
15. Hyp. Selection	.07	.11	.24*	-.06	.27*	1.00									
16. Evaln. of Sel. Obj.	-.06	.24*	.12	.47**	.27*	.05	1.00								
17. Evaln. of Sel. Events	.21	.30*	.30*	.13	.61**	.16	.18	1.00							
18. Evaln. of Sel. Hyps.	.04	.00	.08	.10	.20	.46**	.07	.23*	1.00						
19. Object Generation	-.02	.08	-.04	.03	.18	.09	.28*	.15	.10	1.00					
20. Event Generation	.22	.06	.14	-.06	.15	.10	.11	.08	.26*	.49**	1.00				
21. Hyp. Generation	.20	.11	-.03	.06	.13	.04	.13	.08	.19	.11	.12	1.00			
22. Evaln. of Gen. Objects	.17	.17	.01	-.03	.12	.17	.28*	.20	.06	.79**	.45**	.24*	1.00		
23. Evaln. of Gen. Events	.19	.19	.18	-.10	.06	.15	.07	.17	.18	.40**	.80**	.03	.58**	1.00	
24. Evaln. of Gen. Hyps.	.18	.10	.01	.07	.09	.08	.10	.15	.16	.11	.07	.87**	.24*	.05	1.00

* Denotes significance at 0.05 level.
** Denotes significance at 0.01 level.

TABLE XLI

INTERCORRELATION MATRIX FOR IMPULSIVES ON CRITERION MEASURES

Task	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
10. Tannenbaum Observation	1.00														
11. Pearson Hyp. Sel.	.21	1.00													
12. Pearson Hyp. Form.	.14	.49**	1.00												
13. Object Selection	-.01	.10	-.02	1.00											
14. Event Selection	.04	.16	.09	-.10	1.00										
15. Hyp. Selection	-.17	.00	.15	-.08	.21	1.00									
16. Evaln. of Sel. Objects	.06	-.16	.00	.46**	.01	.05	1.00								
17. Evaln. of Sel. Events	.00	.08	.02	-.08	.60**	.19	.28*	1.00							
18. Evaln. of Sel. Hyps.	.13	-.22	-.23	.15	.02	.51**	.13	.03	1.00						
19. Object Generation	.28*	.12	.08	.12	.07	.00	.06	.08	.06	1.00					
20. Event Generation	.49**	.39**	.10	.12	.03	.02	.12	.15	-.12	.45**	1.00				
21. Hyp. Generation	.15	.12	.01	.22	.33**	.19	.01	.11	.12	.17	.19	1.00			
22. Evaln. of Gen. Objects	.32**	.11	.10	.07	.12	.01	.08	.22	-.10	.83**	.50**	.15	1.00		
23. Evaln. of Gen. Events	.42**	.37**	.11	.07	.15	.03	.17	.27*	-.16	.43**	.85**	.12	.59**	1.00	
24. Evaln. of Gen. Hyps.	.10	.17	.09	.14	.29*	.27*	.01	.19	.07	.17	.28*	.84**	.21*	.21*	1.00

* Denotes significance at 0.05 level.
** Denotes significance at 0.01 level.

Hypotheses Task.

To determine whether any of the differences noted between correlations were significant, a Fisher's Z transformation was executed on each correlation (Ferguson, 1971, pp. 168-171). Fisher values are presented in Table XLII. In a Fisher's Z transformation any values of correlation comparisons exceeding 1.96 are significant at the 0.05 level. Those values exceeding 2.58 are significant at the 0.01 level. Examination of Table XLII reveals that almost all correlation comparisons failed to reach significance at the 0.05 level. Only four comparisons were significant at a 0.05 level or greater. Significant correlation comparisons were obtained between the Pearson Hypothesis Selection Test and both the Evaluation of Selected Objects and Event Generation Tasks, and between the Evaluation of Selected Hypotheses measure and both the Event Generation and Evaluation of Generated Events Tasks. In the opinion of the investigator it is quite possible that the four significant correlation comparisons are artifacts. This view is based on the following evidence:

1. Associated Z values, though significant, were low in magnitude,

2. Correlation comparisons between reflective and impulsive groups involving the Pearson Hypothesis Selection and Evaluation of Selected Hypotheses measures, and all other student variables (except those noted above) were not significant at the 0.05 level, and finally

3. As noted earlier, of the total number of correlation comparisons made (N = 105), the vast majority (N = 101) were not significant at the 0.05 level.

TABLE XLII

FISHER Z VALUES FOR COMPARISON OF STUDENT CRITERION MEASURE CORRELATIONS BETWEEN REFLECTIVE AND IMPULSIVE GROUPS

Task	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
10. Tannenbaum Observation	--														
11. Pearson Hyp. Sel.	0.86	--													
12. Pearson Hyp. Form.	0.17	0.72	--												
13. Object Selection	1.21	0.17	0.00	--											
14. Event Selection	0.46	0.69	1.09	0.80	--										
15. Hyp. Selection	1.38	0.63	0.52	0.11	0.34	--									
16. Evaln. of Sel. Objects	0.69	2.33*	0.69	0.06	1.49	0.00	--								
17. Evaln. of Sel. Events	1.21	1.32	1.67	1.21	0.06	0.17	0.57	--							
18. Evaln. of Sel. Hyps.	0.52	1.26	1.80	0.29	1.04	0.29	0.34	1.15	--						
19. Object Generation	1.78	0.23	0.69	0.52	0.63	0.52	1.32	0.40	0.23	--					
20. Event Generation	1.79	2.02*	0.23	1.04	0.69	0.46	0.06	0.40	2.22*	0.23	--				
21. Hyp. Generation	0.29	0.06	0.23	0.92	1.21	0.86	0.69	0.17	0.40	0.34	0.40	--			
22. Evaln. of Gen. Objects	0.92	1.13	0.52	0.57	0.00	0.92	1.21	0.11	0.92	0.23	0.29	0.52	--		
23. Evaln. of Gen. Events	1.47	1.15	0.40	0.98	0.52	0.69	0.57	0.57	1.97*	0.17	0.29	0.52	0.06	--	
24. Evaln. of Gen. Hyps.	0.46	0.40	0.46	0.40	1.21	1.15	0.52	0.23	0.52	0.34	1.25	0.17	0.17	0.92	--

* Denotes significance at 0.05 level.

It appears therefore, that correlations existing between the criterion variables are generally similar in magnitude for reflective and impulsive groups. In the few instances where comparisons between groups were found to be signifi[illegible]rent, it is possible that differences are artifacts.

Associated Question 2

In order to determine whether common factors existed among the Selection, Generation and Selection and Generation Evaluation Tasks, scores associated with each of these tasks were subjected to Factor Analysis. A Principal Component analysis was performed on reflective and impulsive groups combined (N = 138). The correlation matrix was rotated orthogonally according to Varimax criteria. Six factors identified with Eigenvalues greater than unity were extracted. The Varimax factor rotations of these factors are presented in Table XLIII. The table reveals that:

1. Factor I appears to be defined by four criterion variables: Object Generation (0.79), Event Generation (0.82), Evaluation of Generated Objects (0.84) and Evaluation of Generated Events (0.84). All other variables loaded <0.36.

2. Factor II appears to be defined by two criterion variables: Hypothesis Generation (0.96) and Evaluation of Generated Hypotheses (0.94). All other variables loaded <0.21.

3. Factor III appears to be defined by two criterion variables: Pearson Hypothesis Selection (0.83) and Pearson Hypothesis Formulation (0.84). All other variables loaded <0.28.

4. Factor IV appears to be defined by two criterion variables:

TABLE XLIII

LOADINGS OF CRITERION VARIABLES ON IDENTIFIED COMMON FACTORS (VARIMAX ROTATION)

Variable	Factors						Common-alities
	I	II	III	IV	V	VI	
10. Tannenbaum Observation	0.35	0.20	0.22	0.08	-0.31	-0.14	0.33
11. Pearson Hyp. Sel.	0.15	0.10	0.83	0.10	0.06	-0.04	0.74
12. Pearson Hyp. Form.	0.03	0.02	0.84	0.12	0.02	0.00	0.73
13. Object Selection	-0.01	0.14	0.11	-0.13	0.86	0.02	0.79
14. Event Selection	0.03	0.13	0.14	0.86	-0.01	0.12	0.79
15. Hyp. Selection	0.05	0.06	0.08	0.17	-0.09	0.85	0.76
16. Evaln. of Sel. Objects	0.16	-0.02	0.00	0.24	0.78	0.01	0.69
17. Evaln. of Sel. Events	0.13	0.02	0.09	0.87	0.08	0.07	0.79
18. Evaln. of Sel. Hyps.	0.01	0.08	-0.12	0.03	0.14	0.85	0.76
19. Object Generation	0.79	0.06	-0.15	0.11	0.16	0.01	0.69
20. Event Generation	0.82	0.07	0.23	-0.06	-0.03	0.09	0.74
21. Hyp. Generation	0.09	0.96	0.02	0.07	0.06	0.07	0.94
22. Evaln. of Gen. Objects	0.84	0.13	-0.08	0.16	0.10	-0.04	0.77
23. Evaln. of Gen. Events	0.84	-0.01	0.27	0.01	-0.05	0.07	0.78
24. Evaln. of Gen. Hyps.	0.12	0.94	0.06	0.[illegible]	0.03	0.08	0.91
Percentage Total Variance	19.44	12.71	11.15	11.11	10.16	9.98	74.56

Event Selection (0.86) and Evaluation of Selected Events (0.87). All other variables loaded <0.25.

5. Factor V appears to be defined by two criterion variables: Object Selection (0.86) and Evaluation of Selected Objects (0.78). All other variables loaded <0.17.

6. Factor VI appears to be defined by two criterion variables: Hypothesis Selection (0.85) and Evaluation of Selected Hypotheses (0.85). All other variables loaded <0.13.

From this review it is evident that 14 of the 15 criterion variables loaded heavily on one of the six Factors identified. An exception to this trend was manifested by the Tannenbaum Observation Test which failed to load heavily on any Factor. It is interesting to note that in every instance the six Evaluation Tasks loaded heavily on the same Factor as respective Selection and Generation Tasks. This outcome is not surprising, since in most instances high correlations between Evaluation and respective Selection and Generation Tasks had previously been noted (see Table IV, p. 83 and Table V, p. 84). Again it must be remembered that these correlations may be spuriously high for reasons earlier noted (see p. 57). Four of the six Evaluation tasks loaded heavily on separate Factors. This is understandable in view of the low, usually non-significant, correlations obtained between these Tasks. The two Evaluation Tasks that were moderately highly correlated, the Evaluation of Generated Objects and Evaluation of Generated Events Tasks ($r = 0.59$, significant at 0.01 level), loaded on a common Factor (Factor I).

Four of the six Selection and Generation Tasks loaded heavily on separate Factors, the Object and Event Generation Tasks providing exceptions to this trend by loading together (with respective Evaluation Tasks) on Factor I. Again these outcomes were to be expected in view of the very low, non-significant, intercorrelations of the above noted four tasks, and the substantial correlation obtained between the Object and Event Generation Tasks. The only instance where a Selection and a Generation Task loaded heavily on a similar Factor was in the case of the Pearson Hypothesis Selection and Pearson Hypothesis Formulation Tests loading on Factor III. Again this is not surprising since a substantial correlation of 0.55 existed between these measures.

From these results the following general conclusions may be drawn:

1. Commonalities between all Evaluation Tasks and respective Selection and Generation Tasks are high.

2. Except in the noted cases of the Object and Event Generation Tasks, and the Pearson Hypothesis Selection and Pearson Hypothesis Formulation measures, commonalities between Selection and Generation Tasks are low.

Associated Question 3

The investigator wished to determine whether significant differences existed between reflective and impulsive groups with respect to Factor Scores derived from the six common Factors.

Employing an arbitrarily assigned mean (20.00) and Standard Deviation (5.00) six Factor Scores for each subject (N = 138) were calculated in accordance with the following parameters:

$$F = SR^{-1}Z$$

where F (r x N) = Factor Score matrix

S (n x r) = Factor structure matrix

R (n x n) = Correlation matrix

Z (n x N) = Standardized Score matrix

for n = 15 variables, r = 6 Factors and N = 138 subjects.

Statistical procedures followed in testing Hypotheses H_1-H_{13} were adopted for determining whether significant differences exist between reflective and impulsive groups with respect to each of the six derived Factor Scores. Accordingly Stepwise Regression analysis was performed on Factor Scores in turn, in order to determine whether any of the control variables, Science Knowledge, Age, Nonverbal and Verbal intellectual abilities, Word and Paragraph Meaning abilities contributed significantly to the variance of these scores. In instances where one or more of these variables made significant contribution to the variance, subsequent covariate analysis was executed employing significant contributors as covariates. In instances where no variables contributed significantly to the variance, a t-Test was subsequently performed between reflective and impulsive group means. A review of the procedures and findings pertaining to each of the Factor Scores follows.

<u>Factor I Score</u>

(Factor I is defined as Object and Event Generation Tasks, and Evaluation of Generated Objects and Events Tasks.)

Major results of a Stepwise Regression analysis performed on this Score are presented in Table XLIV. The table reveals that Science

Knowledge was the only variable succeeding in making significant contribution to the variance of the Score. This is to be expected since this variable contributed significantly in each of the Tasks servi g to define Factor I. An analysis of covariance was subsequently performed on Factor I Score using Science Knowledge as covariate. Table XLV reveals that the F-Test for the Science Knowledge covariate was highly significant indicating that the Regression coefficient for the covariate was not zero and that an adjustment was being made to group means. The table indicates that the reflective group mean was adjusted down 0.30 points and the impulsive group mean up 0.30 points. The difference between adjusted group means was found to be 0.46 points, significant at the 0.77 level. Therefore there appears to be no significant difference between Factor I Scores obtained by reflective and impulsive groups, when means are adjusted for performance on the Stanford Ac ment Test (Science).

TABLE XLIV

SIGNIFICANT CONTRIBUTION OF THE VARIABLE SCIENCE KNOWLEDGE WITH FACTOR I AS CRITERION

Contributor	Multiple Regression R	% Variance Accounted For	F	p
X_1 (Science Knowledge)	0.37	13.66	21.5	0.00

Regression Equation:

$$Y = 0.44X_1 + 6.05$$

TABLE XLV

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON FACTOR I SCORES

Source	SS	Df	MS	F	p
Treatment	7.28	1	7.28	0.08	0.77
Covariate: Science Knowledge	1852.42	1	1852.42	21.00	0.00
Error	11909.71	135	88.22		

Bartlett Test for Homogeneity of Group Variance:

$$Df = 1,\ \chi^2 = 0.06,\ p = 0.80$$

Test for Homogeneity of Regression Coefficient:

$$F_{(1,134)} = .01,\ p = 0.95$$

	Unadjusted Means	Adjusted Means
Reflective Group	20.54	20.24
Impulsive Group	19.48	19.78

Factor II Score

(Factor II is defined by Hypothesis Generation and Evaluation of Generated Hypotheses Tasks.)

Stepwise Regression analysis performed on this Score indicated that none of the six variables served as significant contributors to the variance of the Score. This finding is expected since none of the variables contributed significantly to either the Hypothesis Generation or Evaluation of Generated Hypotheses Tasks. Verbal I.Q. was the major contributor (variance accounted for, 0.96%, F = 1.31, p = 0.25). Results of a t-Test performed on Factor II Scores are presented in Table XLVI. The table indicates that the t-Test was not significant p = 0.19). Therefore there appears to be no significant difference between Factor II Scores obtained by reflective and impulsive groups.

TABLE XLVI

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON FACTOR II SCORES

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
21.15	18.88	10.96	8.99	136	1.33	0.19

Factor III Score

(Factor III is defined by Pearson Hypothesis Selection and Pearson Hypothesis Formulation measures.)

Major results of the Stepwise Regression analysis performed on the Score are presented in Table XLVII. The table reveals that Science Knowledge, Non-Verbal and Verbal I.Q., and Paragraph Meaning ability contribute significantly to the variance of this Score. These findings are expected since these same variables contributed significantly to the variances of the two Pearson Tests. As in the case with the Pearson Tests, the Regression Equation indicates that Verbal I.Q. (variable X_3) acted as a suppressor variable. An analysis of covariance was subsequently performed on Factor III Score using the four noted variables as covariates. Table XLVIII reveals that the F-Test for each covariate was highly significant indicating that the regression coefficient for each covariate was not zero, and an adjustment was being made to group means. The table also indicates that the reflective group means were adjusted down 0.85 points and the impulsive group mean up 0.83 points. The difference between adjusted group means was 2.65, significant at the 0.08 level. Therefore there appears to be no significant difference between Factor III Scores, obtained by reflective and impulsive groups, when means are adjusted for performance on Tests associated with Science Knowledge, Non-Verbal and Verbal I.Q., and Paragraph Meaning.

Factor IV Score

(Factor IV is defined by Event Selection and Evaluation of Event Selection Tasks.)

Stepwise Regression analysis performed on the Score indicated that none of the control variables served as significant contributors to the variance of the Score. Paragraph Meaning was the major contributor (variance accounted for, 2.33%, $F = 3.24$, $p = 0.07$). Results of a t-Test performed on this Score are presented in Table IL. The table indicates that the t-Test was not significant ($p = 0.36$).

TABLE XLVII

SIGNIFICANT CONTRIBUTION OF VARIABLES WITH FACTOR III AS CRITERION

Contributors	Multiple Regression R	F_R	% Variance Accounted for	F	p
X_1 (Science Knowledge)	0.34	17.81	11.58	17.81	0.00
X_1, X_2 (Non-Verbal I.Q.)	0.41	13.39	16.55	8.04	0.01
X_1, X_2, X_3 (Verbal I.Q.)	0.47	12.63	22.04	9.44	0.00
X_1, X_2, X_3, X_4 (Paragraph Meaning)	0.50	11.32	25.39	5.97	0.02

Regression Equation:

$$Y = 0.23X_1 + 0.28X_2 - 0.26X_3 + 0.22X_4 + 2.67$$

TABLE XLVIII

ANALYSIS OF COVARIANCE FOR REFLECTIVE AND IMPULSIVE GROUPS ON FACTOR III SCORES

Source	SS	Df	MS	F	p
Treatment	233.53	1	233.53	3.06	0.08
Covariate 1: Non-Verbal I.Q.	1312.17	1	1312.17	17.21	0.00
Covariate 2: Verbal I.Q.	925.99	1	925.99	12.14	0.00
Covariate 3: Science Knowledge	300.44	1	300.44	3.94	0.04
Covariate 4: Paragraph Meaning	433.18	1	433.18	5.68	0.02
Error	10063.63	132	76.24		

Bartlett Test for Homogeneity of Group Variance:

$Df = 1,\ \chi^2 = 0.34,\ p = 0.56$

Test for Homogeneity of Regression Coefficients:

Covariate 1: $F_{(1,128)} = 0.18,\ p = 0.67$

Covariate 2: $F_{(1,128)} = 0.90,\ p = 0.34$

Covariate 3: $F_{(1,128)} = 0.14,\ p = 0.71$

Covariate 4: $F_{(1,128)} = 1.82,\ p = 0.18$

	Unadjusted Means	Adjusted Means
Reflective Group	22.20	21.35
Impulsive Group	17.87	18.70

Therefore there appears to be no significant difference between Factor IV Scores obtained by reflective and impulsive groups.

TABLE XL

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON FACTOR IV SCORES

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
19.21	20.77	10.08	10.00	136	-0.91	0.36

Factor V Score

(Factor V is defined by Object Selection and Evaluation of Object Selection Tasks.)

Stepwise Regression analysis performed on the Score indicated that no control variable contributed significantly to the variance of the Score. Non-Verbal I.Q. was the major contributor (variance accounted for, 1.42%, F = 1.96, p = 0.16). Results of a t-Test performed on the Score are presented in Table L. The table indicates that the t-Test was not significant (p = 0.54). Therefore there appears to be no significant difference between Factor V Scores obtained by reflective and impulsive groups.

TABLE L

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON FACTOR V SCORES

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
20.53	19.49	10.13	9.99	136	0.61	0.54

Factor VI Score

(Factor VI is defined by Hypothesis Selection and Evaluation of Selected Hypotheses Tasks.)

Stepwise Regression analysis performed on the Score indicated that no control variable contributed significantly to the variance of the Score. Science Knowledge was the major contributor (variance accounted for, 0.24%, F = 0.32, p = 0.57). Results of a t-Test performed on the Scores are presented in Table LI. The table indicates that the t-Test was not significant (p = 0.66). Therefore, there appears to be no significant difference between Factor VI Scores obtained by reflective and impulsive groups.

TABLE LI

t-TEST FOR REFLECTIVE AND IMPULSIVE GROUPS ON FACTOR VI SCORES

Reflective Group Mean	Impulsive Group Mean	Std. Devn. Ref. Group	Std. Devn. Imp. Group	Df	t	p
20.39	19.62	10.30	9.84	136	0.45	0.66

Results obtained relating to Associated Question 3 indicate that the reflective group performed superiorly to the Impulsives with respect to Factor Score I, II, III, V and VI. In no instances however, were differences obtained in Factor Scores statistically significant. The impulsive group achieved a higher, though statistically non-significant, Factor Score IV than did the reflective group. It appears therefore, that reflective and impulsive groups achieve similar Factor Scores when these Scores are derived from the noted common factors.

CHAPTER V

SUMMARY AND CONCLUSIONS

INTRODUCTION

In this chapter a summary of the investigation will first be presented. Following presentation of the summary, conclusions regarding decisions to reject or hold tenable the research hypotheses will be presented and results of the investigation discussed. In the final sections the implications that results have for teaching at the elementary school level will be discussed, and recommendations for further research presented.

SUMMARY

The present investigation was designed to compare performances of students classified as Reflective and Impulsive on response uncertainty problem-solving tasks demanding adoption of (a) a guided-discovery approach, and (b) an open-ended approach. Guided-discovery tasks required students to select observations and hypotheses pertaining to problem situations depicted on a film-loop, and to subsequently evaluate those selections made. Open-ended tasks required students to generate multiple observations and hypotheses pertaining to the same problems depicted on the loop, and to subsequently evaluate all generations made. The study also compared the types of questions asked by Reflectives and Impulsives during involvement in the same problem situations. Question-types were classified according to Suchman criteria.

The sample employed in the investigation was drawn from the population of grade six classrooms of schools within the Edmonton Public School System located in middle-income areas. From among ten schools randomly selected from this population, the administrators of eight agreed to have students participate in the investigation. From each of the eight schools involved, students from one complete grade six class participated in the study. By means of these procedures a total sample of 184 students participated. Of these, sixty-eight were subsequently classified according to Matching Familiar Figures Test criteria as being reflective in disposition. Seventy students were classified as being impulsive in disposition.

Since all guided-discovery and open-ended tasks were designed by the investigator, considerable effort was taken to ensure that reliabilities and validities were brought to as high a level as possible. To achieve this objective all guided-discovery and open-ended tasks were piloted on three separate occasions before the main testing program took place. Students participating in the pilot studies were not subsequently involved in the main study. The pilot studies enabled information concerning item difficulty and discrimination values of tasks involved to be collected, and KR_{20} and test-retest reliabilities to be determined. This information in turn enabled subsequent refinements of instrumentation to be made. As part of the refining program a panel of judges was employed to examine all tasks for face and content validity. Any task revisions called for were subsequently effected.

As mentioned earlier (see p. 45) problem episodes forming foci for all guided-discovery and open-ended tasks were presented by means of a film-loop. Due to the nature of the film-loop the content area providing focus for task involvement was restricted in scope. In view of this shortcoming three additional instruments, the Tannenbaum Observation Test, the Pearson Hypothesis Selection Test and the Pearson Hypothesis Formulation Test, were used to assess observation and hypothesizing abilities in different content areas and using different modes of content presentation from those of guided-discovery and open-ended Tasks. These three instruments provided information concerning the extent to which Selection and Generation observation and hypothesizing skills manifested task-specificity.

Another group of instruments used in the investigation consisted of the Stanford Achievement Test (Word Meaning, Paragraph Meaning and Science Batteries) and the Lorge-Thorndike I.Q. Test (Non-Verbal and Verbal Batteries). Several researchers (noted in Chapter II) reported significant correlations between the two R/I variables Number of Errors and Response Latency, with Intellectual ability, Reading ability and Knowledge of Science content. Accordingly it was considered essential that scores from these tests be available for use as covariates in subsequent statistical attempts to remove selection bias between reflective and impulsive groups due to those variables.

All students participating in the study were tested during three separate sessions. During the first session, the MFF Test was administered on an individual basis. During the second session all guided-discovery and open-ended tasks, and the practice and main Suchman Question-Asking Tests were given on a group basis. In the final session, the Tannenbaum Observation Test, the Pearson Hypothesis Selection and Hypothesis Formulation Tests, and the Science Knowledge Test, were given on a group basis. Each session was conducted by the investigator. To avoid problems of readability statements and instructions relating to each task were read aloud by the investigator.

School Cumulative Records supplied recent scores obtained on the Lorge Thorndike Non-Verbal and Verbal I.Q. Batteries, and on the Stanford Achievement Word Meaning and Paragraph Meaning Batteries.

Scores resulting from the guided-discovery and open-ended tasks were subjected in turn to Stepwise Regression Analysis to determine which, if any, of the variables I.Q., Reading ability, Science Knowledge, and Age, contributed significantly to the variance of the measure under investigation. In instances where one or more of the variables made a significant contribution, subsequent covariate analysis was performed on reflective and impulsive group means using the variable(s) that contributed significantly as covariate(s). In instances where no variable made significant contribution to the variance, t-Tests were subsequently performed between reflective and impulsive group means.

Scores associated with the Suchman Question-Asking Test were analysed using a variety of statistical procedures. Covariate analysis was performed using the variable Science Knowledge as covariate to compare the total number of questions asked between reflective and impulsive groups (Stepwise Regression analysis had previously indicated that this variable made significant contribution to the variance of the measure). Comparison of the distribution of the twelve question types obtained between reflective and impulsive groups was made using the Kolmogorov-Smirnov Two Sample Test. Comparisons of the number of Verification and Implication question types asked between Reflectives and Impulsives, when these questions were the first, second, third and fourth respectively to be asked, were made using a χ^2-Test of Independence. In instances where cell frequer were very small (i.e. <5) values for χ^2 were calculated using Y. Correction for Continuity.

The first Associated Question concerning existing correlations between guided-discovery and open-ended Task scores was answered by computing Pearson Product Moment Correlations between measures for both reflective and impulsive groups. Correlations between the two groups were then compared, and a Fisher's Z transformation executed to determine whether differences between correlations were significant.

The second Associated Question con rning the existence of common factors among guided-discovery and open-ended variables, was answered by subjecting associated scores to factor-analysis and rotating the resultant correlation matrix orthogonally according to Varimax criteria.

The third Associated Question, concerning differences between reflective and impulsive groups on derived Factor scores, was answered by adopting statistical procedures similar to those employed to test hypotheses dealing with guided-discovery and open-ended task performances. Thus scores were first subjected to Stepwise Regression analysis. In cases where any of the variables I.Q., Reading ability, Science Knowledge and Age made significant contributions to variances, subsequent covariate analyses were performed on reflective and impulsive group means using significant contributors as covariates. In cases where no significant contributions were made, t-Tests were performed on reflective and impulsive group means.

CONCLUSIONS AND DISCUSSION

A. Research Hypotheses One to Twelve

earch Hypotheses One to Six dealt with the ability of reflective and impulsive groups to select objects, events and hypotheses and to evaluate selections made. Table LII summarizes the decisions the investigator made concerning whether or not to reject these hypotheses. Research Hypotheses Seven to Twelve dealt with the ability of reflective and impulsive groups to generate objects, events and hypotheses, and to evaluate generations made. Table LIII summarizes the decisions the investigator made concerning whether or not to reject these particular hypotheses. As can be seen from Table LII, the investigation did not find any significant differences between Reflectives' and Impulsives' abilities in selecting objects, events and hypotheses. Thus this study indicates that Reflectives and Impulsives perform about as well on tasks demanding the selection of these skill- Both in terms of the type of approach adopted (i.e. a selection approach) and in terms of the

TABLE LII

SUMMARY OF DECISIONS MADE CONCERNING HYPOTHESES ONE TO SIX

Hypothesis		Decisions	
		Tenable	Reject
$H_{1.1}$:	There is no significant difference in ability to select objects (as measured by the Object Selection Task) between reflective and impulsive groups.	+	
$H_{1.2}$:	There is no significant difference in ability to select objects (as measured by the Tannenbaum Observation Test) between reflective and impulsive groups.	+	
H_2:	There is no significant difference in ability to select events (as measured by the Event Selection Task) between reflective and impulsive groups.	+	
$H_{3.1}$:	There is no significant difference in ability to select hypotheses (as measured by the Hypotheses Selection Task) between reflective and impulsive groups.	+	
$H_{3.2}$:	There is no significant difference in ability to select hypotheses (as measured by the P[illegible]rson Hypothesis Selection Test) between reflective and impulsive groups.	+	
H_4:	There is no significant difference in ability to evaluate selected objects (as measured by the Evaluation of Selected Objects Task) between reflective and impulsive groups.	+	
H_5:	There is no significant difference in ability to evaluate selected events (as measured by the Evaluation of Selected Events Task) between reflective and impulsive groups.	+	
H_6:	There is no significant difference in ability to evaluate selected hypotheses (as measured by the Evaluation of Selected Hypotheses Task) between reflective and impulsive groups.	+	

TABLE LIII

SUMMARY OF DECISIONS CONCERNING HYPOTHESES SEVEN TO TWELVE

Hypothesis		Decisions: Tenable	Decisions: Reject
H_7:	There is no significant difference in ability to generate objects (as measured by the Object Generation Task) between reflective and impulsive groups.	+	
H_8:	There is no significant difference in ability to generate events (as measured by the Event Generation Task) between reflective and impulsive groups.	+	
$H_{9.1}$:	There is no significant difference in ability to generate hypotheses (as measured by the Hypothesis Generation Task) between reflective and impulsive groups.	+	
$H_{9.2}$:	There is no significant difference in ability to generate hypotheses (as measured by the Pearson Hypothesis Formulation Test) between reflective and impulsive groups.		+
H_{10}:	There is no significant difference in ability to evaluate generated objects (as measured by the Evaluation of Generated Objects Task) between reflective and impulsive groups.	+	
H_{11}:	There is no significant difference in ability to evaluate generated events (as measured by the Evaluation of Generated Events Task) between reflective and impulsive groups.	+	
H_{12}:	There is no significant difference in ability to evaluate generated hypotheses (as measured by the Evaluation of Generated Hypotheses Task) between reflective and impulsive groups.	+	

particular problem-solving skills involved in these tasks, these results at first appear surprising for the following reasons:

1. Studies reviewed in Chapter II indicated that Reflectives performed superiorly to Impulsives on all tasks that demanded selection of responses from those provided. However it must be noted that all of these studies focussed attention exclusively on performance within specific content domains (subject areas). No study, for example, investigated the abilities of Reflectives and Impulsives to utilize problem-solving skills, as in the case of tasks employed in the present investigation. The fact that a selection approach appears equally effective to the performance of Reflectives and Impulsives in problem-solving settings investigated in this study, highlights the danger of generalizing results obtained from studies involving selection of responses within content domains to those involving selection of responses within problem-solving skill domains.

2. Findings from Siegelman's (1969) study (see p. 35) indicated that Reflectives use more efficient visual search skills than Impulsives, when engaged in solving MFF tasks. To Siegelman, differences in ability in using these skills explain, at least in part, the resultant typically slow and frequently correct responses of reflective students, and the fast, often inaccurate responses of the Impulsives. Results from the present investigation however, fail to support Siegelman's findings. Reflective and impulsive groups performed about as well in selecting objects and events. These results suggest that Reflectives and Impulsives employ equally efficient visual search strategies in the tasks associated with this

investigation. It is possible that differences in search skills are only manifested between Reflectives and Impulsives in instances where tasks demand distinction of very fine detail between variants (as for example in the case of MFF task requirements). Where coarser distinctions between variants are required (as in the case of tasks associated with this study) reflective and impulsive search strategies appear equally efficient. If this proposition is correct, efforts to increase the complexity of detail of variants may serve to accentuate any differences existing in visual search skills of reflective and impulsive students. In concluding the remarks concerning performances on the Selection Tasks, one of the following situations may have been responsible for the similar performances of both groups on these Tasks:

(a) Students of both groups were equally adept in utilizing the Selection approach and equally able in utilizing observation and hypothesizing skills.

(b) Students of one group were more adept in utilizing the Selection approach and less able in utilizing observation and hypothesizing skills, than students from the other group.

In view of the above discussion it would appear that equal perform on the Selection Tasks were due to conditions associated with Situation (a) rather than to those associated with Situation (b). However this is an area that warrants a great deal of further investigation.

As can be seen from Table LIII the investigation failed to find significant differences between Reflectives' and Impulsives' abilities in generating objects and even[illegible] appears that Reflectives and Impulsives perform about as well in generating objects and events. With respect to ability to generate hypotheses, Reflectives performed as well as Impulsives on the Hypothesis Generation Task. Reflectives' performance on the Pearson Hypothesis Formulation Test however, was significantly superior to that of Impulsives. It is interesting to note that in the Pearson Hypothesis Formulation Test, students were required to generate hypotheses in the form of alternative [illegible]wn in on paper. Written generations were not called for. [illegible]s Test demanded the generation of visually-constructed hypotheses, as responses to a visually-depicted problem situation. Requirements of the Hypothesis Generation Task were different. Written responses alone were required. It is probable that the noted differences in performance of Reflectives and Impulsives on each of the generation of hypotheses measures were due to differences in content area, presentation mode and response requirements of each. Both the non-significant correlation existing between these measures and their failure to load heavily on common factors had indicated that they were measuring essentially different qualities. This investigation suggests that performances of Reflectives and Impulsives are significantly different when generating hypotheses in highly visual, non-verbal environments. This proposition is supported by the significantly superior performances of Reflectives over Impulsives achieved on MFF Tasks (see Kagan and Kogan, 1970), tasks which also operate in highly visual, non-verbal environments.

Tables LII and LIII indicate that significant differences did not exist between Reflectives' and Impulsives' abilities in evaluating selected and generated objects, events and hypotheses. Thus this investigation indicates that Reflectives and Impulsives perform about as well on tasks demanding the evaluation of these particular skills. In terms of Siegelman's (1969) study these results are unexpected, since that investigation indicated that Reflectives used more efficient evaluation skills than Impulsives when solving MFF Tasks. No evidence supporting this finding was found in the present investigation. It is possible that differences in these evaluation skills between Reflectives and Impulsives only became apparent in instances where fine evaluations are demanded (as in the case of the MFF Tasks). In situations where considerably less subtle evaluations are required (as in the case of the Evaluation of Selection and of Generation Tasks in which choices have to be made between two given alternatives only) Reflectives' and Impulsives' abilities to evaluate appear equal. It is also possible that differences in evaluation skills are manifested between Reflectives and Impulsives only in instances where evaluation demands are made exclusively within visual environments. In the case of Siegelman's (1969) study, for example, students were required to evaluate visually-depicted phenomena in terms of visually perceived similarities and differences in form. References to non-visual evaluation criteria were not demanded during completion of MFF tasks. In the case of Evaluation of Objects and Events Tasks used in the present investigation, evaluation demands were not solely within the visual field. While students were again required to evaluate visual phenomena, evaluation procedures required relating these phenomena to a depicted problem situation that could be considered somewhat remote from the immediately perceived visual field, since total understanding of the problem situation was not possible merely from visual analysis alone. Thus it may be argued that evaluation tasks that require students to evaluate visually perceived phenomena, in terms of a problem not totally understood by visual processes alone, are demanding more cognitive effort than in evaluation tasks of the MFF type. In the case of tasks involving the evaluation of hypotheses, the above comments may also be appropriate. Because hypothesis evaluation criteria entailed more than mere visual considerations, tasks used in this study demanded more cognitive effort on the part of students than did MFF tasks. These differences in task demands may explain, in part, why differences in ability to evaluate hypotheses were not obtained between reflective and impulsive groups in this study. As in the case of the discussion relating to visual search skills, a great deal of further research is needed to determine whether or not ability to evaluate observations and hypotheses is dependent upon particular evaluation requirements entailed.

In discussion relating to Hypotheses One to Twelve difficulties were encountered in interpreting the performances of Reflectives and Impulsives on tasks involved in this study. These difficulties arose in large part from the lack of research findings in existence pertaining to the particular skills investigated. Research studies that have investigated the general question of Problem-skill transferability are also few in number. This paucity of available material can be explained by the fact that problem-solving skills have only recently become objects of research. The recent interest shown in these skills stems directly from the elevated position they now enjoy in several of the newer science programs (e.g. in SAPA, SCIS and ESS).

Gagné (AAAS, 1965) (see p. 7) claims that problem-solving skills, including those of observation and hypothesizing, manifest high inter-task generality. Some results obtained from the present investigation fail to support this proposition. For example, low, non-significant correlations were obtained between the two measures assessing ability in selecting objects, the Object Selection Task and the Tannenbaum Observation Test, and between the two measures assessing ability in selecting hypotheses, the Hypothesis Selection Task and the Pearson Hypothesis Selection Test. These results indicate that abilities to select objects and hypotheses within the content area and presentation mode pertaining to the film-loop 'The Knife,' are essentially unrelated to abilities to select objects and hypotheses in different content areas using different presentation modes (e.g. those associated with the Tannenbaum Observation Test and with the Pearson Hypothesis Selection Test). Results of factor analysis performed in this investigation further question Gagné's claim that observation and hypothesizing skills exhibit high inter-task generality. Neither the two object selection measures, nor the two hypothesis selection measures, loaded heavily on common factors, thereby indicating a low degree of commonality between respective measures. Reference has already been made to the marked differences in content areas, presentation modes and response requirements of these measures (see p. 170 and p. 171). It is probable that the more tasks resemble one another, in terms of these factors, the greater the li lihood that performance abilities on each will be similar. Evidence supporting this proposition (as well as supporting the transferability claim of Gagné) is supplied by results obtained from the Pearson Hypothesis Formulation Test and the Pearson Hypothesis Selection Test. While response requirements differ between these two measures, both the content and presentation modes involved are identical. It is evident therefore that these measures resemble one another more closely than do either the Object Selection Task and Tannenbaum Observation Test, or the Hypothesis Selection Task and Pearson Hypothesis Selection Test. In contrast to the low non-significant correlations manifested between these measures, the correlation obtained between the two Pearson Tests was moderately high in magnitude ($r = 0.55$, significant at the 0.01 level). Results of factor analysis performed, also indicated a high degree of commonality existing between these measures. Both loaded heavily on Factor III (see Table XLIII, p. 147). Another set of correlations between tasks offers further support for the two propositions noted above. A moderately high correlation of 0.47 (significant at the 0.01 level) was obtained between the two measures assessing ability in generating observations, the Object Generation Task and Event Generation Task. Results of factor analysis performed indicated a high degree of commonality between these Tasks and both loaded heavily on Factor I (see Table XLIII, p. 147).

B. Research Hypotheses Thirteen to Eighteen

These hypotheses relate to ability of reflective and impulsive groups to ask questions according to the format prescribed by Suchman (1962). Table LIV summarizes the decisions the investigator made concerning whether or not to reject these hypotheses. As can be seen from this table, Hypotheses 13 and 14 did not find significant differences between reflective and impulsive groups in terms of total number of questions asked, and in terms of the quantities of the twelve distinct Suchman-type questions asked. It appears that reflective and impulsive groups ask a similar quantity of questions of specific type. Studies by Ault (1973) and Denney (1973) (see p. 37) revealed that Reflectives asked predominantly 'constraint-seeking' questions during problem-solving. These questions indicate evidence of attempts to analyse properties of the pictures provided. Impulsives, on the other hand, ask predominantly 'hypothesis-seeking' questions indicative of consideration of these pictures as wholes, rather than as combinations of parts. The present study did not support these findings. In this investigation the six Suchman question types (Types 1-6), that typify analytic strategies attempting to identify objects, events and general conditions present in the problem situation, were asked with about equal frequency by both reflective and impulsive groups. In addition the six question types (types 7-12), that typify less analytic, more global strategies in attempting to discern relationships between variables entailed, were also asked with about the same frequency by both groups. According to Suchman (1962) the relative frequencies of question types employed reveals

TABLE LIV

SUMMARY OF DECISIONS MADE CONCERNING HYPOTHESES THIRTEEN TO EIGHTEEN

Hypothesis	Decision: Tenable	Decision: Reject
H_{13}: There is no significant difference in the total number of Suchman type questions asked, between reflective and impulsive groups.	+	
H_{14}: There is no significant difference in the distribution of specific Suchman question types asked between reflective and impulsive groups.	+	
H_{15}: With respect to subjects asking one Suchman type question only, there is no significant difference in the number of questions of Verification and Implication type asked between reflective and impulsive groups.	+	
H_{16}: With respect to subjects asking two Suchman type questions, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first and second questions asked, between reflective and impulsive groups.	+	
H_{17}: With respect to subjects asking three Suchman type questions, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first, second and third questions asked, between reflective and impulsive groups.	+	
H_{18}: With respect to subjects asking four Suchman type questions, there is no significant difference in the number of questions of Verification and Implication type pertaining to the first, second, third and fourth questions asked, between reflective and impulsive groups.	+	

the strategies and goals of the problem-solver. Based on this assumption it would appear from the present investigation that Reflectives and Impulsives adopt similar strategies and goals towards problem-solving. It should be noted however, that Suchman's interpretation of problem-solving strategy, in terms of manifested question type, is open to query. It is conceivable that the predominant presence of one particular question form does not, as Suchman contends, necessarily indicate inability or unwillingness to ask questions of a different type. Problem-solvers, for example who ask questions of a global character may have previously analyzed the problem-situation (without public sign of this) and now are engaged in seeking to relate, and to generalize from such analyses made. Thus while Suchman's format, as used in the present investigation, did indicate similar externally manifested strategies on the part of reflective and impulsive groups, caution must be adopted in accepting these findings as indicative of similar general problem-solving strategies between these groups.

Table LIV indicates that significant differences were not found in the number of Verification and Implication questions asked by reflective and impulsive groups, when these questions were the first, second, third and fourth to be asked. Although, as the number of questions asked increased from one to four, the number of Verification types employed increased and the number of Implication types decreased, these changes were manifested by both reflective and impulsive groups. Thus it appears that students who ask several questions, adopt (according to Suchman's claim) essentially analytic strategies, whereas those askin fewer questions tend to adopt both analytic and global strategies to problem-solving. The need to ask several analytic type questions may arise from the fact that each analytic question results in the supply of one specific bit of information only. It is possible that fewer global questions are asked since they result in the supply of a considerably larger amount of information by the relating of several factors within the problem sequence entailed.

C. Associated Questions

Three associated questions were investigated. The first question asked whether or not correlations among all Selection, Generation and Evaluation Tasks were of similar magnitudes for reflective and impulsive groups. Correlations between the two groups were compared by means of a Fisher's Z transformation. The following comparisons were significant at $p < 0.05$: Pearson Hypothesis Selection Test and Evaluation of Selected Objects Task, Pearson Hypothesis Selection Test and Event Generation Task, Evaluation of Selected Hypotheses Task and Event Generation Task, and finally, Evaluation of Selected Hypothe ask and Evaluation of Generated Events Task. As mentioned in Chap IV, it is possible that the significant differences between groups on these variables are artifacts, since all other correlation comparisons investigated (N = 101) were non-significant. The existence of otherwise remarkable similar correlations between variables obtained by reflective and impulsive groups, illuminates what the research hypotheses have show amely that both groups perform as well on almost all tasks employed in this investigation.

The second question asked whether c mon fa existed among Selection, Generation and Evaluation Task s re . swer this question, scores from all tasks were subjectec ctor analysis. The resultant correlation matrix was rotated according to Varimax criteria. Six factors were identified with Eigenvalues greater than unity. If Gagné's proposition that problem-solving skills manifest high inter-task generality is correct, one would expect that measures assessing ability in similar skills would load heavily on common factors. Support for Gagné's proposition however, was not provided in the case of the two selection of hypotheses measures, the Hypothesis Select n Task and the Pearson Hypothesis Selection Test, nor was it provided in the case of the two selection of objects measures, the Tannenbaum Observation Test and the Object Selection Task. As mentioned earlier, the degree of transferability between measures assessing similar problem-solving abilities, is probably dependent upon content area, presentation mode and response requirement factors. More research is needed in this important and scantily-investigated area. In every instance the six Evaluation Tasks loaded heavily with respective Selection and Generation Tasks on common factors. These results would first suggest that Reflectives' and Impulsives' ability to evaluate Selections and Generations made, is highly related to their ability in respective Selection and Generation Tasks. However the heavy loadings obtained may be spurious since the high correlations existing between all Evaluation Tasks and respective Selection and Generation Tasks have previously been questioned, for reasons earlier noted (see p. 57). Thus considerable caution must be exer-d in interpreting these findings.

The third question asked whether significant differences existed between reflective and impulsive groups with respect to factor scores derived from the six common factors. In no instance was a significant difference in factor score obtained between groups. However, in the case of Factor III scores (defined by the Pearson Hypothesis Selection and Pearson Hypothesis Formulation Tests) the differences between adjusted means was significant at the 0.08 level. As noted earlier in Chapter IV, unadjusted reflective means on both Pearson Tests were significantly higher than those of impulsives (see Table VI, p. 89). In the case of the Pearson Hypothesis Test the difference remained at a significant level when adjusted means were compared. In the opinion of the investigator these results do indicate a trend in favour of superior performance of Reflectives over Impulsives on these types of hypothesizing measures.

IMPLICATIONS FOR TEACHING

1. As noted in Chapter II, the accumulation of evidence indicating the general desirability of a reflective disposition over an impulsive one stimulated several attempts to modify impulsive behaviour. In the present investigation the impulsive group performed about as well as the reflective group on all Selection, Generation and Evaluation tasks (except on the two Pearson Hypothesis Tests). In addition, the impulsive group performed as well as the reflective group on the Science Knowledge test (see Table VI, p. 89). It

appears from these findings that the modification of impulsive behaviour would fail to benefit performance on tasks of a similar nature to those involved in this investigation. The present study does not, therefore, support attempts to modify impulsive (or reflective) behaviour in science situations similar to those investigated. Current calls for the general modification of impulsive behaviour appear highly premature. The setting up of problem situations in which selection and generation of objects, events and hypotheses is demanded, and the subsequent requirement that these selections and generations be evaluated (according to criteria similar to those used in this study) would appear to benefit the reflective and impulsive child equally.

2. In this investigation, Reflectives' unadjusted Pearson Hypothesis Formulation and Pearson Hypothesis Selection means were significantly higher than those of Impulsives ($p = 0.01$, $p = 0.03$ respectively). When adjusted means were compared, Reflectives' performance remained significantly superior to that of Impulsives', on the Pearson Hypothesis Formulation Test ($p = 0.04$), but did not remain at a significant level in the case of the Pearson Hypothesis Selection Test ($p = 0.16$). These results suggest that Reflectives perform better than Impulsives on tasks that demand selection and generation of hypotheses in predominantly visual environments. Several investigators (see p. 26) claim that hypothesizing is a highly complex skill demanding the utilization of both logical and non-logical (creative) processes. If it is accepted that the Pearson Hypothesizing Tests in turn demand some degree of creative effort,

then results obtained on the two Tests suggest that Reflectives may perform better than Impulsives on creativity tasks and exercises that are primarily visual in orientation. Liberal provision of opportunity for Impulsives to utilize such creative episodes may serve to reduce any performance differences between reflective and impulsive groups.

3. Reflectives and Impulsives ask similar Suchman types of question during problem-solving episodes. According to Suchman (1962) this would indicate that the problem-solving strategies and goals of these students are similar. Since both groups accordingly appear to adopt equally effective problem-solving strategies, a need for modifying question-asking strategies of one particular group does not appear to exist. Calls for production of specific question types, for example, would be unlikely to influence question-asking strategies of one group, more than of the other. Thus the Suchman Question-asking format seems of limited use for purposes of investigating behaviours of reflective and impulsive students.

4. Gagné's proposition (AAAS, 1965) that problem-solving skills exhibit inter-task generality was only partially supported in the present investigation. It appears that the closer tasks measuring similar skills resemble one another (in terms of content area, presentation mode and response requirement) the greater is the likelihood of a high degree of inter-task generality being manifested. With the exception of the two Pearson Tests, essentially no relationship existed between Selection and Generation Tasks. It appears that abilities to observe and to hypothesize using these particular response formats are unrelated. No attempt should be made to generalize between

selected and generated forms of observation and hypothesizing tasks. Low, non-significant correlations generally exist between Selection of Objects, Events and Hypotheses, and between Generation of Objects, Events and Hypotheses. Only in the case of the Generation of Objects and Generation of Events Tasks were substantial correlations obtained ($r = 0.47$, $p = 0.01$). These results indicate that, with the exception of abilities to generate objects and events, abilities in each of the other skills are essentially unrelated. Thus knowledge of ability in one particular skill provides no information useful in predicting ability in another.

RECOMMENDATIONS FOR FURTHER RESEARCH

During the course of this study it became apparent to the investigator that a number of areas related to the study warranted further investigation.

1. In this study all students were classified as being of a reflective or impulsive disposition, or of neither a marked reflective nor impulsive disposition. No attempt was made to classify students in terms of 'moderate' or 'extreme' reflective and impulsive disposition. If differences exist between reflective and impulsive students in ability to select and generate observations and hypotheses, and in ability to evaluate selections and generations made, then comparison of performances of students of extreme reflective and impulsive disposition may serve to accentuate those differences.

2. This investigation was limited to the grade six level. Use of tasks similar to those employed in this study with students

from other grade levels would enable a more complete picture of reflective and impulsive behaviour to be built. Investigation of performance of Reflectives and Impulsives at the lower elementary grade levels appears a particularly attractive proposition, since extremes in impulsive behaviours are more frequently met, with young children, than with old (Kagan and Kogan, 1970). It is probable that differences in performance between Reflectives and Impulsives are greatest at these levels.

3. Impulsives perform as well as Reflectives on all observation tasks involved in the study. Since quite coarse visual search skills were sufficient to discern between variants incorporated in these tasks, little or no opportunity was provided for exercising finer search skills. Use of more powerful observation tasks that require finer discriminations to be made between variants for successful completion, would serve to indicate whether differences in using more subtle visual search skills occur between reflective and impulsive groups.

4. Results from this study indicate that Reflectives perform superiorly to Impulsives on the Pearson Hypothesis Tests. If, as Martin (1971) claims, hypothesizing ability is dependent in part upon creative ability, then Reflectives (as noted earlier) may perform superiorly to Impulsives on tasks exploiting creative potential. The measurement of Reflectives' and Impulsives' abilities on published creativity tests (particularly those that are highly visual in form) would indicate whether or not postulated performance differences between groups do indeed exist. In addition information

gathered may provide further insight into the general question of hypothesizing/creative ability overlap.

5. As noted earlier in Chapter V (see p. 170) Reflectives performed as well as Impulsives on all Evaluation tasks. These tasks required students to evaluate selected and generated objects, events and hypotheses in terms of the extent to which each plays a role in effecting the depicted problem. The use of other tasks, which require Reflectives and Impulsives to evaluate generated and selected responses according to criteria different to those used in the present investigation, should provide further insight into the general issue of evaluation of selected and generated responses. Effective operation of evaluation tasks used in this investigation required the allocation of previously selected and generated responses to one of two provided evaluation categories. Although trichotomous evaluation keys were shown inappropriate for use at the grade six level (see p. 56) their use with older children would make possible more subtle response evaluations. These modifications, in turn, might serve to accentuate any differences existing in evaluation abilities of Reflectives and Impulsives.

6. Results of this investigation indicate that abilities in selecting, and generating objects, events and hypotheses, depend in large part, upon the particular content area and presentation mode employed. Assessment of abilities in similar selection and generation skills, using different science content domains and different presentation modalities, would provide further information concerning the extent to which these factors influence ability in these skills.

BIBLIOGRAPHY

BIBLIOGRAPHY

AAAS Commission on Science Education. The psychological bases of science—a process approach. Miscellaneous Publication, 1965.

Atkin, J. M. An analysis of the development of elementary school children in certain selected aspects of problem-solving ability. A study of formulating and suggesting tests for hypotheses in science learning experiences. Unpublished doctoral thesis, University of Indiana, Indiana, 1956.

Ault, R. L. Problem solving strategies of reflective-impulsive, fast-accurate and slow-inaccurate children. Child Development, 1973, 44, 259-266.

Barratt, E. S. Factor analysis of some psychometric measures of impulsiveness and anxiety. Psychological Reports, 1965, 16, 547-554.

Buros, O. The sixth mental measurements yearbook. New Jersey: Gryphon Press, 1965.

________. The seventh mental measurements yearbook. New Jersey: Gryphon Press, 1972.

Cathcart, W. G. and Liedtke, W. Reflectiveness/impulsiveness and mathematics achievement. Arithmetic Teacher, 1969, 16, 563-567.

Cattell, R. and Sheier, I. Handbook for the IPAT Anxiety Scale Questionnaire. Illinois: Institute of Personality, 1963.

Chamberlain, T. C. The method of multiple working hypotheses. Science, 1890, 15, 92-96.

Debus, R. L. Effects of brief observations of model behaviour on conceptual tempo of impulsive children. Developmental Psychology, 1970, 2, 22-32.

Denney, D. R. Reflection and impulsivity as determinants of conceptual strategy. Child Development, 1973, 44, 614-623.

Dewey, J. How we think. New York: Heath, 1953.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, New Jersey: Prentice-Hall, 1965.

Egeland, B. Training impulsive children in the use of more effective scanning strategies. Child Development, 1974, 45, 165-171.

Ferguson, G. A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1971.

Ford, G. W., and Pugno, L. The structure of knowledge and the curriculum. Chicago: Rand McNally, 1964.

Gordon, W. J. J. Synetics: The development of creative capacity. New York: Harper and Row, 1961.

Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw Hill, 1956.

Gupta, P. K. Correlations of reflection-impulsivity. Unpublished doctoral thesis, University of Alberta, Edmonton, 1970.

Hess, R. D. and Shipman, V. C. Early experience and the socialization of cognitive modes in children. Child Development, 1965, 36, 869-886.

Hungerford, H. R. and Miles, D. T. A test to measure observation and comparison skills in science. Science Education, 1969, 53, 61-65.

Hunt, D. E. A conceptual level matching model for co-ordinating learner characteristics with educational characteristics. Interchange, 1970, 1(3), 68-80.

Inhelder, B. and Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, 1958.

Jacknicke, K. D A comparison of teacher and student outcomes of science— process approach and an alternative program in selected grade two classrooms. Unpublished doctoral thesis, University of Colorado, Colorado, 1974.

Kagan, J. Impulsive and reflective children: The significance of conceptual tempo. In J. D. Krumboltz (Ed.), Educational process. Chicago: Rand McNally, 1965a.

________. Reflection-impulsivity and reading ability in primary grade children. Child Development, 1965b, 36, 609-628.

________. Individual differences in the resolution of response uncertainty. Journal of Personality and Social Psychology, 1965c, 2(2), 156-160.

________. Developmental studies in reflection and analysis. In A. L. Kidd and J. L. Rivoire (Eds.), Perceptual development in children. New York: International University Press, 1966a.

________. Reflection-impulsivity: The generality and dynamics of conceptual tempo. Journal of Abnormal Psychology, 1966b, 71(1) 17-24.

Kagan, J. and Kogan, N. Individual variations in cognitive processes. In P. H. Mussen (Ed.), Carmichael's Manual of child psychology (Vol. 1). New York: John Wiley, 1970.

Kagan, J. and Moss, H. A. Birth to maturity: A study in psychological development. New York: John Wiley, 1962.

Kagan, J., Pearson, L. and Welsh, L. The modifiability of an impulsive tempo. Journal of Educational Psychology, 1966, 57, 359-365.

Kagan, J., Rosman, B. L., Day, D., Albert, J. and Philips, W. Information processing in the child: Significance of analytic and reflective attitudes. Psychological Monographs, 1964, 78 (1, Whole No. 578).

Keislar, E. R. Teaching children to solve problems: A research goal. Journal of Research and Development in Education, 1969, 3, 3-13.

Kneller, G. F. The art and science of creativity. New York: Holt, Rinehart and Winston, 1965.

Kogan, N. Educational implications of cognitive styles. In G. S. Lesser (Ed.), Psychology and educational practice. Glenview: Scott, Foresman, 1971.

Martin, M. Concepts of science education. Glenview: Scott, Foresman, 1972.

Messer, S. The effects of anxiety over intellectual performance on reflection-impulsivity in children. Child Development, 1970, 41, 723-735.

Messick, S. The criterion problem in the evaluation of instruction: Assessing possible not just probable intended outcomes. In M. C. Wittrock and D. E. Wiley (Eds.), The evaluation of instruction. New York: Holt, Rinehart and Winston, 1970.

Michenbaum, D. and Goodman, J. Reflection-impulsivity and verbal control of behaviour. Child Development, 1969, 40, 785-797.

Moore, W. E. Creative and critical thinking. New York: Houghton Mifflin, 1967.

Mosher, F. A. and Hornsby, J. R. On asking questions. In J. Bruner, R. R. Oliver, P. M. Greenfield et al (Eds.), Studies in cognitive growth. New York: Wiley, 1966.

Mosston, M. Teaching from command to discovery. Belmont: Wadsworth, 1972.

National Society for the Study of Education. Science education in American schools. 46th Yearbook, part 1. Chicago: University of Chicago, 1960.

Nay, N. A. A process approach to teaching science. Science Education, 1971, 55, 197-207.

Osborn, A. F. Applied imagination: Principles and procedures of creative thinking (3rd rev. ed.). New York: Scribner, 1963.

Pearson, D. A. The hypothesizing ability of elementary school students. Unpublished master's thesis, University of Alberta, Edmonton, 1975.

Quinn, M. E. Hypothesis formation can be taught. The Science Teacher, 1972, 3, 30-31.

Quinn, M. E. and Goerge, K. D. Teaching hypothesis formation. Science Education, 1975, 59, 289-296.

Renner, J. W. and Ragan, W. B. Teaching science in the elementary school. New York: Harper and Row, 1968.

Ridberg, E. H., Parke, R. D. and Hetherington, E. H. Modification of impulsive and reflective cognitive styles through observation of film-mediated models. Developmental Psychology, 1971, 5, 369-377.

Rowe, M. B. Teaching science as continuous inquiry. New York: McGraw Hill, 1973.

Schaefer, E. A. and Bayley, N. Maternal behaviour, child behaviour and their intercorrelation from infancy through adolescence. Monographs of the Society for Research in Child Development, 1963, 28(3).

Schwab, J. J. The teaching of science as enquiry. Cambridge: Harvard University Press, 1966.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw Hill, 1956.

Siegelman, E. Reflective and impulsive observing behaviour. Child Development, 1969, 40, 1213-1222.

Souch, S. G. A cross sectional study of reflection-impulsivity with special reference to sex, social class and maternal conceptual systems. Unpublished doctoral thesis, University of Alberta, Edmonton, 1970.

Stanley, J. C. and Hopkins, K. D. Educational and psychological measurement and evaluation. Englewood Cliffs, New Jersey: Prentice Hall, 1972.

Suchman, J. R. The elementary school training program in scientific inquiry, Title VII, Project 216. Urbana, University of Illinois, 1962.

Tannenbaum, R. S. The development of the Test of Science Processes. Journal of Research in Science Teaching, 1971, 8, 123-236.

Ward, W. C. Reflection-impulsivity in kindergarten children. Child Development, 1969, 39, 867-874.

West, M. L. The relationship of the reflection-impulsivity dimension and the quality of intellectual achievement. Unpublished doctoral thesis, University of Alberta, Edmonton, 1973.

Williams, F. E. Foundations of creative problem-solving. Ann Arbor: Edward Brothers, 1960.

Winer, B. Statistical principles in experimental design. New York: McGraw Hill, 1967.

Witkins, H. A. Some implications of research on cognitive styles for problems of education. In M. G. Gottsegen and G. B. Gottsegen (Eds.), Professional school psychology. New York: Grune and Stratton, 1969.

Woodburn, J. and bourn, E. S. Teaching the purs of science. New York Macm lan, 1965.

Yando, R. M. Kag, J. The effects of teacher tempo on the child. Child Deve, 1968, 39, 27-34.

Zelnicker, T., Jeffrey, W. E., Ault, R. L. and Parsons, J. Analysis and modification of search strategies of impulsives and reflectives on Matching Familiar Figures Test. Child Development, 1972, 43, 321-335.

Zelnicker, T. and Oppenheimer, D. Modification of information processing of impulsive children. Child Development, 1973, 44, 445-450.

APPENDICES

APPENDIX A

DATA SHEET FOR THE OBJECT SELECTION TASK AND ASSOCIATED EVALUATION OF SELECTED OBJECTS TASK

NAME ____________________

The following statements are about OBJECTS seen in the film and about OBJECTS not seen in the film. Some statements are completely TRUE in every way. Others are not TRUE in every way. Mark off as I read ONLY those statements that are true in every way.

On the left below some of the objects you marked as present played important parts in making the blade bend upwards. Others did not. Below we have two columns labelled NO EFFECT and SOME EFFECT. For each object you marked PRESENT decide whether it had NO EFFECT or SOME EFFECT in making the blade bend upwards. Put a mark in each correct space. Do not change answers on the left.

	MARK HERE ↓	NO EFFECT	SOME EFFECT
1 Shadow of the man			
2 Piece of tubing			
3 Knife handle made of two halves—one half fixed to one side of the blade, the other fixed to the other side.			
4 White coloured flame			
5 Man with a moustache			
6 Red shiny cloth folded over once			
7 Shiny knife blade			
8 When in an upright position the height of the knife blade was less than the height of the liquid in the container			

DATA SHEET FOR THE OBJECT SELECTION TASK AND ASSOCIATED EVALUATION OF SELECTED OBJECTS TASK

APPENDIX B

DATA SHEET FOR THE EVENT SELECTION TASK AND ASSOCIATED EVALUATION OF SELECTED EVENTS TASK

NAME ____________________

The following statements are about EVENTS seen in the film and about EVENTS not seen in the film. Some statements are COMPLETELY true in every way. Others are not true in every way. Mark off as I read ONLY those statements that are true in every way.

For each event you marked down as present decide whether it had NO EFFECT or SOME EFFECT in making the blade bend upwards. Put your marks in the correct spaces.

	MARK HERE ↓	NO EFFECT	SOME EFFECT
1 The blade was heated on 3 separate times			
2 The blade was wiped on 2 separate times			
3 When removed from the liquid the blade continued to straighten a little			
4 Only one side of the blade was ever heated			
5 The blade straightened out on 2 separate times only			
6 The man always used his right hand to lift the knife			
7 The man always held the cloth using his left hand			
8 On one occasion the man turned over the knife blade			

DATA SHEET FOR THE EVENT SELECTION TASK AND ASSOCIATED EVALUATION OF SELECTED EVENTS TASK

APPENDIX C

DATA SHEETS FOR THE THREE HYPOTHESIS SELECTION SUBTASKS
AND ASSOCIATED EVALUATION OF SELECTED
HYPOTHESES SUBTASK

NAME ____________________

In the film the straight knife blade bent downwards. Read carefully each of the statements below and mark each one that you think gives a REASON why the blade bent down. Each REASON should tell us ONE OR MORE OF THE THINGS that probably were happening to make the blade bend downwards.

Each statement you marked on the left below you thought gave a REASON why the blade bent downwards. For each reason decide whether it gives an A-type or a B-type REASON. An A-type REASON tells us about only ONE of the things that probably was happening to make the blade bend. A B-type REASON tells us at least two things that were probably happening to make the blade bend. For each statement you marked present on the left decide whether it is an A-type r B-type REASON.

The blade bent downwards probably because:

	MARK HERE ↓	A-type REASON	B-type REASON
1 When the blade was heated the top side increased in length more than the bottom side			
2 The handle of the blade was made of material different from that of the blade itself			
3 The blade melted a little			
4 Heat softened the blade and then gravity pulled the blade down			
5 When heated one side of the blade shortened in length more than the other side			
6 One of the sides of the blade shortened in length more than the other side			

DATA SHEET FOR THE HYPOTHESIS SELECTION SUBTASK 1 AND ASSOCIATED EVALUATION OF SELECTED HYPOTHESES SUBTASK 1

NAME ______________________

In the film the knife blade bent upwards. Read the following statements carefully and mark off any statement that you think gives a REASON why the blade bent upwards. Each reason should tell us about at least ONE of the things that was probably happening to make the blade bend upwards.

Each statement to the left you thought gave a REASON why the blade bent upwards. For each statement you marked decide whether it is an A-type REASON or a B-type REASON. An A-type REASON tells us about only ONE of the things that probably was happening while a B-type REASON tells us at least two things that were probably happening to make the blade bend upwards.

The blade bent upwards probably because:

	MARK HERE ↓	A-type REASON	B-type REASON
1 Only the bottom side of the blade was heated			
2 Heat softened the blade			
3 The blade was very long			
4 Just before it bent upwards the blade had been turned over			
5 The bottom side of the knife increased in length more than the top side			
6 When heated both sides of the blade changed in length by the same amount			

DATA SHEET FOR THE HYPOTHESIS SELECTION SUBTASK 2 AND ASSOCIATED EVALUATION OF SELECTED HYPOTHESES SUBTASK 2

NAME ______________________

Soon after bending downwards the blade straightened out. After bending upwards the blade also straightened out. Mark off any statements below that you think give a REASON why the blade probably straightened out.	Each statement you marked to the left you thought gave a REASON why the blade straightened out. For each reason you had marked of decide whether it is an A-type reason or a B-type reason.

The blade straightened out probably because:	MARK HERE ↓	A-type REASON	B-type REASON
1 The blade weakened and then gravity pulled the blade straight			
2 The length of the blade remained the same			
3 The cool liquid kept the length of the blade the same			
4 One side of the blade shortened in length more than the other side			

DATA SHEET FOR THE HYPOTHESIS SELECTION SUBTASK 3 AND ASSOCIATED EVALUATION OF SELECTED HYPOTHESES SUBTASK 3

APPENDIX D

DATA SHEET FOR THE OBJECT GENERATION TASK AND ASSOCIATED EVALUATION OF GENERATED OBJECTS TASK

NAME ______________________

In the space below write down every OBJECT no matter how small that you remember seeing in the film. An object is something that you can remember seeing in the film. A good object is KNIFE HANDLE because this is an object that we saw in the film. If we wrote down THE MAN HELD THE KNIFE HANDLE it is true that the man did hold the handle but that statement would be an EVENT or something that happened and is not an object. Write down only OBJECTS. One object has been written in for you. Use a new line for each object that you think of.

Some of the objects you wrote down on the left played an important part in making the blade bend upwards. Other objects did not play important parts. Down below there are two columns labelled NO EFFECT or SOME EFFECT in making the blade bend upwards. KNIFE HANDLE had no effect in making the blade bend up so we put a mark in the NO EFFECT space. Put a mark in the correct space for each object. Do not change answers on the left.

	NO EFFECT	SOME EFFECT
KNIFE HANDLE	X	
1		
2		
3		
4		
5		
6		
7		
8		
9		

DATA SHEET FOR THE OBJECT GENERATION TASK AND ASSOCIATED EVALUATION OF GENERATED OBJECTS TASK

APPENDIX E

DATA SHEET FOR THE EVENT GENERATION TASK AND ASSOCIATED EVALUATION OF GENERATED EVENTS TASK

NAME ______________________

In the spaces below write down all the different EVENTS that you can remember seeing in the film. An event is something you saw happening. A good event would be THE MAN HELD THE KNIFE HANDLE because this was an event that happened in the film. If we wrote down KNIFE HANDLE this is an OBJECT and is NOT an EVENT. Write down only events. One event has already been written in for you. Use a new line for each event you think of.

For each event you wrote down on the left decide whether it had NO EFFECT or SOME EFFECT in making the blade bend upwards. THE MAN HELD THE KNIFE HANDLE had no effect in making the blade bend so we put a mark in the NO EFFECT column. Put a mark in the correct spaces for each event. DO NOT change answers on the left.

	NO EFFECT	SOME EFFECT
THE MAN HELD THE KNIFE HANDLE	X	
1		
2		
3		
4		
5		
6		
7		
8		
9		

DATA SHEET FOR THE EVENT GENERATION TASK AND ASSOCIATED EVALUATION OF GENERATED EVENTS TASK

APPENDIX F

DATA SHEETS FOR THE THREE HYPOTHESIS GENERATION SUBTASKS AND ASSOCIATED EVALUATION OF GENERATED HYPOTHESES SUBTASKS

NAME ____________________

In this part of the film the straight knife blade bent downwards. In the spaces below write down as many different REASONS that you can think of that might explain why the blade bent downwards. Each reason should tell us about AT LEAST ONE of the things that was probably happening to make the blade bend downwards. One reason has been written in for you.

Each reason you gave on the left as to why the blade bent downwards may be either an A-type REASON or a B-type REASON. An A-type REASON tells us only about ONE of the things that probably was happe ng to make the blade bend down. A B-type REASON tells us at least TWO things that were probably happening to make the blade bend down. THE BLADE WAS HEATED only tells us about ONE thing that happened so we put a mark in the A-type REASON space. For each reason you gave put a mark in the correct space.

The knife blade bent downwards probably because:	A-type REASON	B-type REASON
THE BLADE WAS HEATED	X	
1		
2		
3		
4		
5		
6		
7		
8		

DATA SHEET FOR THE HYPOTHESIS GENERATION SUBTASK 1 AND ASSOCIATED EVALUATION OF GENERATED HYPOTHESES SUBTASK 1

NAME ______________________

In this part of the film the knife blade bent upwards. In the spaces below write down as many different reasons as you can think of, that might explain why the blade bent upwards. Each REASON should tell us about at least ONE of the things that probably was happening to make the blade bend upwards. One reason has been written in for you.

The knife blade bent upwards probably because:

For each REASON you gave as to why the blade bent upwards decide whether it is an A-type REASON or a B-type REASON. Remember an A-type REASON tells us only one of the things that was probably happening, whereas a B-type REASON tells us TWO or more things that were probably happening to make the blade bend upwards. Put a mark in the correct spaces for each reason.

	A-type REASON	B-type REASON
THE BLADE WAS HEATED	X	
1		
2		
3		
4		
5		
6		
7		
8		
9		

DATA SHEET FOR THE HYPOTHESIS GENERATION SUBTASK 2 AND ASSOCIATED EVALUATION OF GENERATED HYPOTHESES SUBTASK 2

NAME ____________________

Soon after bending downwards the knife blade straightened out. Soon after bending upwards the blade also straightened out. In the spaces below write down as many REASONS as you can think of that might explain why the blade straightened out. Each REASON should tell us about at least ONE of the things that probably happened to make the blade straighten out.

For each reason you gave as to why the blade straightened out decide whether it is an A-type REASON or a B-type REASON. Put a mark in the correct space.

The blade straightened out probably because:	A-type REASON	B-type REASON
1		
2		
3		
4		
5		
6		
7		
8		
9		

DATA SHEET FOR THE HYPOTHESIS GENERATION SUBTASK 3 AND ASSOCIATED EVALUATION OF GENERATED HYPOTHESES SUBTASK 3

APPENDIX G

ITEM DIFFICULTY INDEX (p) VALUES AND ITEM DISCRIMINATION INDEX (D) VALUES ON GUIDED-DISCOVERY TASKS AND GUIDED-DISCOVERY EVALUATION TASKS (FINAL EDITION)

p AND D VALUES ON GUIDED-DISCOVERY TASKS (FINAL EDITION)

Guided-Discovery Task	Item	p value	D value
Selection of Objects	1	.41	.57
	2	.48	.58
	3	.61	.69
	4	.59	.79
	5	.52	.59
	6	.62	.73
	7	.60	.65
	8	.62	.66
Selection of Events	1	.83	.92
	2	.81	.82
	3	.51	.61
	4	.57	.63
	5	.81	.96
	6	.70	.91
	7	.70	.77
	8	.64	.62
Selection of Hypotheses (1)	1	.37	.68
	2	.28	.74
	3	.54	.70
	4	.59	.73
	5	.67	.77
	6	.76	.79
Selection of Hypotheses (2)	1	.46	.57
	2	.20	.66
	3	.60	.73
	4	.30	.86
	5	.31	.72
	6	.51	.77
Selection of Hypotheses (3)	1	.39	.82
	2	.51	.83
	3	.25	.88
	4	.28	.96

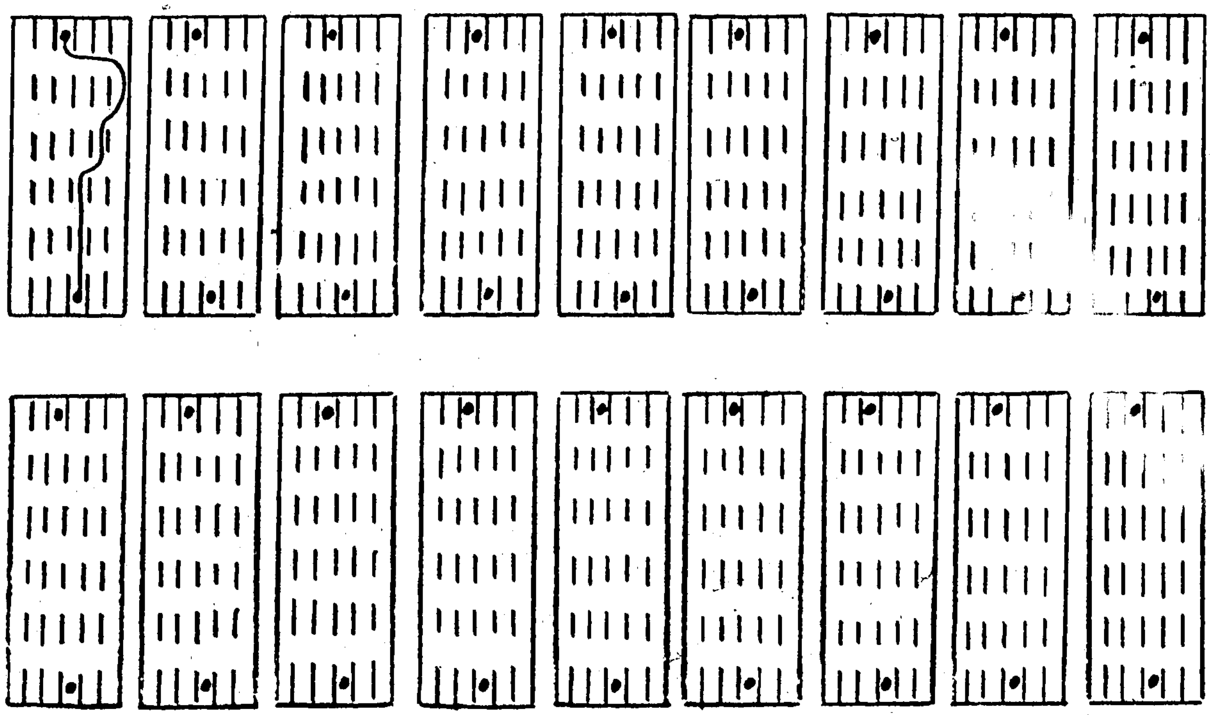
p AND D VALUES ON GUIDED-DISCOVERY EVALUATION TASKS
(FINAL EDITION)

Guided-Discovery Evaluation Task	Item	p value	D value
Evaluation of Selected Objects	1	.42	.60
	2	.25	.73
	3	.46	.65
	4	.49	.72
	5	.46	.77
	6	.44	.65
	7	.40	.59
Evaluation of Selected Events	1	.55	.75
	2	.62	.85
	3	.65	.85
	4	.64	.90
	5	.36	.61
Evaluation of Selected Hypotheses	1	.46	.81
	2	.41	.61
	3	.49	.60
	4	.36	.88
	5	.25	.82
	6	.37	.63
	7	.29	.64

APPENDIX H

DATA SHEET FOR THE PEARSON HYPOTHESIS FORMULATION TEST

NAME ______________________________



(1/2 of Natural Size)

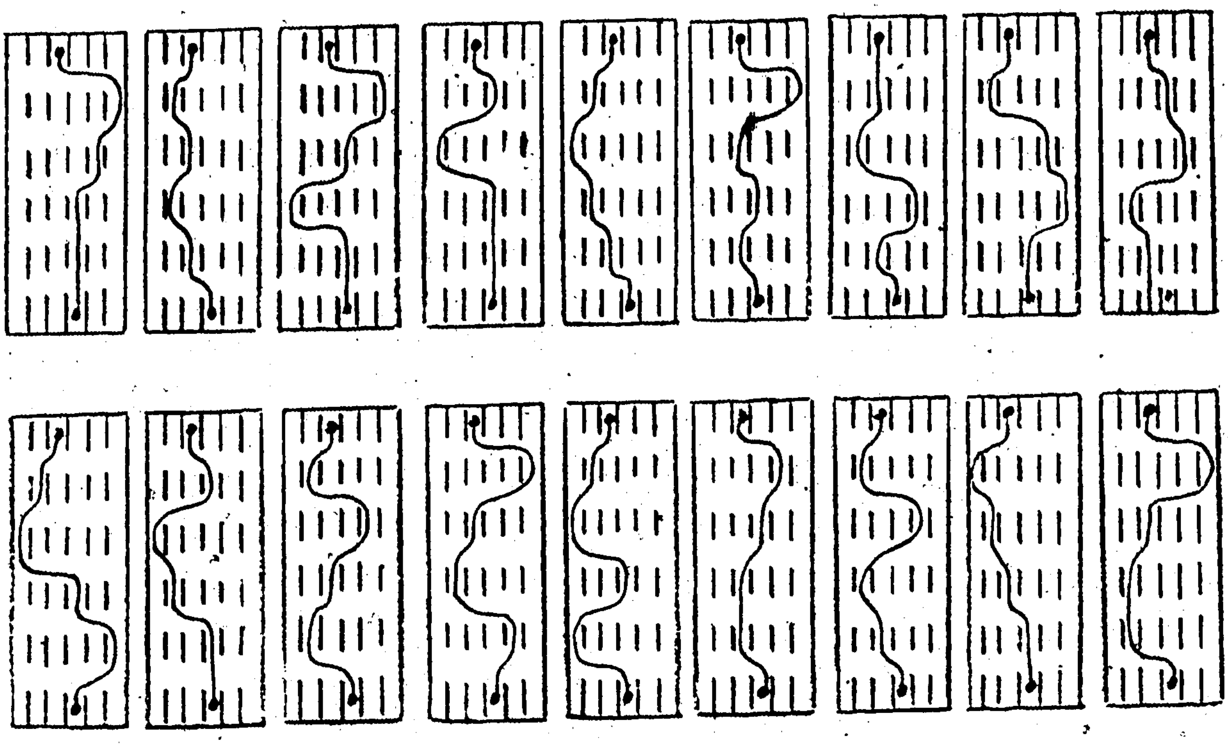
On the sheet you see little pictures just like the large drawing I have here on the board. Each picture is exactly the same as another. In each picture you can see a ball at the top and a ball at the bottom. In a minute I want you to draw in as many routes as you can think of, for getting the ball from the top position here to the bottom position here. Remember you can only move the ball one channel to the right or left at any of the gaps, or three channels to the right or left at any of the gaps. If you make a mistake just put a mark through the picture and go on to the next picture. Are there any questions? You may begin.

DATA SHEET FOR THE PEARSON HYPOTHESIS FORMULATION TEST

APPENDIX J

DATA SHEET FOR THE PEARSON HYPOTHESIS SELECTION TEST

NAME ______________________



(1/2 of Natural Size)

On the sheet you can see lots of little pictures showing a ball at the top and a ball at the bottom. This time in each picture routes have been drawn in, for getting the ball from the top position to the bottom position. Some of the routes drawn in are CORRECT—they get the ball from top to bottom moving only one channel to the right or to the left at a gap, or three channels to the right or left. Other routes are INCORRECT—they don't get the ball correctly from top to bottom. Sometimes they move the ball two or four channels to the right or left at a gap. Put a mark against each picture you think gives a correct way for gettinc the ball from top to bottom. Are there any questions? You may begin.

DATA SHEET FOR THE PEARSON HYPOTHESIS SELECTION TEST

APPENDIX K

DATA SHEET FOR THE PRACTICE QUESTION-ASKING TEST
PERTAINING TO THE FILM-LOOP 'BOILING COFFEE'

NAME ______________________________

In the film, we saw that soon after the coffee boiled, the man drank the coffee. Maybe you are not sure how anyone can drink coffee so easily after it has boiled. To understand better how the man was able to do this so easily, you probably need to find out more about some of the things you saw happening in the film. You can do this by writing questions to me about anything you saw in the film. There is just one important thing you must remember when you write each question. I must be able to answer each question with either a YES or a NO. Here are some examples of good questions that I can answer with a YES or a NO:

'Had the coffee really boiled?'

'Did the man really drink the coffee?'

These questions are good because I can answer them with a YES or a NO. Here is an example of a poor question because I cannot answer it with a YES or a NO:

'What was the thing placed over the coffee?'

In the space below write down as many YES-NO questions as you wish about anything in the film that might help you understand how the man was able to drink the coffee after it had just boiled.

____

____

____

____

____

____

____

____

DATA SHEET FOR THE PRACTICE QUESTION-ASKING TEST

APPENDIX L

DATA SHEET FOR THE MAIN QUESTION-ASKING TEST PERTAINING TO THE FILM-LOOP 'THE KNIFE'

NAME ______________________________

In the second part of the film the knife blade bent upwards. Maybe you are not quite sure why this happened. To understand better why the blade bent upwards you maybe need to find out more about some of the things you saw happening in the film. You can do this by writing YES-NO questions to me about anything you saw in the film. Make sure I can answer each of your questions with a YES or a NO. Two examples of YES-NO questions have already been written in.

WAS THE KNIFE BLADE MADE OF STEEL?
WOULD THE BLADE HAVE BENT IF NO HEAT HAD BEEN USED?
1
2
3
4
5
6
7
8
9

DATA SHEET FOR THE MAIN QUESTION-ASKING TEST

APPENDIX M

EXAMPLES OF SUCHMAN QUESTION-TYPES PERTAINING TO THE FILM-LOOP 'THE KNIFE'

EXAMPLES OF SUCHMAN QUESTION-TYPES PERTAINING
TO THE FILM-LOOP 'THE KNIFE'

Categorical verification: nominal: 'Was that blade a knife?'

Categorical verification: normative: 'Was that an ordinary knife?'

Analytical verification: condition-descriptive: 'Was the blade hot when it was bent?'

Analytical verification: condition-comparative: 'Was the blade longer after it was heated than before?'

Analytical verification: structural-component: 'Was the blade half one metal and half another?'

Analytical verification: properties check: 'Can the burner melt iron?'

Abstract-conceptual: diffuse: 'Did the liquid have anything to do with the blade bending?'

Abstract-conceptual: directed: 'Did the liquid reduce the temperature of the blade?'

Concrete-inferential: elimination: 'If the liquid was not used at all would the blade still straighten out?'

Concrete-inferential: substitution: 'If the blade was held at an angle in the liquid instead of in an upright position, would the blade still straighten out?'

Concrete-inferential: addition: 'If you put oil on the blade would it still bend upwards?'

Concrete-conceptual: 'Was it necessary to wipe the blade in order to make it bend upwards?'