THE CONSTRUCTION OF A GROUP TEST OF COGNITIVE PROCESSES FOR USE IN EDUCATION

by

Douglas Vivian Bromley

Submitted in partial fulfillment of the requirements for the degree of

Master of Science

in the

Department of Psychology

University of Natal

1980

Durban

1980

DECLARATION

The whole thesis, unless specifically indicated to the contrary in the text, is the candidate's own original work.

ACKNOWLEDGEMENTS

I would like to thank:

- Mr A.P. Moerdyk, my supervisor, for his assistance,
- Mr C.O. Murray who helped with the experimental design and processed the results,
- The Director of Education, Natal Education Department, for study leave and for allowing me access to subjects for testing,
- The Principals of the schools where testing was carried out,
- My colleagues, Mr D.R. Goedeke, who gave valuable advice regarding Physical Science aspects and Mr D.J.H. Smith for many worthwhile comments,
- Mrs C. Dunscombe for her careful typing and the interest she showed in this work.

I would also like to acknowledge the assistance of the following who, during the research, sent and made helpful comments on tests:

- Prof. A.E. Lawson, University of California, for copies of the *Biology reasoning test* and the *Longeot examination*.
- Prof. E.A. Peel, University of Birmingham, for revised versions of the *Predilection for generalising and abstracting test*.
- Prof. R.B. Sund, University of Northern Colorado, for helpful comments on the *Logical reasoning test*.
- Prof. R.P. Tisher, University of Monash, for the *Understanding in science test* and for directing my attention to the *Logical reasoning test*.

I also wish to acknowledge the contribution which my wife and children made to this research. Without their support it would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF DIAGRAMS

DIAGRAM

LIST OF GRAPHS

GRAPH

LIST OF PLATES

ABBREVIATIONS

ASEP: Australian Science Education Project

CI: Concept Identification

df: degrees of freedom

IQ: Intelligence Quotient

MA: Mental Age

SD: Standard Deviation

WISC: Wechsler Intelligence Scale for Children

I

PREFACE

During the past fifteen years, psychology has been infused with a vigorous new interest in the mentalistic problems which sparked the birth of scientific psychology in the nineteenth century. (Paris, 1975, p. 11)

There is little doubt that the mind is back in style. (Neimark and Santa, 1975, p. 173)

Another consequence for intelligence testing of the split between experimental and applied psychology is that, relative to the total amount of research on intelligence tests and their predictive power, very little consideration has been devoted to the psychological processes involved in attaining the correct answers (or in failing to attain them). (Butcher, 1968, pp. 73-74)

... I would entirely agree that there is room for other supplementary tests to tell us more about children's cognitive styles and strategies, and specific learning disabilities, if someone would invent them. (Vernon, 1979, p. 11)

There is at present but one theoretical description of the nature and organisation of adolescent thought: Piaget's elegant and comprehensive treatment of formal operations (Inhelder and Piaget, 1958). (Neimark, 1975, p. 542)

With the increasing interest in application of Piagetian theory for curriculum design and evaluation, the use of Piagetian formal operational tasks to quantify gains in intellectual development has and will continue to become more widespread .... (Lawson, Nordland and De Vito, 1974, p. 267)

We know, however, that the study of the child and the adolescent can help us understand the further development of the individual as an adult.... (Piaget, 1972, p. 12)

II

ABSTRACT

While there is increasing emphasis in education on the learning of intellectual processes, relatively little attention has been given to the rigorous assessment of these processes.

An attempt was made to construct a group test which measured both specific processes of thinking as well as the general level of thinking attained at adolescence. Test items were modelled on Piagetian tasks as described by Elkind (1961b), Lawson and Renner (1974) and Shayer et al. (1976).

The test as a whole was considered to have acceptable face and content validity. Most items, as well as the test as a whole, showed low, but acceptable construct validity for a research instrument.

The reliability of the test in its present form was unacceptably low.

Further development of the test is discussed as well as the implications which were raised for education.

III

INTRODUCTION

1. PROCESSES IN EDUCATION

Education today stresses the learning of intellectual processes (Nay, 1971). A curriculum illustrating the stress laid on processes is *Science - a process approach* (1968) produced for the Elementary School by the American Association for the Advancement of Science. This curriculum promotes the learning of the processes of science,e.g. observing, measuring, controlling variables, etc. It was considered by the course designers that the learning of processes was more important than the learning of content and the achieving of correct factual answers. The primary reason for the stress on processes is that because of their generality processes promise to transfer more readily to a wide variety of areas. This is in contrast to content which is more likely to be subject specific.

However despite the stress on processes there has not been a concommitant increase in rigorous testing and the production of valid and reliable measures of processes. Most educational tests are more concerned with the measurement of products than with the measurement of processes (Butcher, 1968; Vernon, 1979); e.g. if ability to solve verbal analogy problems is being tested the stress is usually on deriving the correct answer (the product) rather than recording in some way the means by which the answer was obtained (the process).

This is not to say that processes have been entirely neglected.

- Many teacher-tests and examinations have looked at the processes raised by Bloom's taxonomy of skills in the cognitive domain (Bloom et al., 1956), e.g. the ability to apply knowledge (Crossley, 1979).

- Essay questions have always been used to test the process of sustained verbal reasoning.

- Individual intelligence tests e.g. the Wechsler Intelligence Scale for Children (WISC) and the Stanford-Binet Intelligence Scale have allowed the experienced psychologist to make inferences regarding the level of certain of the testee's processes (Butcher, 1968).

However while these methods may have merit in individual situations, there would be much benefit in producing a valid and reliable test which was independent of the idiosyncrasies of specific situations.

It is acknowledged that although it is useful to use the product - process distinction, there is an obvious interdependence between product and process (Philp and Kelly, 1974). Generally however products are more easily observable than processes.

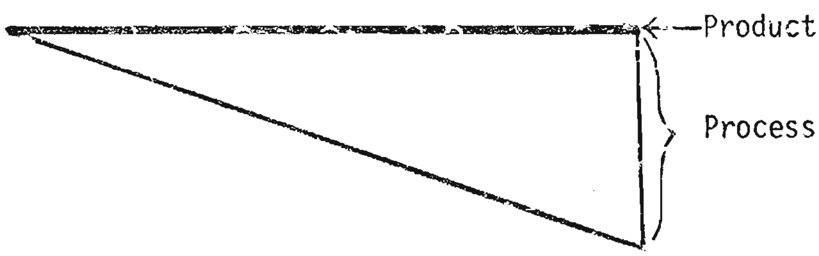
One way to make processes more evident is to ask the subject to describe his thinking. Unfortunately this introspective method fell into disrepute with the advent of behaviourism and Butcher (1968) and Whimbey (1975) suggest that this is one of the major reasons for the current neglect of cognitive processes.

2. A MODEL OF THE RELATIONSHIP BETWEEN PRODUCT AND PROCESS

DIAGRAM 1 below attempts to clarify the relationship between product and process test items. All test questions require a product-type answer represented by the heavy line. The thought processes lying behind a product-type answer may be infinitely complex. The shaded area represents the quantity of process *made observable* when an answer is produced.

DIAGRAM 1

MODEL INDICATING RELATIONSHIP BETWEEN PRODUCT AND OBSERVABLE PROCESS IN A TEST ITEM



Product answers which require fairly automatic responses e.g. 2 X 2 = ? give an immediately observable product but reveal little of the underlying processes and would be situated at the left of the product line. If error analysis (Behr, 1975) is carried out on the answers the product would move towards the right. Product answers accompanied by explanations which reveal much regarding the thought processes involved will fall at the extreme right of the product line. Most items from published educational tests would fall at the left hand side of the product line. However there are some tests where processes of various types have been rendered observable, e.g. the *Frostig developmental test of visual perception.* (Frostig, Lefever and Whittlesey, 1966) is used to measure the component processes of visual perception e.g. the ability to differentiate figure from ground.

Another process - oriented test is the *Illinois test of psycholinguistic abilities* (Kirk, McCarthy and Kirk, 1968) which measures higher cognitive processes, e g. verbal expression - the ability to express meaning through the use of spoken language. This is in contrast to the *Frostig developmental test of visual perception* which is concerned with lower level perceptual processes.

There are however few tests concerned with logical processes at adolescence. Some which have been constructed are discussed in V, pp. 36-40.

It was decided that there was value in working in this area and attempting to construct a valid and reliable test to measure cognitive processes at adolescence - with the knowledge that what is applicable to adolescent thought is to a large measure applicable to adult thought (Piaget, 1972).

3. THEORETICAL BASIS FOR THE RESEARCH

Before a serious attempt can be made to construct a test at adolescent level a suitable theoretical basis must be chosen. Two broad approaches are possible.

- A number of different models can be used. Each model would be regarded as being particularly appropriate for the construction of items measuring specific processes.

- A wholistic theory, which attempts to deal with all intellectual processes, can be used. Where applicable reference can be made to one or more circumspect models as a means of giving the theory even more generality.

The latter approach was chosen.

The most detailed treatment of the processes of thinking and cognitive development is undoubtedly that of Piaget and his associated researchers. Modgil (1974) lists 112 publications emanating from Piaget and his students which span the years from 1923 - 1972.

Furthermore there is only one full description of the nature of adolescent thought - as found in Inhelder and Piaget's 1958 publication[1] (Neimark, 1975). Their description of the processes involved in the solving of science problems is so detailed that it even reaches the level of logical operations.

Although lacking in scope the concept identification (CI) model of Bruner et al. (1955) can be used to infer greater generality. Although this model has some short-comings (see Carroll, 1964; Stones, 1966; Paris, 1975), it is famous for its description of the various strategies which subjects employ when they are solving CI problems. It is considered appropriate to relate this model to certain of Piaget's formulations regarding adolescent thought.

Piaget's system offers the advantage that if the test is constructed within his system it can invoke his stage theory. This promises to compensate for any undue fractionation of the subject's thinking, the dangers of which in education are discussed by Mann and Phillips (1967). Because Piaget has a wholistic model of the formal stage it is possible to predict that because

---

[1] Although *The growth of logical thinking from childhood to adolescence* is co-authored by Inhelder and Piaget and although neo-Piagetians have introduced theoretical modifications the theory will, at times, be attributed to Piaget alone. This is done in the interests of brevity and readability.

a pupil has reached this stage he will display a number of interrelated processes of thinking. e.g. he will be able to practise sustained logical reasoning, generate new combinations of variables,etc. While the entry into the stage is not argued to be complete in every respect, knowledge that a pupil is in a particular stage gives evidence that a certain configuration of processes and concommitant emotional behaviour will probably be within his reach.

4. THE TEST ENVISAGED AND ITS PROPOSED USES

While it is possible for researchers to administer Piaget's tasks individually most class teachers have not the background or time to devise and administer the tasks and then to interpret the results. If a test is to be devised it must be of such a nature that it can be used to assess the cognitive development of large numbers of pupils simultaneously so that time and effort may be economically used.

Attempts to use Piaget's tasks in the construction of a valid and reliable group test look promising. Longeot (1962, 1965), Tisher (1971), Tisher and Dalc 1975a; 1975b),Burney (1976), Shayer et al. (1976) and Shayer and Wylam (1978) are researchers who, using Piaget's tasks, have attempted to devise group tests capable of differentiating pupils who have reached the stage of formal operations from those who are at other levels of cognitive development.

The proposed test could have a number of uses: e.g.

- A contemporary paradigm of great importance in education is the diagnostic remedial model (Bateman, 1971). In this approach, tests (or detailed observations) are used to specify as precisely as possible what processes are deficient. Using this information a remedial programme aimed at corre ting the deficiencies is implemented.

A number of diagnostic tests are accompanied with advice on remediation, e.g. the *Frostig developmental test of visual perception* (Frostig, Lefever and Whittlesey, 1966) can be followed up by the use of the *Pictures and patterns* programme (Frostig, Horne and Miller, 1972) which attempts to remediate deficient areas of visual perception. At a higher cognitive level researchers have had some success in remediating deficits[1] in the ability to control variables. e.g. Case and Fry (1973), Breddermann (1973), Lawson and Wollman (1976) and Wollman and Lawson (1977), showing that the diagnostic-remedial paradigm can be applied to formal intellectual processes. It is hoped that the planned test will be able to help in identifying deficient intellectual processes so that remediation may be implemented.

- The proposed test could be used to identify a pupil's general cognitive orientation. This would allow educators to match instructional attention to the pupil's level of development (Bart, 1972; Lawson and Renner, 1975; Sayre and Ball, 1975; Tisher and Dale, 1975a).

- If the stage at which the pupil is functioning is known instructional treatments may attempt to move the pupil to a higher level of functioning (Lawson and Blake, 1976).

---

1 The word "deficits" is used in a broad sense to include both older disadvantaged pupils who were clearly educationally retarded and younger pupils who did not show widespread educational retardation.

5. CONCLUSION

An attempt will be made to construct a valid and reliable process test based on the formulations of the Piagetian school and where possible related to the findings of CI research. It will give information on the processes and general orientation attained by adolescents so that educational treatments may be planned accordingly.

The following section will evaluate the formulations of Piaget and the CI theorists as a basis for the test.

IV

THEORETICAL BASIS FOR THE TEST

1. PIAGET'S SYSTEM OF GENETIC EPISTEMOLOGY

According to Piaget (1973):

> Genetic epistemology deals with the formation and meaning of knowledge and with the means by which the human mind goes from a lower level of knowledge to one that is judged to be higher. (p. xlii)

Genetic epistemology can therefore be argued to be a very suitable discipline to supply information regarding intellectual processes and the overall level of intellectual development attained.

However no systematic attempt will be made to describe Piaget's theories; this has already been carried out by a host of most able commentators e.g. Flavell (1963), Beard (1969), Ginsberg and Opper (1969) and Sund (1976). Only those features of the theory which can be utilized in the construction of the test will be discussed.

a. The processes of thought

It has already been noted that Piaget has produced the only detailed account of the nature of adolescent thought (Neimark, 1975). His account includes a general description of formal reasoning, details of intermediate level processes, e.g. the controlling of variables, and an analysis of the logical processes postulated. Thus his description provides information at various levels - allowing a test designer to pick the material he will use.

Piaget's most penetrating description of processes has been at the logical level. He actually believes the thinker "thinks in logic". He states (1973):

> The fundamental hypothesis of genetic epistemology is that there is a parallelism between the progress made in the logical and rational organisation of knowledge and the corresponding formative psychological processes. (p. xlii)

Thus it is not surprising that logic plays such an important part in his writings.

There are two basic components of Piaget's system of logic. One is the combinatorial system; the other is the INRC group of operations. Parsons (1958) speaks of the combinatorial system (also called "the structured whole" as follows:

> In sum, the structured whole, by virtue of which the subject is able both to combine parts into a whole and to separate them from it, might be impressionistically characterised as a sort of mental scaffolding held up by a number of girders joined to each other in such a way that an agile subject can always get from any point - vertically or horizontally - to any other without trapping himself in a dead end. (p.xx)

The INRC group (Identity, Negation, Reciprocity and Correlativity) are operations which, if the analogy is extended, could be regarded as skills which allow the subject to move around the system of girders.

The generality of Piaget's model promises that these processes are potentially present in all formal thinking - regardless of the content or area being considered. However Piaget's view that the thinker thinks in logic is not without problems.

If one considers that all thinkers think in a perfectly logical fashion - and especially in terms of the combinatorial system and the INRC group of operations - one is simply not taking into account the many alternative and highly idiosyncratic methods which people use when thinking (Paris, 1975). Furthermore there is a danger that one's understanding of cognitive processes may be hampered because one is trying to fit observable behaviour into a preconceived model. Thus Parsons (1960), a logician, considers that Piaget's model is too restrictive and cannot be applied to all thinking. He suggests that Piaget's logico-mathematical model can only readily be used in certain experiments "of a clear-cut and simple structure."(p. 82)

Logician Ennis (1975) has another type of criticism. He has actually worked through Piaget's logic and he claims that Piaget's logic contains faults. He writes:

> Since Piaget uses logic to judge the adequacy of children's thinking, as well as to attempt to describe the thinking, these inadequacies in the logic are a significant flaw. (p. 38)

Furthermore few researchers have actually worked through Piaget's data or attempted to replicate his experimental work. A notable exception is Bynum et al. (1972) who have published a re-analysis of the protocols of Inhelder and Piaget's (1958) experiment with "invisible magnetism".[1] They note that the only evidence for all sixteen binary propositions being used comes from Gou's protocol. Their re-analysis of the same protocol yielded only eight of the sixteen operations claimed. When the same basic research group (Weitz et al., 1973) carried out a replication experiment (with fifty-seven subjects) only five of the operations were noted.

---

[1] See Inhelder and Piaget (1958). Chapter 6. The role of invisible magnetization and the sixteen binary propositional operations. pp. 93-104.

It is acknowledged that Piaget's views are at times suspect. However his attempt to find a wholistic theory of thinking has, despite its flaws, laid a foundation which can be built upon; e.g. as done by Ennis (1978). It seems that while some of the processes which Piaget has identified may be investigated in a test, there is no need to assume that his system is faultless and that adolescent thinking is limited to one model, which does not allow for individual variations.

b. The Stage of formal operations

Under optimum conditions the child can enter the stage of formal operations at about 11 - 12. Piaget (1972) writes:

> ... from 11 - 12 years to 14 - 15 years a whole series of novelties highlights the arrival of a more complete logic that will attain a state of equilibrium once the child reaches adolescence at about 14 - 15 years. (p.

Flavell (1963) describes the stage of formal operations in these terms:

> We see, then, that formal thought is for Piaget not so much this or that specific behaviour as it is a generalised *orientation*, sometimes explicit and sometimes implicit, towards problem-solving: an orientation towards organising data (combinatorial analysis), towards isolation and control of variables, towards the hypothetical, and towards logical justification and proof. (p. 211)

Piaget's stage of formal operations is a stage when a set of related thinking processes are used. The advantage of having a test which will put pupils into a stage is that it is able to draw attention away from a highly analytical view of abilities towards a view where many related abilities are considered at the same time. If a group of related abilities are considered at the same time there is very likely to be some tie-up with emotional factors - which are of critical importance in most educational-learning situations.

The relationship of the affective dimension and the cognitive dimension has been noted by Piaget(1958), and by neo-Piagetians e.g. Blasi and Hoeffel (1974). Thus Piaget's view of a stage has the advantage that it allows the adolescent to be seen *in toto.*

The existence of the formal stage has been queried. However there is evidence that formal thinking exists and that it is different from concrete thinking from which it grows. Significant correlations have been found between tasks which are supposed to measure formal thought suggesting that the tasks do in fact rest on a common cognitive basis (Lovell, 1971; Lawson, Nordland and De Vito, 1975).

When formal tasks have been factor analysed it has been reported that formal operational thinking is unifactorial: e.g. Lovell and Butterworth (1966), Lovell and Shields (1967), Bart (1971). When Lawson and Renner (1974) and Lawson and Nordland (1976) working with a battery of concrete and formal tasks, carried out principal component analysis, they found evidence for two fairly distinct types of thought in these tasks, viz. concrete and formal, thus corroborating Piaget's views.

But while there may be evidence for two types of thought there is considerable evidence against a *sudden* transition from the stage of concrete operations to the stage of formal operations. Dale (1970) writing on the "chemicals problem"[1] remarks:-

> The investigation produced no evidence for sharp transition from concrete to formal thinking at age 11 - 12 years. It appears that there is a gradual, almost linear, increase in ability to solve this

1 See Inhelder and Piaget (1958). Chapter 7. Combinations of colored and colorless chemical bodies. pp. 107-122.

> particular problem and in the ability to construct and systematically test combinations. The latter ability begins at just below age 6 years and appears to be still increasing at age 15 years. Ability to completely solve the problem began at age 10 years approximately and is still increasing at 15 years . (p. 285)

Somerville (1974) in her exhaustive study of the"pendulum problem"[1]produced a total of nine sub-stages en route from the beginning of concrete thought to late formal thought. Furthermore each sub-stage has up to four sets of criteria for recognizing it (with alternative sets of criteria). An unbiased viewing of this scoring system would suggest that there is no sudden transition from concrete to formal thinking: in fact there is a continuous scale.

Wollman (1977) gives a sensible interpretation of the evidence:

> The distinction between Piaget's concrete and formal stages of intellectual development may not be as sharp as is suggested by his data .... It thus becomes possible to assess a student's performance, not as either concrete or formal, but rather as lying on a continuous scale extending from relatively concrete to relatively formal . (p. 385)

Perhaps the discrepancy between the above observations and Piaget's own arises because Piaget does not use enough carefully sampled subjects and does not standardise his interview technique. Piaget uses a method of interviewing which he calls the *méthode clinique* . Working with pupils individually he sets them problems with fairly commonplace materials. The child's answers

---

[1] See Inhelder and Piaget (1958). Chapter 4. The oscillation of a pendulum and the operations of exclusion. pp. 67-79.

are followed up with detailed questioning (Inhelder and Piaget, 1958). Piaget is probably more interested in the wrong answers than in the right answers, for it is these which show up the characteristics of the child's thought. The *méthode clinique* undoubtedly produces great insight into a few pupils' thinking but does not necessarily lead to valid generalisations.

Because of the lack of rigour in his research many of Piaget's views remain essentially hypotheses waiting to be tested by energetic experimentalists - possibly in simpler experimental situations, (as done by O'Brien and Shapiro, 1968 and Shapiro and O'Brien, 1970). However once this is acknowledged, Piaget's formulations are very respectable: they should however not be presented as a rigid system of well-nigh proven discoveries.

What can be accepted is that:

- there is evidence for at least two types of thinking , viz. concrete and formal thinking.

- a correlation exists between success in the different tasks which measure formal thinking.

- as children increase in age beyond adolescence there will be an overall increase in the proportion of instances in which children show formal thought. Criteria of $66^2/_3\%$ (Shayer et al., 1976) and 80% (Sayre and Ball, 1975) have been used in tests to allow for the fact that formal thinking is not always used - even by pupils supposed to be in the formal stage.

c. Genetic Epistemology and education

Another reason why Piaget's formulations offer a suitable basis for the test is because they allow for the influence of education, provided due note is taken of the role of maturation. Inhelder and Piaget (1958) comment:

> ... the maturation of the nervous system can do no more than determine the totality of possibilities and impossibilities at a given stage. A particular social environment remains indispensable for the realisation of these possibilities. It follows that their realisation can be accelerated or retarded as a function of cultural and educational conditions. (p. 337)

Elsewhere Piaget (1964) comments on the importance of a child having the structures characteristic of a certain level of cognitive development before he is taught particular subject-matter:

> The child can receive valuable information via language or via education directed by an adult only if he is in a state where he can understand this information. That is, to receive the information, he must have a structure which enables him to assimilate this information. This is why you cannot teach higher mathematics to a five year old. He does not have structures which enable him to understand. (p.180)

This implies that teachers should be able to assess the stage of cognitive development which their pupils have reached.

To some extent this is acceptable educational practice.

Tisher and Dale's (1975) *Understanding in science test* has been used in Australia to identify those pupils who need a more concrete apprach to their school science. The Australian Science Education Project (ASEP) has three types of materials matched for three levels of student cognitive development (Lucas, 1972). Elsewhere the following description is given:

> Stage 1 materials are suitable for students at Piaget's concrete stage of development when thinking is dependent on the presence of concrete objects and examples.
> Stage 2 materials are for students in transition from the concrete to the formal stage.
> Stage 3 materials are for students at Piaget's formal stage when there is freedom from dependence on concrete examples, and hypothetical situations can be considered (Tisher and Dale, 1975a p. 3).

However most school curricula, such as the ASEP, while giving some attention to matching academic demands and pupils' level of cognitive development, also attempt to move pupils on to use more advanced thinking.

While maturation and random experiences, which Piaget stresses (Brainerd, 1978), do have a part to play, educators also have a responsibility to plan experiences so that the pupils learn the most effective processes of thought which their maturing nervous systems allow them to use.

It must be pointed out that the relationship between learning and cognitive development is complex. Usually learning is viewed as applying to more specific information or processes which do not necessarily involve an increase in intellectual ability. On the other hand the term "cognitive development" *implies* a wide ranging permanent change in thinking usually accompanied by some physical maturation. Once cognitive development has taken place it is assumed that subsequent thought at the same level of difficulty will be facilitated.

However if *maturation* and *learning* take place at the same time then there seems to be no reason why the process cannot be regarded as cognitive development[1].

Thus it can be argued that pupils who are *maturing or have matured* sufficient-ly can *learn* to use both specific operations and adopt a general formal orientation. Despite Piaget's reluctance to encourage the teaching of formal operations (Gaudia, 1974) it does not seem that there is any serious incompatibility between Piaget's views and those of contemporary education.

However Gaudia (1974) claims that as far as education is concerned Piaget's system is accused by what it leaves out. He writes that a great system such as Piaget's which only gives cursory attention to education, is lacking. Gaudia writes:

> ... Piaget's own writings on that subject (education) probably make up less than one tenth of 1 per cent of his total literary output. (p.482)
>
> ....
>
> We get the impression of a reluctant pedagogue who would much rather concern himself with philosophy, logic, development, mathematics, and biology, and his reluctance is manifested in the imbalance between his educational and his noneducational writings. (p. 483)

---

[1] From current usage it seems that the word "intellectual" is used in the same way as "cognitive" and "growth" in the same way as "development". This gives four combinations which do not substantially differ in meaning, viz. intellectual growth, cognitive growth, intellectual development and cognitive development.

However few researchers try to build total systems and Piaget's is more inclusive than most. Educators have to look for the implications and use his theory to amplify their own models. The great value of Piaget lies in the fact that he has provided a theory which can be assimilated into educational practice. A number of curricula have used aspects of Piaget's system. ASEP has already been mentioned. Amongst others, Sund (1976) lists the Biological Science Curriculum Study course for the Middle School, Science 5/13 and the Early Childhood Curriculum of Celia Lavatelli. Many authors such as Sime (1973) and Mc Nally (1974) have attempted to show the relevance of Piaget's work for education.

In an interview with Evans (1973) Piaget had this to say:

> Oh, I am convinced that what we have found can be of use in the field of education, in going beyond learning theory, for instance, and suggesting other methods of learning. I think this is basic. But I am not a pedagogue myself, and I don't have any advice to give to educators. All we can do is provide some facts. (p. 51)

Piaget has certainly provided some facts. He has also provided useful theory which will be used in the construction of this educationally oriented test.

d. <u>Relationship of Piagetian measures with age, intelligence and academic Achievement</u>

It is most important that the relationship between Piagetian measures and measures of cognitive development accepted in education e.g. age, IQ and academic achievement are investigated.

If there is a suitably positive relationship between Piagetian measures and measures of age, intelligence and academic achievement test results it will be permissible to argue that Piaget's theory is relevant to education. Furthermore it can be argued that Piagetian measures used in previous research validly measure the trait of increased cognitive development with age, and the abstract constructs of intelligence (broadly defined) and ability to achieve academically.

i. Age

Piaget has repeatedly noted that there is a relationship, however flexible, between increasing age and increased cognitive development (Inhelder and Piaget, 1958). Thus if a large number of pupils are given Piagetian tasks at an appropriate level of difficulty there should be a significant relationship between increased age and increased scores on Piagetian measures. Evidence showing that the percentage of pupils attaining success on formal operational tasks increases with age comes from a large number of sources and has been tabulated by Blasi and Hoeffel (1974).

Blasi and Hoeffel give the percentage incidence of formal operational thinking for different tasks in the following age groups, -11, 11-14, 15-18 and 18+. Twenty-seven studies allow comparison of the incidence of formal operational thinking between groups. *All these studies show an increase in the percentage of formal operational thinking with age.*

Some of the differences between studies dealing with the same task (e.g. percentage formal thinkers in chemicals task with 11-14 year olds varies from 0% to more than 50%) may be due to the lack of standardisation in the presentation and evaluation of the tasks, some to cultural differences, some to differences in intelligence and some to differences in educational experiences.

A striking observation is that contrary to the impression given in Inhelder and Piaget (1958) not all adolescents - or adults - enter the stage of formal operations. Nevertheless the trend to increased success with age is present - a trend which suggests that Piaget's tasks are valid measures of cognitive development.

ii. Intelligence

Piaget uses the concept of intelligence. However his view of intelligence is somewhat different from a typical psychometric view of intelligence.

The most important difference is that Piaget gives definite descriptions of the underlying processes which may be expected at each stage of intellectual development. Following the publication of Inhelder and Piaget's (1958) work, Bruner (1959) praised Piaget's views thus:

> The psychologist for his part welcomes the qualitative character of logic, since it facilitates the analysis of the actual structures underlying intellectual operations as contrasted with the quantitative treatment of their behavioral outcome. Most "tests" of intelligence measure the latter, but our real problem is to discover the actual operational mechanism which govern such behaviour and not simply to measure it . (p. 363)

Another difference between Piagetian measures and intelligence tests is that intelligence tests are norm referenced, and as such are more sensitive to individual differences in ability. Piagetian measures however are criterion referenced as they give a definite indication regarding the type of behaviour expected at certain points in cognitive development.

However, despite the fact that Piaget's view of intelligence is somewhat different from a typical psychometric view there should still be some correlation between scores on typical intelligence tests and Piagetian measures. Piagetian tasks and most psychometric test items are both concerned with what could broadly be called problem-solving and reasoning.

Research does in fact indicate a moderate correlation between scores on Piagetian tasks and Intelligence Quotient (IQ) scores. Kohlberg and Gilligan (1971) note that these correlations are in the 50's . Elkind (1961b, p.558)studying quantity concepts, found a low but statistically significant correlation between IQ measures and attainment of an "abstract conception of volume". Shayer et al. (1976) found a correlation of between 0,55 and 0,63 between the cognitive level (i.e. overall classification) attributed to pupils and their non- verbal intelligence. Bart (1971, p. 76) notes that "the component indicating formal thought correlated modestly with the measure of verbal intelligence". i.e. r = 0,467.

Other researchers e.g. Mealings (1963), Goodnow and Bethon (1966) have worked with Mental Age (MA) instead of IQ. As in the case of IQ's correlations between Piagetian task measures and MA's have also been moderate. Furthermore the fact that Piagetian tasks have been included in the recently produced New British Individual Scale (Warburton, 1970) is evidence that Piagetian tasks measure at least in part what psychometric tests measure.

Overall it can be said that Piaget's qualitative description of intelligence is most useful. The statistical relationship between Piagetian measures and psychometric measures indicates that the constructs being measured by these two different methods are actually similar. Piaget's tasks can be said to validly measure intelligence.

iii. <u>Academic achievement</u>

Although Piaget has not commented on academic achievement a number of neo-Piagetians have researched the relationship between academic achievement and performance on Piagetian tasks.

Performance on Piagetian tasks has been found to show significant positive correlations with achievement when measured by school grades (Sayre and Ball, 1975), and when measured using standardised tests (Lawson, Nordland and D Vi 1975). Sayre and Ball (1975) have shown that Junior High School and Senior High School Pupils who were classed as formal thinkers received significant -ly higher science grades than non-formal students. Lawson et al. (1975) found significant correlations between a wide range of standardised achievement tests (e.g. College Entrance Examination Board Achievement Test in English, Mathematics and Science) AND Piagetian task measures.

Since there is a significant relationship between Piagetian measures and achievement, Piaget's theorising can be argued to offer a useful theoretical basis for the design of a test which will be measuring processes which play an important part in academic achievement. Piagetian tasks can also be claimed to validly measure the trait of "ability to achieve academically".

IDENTIFICATION (CI) MODEL

sks to be used in the test is the pendulum problem (Inhelder 958). This task will later be related to the CI model.

CI problem an array or a deck of cards is used. Each card has dimensionalised attributes e.g. colour, shape, number of figures borders. Each attribute is represented by one of a number of colour may be red, blue or green or number of borders may be hree. The subject is then required to do *one or both* of the ygood and Bourne, 1965):

the defining attribute values - usually two - which are ristic of the concept (which may be given a name such as "BIF"), IF may be any card which has e.g. TWO figures *AND* TWO borders

the rule indicating the relationship between the attributes. xample above the rule is that of conjunction - one attribute her attribute must be present.

select single cards from an array and state whether any are non-exemplars of the concept. This is known as the selection they may be given a card from a deck and asked to indicate card shown is an exemplar or a non-exemplar of the concept. d the reception paradigm. It is not intended to review all With Paris (1975) it is considered that the description of amongst the most useful research done in this field. Strategies ered as one type of process.

a. CI strategies

A number of strategies can be used to solve CI problems. Bruner et al. (1956) described four strategies, viz. conservative focussing, focus gambling, successive scanning and simultaneous scanning. While Butcher (1968), Laughlin (1973) and Neimark and Santa (1975) have pointed out that the difference between these strategies is not as great as was originally thought, there is merit in considering them as they represent one view of the processes involved in problem-solving.

In *conservative focussing* with the selection paradigm the subject chooses an instance so that only one attribute is varied at a time; but in doing this he must perform a number of matching operations to ensure that the rest of the attributes are kept constant. Greater inferential or logical demands are thus being made. In the reception paradigm the subject compares cards which have only one differing attribute. Conservative focussing is according to Bruner et al. (1956), the most successful strategy. It has low memory requirements but as mentioned has high inference strain. Further evidence that it is a more efficient strategy is provided by Laughlin (1973). Conservative focussing appears to be a special case of the control of variables - an important formal process.

In *focus gambling* two attributes are changed with each trial. As the name suggests the subject is gambling. If the gamble is successful the status of the card will not change thereby rapidly giving information on two attri -butes. If the status of the card does change the student has to redo the step he attempted to leave out. Overall he will have carried out the same number of choices as would have been required using conservative focussing. This strategy has not received the attention which the other strategies have received and Laughlin (1973) shows that in many experiments it cannot be differentiated from scanning.

Scanning is essentially hypothesis testing behaviour where the accent is not on controlling variables.

In *successive scanning* the student only considers his current hypothesis. Hence he does not need to match the attributes other than the one he is currently testing. It is the least demanding strategy but is also the least efficient. *Simultaneous scanning* is similar to successive scanning in that hypotheses are tested. The difference is that in simultaneous scanning a number of hypotheses are tested with every card choice while in successive scanning only one hypothesis is tested at a time. Because of its tremendous memory requirements few subjects use simultaneous scanning. Because a number of attributes are kept in mind at the same time it is actually similar to conservative focussing and may perhaps be regarded as formal in nature. On the other hand successive scanning appears to be concrete operational behaviour in Piaget's system. The following example should support this latter statement.

Somerville (1974) writes regarding the identification of the role of the length of the string in the pendulum experiment:

> The situation is complicated by the fact that the role of the length of the string may be discovered by the concrete operational method of establishing a correspondence between two ordered variables. (p.267)

i.e. noticing that the longer the string the slower the pendulum swings to and fro. This procedure is the same as that used in successive scanning when a judgement is made merely because one attribute is present, even though no attention has been paid to other attributes e.g.

> A triangle is present and the instance is positive. Therefore a triangle is part of the concept .

Later analysis of the pendulum task will show that the subject's approach to the solving of the problem can be conceptualised in CI terms. Since strategies have attracted considerable interest it will be possible to relate the findings of the test to research in this area.

b. Relationship between use of strategies AND age, intelligence and academic achievement

If research has shown that there is a suitably positive relationship between CI measures and measures of cognitive development accepted in education i.e. age, IQ and academic achievement then it can be argued that CI theory is useful in constructing a test of processes of importance in education. As with Piagetian measures it can also be argued that CI measures (as used in previous research) have validly measured associated traits and constructs.

i. Age

Anderson (1965, 1968) has shown that 6-year-old pupils are able to focus. However Eimas (1969, 1970) has indicated that younger children can only focus in very simple situations. There is in fact a progression in focussing ability with increase in age (Yudin and Kates, 1963; Eimas, 1969,1970).

Focussing is consistently used at about 15 years of age (Boute cited by Butcher, 1968) and in adulthood (Eimas, 1969, 1970). Certainly the trend to increased success with age is present - a trend which suggests that CI strategies are valid measures of cognitive growth.

ii. Intelligence

Observations regarding CI strategies and intelligence come from Wason (1968) and Wetherick (1969) who note that even very intelligent subjects may scan if they consider scanning appropriate. Wetherick (1969) points out that in real-life situations there are so many variables affecting our everyday decisions, that it is often impossible to focus. So scanning, based on probability, but without any certainty, is used. This observation makes it clear that any test measuring a student's ability to focus must present the task in such a way that the student is encouraged to show whether he *can* focus - and not in such a way that he is left with the impression that any strategy may be used.

Yudin and Kates (1963) and Yudin (1966) show that MA is an important factor to be considered when studying the use of effective strategies. The fact that pupils with higher MA's use"ideal strategies"suggests that the measurement of CI strategies can well be included in a test of cognitive development This relationship supports the view that CI measures are valid measures of the construct of mental age which is of course intimately related to the construct of intelligence.

iii. Academic achievement

No literature could be located which looked at the relationship between strategies used and academic achievement.

3. CONCLUSIONS

Piaget's genetic epistemology should provide useful information when constructing a process test for adolescence since it is concerned with individual processes, and also with the characteristics of the stage of formal operations. Furthermore it has been shown that Piaget's system is relevant to education.

The CI model is restricted in scope, but allows a detailed analysis of the strategies used when considering a number of combinations.

In most cases measures obtained from both research areas show suitably positive relationships with accepted measures of cognitive development (e.g. age, IQ, academic achievement).

The next section will discuss the use of Piaget's tasks in group tests.

V

THE USE OF PIAGETIAN TASKS IN GROUP TESTS

1. PROBLEMS ASSOCIATED WITH THE USE OF PIAGETIAN TASKS IN A GROUP TEST

Piaget has never used group tests. Thus the advisability of using Piagetian tasks in a group test must be evaluated.

A number of issues are raised:

- The role of questioning and protocol
- The role of manipulation of objects
- The social situation of the test.

a. <u>The role of questioning and protocol</u>

Piaget conducts individual interviews in which he pursues the reasoning behind answers which are wrong or not clear. This is a most necessary procedure if one wishes to gain greater insight into patterns of thought. The necessity of using protocol analysis in research into thinking is also noted by two researchers in the CI field viz. Eifermann (1965a;1965b) and Giambra (1971).

Ayers (1971) notes:

> Undoubtedly the *méthode clinique* was suited to Piaget because of its great flexibility in probing for thought patterns and structures, and without it, it is doubtful that Piaget would have developed the theory for which we are so indebted. (p. 76)

However since Piaget's research group has isolated the precise operations and relevant features of the formal stage it does not "seem necessary for research replicating his experiments to probe every evasive or wrong answer" (Ayers, 1976, p. 76).

Thus it is argued that a test which is merely embodying in a group format what has been discovered already does not need to have the freedom which Piaget's original research method engenders.

There are of course limitations to the use of protocol analysis. Considerable personal skill is involved in interpreting protocol and there is always a degree of subjectivity. luddenham (1971) comments:

> Most investigators have followed Piaget in utilizing interrogation, even though the questions were sometimes read from a standard list. This approach standardises the examiner's questions but not the subject's answers. Scoring entails a degree of subjectivity in classifying responses, and almost forces resumption of the *méthode clinique* to clarify obscure or incomplete explanations by the subject. (p. 66)

Vocalization during problem solving has been shown to have a facilitative effect in a wider range of situations. Gagnê and Smith (1962), Bower and King (1967), Anderson (1974), Durling and Schick (1976) all report increased performance in problem-solving when vocalisation is used. Byers and Davidson (1967) and Dominowski (1973) found that written hypotheses (i.e. written protocol) in a CI problem situation also facilitated performance.

It would, presumably, be advantageous to ask for written protocol in a group test using Piagetian tasks as this would make the test situation a little more like the original Piagetian situation - and play some part (however small) in clarifying thinking.

b. The role of manipulation of objects

Despite an extensive search of the literature no evidence has been found that at adolescence the actual process of motor manipulation (per se) is important in the learning processes. It will of course play a part if the information is not available in another form,e.g. if the student has no other way of obtaining information on the weight of an object.

The visual modality tends to be dominant over the others. Thus Rock and Harris (1967) note:

> ... when a subject's sense of touch conveys information that disagrees with what he is seeing, the visual information determines his perception. (p. 104)

Thus visual information is considered to be far more important than any information gained by manipulation of objects. It follows that the lack of information from manipulation is not considered a defect of the group test situation as apparatus will be displayed in a manner which allows all pupils to be in possession of relevant visual information. In the test a few pupils will be asked to confirm the experimenter's "same" judgements e.g. they will be required to confirm that two plasticine spheres are equally heavy. Thus necessary information not available from manipulation will also be presented through the aural channel.

There is also theoretical justification from Piaget's model for not allowing object manipulation as a part of the experiment. At the stage of formal operations possibility becomes more important than reality. Therefore judgements must be made "in the head" (Goodnow and Bethon, 1966,p.581) and not with concrete props. Thus a test of formal operations should give minimal opportunity for the subject to get answers from the concrete material;(See also

Somerville, 1974).

In conclusion, the fact that the subjects in a group test do not receive individual apparatus is not regarded as a limitation which might invalidate the measuring of processes or general orientation.

c. The social situation of the test

By using a group situation the social circumstances of the test are changed. This could offer advantages to pupils who are anxious in a one-to-one test situation. Lawson and Renner (1974) note that:

> Some subjects may be perfectly capable of formal thought but do not exhibit it in the interview situation because they are in a strange situation and intellectually "freeze up". (p. 557)

Thus the group test format *could* be advantageous for anxious pupils. The reverse position would hold for any pupil who required the extra motivation which could be induced in a one-to-one testing situation. It was not anticipated that many of the proposed subjects would fall into this latter category. It is therefore argued that the group situation is not in itself a handicapping situation and the advantages accruing to anxious pupils might allow maximum manifestation of ability.

2. GROUP TESTS USING PIAGETIAN TASKS DESIGNED FOR USE WITH ADOLESCENTS

A number of tests will be discussed. Available test statistics will be presented.

a. Understanding in Science Test (Tisher and Dale, 1975a, 1975b; R.P. Tisher, pers. comm.)

---

[1] Professor R.P. Tisher, Faculty of Education, Monash University, Clayton,

This test is based on a test described by Tisher (1971) which was also used by Cropley and Field (1969) and Field and Cropley (1969). It differentiates between early concrete, late concrete, early formal and late formal using a criterion system. It does not consider individual processes.

The test uses four demonstration experiments based on Inhelder and Piaget's (1958) tasks: the bouncing ball, equilibrium in the balance, communicating vessels and projection of shadows. A total of twenty-four multichoice items are divided amongst these four experiments. Fourteen items are aimed at the concrete level and ten at the formal level. Formal questions score two points and concrete items one point each. Validity was established by correlating the performance of fifty-seven pupils in a group situation and in an individual interview situation. Pupils were considered formal in the interview situation if they performed formally on two of the three tasks used. Significant agreement was found between stage classification using the two methods ($p<0,005$).

b. Class-tasks described by Shayer et al. (1976), Shayer and Wylam (197

Shayer et al. introduced the term "class-task" to describe their method of testing. They used three tasks - each of which "did not exceed the duration of a double lesson". Teachers administered the tests which were largely objective in design. The testing was aimed at establishing the distribution of Piagetian stages in the British school population.

The three tasks were constructed in such a way that if all three tasks were used pupils could be categorized at any point from Stage 1 (pre-operational) to Stage 3B (late formal operational). This means that general orientation could be established. Although processes are mentioned they are not stressed.

Validity was established by correlating the performance of pupils in the class-task situation with the pupils' performance in individual interviews - some of which were conducted by Lovell[1]. Correlations were:

TaskI - (Spatial concepts) 7 pupils, $r = 0,85$

TaskII - (Quantities) 4 pupils, no correlation given

TaskIII- (Pendulum problem) 54 pupils, $r = 0,51$.

Reliability: $r = 0,80$(KR20).

c. Logical Reasoning Test (Burney, 1974; R.B. Sund, pers. comm.[2])

Burney used modified forms of some of Tisher's items. Other items include propositional logic, analogies, etc. The test puts subjects into pre-operational, concrete and formal stages on the basis of their logical reasoning scores.

d. The Longeot Examination (Longeot, 1962, 1965)

The English translation is by Sheehan, 1971 (A.E. Lawson, pers. comm.[3]).

The examination has a multi-choice format. It classifies subjects as concrete formal operational - in this way drawing attention to the general level of reasoning used by the subject. However by using the following subdivisions the examination is concerned with more specific processes.

---

1 Lovell (1961) was involved in early replications of Inhelder and Piaget's research.

2 Prof. R.B. Sund, Department of Science Education, The University of Northern Colorado, Greeley, Colorado, U.S.A., 80639.

3 Prof. A.E. Lawson, Lawrence Hall of Science, University of California, Berkeley, California, USA, 94720.

Part I tests class inclusion and serial ordering

Part II tests proportionality

Part III tests propositional reasoning

Part IV tests combinational reasoning

Nevertheless separate test statistics are not provided.

The KR20 for the whole examination is 0,98.

Validity was established using scalogram techniques (Longeot, 1965).

Because it takes about 2 hours to administer the tests, shortened forms have been used by Lawson and Blake (1976). These shortened forms yield a KR20 of 0,85.

Concurrent validity was assessed by Lawson and Blake (1976) using a group of Piagetian tasks administered individually. Agreement between classification of subjects was significant ($p<0,02$).

e. <u>Subject-matter tests (Lawson and Renner, 1975)</u>

Lawson and Renner (1975) report research into subject-matter tests in physics, chemistry and biology. The biology test is called the *Biology reasoning test* (Lawson, 1977; A.E. Lawson, pers. comm.[1]).

These tests are simultaneously both Piagetian tests and achievement tests as they test concepts which have been taught, but the questions have been set in such a way that they also put subjects into cognitive levels. Some of the questions test processes e.g. the *Biology reasoning test* has a

---

[1]Prof. A.E. Lawson, Lawrence Hall of Science, University of California, Berkeley, California, USA, 94720.

question which tests the ability to control variables. However there are no test statistics on individual processes. Fifteen questions are set at the concrete level and fifteen questions are at the formal level. Face validity was established by six prominent science educators familiar with science education and Piagetian theory and content validity was established by classroom teachers. Concurrent validity was acceptable. Reliability coefficients using the Spearman - Brown split-half technique were:

Physics: $r = 0{,}59$;

Chemistry: $r = 0{,}71$;

Biology: $r = 0{,}76$.

Although the reliability of the Physics test is somewhat low the tests show that Piagetian theory can be applied to yet another type of group test, viz. achievement tests.

f. Subject matter tests (Bart, 1972)

Bart (1972) describes the design and validation of three formal reasoning instruments concerned with respectively English Literature, Biology and History. Bart gives a very thorough discussion of validation. Content validity was established by asking two trained high school teachers to act as judges. They had "substantial content validity". Concurrent validity was established by correlating the test with four individually administered tests. They had "modest concurrent validity". Construct validity was found to be "limited". However all in all Bart considers the tests "somewhat successful". (p. 669)

3. CONCLUSION

It has been argued that the differences between the individual and group testing formats should not cause any major difficulties. The fact that Piaget's tasks have been used in a number of group tests, with apparent success, is used as evidence to support the view that a process test can be designed using Piagetian tasks.

VI

THE DESIGN OF THE TEST

1. FORMAT OF THE TEST

After reviewing the literature on group tests it was decided that Shayer et al.'s (1976) class-task approach held promise. This also appeared to be the method used by Elkind (1961b, 1962), although he did not call it by this name. According to Shayer et al. their procedure falls midway between Piaget's original method and a strictly psychometric approach. In the class-task method of testing the class teacher administers the test. Shayer et al. point out that this usually encouraged a relaxed atmosphere where the skill of the teacher helped in the smooth running of the test. Neale (1968) also speaks in favour of using a procedure somewhere in between the statistical and the clinical:

> Too often one is disposed to conceive of statistical and clinical approaches as being at opposite poles in scientific research. I believe that this defect in the teaching of psychology prevents research workers seeing these approaches as complementary and equally valuable so that investigation of dyslexia and perhaps other aspects of the behavioural sciences have suffered. (p. 38)

If one puts Piaget's interview and a psychometric test at opposite ends of a continuum then the class-task would be found in the middle. The research reported herein would fall nearer the "psychometric test" end of the continuum. See DIAGRAM 2 below.

DIAGRAM 2

MODEL SHOWING RELATIONSHIP BETWEEN PIAGETIAN INTERVIEW AND PSYCHOMETRIC TEST

Piagetian interview	This research ↓ ——— ↑ Shayer et al.'s class-task	Psychometric test

In the present research, the experimenter (who has experience in High School science teaching) will conduct *all* testing. Because of this and the fact that the testees are *not known* to the researcher the situation is slightly more standardised than in Shayer et al.'s procedure and is thus slightly nearer the psychometric test end of the continuum. The main factor which will differentiate this test from a typical psychometric test is that demonstrations will be conducted throughout the test. Although every attempt will be made to standardise instructions, demonstrations do encourage class participation and therefore spontaneous questions. Every effort will be made to ensure that no extra help is given to any one test group. Overall it is argued that the class-task format does not present any serious difficulties especially in view of the fact that items used in the test have been carefully chosen because they were judged suitable for the class-task situation.

2. VALIDITY

According to Anastasi (1976) there are a number of types of validity, viz. face validity, content validity, criterion validity (which includes concurrent validity and predictive validity) and construct validity.

a. Face validity

Face validity is concerned with whether the test looks valid to the testees and any other untrained personnel involved (Anastasi, 1976). As no untrained personnel were involved the researcher had only to be concerned with the testees' perception. Face validity was established by the researcher drawing on his knowledge of the teaching situation in Standards 6, 7 and 8, and also the use of a pilot study where pupils' reactions were observed.

b. Content validity

According to Anastasi (1976):

> Content validity involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured. (pp. 134-135)

In this study content validity was judged by the researcher. Within the confines of the time available attention has been directed to testing processes at different levels. SECTION A will test the process of reasoning involved: i.e. was it predominantly formal or concrete? SECTION B will test the more specific processes involved in the pendulum problem e.g. the logical skills involved in manipulating variables and the strategies (as defined in CI studies) which are used to do this.

It could be argued that the findings regarding processes apply only to Physics because the test does not include content from other areas. However it should be noted that Inhelder and Piaget's (1958) research into formal operations was mainly in the field of Physics and yet the wider support for their findings is now being provided by researchers in a variety of subjects: e.g. Hallam (1969) in History, Ellis (1977) in Literature.

The reason why content from other areas will not be used is because pupils may have widely differing standards of knowledge in different content areas. It is possible for this factor to introduce unnecessary variation in scores if the test samples from different content areas. No final content validity check was made (as is often the case, e.g. Lawson and Renner, 1975) as there was no-one known to the researcher with a thorough background in the various theoretical areas covered by this research.

However SECTION A of the test had a strong resemblance to the Study 1 of Lawson and Renner (1974) as well as Task II of Shayer et al. (1976). Some of the items resembled those used by Elkind (1961b, 1962). SECTION B had a strong resemblance to Task III of Shayer et al. (1976). Since these studies were conducted by experienced personnel and appeared successful the resemblance cited above argues for the content validity of the test.

c. Concurrent validity

According to Anastasi (1976) concurrent validity and predictive validity can be classed together as criterion-related validity. She differentiates between these two types of validity on the basis of the time interval between the initial test and the assessment in the criterion situation. Bart (1972) favours the view that concurrent validity is found when a test's findings regarding a particular trait are supported by direct testing of that trait e.g. the stage classification given by his paper and pencil test was supported by the stage classification produced by direct testing with individually administered Piagetian interviews.

In this research no attempt will be made to assess true concurrent validity by correlating results of tasks administered in this group test with tasks administered individually. There were a number of reasons for deciding not to administer individual tasks:

- The tasks to be administered in this group test are typical Piagetian science tasks, and not tests using content from literature etc. as used by Bart (1972). It is argued that the format and structuring of the material is only slightly different from that which would be used in an individual interview. It is doubted whether the impact of this slight difference will have a major effect on the results. Furthermore it is doubted whether the impact of this slight difference could be assessed by the very variable interview method; i.e. it has been shown that Piaget's procedures lack standardisation and that there is little detail available describing his precise methods (Weitz et al., 1973). There is thus no guarentee that the individual interview will be a true portrayal of Piaget's method.

- Lovell was called in to do the individual testing for two of Shayer et al.'s (1976) class-tasks. This fact underlines the personal skill involved in carrying out individual tests. Lovell has been in the forefront of British attempts to replicate Piaget's studies (see Lovell, 1961). To the best of this researcher's knowledge there is no-one in the area with any precise skill in carrying out Piagetian interviews who could have been called in to give personal help. Ideally such a person should have studied under Piaget.

- The number of subjects who could have been meaningfully tested would be very small. Lovell tested 7 and 4 pupils for Tasks I and II respectively (Shayer et al., 1976). Validation using small numbers such as this is of limited use.

- If the same tasks are administered to pupils, first in a group situation and then in an individual interview, learning effects may be found. Shayer et al. (1976) bear out this fact. Lawson, Nordland and De Vito (19 report a learning effect when individual tasks were repeated. The learning effect provides a difficulty which weakens the case for retest -ing.

- One objective in designing the test is to investigate the possibility of measuring specific processes of importance in education. No other suitable tests of named processes at the adolescent level were located.

- The use of another Piaget-based *group* test to assess cognitive level was rejected. These group tests are still one step away from Piaget's individual interview and represent an interpretation of Piaget's method.

d. Construct validity

Another method of assessing validity is to investigate construct validity. According to Anastasi (1976):

> The construct validity of a test is the extent to which a test may be said to measure a theoretical construct or trait.... Any data throwing light on the nature of the trait under consideration and the conditions affecting its development and manifestations are grist for this validity mill. (p. 151)

Anastasi draws attention to the fact that *age differentiation* and *correlation with other tests* are important criteria for establishing the construct validity of a test.

Individually administered Piagetian tasks measures have already been shown to correlate positively with age measures and test results (intelligence and academic achievement tests). This suggests that the four measures viz. individually administered Piagetian task measures, age measures, IQ and academic achievement results are measuring more or less the same construct, even if it is vaguely defined. This construct could be called *state of cognitive development*. The four measures can be said to be in the same domain because they measure the same construct (Nunnally, 1967).

If another measure is introduced - in this case a measure derived from the planned test - and its scores correlated significantly with three of the variables being considered, viz. age, IQ and academic achievement, it can be assumed to be in the same domain as these variables. It can be further assumed that it measures the same construct i.e. state of cognitive development.

The fact that no correlation showing the relationship between the newly introduced group test measure and individually administered Piagetian task measures is available does not present a serious difficulty. There are sufficient intercorrelations available between variables in the same domain for an assessment to be made of the validity of the test.

If the significant positive correlations between the group test results and the variables of age, IQ and academic achievement are found, then it can be argued that the test has construct validity i.e. that it validly measures the same construct as individually administered Piagetian tasks.

The construct validity of the test items can also be investigated by intercorrelating all items. Items which are unequivocally at the same level e.g.

formal level should show significant relationships indicating that they are measuring the same trait, viz. ability to think formally.

The construct validity of the test can also be investigated using principal components analysis. The finding of two factors in a principal components analysis would support the view that the test is measuring concrete and formal thought - two important Piagetian constructs.

The construct validity of the test as a whole can be assessed by investigating the relationship between cognitive stage
AND age, IQ, and academic achievement.

Positive relationships would suggest that the test has construct validity, i.e. if the older pupils and those that gained higher test scores, tended to be placed in the formal stage it would suggest that the test was validly placing pupils in that cognitive stage.

3. RELIABILITY

Reliability refers to the consistency with which a test measures a given trait or traits (Anastasi, 1976).

As the test is relatively short, with a total of fifteen items, only the reliability of the test as a whole will be assessed. The reliability of brief sub-scales e.g. the three items which could be formed into a sub-scale measuring the process of controlling variables, will not be assessed.

Generally test reliability coefficients in the region of 0,80 and 0,90 are acceptable (Anastasi, 1976). Despite the fact that the test is short and the method of presentation less controlled than that used in a normal paper and pencil test, it will be evaluated according to this standard.

Protocol produced while doing Piagetian tasks needs unambiguous marking if reliability is not to be adversely affected. For this reason it has been attempted to design questions which require simple protocol which can be marked using a fairly objective mark-scheme where marks for each item are of equal value and are simply summed.

4. THE TEST

An intact copy of the test is found in APPENDIX A. *It is essential to refer to this copy when reading through this section.* Another copy integrated with the tester's commentary and a detailed mark scheme is available in APPENDIX B.

SECTION A deals with weight - volume - displacement - density - Archimedes Principle. SECTION B deals with the pendulum problem.

Only one score will be reflected for each item in SECTION A. In items A1 - A4 the bulk of each score would be made up of marks allocated for the process of reasoning: only one mark in each item will be allocated for the production of the correct product. Item A5 does not have a product answer and therefore the entire score for this item would be a measure of the reasoning process employed.

In SECTION A items will be classified according to whether they measured *predominantly* concrete or formal reasoning. This point is important because researchers differ on whether these tasks are concrete, formal or transitional between concrete and formal.

No attempt will be made to separate stages into sub-stages e.g. early formal (3A), etc. An inspection of Inhelder and Piaget (1958) and Somerville (1974) shows how difficult it is to assign answers to sub-stages; e.g. Inhelder and Piaget (1958) consider the operation of exclusion to be at the 3B sub-stage while Sommerville considers that it can occur at the 3A sub-stage.

In order to derive maximum measurement of reasoning and maximum distribution of scores some reasons not mentioned by Piaget will be given marks, provided it is considered that they contribute to the reasoning process.

SECTION A

Item 1: The conservation of weight

This item was originally discussed by Piaget and Inhelder (1974). It has been used by, amongst others, Elkind (1961a, 1961b, 1962), Lawson and Renner (1974), Nordland et al. (1974) and Lawson (1975b).

Eight marks were allocated for the answer, of which one was given for giving the correct product answer and seven were given for the reasoning process involved.

A prerequisite for attempting to reason in this item is that the subject ignores the perceptual impressions provided by the distorted shape. Failure to do this will place the subject in a pre-operational category.

According to Elkind (1967) this item requires the subject to judge the equivalence of the two objects (equivalence conservation). However conservation of equivalence cannot be established by direct comparison of the dimensions of the objects after deformation. Therefore Elkind argues that this item is really testing a pattern of reasoning which he calls identity conservation. DIAGRAM 3 below indicates the process of reasoning used.

DIAGRAM 3

PROCESS OF REASONING REQUIRED IN ITEM A1[1]

Subject judges s (standard) = v (variable)

$\downarrow$ is changed to

v_1

Subject covertly infers

$v = v_1$ (identity conservation)

Therefore subject using identity conservation judges

$s = v_1$ (equivalence conservation)

The reasoning is believed by Piaget and Inhelder (1974) to be concrete in nature. Principal components analysis (Lawson and Renner, 1974) also suggests that this item tests concrete thought.

The process tested by this item is concrete reasoning.

---

[1] After Elkind (1967, p. 16)

Item 2: The conservation of volume using plasticine spheres

The original version of this item is found in Piaget and Inhelder (1974). It has been used by, amongst others, Elkind (1961b, 1962), Towler and Wheatley (1971), Lawson and Renner (1974), Nordland et al. (1974), Lawson, Nordland and De Vito (1975), Lawson and Nordland (1976) and Shayer et al. (1976).

Piaget and Inhelder (1974) remark regarding the use of displacement:

> Trial and error has, in fact, convinced us that the best means of assessing the child's approach to volume is to rely on the amount of water displaced by a clay ball, etc . (p. 47)

As presented, this item also used the words "room" and "space" in the initial confirmation of equivalence of volume (following Elkind, 1961b). It was considered that to introduce the question in such a way that pupils were asked to consider both the terms *room* and *space* as well as to think of *displacement of water*, would ensure understanding of what was required.

Seven marks were allocated to this answer of which one was given for the correct product answer and six marks were given for the reasoning process involved.

The subject must, as in the previous item, ignore perceptually dominant information and rely on reasoning.

Following Elkind (1967)it is argued that this example requires a pattern of reasoning which will establish the identity of the volume of the distorted object to itself when undistorted. See DIAGRAM 4 below.

DIAGRAM 4

PROCESS OF REASONING REQUIRED IN ITEM A2[1]

Subject judges s (standard) = v (variable)

$\downarrow$ is changed to

v_1

Subject covertly infers

$$v = v_1 \quad \text{(identity conservation)}$$

Therefore subject using identity conservation judges

$$s = v_1 \quad \text{(equivalence conservation)}$$

A further process tested in this item is the ability of the subject to establish the relationship between an object and its surroundings.

It appears that the conservation of equivalence using the conservation of identity argument is concrete in nature as in the previous example. But the establishing of a relationship between an object and its surroundings has been argued to involve formal thinking (Towler and Wheatley, 1971 citing Piaget et al., 1960).

Shayer et al. (1976) regard the task as concrete, Lawson and Renner (1974) regard it as an 2B/3A (i.e. transitional)task while Lawson and Nordland (1976) consider it to be at the 3A (early formal) level. It has been decided that in this research this task will be classified as concrete.

The process tested by this item is concrete reasoning.

---

1 After Elkind (1967, p. 16)

Item 3: The conservation of density

This item is concerned with density (Inhelder and Piaget, 1958), although the problem itself is original to this test. All pupils used in this study work through a syllabus which includes density. In dealing with density they learn the *rule* that objects more dense than water sink. *Thus they do not require an understanding of Archimedes Principle to complete this item successfully.*

The question was presented in terms of "float and sink" to avoid using the word density and thus possibly focussing their attention directly on learned school material.

Five marks were allocated for this item of which four were allocated for the reasoning process. The item requires the subject to use ratio and proportion to decide on which block is more likely to sink (Brainerd, 1971; Brainerd and Allen, 1971).

Suppose the ratio of mass/weight[1] to volume in the blue block is 1:1. As the red block is four times as large as the blue block but only twice as heavy as the blue block the ratio of mass/weight to volume in the red block is 2:4 or 1:2. Therefore when the mass/weight to volume ratio in the two blocks is compared it is found that the blue block is proportionately more dense than the red block. The blue block is thus more likely to sink.

The process tested by this item is formal reasoning.

---

[1] In order to avoid confusion regarding *mass* and *weight* the terms were not used in the test. Thus the term *mass/weight* will be used in the explanation of this item.

Item 4: The conservation of volume using metal cylinders

This item is based on the task presented by Piaget and Inhelder (1974) as interpreted by Lawson, Blake and Nordland (1974). It has been used by, amongst others, Lawson, Nordland and De Vito (1974), Nordland et al. (1974), Lawson (1975b), Lawson, Nordland and De Vito (1975), Lawson and Renner (1975) and Shayer et al. (1976).

The item requires that the subject arrives at the equivalence of the volume of the cylinders and then correctly relates the volume and not the weight of the objects to the outer medium. The reasoning involved is considered to be formal by Lawson, Blake and Nordland (1974) because the subject must establish the relationship between an object and an outer medium - which is an abstract relationship. He must then also separate two factors, weight and volume, and ignore one factor, i.e. weight, which is perceptually dominant, and focus on a less prominent factor, i.e. the volume of the cylinders.

Shayer et al. (1976) regard the item as a 2B/3A item, presumably basing their judgement on the fact that Piaget and Inhelder (1974) regarded a reflective corrected answer - which one cannot prevent in a paper and pencil test - as evidence of a transitional stage.

For the purpose of placing subjects in stages in this research, Lawson, Blake and Nordland (1974) will be followed and this item will be regarded as a formal item.

The process tested by this item is formal reasoning.

Item 5: Archimedes Principle

This item is based on Inhelder and Piaget (1958), who regard it as being at the 3B level. They make the point that they require the answer to be produced spontaneously. This is the reason why it was considered that directed questioning was not necessary in this question: pupils were simply asked to "Explain as fully as you can."

In discussing Archimedes Principle, Piaget symbolises the understanding of the relationship between small/large, light/heavy, float/sink in terms of proposition logic. But before the process of reasoning can begin the subject must be able to imagine a hypothetical unit of water equal in volume to the object under consideration - a formal skill.

Inhelder and Piaget (1958) write regarding the logical process used:

> If we let p be the assertion that a given object floats and $\bar{p}$ the assertion that it does not, q the assertion that its volume is equal to a certain quantity of water, r the assertion that it is lighter than that quantity of water, and $\bar{r}$ the assertion that it is heavier, the relationship which the subject establishes is the following:
>
> $p.q.r \vee \bar{p}.q.\bar{r}$ (4a)
>
> which is in fact the schema of proof based on the assumption "all other things being equal". (p. 41)

In simpler language: If a block of the same volume (q) as a hypothetical block of water is lighter (r) than that block of water it will float (p). If a block of the same volume (q) as a hypothetical block of water is heavier ($\bar{r}$) than that block of water it will sink ($\bar{p}$). The subject must keep one variable constant and simultaneously manipulate two other variables which move in linked fashion i.e. he must keep volume constant and manipulate weight and float/sink. This is a distinctly formal process.

The process tested by this item is formal reasoning .

Section B

This set of questions is based on the pendulum problem (Inhelder and Piaget, 1958) as adapted by Shayer et al. (1976) for group testing.

Other experimenters who have used this task in their research are: Lovell (1961), Lengel and Buell (1972), Siegler et al. (1973), Lawson, Nordland and De Vito (1974), Somerville (1974), Lawson and Wollman (1976), Wollman and Lawson (1977).

The pendulum problem tests both the processes of concrete and formal reasoning. However not only these processes are tested; in addition it is possible to test the subject's use of more precise processes e.g. the concrete level process of identifying variables and the formal level process of controlling variables. These more precisely defined processes were called intermediate level processes in IV p. 12.

However Inhelder and Piaget (1958) actually consider the processes involved in terms of propositional logic. They write:

> Let p be the statement that there is a modification in the length of the string and $\bar{p}$ the absence of such modification; q will be the statement of a modification of weight and $\bar{q}$ the absence of any such

> modification; likewise r and s state modifications in both the height of the drop and the impetus and $\bar{r}$ and $\bar{s}$ the invariance of these factors. Finally, x will be the proposition stating a modification of the result - i.e. of the frequency of the oscillations - and $\bar{x}$ will state the abscence of any change in frequency. (p. 76)

In the class-task planned s and $\bar{s}$ can be excluded from the reasoning as impetus was omitted from the problem.

In the pendulum problem the subject must operate mentally with a number of problems. Using the INRC group he must relate these combinations to his needs.

He may wish to vary only one variable to examine an effect; e.g. if he wishes to find out the effect of the length of the string he must be able to generate the combination $p\bar{q}\bar{r}x$ which allows him to conceive of the situation where the length of the string is changed and the speed at which the pendulum swings to and fro is also changed. If he can conceive of this combination he should be able to change the *one, correct,* variable.

Later he must also be able to draw correct conclusions,e.g. he must be able to reject those combinations which are not possible in terms of the data he has been given: e.g. he must reject as a possibility

$p\bar{q}\bar{r}\bar{x}$

which indicates that a change in the length of the string, while the other factors are kept constant, will not affect the speed at which the pendulum swings to and fro. He must do this because he has noted

$p\bar{q}\bar{r}x$

which indicates that a change in the length of the string *did affect the speed at which the pendulum swung to and fro.*

Before the subject can use the formal processes described by Piaget he must of course be able to ignore the perceptual impressions given by the size of the weight and the height of the drop, i.e.:

that the large weight will drop more quickly therefore making it swing to and fro more rapidly;

that the higher the weight the more it will accelerate therefore making it swing to and fro more rapidly.

As indicated earlier the pendulum problem can be related to the paradigms and strategies of CI research. In order to do this it is necessary to transpose the pendulum problem into typical CI format.

This could be done as follows:

a fast swinging pendulum could be regarded as a positive exemplar

a slow swinging pendulum could be regarded as a negative exemplar

a long string could be represented by a double border

a short string could be represented by a single border

and so on. If transposed into CI form the pendulum problem as presented here would be requiring the identification of the attribute(s) and the rule. It would differ from most CI problems in that the subject is not directly told that there is a clearly defined number of attributes making up the concept, although it is commonly assumed that only one factor is involved.

As only one attribute is responsible the concept is called an unidimensional concept (Dominowski, 1973) or a univariate concept (Neisser and Weene, 1962), and the rule is the affirmation rule (Haygood and Bourne, 1965). With the foregoing discussion in mind items, sometimes grouped according to the processes they measure, will be considered in terms of both Piaget's model and the CI model.

Item B1i,ii,iii

This item concerning the identification of possible variables is not represented in CI research where the variables are carefully dimensionalised and given to the subject. In terms of the Piagetian model this item tests the *process of identifying variables.*

Inhelder and Piaget (1958) indicate that subjects in the concrete stage can separate the variables involved (They are supported by Lawson, Nordland and De Vito, 1974 and Somerville, 1974).

Items B2, B3, B4

In terms of Inhelder and Piaget (1958) these three questions try to measure the *process of controlling variables* in an experimental situation. Is only one variable changed? Is the correct variable changed? (Somerville, 1974).

Ideally the subject would have to generate his own combinations. In a pilot study this was attempted but pupils could not understand the format and carry out the instructions. It is considered that despite the fact that subjects do not have to generate combinations their ability to control variables would be tested in this recognition and choice situation.

In terms of the CI model items B2, B3 and B4 require that the subject selects an example which varies by only one attribute from either of the initial exemplars presented. In other words the subject is selecting an example with which he can focus on either of the initial exemplars. As no feedback was given after each choice there would be no likelihood of learning to focus because of feedback given. In this regard there is *some* similarity to

Levine's (1966) "blank trials" procedure where the subject must make, say, ten sequential card choices without feedback being given so that the experimenter can probe the strategies the subject is using without the confounding effect of feedback.

Items B5, B6, B7 and items B8, B9, B10

In terms of Inhelder and Piaget's model B5, B6 and B7 test the *process of drawing conclusions*. Specifically item B5 tests the *process of identifying the causative factor*. Items B6 and B7 test the *process of excluding the irrelevant factor*.

Items B8, B9 and B10 also test a process concerned with the drawing of conclusions, viz. *method of drawing conclusions*.

In terms of the CI model items B5, B6 and B7 are in the reception mode because the subject is not in control of the combinations presented (Smalley, 1974). The experimental results in the table are structured in such a way that conservative focussing can be used to identify the causative factor (the length of the string) and exclude the irrelevant factors (the size of the weight and the height of the drop). However it is a relatively easy matter to identify the causative factor as the length of the string by successive scanning (which could eventually summate to form simultaneous scanning since the combinations are written down).

While either conservative focussing[1] or simultaneous/successive scanning may

---

1 Focus gambling could well be attempted but because the instances are written down and therefore available simultaneously it has no advantage.

be used to derive the answer for item B5, item B8 asks that the *best* method of identifying the causative factor in B5 is given. Therefore the answer required is:

"Compare experiments done by Group 1 and Group 3."

Since it is never possible to prove anything using the inductive method, successive/simultaneous scanning cannot be the best method. Therefore any of the other distractors which represent instances of hypotheses testing cannot be regarded as correct. The subject who understood - even without being able to explain it - the importance of controlled experimentation and who could manipulate the variables should have been in no doubt as to which combination was best. It is realized that it might be argued that "best" is a word with many meanings; however brief instructions were deemed to be most important.

Item 6 could be solved by either conservative focussing or successive/simultaneous scanning. Item 9 again asks for the *best* method. Therefore the answer required is:

"Compare experiments done by Group 1 and Group 4."

which is an example of conservative focussing.

Item 7 does present an instance where one conservative focus could not provide the answer. In this instance *only a series of conservative focusses would provide the answer with certainty.* i.e. first it was necessary to focus to establish the role of the length of the string (by comparing the experiments conducted by Group 1 and Group 3). This information had to be remembered. Then it was necessary to focus to establish the role of the weight (by comparing the experiments conducted by Group 1 and Group 4).

This information had to be remembered. Then allowing for the effect of the short string and the non-effect of the weight the experiments of Group 2 and Group 3 had to be compared. The height of the drop did change, and the speed at which the pendulum swung to and fro did not change. Therefore using a final focus manoevre it could be shown that the height of the drop does not affect the speed at which a pendulum swings to and fro. Thus the best answer for question 10 is:

"One must use a combination of the above methods."

(Although not asked for, these would be:

"Compare experiments done by Group 1 and Group 3

Compare experiments done by Group 1 and Group 4

Compare experiments done by Group 2 and Group 3").

This sort of focus manoevre could correctly be argued to be simultaneous scanning because each variable was being systematically tested across a number of instances in which more than one factor was being changed and memory was being used to remember the role of each factor. However it has been pointed out by Laughlin (1973), Neimark (1975) and by both Forrest and Boote (cited in Butcher, 1968) that there is no clear cut division between these two categories.

Thus the distinction between conservative focussing and simultaneous scanning is really academic. The point worth stressing is that this type of reasoning involves the simultaneous control of four variables - clearly a type of formal reasoning.

5. CONCLUSIONS

The test, as designed, is based on Piaget's tasks. The work of Elkind (1961b), Lawson and Renner (1974) and Shayer et al. (1976) has been of particular help in constructing items. SECTION B has also been interpreted in terms of CI strategies.

Items in SECTION A require predominantly concrete or formal reasoning. Items in SECTION B measure more precisely defined processes e.g. the ability to control variables (or focus).

The test's validity and reliability will be assessed in the following section.

VII

EXPERIMENTAL INVESTIGATION INTO THE VALIDITY AND RELIABILITY OF THE TEST

1. HYPOTHESES

a. Individual items

In order to assess the construct validity of each item the following hypotheses were formulated:

- *There will be a significant relationship between item scores AND age; IQ; form; English marks; Mathematics marks; Science marks.*
- *There will be significant positive correlations between scores on items which are agreed to be at the same level of cognitive development.*
- *When item scores are subjected to principal components analysis they will load mainly on two factors which may be interpreted as concrete and formal thought.*

b. Test as a whole

In order to assess the construct validity and reliability of the test as a whole the following hypotheses were formulated:

- *There will be a significant relationship between performance on the test as a whole AND age; IQ; English marks; Mathematics marks; Science marks.*
- *The test as a whole will have a reliability coefficient of 0,80 or more.*

2. SUBJECTS

These were drawn from three form levels within a state boys' school.

The study was restricted to boys because some researchers report that boys show superior performance in Piagetian-based class-tasks (Elkind, 1961b, 1962 ; Shayer and Wylam, 1978). Girls' improvement in performance also tended to level out before boys (Shayer and Wylam, 1978). In individual interviews boys were found to be superior to girls (Dale, 1970; Lawson, 1975). Even though Case and Collinson (1962) and Tisher (1971) report no difference in male and female performance on Piagetian tasks the intrusion of sex differences is a possibility and it was therefore decided to restrict the research to boys.

Furthermore the researcher-cum-tester wished to avoid any depression of performance such as has been found when female subjects were administered a test by a male tester (Brekke and Williams, 1973).

Whole class units as present on the day of testing were used. It must be noted that in the school concerned classes are streamed according to school achievement.

At Form 2 level two classes were used i.e. 2B and 2G (n = 46). At Form 3 level two classes were used i.e. 3B and 3F (n = 48). In both Form 2 and Form 3 the second highest and the second lowest classes were used to get a range of ability and to avoid those at the extremes of scholastic attainment which would be presented by the top and bottom classes. In Form 4 the B class and the E and F classes were used (n = 63). Two lower level classes were used because there were insufficient students taking Physical Science in the F class: many took Biology only and thus did not fulfill the condition for contributing to the Science mark.

3. METHOD AND RESULTS

Tests were administered in three school morning sessions at convenient times between 9 am and 12 noon. The researcher administered the test in the manner indicated in APPENDIX B. The three test days were within a ten day period.

a. Data concerning subjects

i. Date of birth

Pupils indicated their date of birth on their answer sheets. This was then approximated to the nearest month i.e. dates of 15 and less were regarded as signifying the month in which the pupils were born. Dates of 16 and more were regarded as signifying birth in the next month. This information was used by the computer programme to produce an age as at 1979-01.

ii. Intelligence Quotient

The IQ used was that obtained by the school. All pupils should have their IQ assessed by means of the New South African Group Test (NSAGT) (1965b) in Standards 5 and 7 (also called Forms 1 and 3). Thus IQ's should therefore not have been more than two years old. This conforms to the HSRC specification that IQ's derived more than two years before should not be used (HSRC, 1965a).

iii. English, Mathematics and Science marks

These were to be used for promotion purposes at the end of 1978. They were calculated by finding the mean for a number of tests administered during the year.

A short explanation regarding the composition of Science marks is necessary. In Forms 2 and 3 the Science mark is a General Science mark made up of Physical Science and Biology. The Science marks for Form 4 are Physical Science marks. Pupils taking Biology in Form 4 instead of Physical Science were excluded. The reason for this was that Biology is in some cases - depending on the work studied - very different from Physical Science in its intellectual demands (Moerdyk, 1973). It was felt that the Science mark should have at least some Physical Science in it so that it might act as a better indicator of ability in Science.

iv. Form

This measure was included in later calculations as a measure of chiefly age, but one which would also take into account ability to achieve at school.

Means and standard deviations describing the parameters noted above are given in TABLE 1 below.

TABLE 1

MEAN SCORES AND STANDARD DEVIATIONS FOR INFORMATION ON SUBJECTS BY FORMS

Variable	Entire		Form 2		Form 3		Form 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age-years	15,11	0,98	14,00	0,52	14,95	0,40	15,98	0,58
IQ	112,99	12,20	111,96	12,98	114,60	13,45	112,55	10,68
English %	55,76	9,07	61,56	7,65	56,72	9,94	51,06	6,51
Maths. %	51,95	20,21	57,22	21,29	57,49	16,72	44,17	19,49
Science %	58,34	13,62	61,36	10,81	60,32	12,57	53,47	15,80

b. Data concerning scores on the test

The test was marked according to the mark scheme presented in APPENDIX B. Means and standard deviations for the test items are given in TABLE 2 below.[1]

---

1 TABLE 2 gives the names of the Piagetian tasks and the processes which they represent:thereafter some of the tables presented will use item names based on either the names of the tasks or the processes.

TABLE 2

MEAN SCORES AND STANDARD DEVIATIONS FOR SCORES ON ITEMS

Item	Task	Process	Mean	S.D.
Section A				
A1	Conservation of weight	Concrete reasoning	2,61	1,28
A2	Conservation of volume - plasticine spheres	Concrete reasoning	1,64	1,33
A3	Density	Formal reasoning	0,66	1,16
A4	Conservation of volume - metal cylinders	Formal reasoning	1,83	1,15
A5	Archimedes Principle	Formal reasoning	0,83	0,68
Section B	Pendulum problem			
B1i		Identifying variable	0,99	0,11
B1ii		Identifying variable	0,85	0,36
B1iii		Identifying variable	0,82	0,39
B2		Controlling variables	0,48	0,50
B3		Controlling variables	0,48	0,50
B4		Controlling variables	0,51	0,50
B5		Identifying causative factor	0,72	0,45
B6		Excluding irrelevant factor	0,26	0,43
B7		Excluding irrelevant factor	0,83	0,28
B8		Method of identifying causative factor	0,64	0,48
B9		Method of excluding irrelevant factor	0,58	0,50
B10		Method of excluding irrelevant factor	0,32	0,47

c. Investigations

Six investigations will now be reported. Each investigation has been conducted with reference to one of the aforementioned hypotheses.

INVESTIGATION 1

It is hypothesized that there will be a significant relationship between item scores.
AND age; IQ; form.

In terms of the chi-squared test (χ^2) there will be a significant difference between the expected and the observed distribution of item scores in terms of each of the variables: age; IQ; form.

It is predicted that the significant differences will be congruent with the view that performance on Piagetian tasks generally increases as age, IQ, and form increase.

Method

In order to perform the χ^2 test the variable values were grouped as follows:
Ages were grouped at six monthly intervals from 13 years to 17 years 6 months.
IQ's were grouped in intervals of 10 IQ points from 84 to 144.
The subjects were considered by their forms, viz. 2, 3 and 4.

Results

The results of the test are shown in TABLE 3 below.

TABLE 3

ITEMS SHOWING A SIGNIFICANT DIFFERENCE BETWEEN OBSERVED AND EXPECTED FREQUENCIES

Item	Name	Age	IQ	Form
A1	Conservation of weight			$p<0{,}05$
A2	Conservation of volume - plasticine spheres			$p<0{,}05$
A3	Density	$p<0{,}05$		
A4	Conservation of volume - metal cylinders		$p<0{,}01$	
A5	Archimedes Principle	$p<0{,}05$	$p<0{,}001$	$p<0{,}05$
B1i	Identifying variable			
B1ii	Identifying variable			
B1iii	Identifying variable			
B2	Controlling variables			
B3	Controlling variables			
B4	Controlling variables			
B5	Identifying causative factor			
B6	Excluding irrelevant factor	$p<0{,}002^{+}$		
B7	Excluding irrelevant factor			$p<0{,}05^{+}$
B8	Method of identifying causative factor		$p<0{,}05$	$p<0{,}05$
B9	Method of excluding irrelevant factor		$p<0{,}005$	
B10	Method of excluding irrelevant factor			$p<0{,}05$

+ Not in predicted direction

Discussion

It will be noted that items A1 - A5 and B8 - B10 show a significant relationship at the 5% level in the predicted direction with at least one of the criteria of age, IQ and form.

It will be argued that for an initial experimental test version this level of significance indicates an acceptable level of relationship. This will in turn argue for these items' ability to measure the construct of *Piagetian cognitive development*. However some items do not show any significant relationship and in two cases where a significant relationship is present it is not in the predicted direction. These items will now be discussed.

B1i,ii,iii is a fairly easy concrete item . All pupils should have reached the concrete level by the time they are in Form 2. Thus it can be argued that a great increase in performance on this item with increasing age and higher IQ should not be expected. This can be used to argue in favour of the construct validity of the item.

As far as B2, 3 and 4 are concerned there is a one-in-four chance of guessing each of these items correctly. Guessing would interfere with any relationship between an item score and variables such as age, IQ or form. Grouping of items should restrict the effect of guessing to an acceptable level and this will be done in the following investigation.

B5 is a concrete item and should therefore be expected to be within the reach of all the pupils of the age studied. In CI terms it may be solved by using successive scanning which has been argued to be a concrete operational skill.

As no increase in performance was found this can be used to argue for the construct validity of the item.

There is a one-in-five chance of guessing B6 and B7 correctly. However a far more likely cause of the significance being in the non-predicted direction is that the pupils tested on the second and third days of testing (viz. Form 3 and Form 2 respectively) were helped by another person who told them the easily remembered rule: "It's only the string that counts".

In order to reduce the possibility of prior knowledge of a rule interfering with the measurement of processes the test had in fact been designed with
B5 and B8,
B6 and B9, and
B7 and B10
being answers and reasons respectively. If the subject was simply remembering the rule and was not able to carry out the required process then he would have difficulty providing the reason. If the items were grouped into new items and scored the effect of prior knowledge would be reduced.

INVESTIGATION 2

The items in Section B were grouped into the following new items as discussed above, viz.:
1 i, ii and iii
2, 3 and 4
5 and 8
6 and 9
7 and 10

With grouping the new items[1] became scored variables. Correlation calculations could thus be performed.

It is hypothesized that there will be a significant relationship between item scores
AND Age; IQ; Form; English marks; Mathematics marks; Science marks.

In terms of the Pearson product-moment statistic this would mean that the *correlations* would be significant. Naturally higher correlations will argue more forcefully for the test's construct validity.

Method

Pearson product-moment correlations were performed between item scores and the six variables listed above.

Results

A correlation matrix was obtained. See TABLE 4 below.

---

[1] When items are regrouped the new *item group* will for the sake of brevity simply be referred to as an item.

TABLE 4

CORRELATION MATRIX FOR ITEM SCORE, AGE, IQ, FORM AND SCHOOL MARKS

Item	Age	IQ	Form	English	Mathe-matics	Science
A1	-0,03	0,18	0,00	0,22+	0,17	0,10
A2	-0,10	0,30++	-0,10	0,35++	0,26++	0,24+
A3	0,06	0,28++	0,07	0,25+	0,26++	0,25+
A4	-0,07	0,39++	0,01	0,25+	0,23+	0,14
A5	0,10	0,31++	0,16	0,22+	0,24+	0,29++
B1i,ii,iii	-0,05	0,25+	-0,06	0,19	0,28++	0,10
B2,3,4	-0,07	0,16	0,02	0,09	0,15	0,20+
B5,8	0,17	0,19	0,13	0,07	0,19	0,19
B6,9	-0,12	0,29++	-0,09	0,25++	0,20+	0,15
B7,10	-0,03	0,21+	-0,04	0,18	0,14	0,16
Age		-0,17	0,85++	-0,52	-0,36	-0,33
IQ			0,01	0,58++	0,55++	0,44++
English					0,74++	0,70++
Mathe-matics						0,82++

+ $p<0,05$ (Level of significance for positive correlations according to
++ $p<0,01$ two-tailed test, Ferguson, 1959.)

Discussion

Correlations of item scores with age and form show non-significant correlations. At first inspection this is contrary to what is expected in terms of the Piagetian model. However since the sample is a convenience sample it is possible that sampling bias has occurred in the selection of pupils. In individual forms the older pupils are more likely to be the low achievers. Therefore it is probable that more of the older pupils in the sample are less able intellectually.

The correlations of item scores with IQ scores are significant at the 5% level in the majority of cases. This fact argues for the construct validity of these items. i.e. they measure intelligence (broadly defined). The sampling bias discussed above should not interfere with these correlations because IQ takes age into account.

Items A2, A3, A4, A5, B6,9 show at least three significant correlations with the indicators of cognitive development at the 5% level suggesting that these items do in fact measure the advanced thinking required of pupils in high schools i.e. formal thought.

As item B6, 9 measures focussing and as previous research has shown a significant correlation between focussing and IQ, the significant correlation existing between B6, 9 and IQ argues in favour of the item's ability to measure focussing.

The lesser number of significant correlations associated with A1 and B1i,ii,iii is to be expected as they measure concrete thought which is not as strongly related to the criterion measures of age, IQ and school marks.

It is not clear why formal items B2,3,4 and B7,10 show altogether only one significant correlation. They may lack the ability to validly measure the process of controlling variables (focussing).

INVESTIGATION 3

It is hypothesized that there will be significant positive correlations between scores on items which are agreed to be at the same level of cognitive development.

Method

Items were grouped into the following new items. B1i,ii,iii and B2,3,4 remained as originally grouped. It was decided to follow Shayer et al. (1976) in the grouping of items B5-10. B5, B6 and B7 were thus grouped as an item concerned with drawing of conclusions. Theoretically there is some difficulty in that B5 is a concrete item whereas B6 and B7 are formal items. A further difficulty is that the construct validity of items B6 and B7 was doubted.

B8, B9 and B10 were grouped as an item representing the method of drawing conclusions (Shayer et al., 1976). In this case all three items were at the formal level.

The items were then intercorrelated.

Results

See correlation matrix in TABLE 5 which follows.

TABLE 5

CORRELATION MATRIX OF SCORED ITEMS

Item	A1	A2	A3	A4	A5	B1 i,ii,iii	B2,3,4	B5,6,7	B8,9,10
A1		0,25+	0,07	0,18	0,12	0,13	-0,04	0,05	0,04
A2			0,19	0,34++	0,22+	0,18	0,04	0,11	0,01
A3				0,21+	0,31++	0,11	0,19	0,08	0,30++
A4					0,17	0,31++	0,08	0,04	0,16
A5						0,16	0,13	-0,02	0,25+
B1i,ii,iii							0,11	-0,04	0,16
B2,3,4								0,10	0,21+
B5,6,7									-0,02

\+ $p<0,05$ Level of significance according to two-tailed test (Ferguson,1959).

++ $p<0,01$

Discussion

An inspection of the correlation matrix reveals a number of correlations which are significant at the 5% level. However these correlations are still low with only two of them in the 30's. Nevertheless in some cases the correlations argue for the construct validity of the items, while those existing between items which are not clearly concrete or formal do not give any indication in favour or against the hypothesis.

In favour of the construct validity of the test are the following correlations

A3 and B8,9,10 ($p<0,01$)

A5 and B8,9,10 ($p<0,05$)

Since these items are unequivocally formal *and the content of these items is different* the significant correlations suggest that they are validly measuring the shared underlying processes. As this is Piaget's view it is argued that these test items are validly measuring formal thought.

The correlations between

A3 and A5 ($p<0,01$)

B2,3,4 and B8,9,10 ($p<0,05$)

can be related to Piaget's view that once a subject reaches the formal stage he will be able to do most tasks at that level. As these four tasks are the only tasks in the test, which according to research, are unequivocally formal, the presence of these two significant correlations suggests that the test is validly measuring formal thought - although it could be argued that the significant correlations were being brought about by similar content. The remaining significant correlations do not reflect on the hypothesis in either way because at least one item in each pair is not clearly concrete or formal.

A correlation matrix using dichotomised[1] item values was also conducted but since it revealed a very similar pattern of significant correlations it has not been included.

INVESTIGATION 4

It is hypothesized that when item scores are subjected to a principal components analysis they will load mainly on two factors which may be interpreted as concrete and formal thought.

Method and results

In order to determine the factor structure of the nine items the item scores were subjected to a principal components analysis. Eigenvalues greater than 1 were extracted. Three factors were found. See TABLE 6 for factor loadings on each item.

---

[1] A score of 2 for each of these items was accepted as the criterion of success. The statistics derived in this manner will be noted in Table 15 in VIII p. 109.

TABLE 6

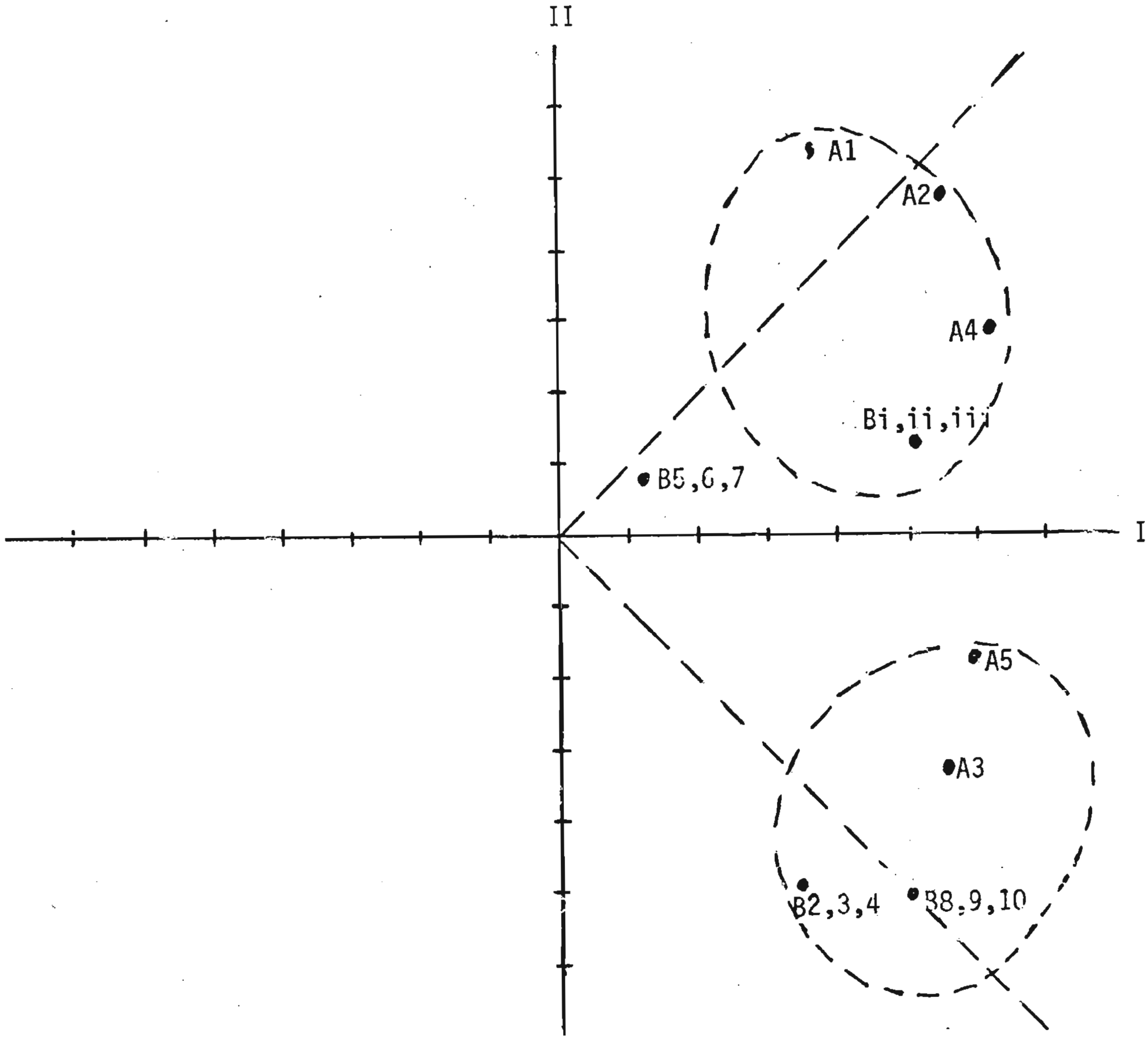
PRINCIPAL COMPONENTS ANALYSIS SHOWING FACTOR LOADINGS

Item	Name	Principal components I 24,5% of variance	II 14,4% of variance	III 12,1% of variance	Percent Communality
A1	Conservation weight	0,37	0,54	0,01	42,67
A2	Conservation volume - plasticine spheres	0,56	0,49	0,18	57,99
A3	Density	0,60	-0,32	0,15	48,97
A4	Conservation volume - metal cylinders	0,62	0,29	-0,10	48,65
A5	Archimedes Principle	0,59	-0,18	-0,12	39,81
B1i,ii,iii	Identifying variables	0,51	0,14	-0,36	40,97
B2,3,4	Controlling variables	0,35	-0,50	0,29	45,95
B5,6,7	Drawing conclusions	0,12	0,08	0,87	78,73
B8,9,10	Method of drawing conclusions	0,50	-0,52	0,17	54,67

Inspection of the above matrix suggested that Factor 1 and Factor 11, which account for 38,8% of the variance, may be orthogonal. When plotted they had the appearance indicated in GRAPH 1 below.

GRAPH 1

REPRESENTATION OF PRINCIPAL COMPONENTS ANALYSIS



The configuration of item scores, viz. two main clusters encircled by the dotted lines suggested that the two factors may represent concrete and formal reasoning.

All items except B5,6,7 loaded significantly on at least *one of* Factor 1 and II. Since the Piagetian model predicts that two types of reasoning are involved in these tasks (viz. concrete and formal) it was supposed that

Factor 1 might represent formal reasoning and Factor 11 might represent concrete reasoning. Item B5,6,7 did not load significantly on either Factor I or II and therefore does not seem to involve formal or concrete reasoning to any significant degree. B5,6,7 was the *only* item to load significantly with Factor III. This suggests that this item is measuring a non-Piagetian factor.

The principal components analysis was then repeated. Only two factors were extracted. They were then rotated. See TABLE 7 below.

TABLE 7

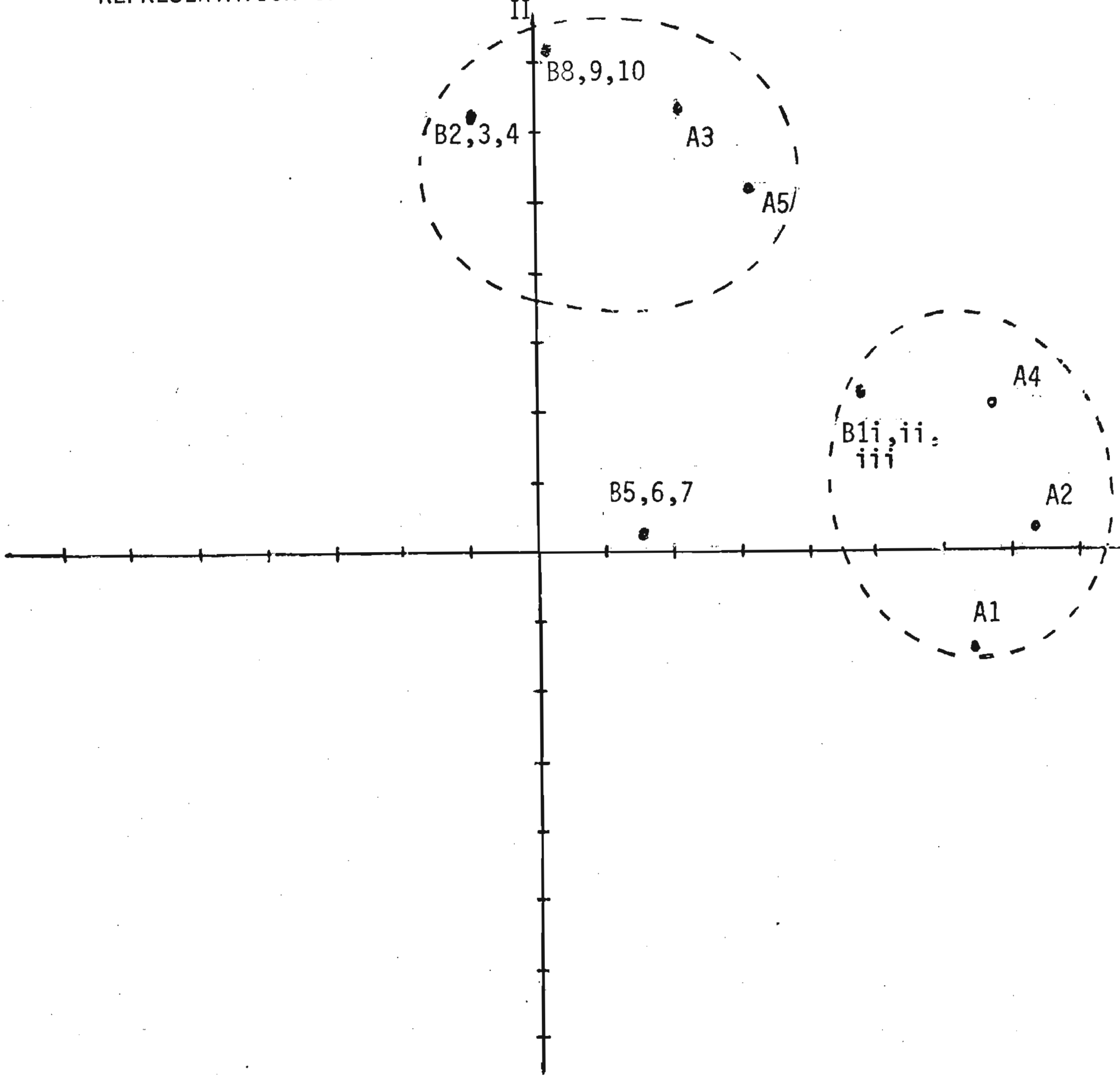
PRINCIPAL COMPONENTS ANALYSIS AFTER EXTRACTION OF TWO MAJOR FACTORS AND ROTATION

Item	Name	Principal components I	II
A1	Conservation weight	0,64	-0,14
A2	Conservation volume - plasticine spheres	0,74	0,03
A3	Density	0,22	0,65
A4	Conservation volume - metal cylinders	0,66	0,21
A5	Archimedes Principle	0,31	0,54
B1i,ii,iii	Identifying variables	0,47	0,25
B2,3,4	Controlling variables	-0,09	0,60
B5,6,7	Drawing conclusions	0,15	0,03
B8,9,10	Method of drawing conclusions	0,01	0,72

A graphical presentation of this data is shown in GRAPH 2 below.

GRAPH 2

REPRESENTATION OF PRINCIPAL COMPONENTS ANALYSIS AFTER ROTATION



Discussion

While the factor loadings are not clear cut items A1, A2, A4, A5 and B1i,ii, iii have significant loadings on Factor I (arguably the concrete factor. Items A3, A5, B2,3,4 and B8,9,10 have significant loadings on Factor II (arguably the formal factor).

It will be noted that A5 loads significantly on both Factor I and Factor II but the major loading is on the formal component. This may be because the mark scheme for A5, in attempting to give maximum distribution of scores, allowed subjects to score one point with concrete level statements. The presence of marks yielded by concrete statements appears to have been reflected in the principal components analysis.

Item A4 has, in contrast to neo-Piagetian research (Lawson, Blake and Nordland, 1974), loaded more on the concrete factor than on the formal factor. Again with this item marks could be obtained with concrete statements. It is an item which does not require statements which are uniquely formal. According to Lawson, Blake and Nordland (1974) an answer is judged to be formal when a subject has provided *two* answers, viz. subject separates the variables weight and volume *and* shows that water displacement is dependent on volume and not weight. However it must be noted that if these answers are considered *individually* it could be argued that they are concrete answers. It is therefore not surprising that this item loads on the concrete factor.

Since the items tended to load on two factors which appear to represent concrete and formal thought it can be concluded that principal component analysis supports the view that most items in the test validly measure the two constructs of concrete and formal thought - which are characteristic of the Piagetian tasks used. In view of the suspect construct validity of item B5,6,7 this item will be discarded in the following two investigations.

INVESTIGATION 5

It is hypothesized that there will be a significant relationship between performance on the test as a whole
AND age; IQ; English marks; Mathematics marks; Science marks.

Method and results

Subjects will first be grouped according to cognitive stage reached. An analysis of variance will then be performed. A significant F ratio will indicate a significant relationship between performance on the test as a whole (as indicated by the cognitive stage reached) and the accepted indicators of cognitive development mentioned above. This will support the argument for the test's construct validity. The correlation η will also be derived in the analysis of variance.

Inspection of Piagetian and neo-Piagetian research suggested that it would be possible to classify the items as in TABLE 8 below.

TABLE 8

CLASSIFICATION OF ITEMS BASED ON PREVIOUS RESEARCH

Concrete	Formal
A1	A3
A2	A4
B1i,ii,iii	A5
	B2,3,4
	B5,6,7
	B8,9,10

Four classes were then instituted and subjects were classified according to the method shown in TABLE 9 below. The number of concrete items correct and the number of formal items correct were summed for each subject.

TABLE 9

CLASSIFICATION KEY FOR ALLOCATING SUBJECTS TO COGNITIVE STAGES BASED ON RESULTS OF PREVIOUS RESEARCH

Stage 1 - Pre-operational	Scores of 0 or 1 on concrete items
Stage 2 - Concrete	Score of at least 2 on concrete items
Stage 3 - Transitional between concrete and formal	Score of 3 on concrete items AND 1 or 2 or 3 on formal items
Stage 4 - Formal	Score of 3 on concrete items AND at least 4 on formal items

If a criterion was not reached, further correct items were not taken into account.

There was no significant relationship at the 5% level between this system of classification AND age, IQ, English marks, Mathematics marks and Science marks.

This might be due to the system of categorization being used as it had been shown that there were at least some significant relationships between individual item scores and the criteria listed above.(See TABLE 4, p.77) showing significant distributions and TABLE 5, p.80 showing significant correlations).

The major problem with the system of categorization used was that subjects who did not reach lower criteria sometimes went on to get a number of more difficult items correct:

e.g. Subject number 63 who got 1/3 concrete items correct AND 4/6 formal items correct was placed in Stage 1 (pre-operational) because he could not reach the criterion of 2/3 concrete items correct and therefore could not be placed in a higher stage. It is worth noting that this subject actually reached the criterion of 4/6 for Stage 4 (formal thought) but was denied access to it because he did not reach the lower criterion for Stage 2 (concrete thought).

Therefore another system was devised. Because of the difficulty of knowing precisely how Piaget marked his items it was decided that it would be permissible to take into account the results of the principal components analysis. In the principal components analysis the items loaded significantly as shown in TABLE 10 below.

TABLE 10

TABLE SHOWING FACTORIAL LOADING OF EACH ITEM

Item	Factorial loading
A1	Concrete
A2	Concrete
A3	Formal
A4	Concrete
A5	Formal and concrete but predominantly formal
B1i,ii,iii	Concrete
B2,3,4	Formal
B5,6,7	Neither concrete nor formal
B8,9,10	Formal

Items were then grouped as shown in TABLE 11 below. B5,6,7 was rejected.

TABLE 11

CLASSIFICATION OF ITEMS BASED ON PRINCIPAL COMPONENTS ANALYSIS

Concrete	Formal
A1	A3
A2	A5
A4	B2,3,4
B1i,ii,iii	B8,9,10

Three stage groups were then instituted as indicated in TABLE 12 below.

TABLE 12

CLASSIFICATION KEY FOR ALLOCATING SUBJECTS TO COGNITIVE STAGES BASED ON PRINCIPAL COMPONENTS ANALYSIS

Stage 1 - Pre-operational	Did not reach criterion of 3/4 concrete items AND did not reach criterion of 3/4 formal items
Stage 2 - Concrete	Reached criterion of 3/4 concrete items but did not reach criterion of 3/4 formal items OR Reached criterion of 3/4 formal items but did not reach criterion of 3/4 concrete items
Stage 3 - Formal	Reached criterion of 3/4 concrete items AND reached criterion of 3/4 formal items

This method was accepted as a possible method of classifying subjects so that their stage - classification might be submitted to a validity check. Four pupils reached the formal criterion but did not reach the concrete criterion: as indicated they were classified as concrete thinkers. An analysis of variance was then performed with the stage classification against the variables of age, IQ, English, Mathematics and Science marks. Results are given in TABLE 13 below.

TABLE 13

ANALYSIS OF VARIANCE DATA

Variable	Groups	Sum of squares	df	Mean squares	F ratio	Significance level	Correlation (n) with criterion
Age	Between groups	5,74	2	2,87	3,06	p=0,05	0,20
	Within groups	140,52	150	0,94			
	Total	146,26	152				
IQ	Between groups	3053,73	2	1526,87	11,67	p<0,001	0,36
	Within groups	20147,27	154	130,83			
	Total	23201,00	156				
English	Between groups	1366,60	2	683,30	9,18	p<0,001	0,33
	Within groups	11465,69	154	74,45			
	Total	12832,29	156				
Mathematics	Between groups	5047,03	2	2523,51	6,63	p<0,01	0,28
	Within groups	58270,57	153	380,85			
	Total	63317,59	155				
Science	Between groups	1375,87	2	687,93	3,86	p<0,05	0,23
	Within groups	24225,25	136	178,13			
	Total	25601,11	138				

Discussion

The levels of significance suggest that once allowance is made for the fact that A4 appears to measure concrete thought and that B5,6,7 does not appear to measure concrete or formal thought there is evidence of construct validity.

However the correlations (n) between, on the one hand cognitive stage, and on the other age, IQ and school marks are fairly low.

Nevertheless the test appears to some degree to be measuring Piagetian cognitive stages as it has already been shown in previous research that there are positive relationships between these stages and age, IQ and academic achievement.

INVESTIGATION 6

It is hypothesized that the test as a whole will have a reliability coefficient of r = 0,80 or more.

The more items making up a test or sub-test which measure a particular trait, the more reliable a test will be. Because the whole test is fairly short it will only be possible to calculate the reliability coefficient as a whole and no reliability coefficients can be calculated for the very short sub-scales such as those in Section B which measure respectively

Separation of variables

Control of variables

Drawing of conclusions

Justifying of conclusions.

Each of these sub-scales is made up of only three items.

Method and Results

The reliability of the test was assessed using three methods: Spearman - Brown Split - half method (using the Pearson product-moment formula), the KR20 formula and Hoyt's analysis.

i. Spearman - Brown Split - half method

This method was judged appropriate because the data fulfilled two important conditions viz.

- it must be possible to split the test into two halves made up of items matched according to difficulty and
- subjects must have the opportunity to complete all items i.e. the test must be a power test and not a speed test.

Because of the peculiar nature of the test, the items were not numerically arranged in ascending order of difficulty. Therefore it was not possible to construct two parallel sub-tests by simply dividing the items on an odd-even basis. Furthermore since no previous study could yield precise information on the relative difficulty of the items being used in this test it was considered appropriate to use information derived from the principal components analysis. Using this information the test was therefore divided into two parallel sub-tests. The one sub-test was made up of A2, A3, B1, B2,3,4 and the other was made up of A1, A4, A5 and B8,9,10. As indicated previously B5,6,7 was discarded as it did not appear to have an acceptable level of construct validity.

The Pearson product-moment correlation was then calculated. This yielded a reliability coefficient of 0,47. This fell far short of the desired reliability coefficient of 0,80.

ii. The KR20 formula

Instead of splitting the test in two parallel forms this method splits a test into "n parts of one item each" (Guilford, 1954, p. 380): thus it rests on the assumption that the items of a test are all parallel items. A further assumption is that the test under consideration is unifactorial. Clearly the present test does not meet these conditions; the items are not of equal difficulty and the items have been shown to load on two factors. However Guilford notes that despite the fact that few tests meet these criteria the KR20 formula is widely used. After reviewing evidence he concludes that even when the limiting assumptions of the test are violated the KR20 formula can give fairly accurate results.

The KR20 formula follows (Guilford, 1954, p. 380, Formula 14,5):

$$r_{tt} = \left(\frac{n}{n-1}\right)\left(\frac{\sigma^2 t - \Sigma pq}{\sigma^2 t}\right)$$

where

r_{tt} = reliability coefficient

n = number of items in the test

p = proportion of correct responses to each item in turn

q = $1 - p$

$\sigma^2 t$ = total variance of test

Applying the formula:

$$r_{tt} = \left(\frac{n}{n-1}\right)\left(\frac{\sigma^2 t - \Sigma pq}{\sigma^2 t}\right)$$

$$= \left(\frac{9}{8}\right)\left(\frac{2,83 - 1,58}{2,83}\right)$$

$$= 0,50$$

This coefficient again falls far short of the desired coefficient.

iii. Hoyt's analysis

The following formula was used (Guilford, 1954, p. 384 Formula 14.11):

$$r_{tt} = 1 - \frac{V_r}{V_e}$$

where

r_{tt} = reliability coefficient

V_r = variance for remainder sum of squares

V_e = variance for examinees

The various sums of squares are computed using formulae provided by Guilford (1954) p. 384.

Applying the formula:

$$r_{tt} = 1 - \frac{V_r}{V_e}$$

$$= 1 - \frac{0,16}{0,31}$$

$$= 0,48$$

This again falls short of the desired reliability coefficient.

Discussion

It will be noted that the above three reliability coefficients were in the 0,47 - 0,50 range. These coefficients are at first inspection unsatisfactory. However it must be noted that the test is relatively short. A lengthening of the test would result in a greater measure of reliability. This is because with increase in length, provided the new items resemble the old in form, content and difficulty, the true variance will increase at a more rapid rate than the error variance (Guilford, 1954).

Possible sources of error which may have been responsible for a lowering of the reliability are:

- Although every effort was made to standardise instructions there were the inevitable difficulties in answering spontaneous questions in a helpful manner which at the same time did not reveal the answer. Even though all pupils indicated they could see the demonstrations, pupils at the front desks probably had a better view of the proceedings.

- The instructions for the test indicate that
  "Once you have put an answer in you may not change it".
  In the first administration, as a result of requests, this was modified by oral instructions to
  "Once you have turned over the page you may not change an answer".
  This was repeated before the turning over of each page. However since this was in contradiction to the written instruction this may have produced some uncertainty.

- While it is not considered that these had a serious effect two inconsistencies were found. B5 and B7 begin the fourth choice statement with the words, "All three factors ...". B6 begins this statement with the words, "All the factors ...". B2 uses one long opening sentence. B3 and B4 show the same sentence divided into two.

However as indicated earlier lengthening the test could raise the reliability coefficient. Guilford (1954, p. 391) gives a formula which will indicate the ratio (of the new test length to the original test length) necessary to achieve a specified reliability in the new test. A reliability of 0,48 will be accepted as the reliability of the old test.

The formula is:

$$n = \frac{r_{nn}(1 - r_{tt})}{r_{tt}(1 - r_{nn})}$$

where

n = number of times the test must be lengthened

r_{nn} = reliability coefficient which is required

r_{tt} = reliability of the original test

Applying the formula to the reliability under consideration, it will be assumed that a reliability of 0,80 is desirable.

$$n = \frac{r_{nn}(1 - r_{tt})}{r_{tt}(1 - r_{nn})}$$

$$= \frac{0{,}80(1 - 0{,}48)}{0{,}48(1 - 0{,}80)}$$

$$= 4{,}33$$

i.e. The test would have to be 4,33 times as long as the present test.

However the time which would be needed to administer this lengthened test would rule out its practical usefulness. The present test took approximately 45 minutes to administer. It would therefore take approximately 3 hours 15 minutes to administer a test of this length.

Therefore it is suggested that the format must change. Instead of a class-task format the task would have to be redesigned to take on a more normal paper-and-pencil format. In this way the optimum number of items could be dealt with in a shorter period of time.

It could well be found that a more structured paper and pencil format will also lead to a relative decrease in error variance when compared with the class-task format. This would in turn mean that acceptable reliability would be obtained with a shorter test than the calculation above indicates.

4. CONCLUSION

a. Individual items

TABLE 14 below gives the conclusions regarding the construct validity of items (grouped according to the processes they measure), assessed as very acceptable, acceptable, minimally acceptable or not acceptable. It must be stressed that these conclusions are subjective and are not calculated according to any formula. *They are also made subject to the condition that the test is at present an experimental instrument.*

Processes for SECTION A items are those indicated after taking the results of the principal components analysis into account.

TABLE 14

CONSTRUCT VALIDITY OF ITEMS

Item	Task	Process	Construct validity
Section A			
A1	Conservation of weight	Concrete reasoning	Very acceptable
A2	Conservation of volume - plasticine spheres	Concrete reasoning	Very acceptable
A3	Density	Formal reasoning	Very acceptable
A4	Conservation of volume - metal cylinders	Concrete reasoning	Very acceptable
A5	Archimedes Principle	Formal reasoning	Very acceptable
Section B	The pendulum problem		
B1i,ii,iii		Identifying variables	Acceptable
B2,3,4		Controlling variables	Minimally acceptable
B5,6,7		Drawing conclusions	Not acceptable
B8,9,10		Method of drawing conclusions	Very acceptable

b. Test as a whole

After item B5,6,7 was excluded and the results of the principal component analysis were taken into account the construct validity of the test was found to be acceptable for an experimental instrument.

The use of the CI model appears to have added an extra dimension to the study of the variables in the pendulum problem. However because there was relatively little existing research on the relationship of CI theory to the measures of cognitive development used in this study, it was difficult to integrate the CI model when the results were discussed.

Judged according to the desired reliability coefficient the reliability of the test was not acceptable. Comments have been made regarding the possibility of raising the reliability coefficient.

The possible future development of the test will be discussed in the next section.

VIII

FUTURE RESEARCH AND IMPLICATIONS FOR EDUCATION

1. THE POSSIBLE DEVELOPMENT OF THE TEST

Clearly the test is at this stage only a research instrument. The following points need discussion.

a. The theoretical basis of the test

In constructing this test Piaget's system has provided the theoretical basis. It appears that it is not necessary to restrict the development of the test to Piaget's formulations. An important step in developing the test would be to ascertain, using task analysis (Gagné,1970; Linke,1975; Levine and Linn,1977), what processes were needed for carrying out a variety of school tasks. The test could then be reformulated so that it was able to measure these processes. Thus future research should use whatever formulations were able to shed light on processes regarded as important. This is especially necessary if one is to develop reliable sub-tests to measure particular processes.

A model which might be used to advantage is that of Gagné (1970). Gagné's model is hierarchical in nature. The subordinate behaviours are those which must be learned first; later the superordinate behaviours must be learned. In contrast to Piaget, Gagné's model is by nature very flexible. In fact its major contribution lies in its method of representing behaviours in hierarchies - more than any precise information it might impart regarding learning. Some of Gagné's hierarchies are merely proposed hierarchies awaiting validation (Linke, 1975). Some of Piaget's precise information could in fact be hierarchically organized.

Towler and Wheatley (1971) implicitly draw attention to this relationship between Gagné and Piaget's views:

> An examination of the subjects' responses in this study have led the author to suspect that part of the problem, at least, lies in inadequately formed concepts of atomism, since nearly all of the erroneous explanations centre on molecules, density, or the surface area of the clay. (p. 269)

This represents an explanation of superordinate failure in terms of possible subordinate failure - an hierarchical representation.

Case (1972a, 1972b, 1974) has attempted to devise a model which integrates Gagné's and Piaget's views - showing that the two models are compatible.

Another potentially useful model is that of Ausubel (1968). Novak (1974, 1975) proposes that Ausubel's theory offers a more parsimonious interpretation than does Piaget's model of the data provided by Lawson, Nordland and De Vito (1974) and Lawson, Blake and Nordland (1975).

Guilford's (1967) model of intelligence provides many possible processes which could be incorporated into a test.

It is proposed that a future process test should ignore theoretical consistency and be pragmatic. Items should be included because they are able to test processes identified as important and are of suitable difficulty, have high discrimination indices, etc.

b. Alternative method of testing processes

It has been argued that a formal test would be useful. However an alternative method would be to link testing with the curriculum in such a way that information on the attainment of processes was collected throughout the school year. This approach would be in keeping with the current move away from formal testing to continuous less formal methods of assessment.

However such a method would be difficult to implement unless a large scale curriculum project was instituted, similar to that reported by Lucas (1972) in the field of Science Education in Australia.

After the course content, e.g. in the field of Physical Science, had been chosen a task analysis would be carried out to ascertain what processes were required for the successful completion of the course. Then curriculum planners would draw up instructional programmes. These would include material designed to promote learning of both content and process. Tests would then be drawn up. Each question would be designed in such a way that it would test not only content but also definite processes.

Each test would be backed by a duplicate. Schools would be asked to administer the tests and to submit the duplicate copies to the curriculum designers. The original tests would be given back to the pupils and if desired the marks could be used for in-school evaluation.

Using information derived from these tests hierarchies of processes could be drawn up using the method described by Linke (1975). Where sections were found to cause difficulty remedial programmes could be devised so that pupils not reaching criterion on the evaluation instrument could be given extra help.

This approach has a number of strong points.

- It allows teachers to practise their class-based professional skills instead of putting a tremendous amount of effort into preparing unco-ordinated programmes and tests.

- Curriculum aids would be prepared by those with the time and training to do so. The personnel drawing up the tests would be skilled in the areas of psychometrics and educational psychology. Teachers would assist as co-researchers by advising and applying the material. Naturally the whole approach would hinge on a spirit of co-operation between the research personnel and the teaching profession.

- The content would remain integrated. In the writer's opinion the major defect in *Science - a process approach* was that in order to teach processes in an hierarchical order content was dismembered into a series of disjointed exercises. Only a very formal thinker would be able to follow the sequence. It is proposed that the content sequence be arranged in as interesting a form as possible and processes be investigated where they occur in the natural flow of content.

- Apart from the tremendous benefit to teachers and teaching, research of great import could be carried out if this method could be instituted. If the procedure was a success in one subject it could be carried out in other subjects where the content lent itself to this approach.

However it must be noted that a school-based approach such as that described above is subject to far more intruding factors than a validated, reliable and possibly standardised test of processes similar to that which it was researched in this study.

2. IMPLICATIONS STEMMING FROM PUPILS' PERFORMANCE ON TEST ITEMS

Even after accepting the limitations of the test the number of pupils not achieving success on acceptably valid individual formal items does suggest that there is considerable room for cognitive development in the pupil population.

Other researchers have also noted the large number of adolescent pupils who have not attained formal thought. The following table gives information on the percentages of pupils attaining success in various tasks in other investigations. Only studies using more than 100 subjects and giving (or allowing the deduction of) percentages were included in the table.

TABLE 15

PERCENTAGE SUCCESS ON TASKS IN PRESENT STUDY WITH COMPARATIVE PREVIOUS RESULTS

Present study			Previous studies						
Item	Task/Process	% success+	% success	Sex	Mean age++	Age range++	Researchers	Country	Individual(I) Class-task(C)
Section A									
A1	Conservation weight	86	83	Male	15,1 years	Approx.6 years	Elkind(1961b)	USA	C
A2	Conservation volume -plasticine spheres	51	54	Male	15,1 years	Approx.6 years	Elkind(1961b)	USA	C
A3	Density	17,2	28,4	Male	15 years 3 months	15 years 0 months-15 years 6 months	Shayer and Wylam(1978)	UK	C
			33+++	Male & female	College freshmen		McKinnon and Renner(1971)	USA	I
A4	Conservation volume-metal cylinders	68,2	40,5	Male	15 years 3 months	15 years 0 months-15 years 6 months	Shayer and Wylam(1978)	UK	C
A5	Archimedes Principle	14,6	33+++	Male & female	College freshmen		McKinnon and Renner(1971)	USA	I
Section B	Pendulum problem	39	53	Male & female		16-20 years	Kohlberg and Gilligan(1971)	USA	I
			48	Male & female	College freshmen		McKinnon and Renner(1971)	USA	I
			61	Male & female		13 years 11 months-14 years 5 months	Somerville (1974)	Australia	I
			33,7	Male	15 years 3 months	15 years 0 months-15 years 6 months	Shayer and Wylam(1978)	UK	C

\+ A score of 2 was accepted as the criterion of success for individual items in <u>SECTION A</u>. A score of 2 in two of :-

controlling variables,
drawing conclusions,
method of drawing conclusions,

was regarded as the criterion for success in the pendulum problem. (SECTION B)

++ Where available

+++ McKinnon and Renner's task appears to test both Density and Archimedes Principle as represented by Items A3 and A5 respectively.

Inspection of the table does indicate that other researchers have also reported that a large number of pupils/students (from 12 years 6 months onwards) do not solve formal problems.

What should educationalists do once they have found that so many pupils do not reason formally?

They could argue that many pupils could probably have used formal operations - but in the testing situation, with unfamiliar content, they regress to concrete thinking (Chiapetta,1976). While this may play a part in some studies the pupils used in this study have been exposed to science in some form since the Std 2 level. The use of a class-task probably also relieved their anxiety. This makes it unlikely that the regression effect has had any major distorting effect on the results. It is more likely that pupils everywhere really do have difficulty with formal reasoning.

Alternatively they could accept that one must wait for maturation and random experiences to take their natural path. Maturation clearly is one factor which must be taken into consideration. However recognition of its importance should not lead to the fatalistic view that intellectual ability will unfold willy-nilly. Varied experience is also necessary (Gourlay,1978).

It is interesting to note Vernon's (1979) comments on the growth of psychometric intelligence:

> Intelligence refers to the general reasoning and other cognitive capacities which are developed largely by stimulation in the home and in leisure hours or peer-group activities,whereas achievement refers to the more specialised performance in school subjects which

> depends greatly on the quality of teaching and on children's motivation to learn. (p. 12)

It is difficult to understand why the school should not be given a great deal of responsibility for developing what may be called intelligence, intelligent behaviour or intellectual ability since there is much evidence that special intervention can bring about more effective thinking (Whimbey 1974, 1975).

If the school does become involved two approaches seem possible. The first approach deals with the learning of specific processes of thinking while the second deals with the raising of the general level of intellectual functioning. In the discussion which follows evidence from a wide variety of sources will be taken into account and the discussion will not be limited to Piagetian and CI models.

a. Learning of specific processes

It has been shown that specific processes can be learned if instruction is appropriate.

According to Raven (1974) more attention should be given to the incorporation of logical processes in the school curriculum. He writes:

> During the twelve to fourteen years that it has taken these (logical) abilities to develop, most schools have only incidently incorporated the use of these logical operations into their instructional strategies. Much work is spent on the acquisition of symbols and vocabulary but little time is spent on how these symbols are put together. (p. 260)

Raven (1974) has attempted to teach logical processes. He claims to have taught the tautology operation in the context of the subject matter of Density and Archimedes Principle.

As it will prove to be extremely difficult to work at this level it may be necessary to deal with intermediate level processes: e.g. controlling variables (Bredderman, 1973), ratio and proportion (Brainerd, 1971; Brainerd and Allen, 1971).

Lawson (1975) gives a series of steps which he considers will lead to the acquisition of the formal process: controlling variables. He describes the procedure as follows:

> .... learning begins with physical experience with objects. This experience provides the student with a mental record of what he has done and seen. Symbolic notation is then introduced which aids in the identification of patterns in the experiences. Finally, additional experiences that involve the same conceptualization are provided along with the repetition of the invented symbolic notation to allow the student to self regulate and abstract the formal pattern from the particular situations. (p. 419 and p. 429)

In the CI field Anderson (1965, 1968), Wells and Watson (1965), McKinney (1972), working with 6-year-olds, adults and mental-retardates respectively have shown that strategies involved in solving CI problems can be taught. Siegler and Liebert (1975) using tree diagrams and Siegler and Atlas (1976) using an algorithm have also shown how the use of certain learned strategies can aid problem-solving.

However the importance of maintaining affective involvement in the material cannot be overemphasized. A series of disjointed logical exercises which violate a pupil's interest in content and inhibit personal involvement are doomed. (See Mann and Phillips,1967).

b. Raising the general level of intellectual functioning

The following methods could be attempted:

- Piaget (1950) suggests that discussion has a large part to play in encouraging the appearance of formal operations. Outside the Piagetian model Abercrombie (1960) with medical students and Bloom and Broder (1950) with underachieving College students argue that discussion improves the quality of thinking. It is not accidental that exclusive educational institutions boast low student-staff ratios which allow more discussion between students and staff.

- Piaget (1972 cited by Modgil, 1974) suggests that:

> what is needed at both the university and secondary level are teachers who indeed know their subject but who approach it from a constantly interdisciplinary point of view, i.e. knowing how to give general significance to the structures they use and to reintegrate them into overall systems embracing other disciplines. (pp. 273-274)

This view is agreed with in principle: however this is the age of specialisation and broad interests are selected against. High school teachers find that keeping up to date with one subject is sufficiently demanding. However if the curriculum developers are able to perform the integration and give guidance it may well become possible for persons initially lacking the interdisciplinary approach Piaget speaks of to implement his views. Lucas (1972) shows how the Australian Science Education Project has indeed managed to bridge knowledge from traditionally separate subjects by using prepared programmes.

- Lawson and Wollman (1977) argue that successful learning takes place in a situation which imitates what is in fact the historical process of discovery. In this situation the learner notices interesting occurrences, follows them up and experiences personal involvement. In the classroom this method is called discovery learning (Ausubel,1968) or inquiry learning (McKinnon and Renner,1971) and has been claimed to be associated with both a rise in IQ and a movement towards the use of formal operations (Marek and Renner,1979).

- A subject called *thinking* could be taught (De Bono,1972). After reading De Bono's literature it is considered that De Bono has not related his theorising to any accepted body of knowledge concerned with thinking and is too idiosyncratic to be of general use.

- Mays (1965) suggests that logic be taught at school. He argues that mathematics is taught: logic has just as much right to be taught as has mathematics. At first this seems a good idea. However it would only really be of use if transfer could be encouraged. School mathematics is a subject using many formal operations. However it is presented as a subject in its own right and not as a vehicle for the teaching of thinking. If the same happens to logic, the school will simply have a subject of no direct use in its already overloaded curriculum.

- Whimbey (1975) suggests that if reading is regarded as thinking in response to printed stimuli, then reading comprehension exercises offer a socially acceptable and efficient method of practising the processes of adult thinking.

- Certain "study habits" (a popular term for a wide variety of strategies and processes e.g. Goldman and Hudson, 1973) are characteristic of poor comprehension and reasoning ability. Whimbey (1974) describes these processes as follows:

  (1) inadequate attention to the details of the problem to be solved,

  (2) inadequate utilisation of prior knowledge that would help in solving the problem, and

  (3) absence of sequential step-by-step analysis of the relationships among the ideas involved. (p. 50)

Bloom and Broder (1950) note that low aptitude students are inclined to take the view that answers must be immediately available - if they are not then nothing can be done to help solve the problem presented.

Remedial attention should lead to enhanced functioning and a greater respect for the process of reasoning.

3. CONCLUSION

Although the attempt to develop a valid and reliable test of intellectual processes was of limited success it is concluded that research in this area is highly relevant to education at the present time. There is a growing conviction that pupils can learn to use processes, strategies and intellectual skills which will increase their general level of intellectual functioning.

Available tests do not give accurate measures of processes and it remains for someone to produce an acceptable process test - or series of tests - which can measure processes of importance in education. Ideally, this test would be related to a curriculum project in which a suitably pragmatic model of desired processes was constructed and programmes aimed at bringing about the learning of these processes were brought into being. With the emphasis on life-long intellectual growth models of adolescent intellectual functioning would probably be readily applicable in the emergent field of adult education.

It is hoped that the skills of psychologist and educator will both be used for the benefit of the future development of the individual.

IX

REFERENCES

Abercrombie, M.L.J. *The anatomy of judgement.* London: Hutchinson,1960.

American Association for the Advancement of Science. *Science - a process approach.* New York: Xerox Corporation, 1968.

Anastasi, A. *Psychological testing.* New York: MacMillan, 1976.

Anderson, J. Visualisation and verbalisation as mediators of thought. *Speech Monographs,* 1974, 41, 408-412.

Anderson, R.C. Can first graderslearn an advanced problem-solving skill? *Journal of Educational Psychology,* 1965, 56, 283-294.

Anderson, R.C. Part-task versus whole-task procedures for teaching a problem-solving skill to first graders. *Journal of Educational Psychology,* 1968, 59, 207-214.

Ausubel, D.P. *Educational psychology: a cognitive view.* New York: Holt, Rinehart and Winston, 1968.

Ayers, J.D. Comments on Tuddenham's paper. In D.R. Green, M.B. Ford and G.B. Flamer (Eds.), *Measurement and Piaget.* New York: McGraw Hill, 1971.

Bart, W.M. The factor structure of formal operations. *British Journal of Educational Psychology*, 1971, 41, 70-77.

Bart, W.M. Construction and validation of formal reasoning instruments. *Psychological Reports*, 1972, 30, 663-670.

Bateman, B. Three approaches to diagnosis and educational planning for children with learning disabilities. In D.B. Ryckman (Ed.), *Learning disabilities; theoretical approaches*. New York: MSS Educational Publishing Co., 1971.

Beard, R.M. *An outline of Piaget's developmental psychology*. London: Routledge and Kegan Paul, 1969.

Behr, A.L. Diagnosing learning deficiencies in the New Mathematics - part II. *Spectrum*, 1975, 13, 33-36.

Blasi, A. and Hoeffel, E.C. Adolescence and formal operations. *Human Development*, 1974, 17, 344-363.

Bloom, B.S. (Ed.), *Taxonomy of educational objectives. Handbook I, cognitive domain*. New York: David McKay, 1956.

Bloom, B.S. and Broder, L. *Problem solving processes of college students*. Chicago: University of Chicago Press, 1950.

Bower, A.C. and King, W.L. The effect of number of irrelevant stimulus dimensions, verbalisation, sex on learning. *Psychonomic Science,* 1967, 8, 453-454.

Brainerd, C.J. The development of the proportionality scheme in children and adolescents. *Developmental Psychology,* 1971, 5, 469-476.

Brainerd, C.J. Learning research and Piagetian theory. In L.S. Siegel and C.J. Brainerd (Ed.), *Alternatives to Piaget: critical essays on the theory.* New York: Academic Press, 1978.

Brainerd, C.J. and Allen, T.W. Training and generalization of density conservation: effects of feedback and consecutive similar stimuli. *Child Development,* 1971, 42, 693-704.

Bredderman, T.A. The effects of training on the development of the ability to control variables. *Journal of Research in Science Teaching,* 1973, 10, 189-200.

Brekke, B. and Williams, J.D. Conservation and sex. *Perceptual and Motor Skills,* 1973, 37, 14.

Bruner, J.S. Inhelder and Piaget's:"The growth of logical thinking" I.A psychologist's viewpoint. *British Journal of Psychology,* 1959, 50, 363-370.

Bruner, J.S., Goodnow, J.J. and Austin, G.A. *A study of thinking.* New York: Wiley, 1956.

Burney, G.M. *Logical reasoning test.* In R.B. Sund *Piaget for educators.* Columus, Ohio: Merrill, 1976.

Butcher, H.J. *Human intelligence.* London: Methuen, 1968.

Byers, J.L. and Davidson, R.E. The role of hypothesizing in the facilitation of concept attainment. *Journal of Verbal Learning and Verbal Behaviour,* 1967, 6, 595-600.

Bynum, T.W., Thomas, J.A. and Weitz, L.J. Truth functional logic in formal operational thinking: Inhelder and Piaget's evidence. *Developmental Psychology,* 1972, 7, 129-132.

Carroll, J.B. Words, meanings, concepts. *Harvard Educational Review,* 1964, 34, 178-202.

Case, D. and Collinson, J.M. The development of formal thinking in verbal comprehension. *British Journal of Educational Psychology,* 1962, 32, 103-111.

Case, R. Learning and development: a Neo-Piagetian interpretation. *Human Development,* 1972a, 15, 339-358.

Case, R. Validation of a Neo-Piagetian mental capacity construct. *Journal of Experimental Child Psychology,* 1972b, 14, 287-302.

Case, R. Structures and strictures: some functional limitations on the course of cognitive growth. *Cognitive Psychology*, 1974, 6, 544-574.

Case, R. and Fry, C. Evaluation of an attempt to teach scientific inquiry and criticism in a working class high school. *Journal of Research in Science Teaching*, 1973, 10, 135-142.

Chiapetta, E.L. A review of Piagetian studies relevant to science instruction at the secondary and college level. *Science Education*, 1976, 60, 253-261.

Cropley, A.J. and Field, T.W. Achievement in science and intellectual style. *Journal of Applied Psychology*, 1969, 53, 132-135.

Crossley, B. Some classroom applications of an ability classification based on the Bloom taxonomy. *Mentor*, 1979, 61, 11-14.

Dale, L.G. The growth of systematic thinking: replication and analysis of Piaget's first chemical experiment. *Australian Journal of Psychology*, 1970, 22, 277-286.

De Bono, E. *About think*. London: Jonathan Cape, 1972.

Dominowski, R.L. Requiring hypotheses and the identification of unidimensional conjunctive and disjunctive concepts. *Journal of Experimental Psychology*, 1973, 100, 387-394.

Durling, R. and Schick, C. Concept attainment by pairs and individuals as a function of vocalisation. *Journal of Educational Psychology,* 1976, 68, 83-91.

Eifermann, R.R. Response patterns and strategies in the dynamics of concept attainment behaviour. *British Journal of Psychology,* 1965a, 56, 217-222.

Eifermann, R.R. Selection strategies in concept attainment: a re-examination. *Scripta Hierosolymitana,* 1965b, 14, 205-215.

Eimas, P.D. A developmental study of hypothesis behavior and focusing. *Journal of Experimental Child Psychology,* 1969, 8, 160-172.

Eimas, P.D. Effects of memory aids on hypothesis behavior and focusing in young children and adults. *Journal of Experimental Child Psychology,* 1970, 10, 319-336.

Elkind, J. Children's discovery of the conservation of mass, weight and volume. *Journal of Genetic Psychology,* 1961a, 98, 219-227.

Elkind, D. Quantitative conceptions in junior and senior high school students. *Child Development,* 1961b, 32, 551-560.

Elkind, D. Quantity conceptions in college students. *Journal of Social Psychology,* 1962, 57, 459-465.

Elkind, D. Piaget's conservation problems. *Child Development*, 1967, 38, 15-27.

Ellis, J.I. Adolescent awareness of stability and change in literature. *Educational Review*, 1977, 29, 241-253.

Ennis, R.H. Children's ability to handle Piaget's propositional logic: a conceptual critique. *Review of Educational Research*, 1975, 45, 1-41.

Ennis, R.H. Conceptualization of children's logical competence: Piaget's propositional logic and an alternative proposal. In L.S. Siegel and C.J. Brainerd (Eds.), *Alternatives to Piaget: critical essays on the theory*. New York: Academic Press, 1978.

Evans, R.I. *Jean Piaget: the man and his ideas*. New York: Dutton, 1973.

Ferguson, G.A. *Statistical analysis in psychology and education*. New York: McGraw Hill, 1959.

Field, T.W. and Cropley, A.J. Cognitive style and science achievement. *Journal of Research in Science Teaching*, 1969, 6, 2-10.

Flavell, J.H. The developmental psychology of Jean Piaget. *Princeton: Van Nostrand*, 1963.

Frostig, M., Horne, D. and Miller, A.M. *Pictures and patterns*. Chicago: Follett Publishing Company, 1972.

Frostig, M., Lefever, W. and Whittlesey, J.R.B. *Frostig developmental test of visual perception.* Palo Alto, California: Consulting Psychologists Press, 1966.

Gagné, R.M. *The conditions of learning.* New York: Holt, Rhinehart and Winston, 1970.

Gagné, R.M. and Smith, E.C. A study of the effects of verbalisation on problem solving. *Journal of Experimental Psychology,* 1962, 63, 12-18.

Gaudia, G. The Piagetian dilemma: what does Piaget really have to say to teachers? *The Elementary School Journal*, 1974, 481-492.

Giambra, L.M. Selection strategies for eight concept rules with nonexemplar start cards. *Journal of Experimental Psychology,* 1971, 87, 78-92.

Ginsberg, H. and Opper, S. *Piaget's theory of intellectual development.* Englewood Cliffs: Prentice Hall, 1969.

Goldman, R.D. and Hudson, D.J. A multivariate analysis of academic abilities and strategies for successful and unsuccessful college students in different major fields. *Journal of Educational Psychology,* 1973, 65, 364-370.

Goodnow, J.J. and Bethon, G. Piaget's tasks: the effects of schooling and intelligence. *Child Development,* 1966, 37, 573-582.

Gourlay, N. Heredity vs environment: the effects of genetic variation with age. *British Journal of Educational Psychology*, 1978, 48, 1-21.

Guilford, J.P. *Psychometric methods.* New York: McGraw Hill, 1954.

Guilford, J.P. *The nature of human intelligence.* New York: McGraw Hill, 1967.

Hallam, R. Piaget and the teaching of History. *Educational Research*, 1969, 12, 3-12.

Haygood, R. and Bourne, L. Attribute - and rule - learning aspects of conceptual behaviour. *Psychological Review*, 1965, 72, 175-195.

Human Sciences Research Council. *Manual for New South African group test.* Pretoria, 1965.

Human Sciences Research Council. *New South African group test*, Pretoria, 1965.

Inhelder, B. and Piaget, J. *The growth of logical thinking from childhood to adolescence.* London: Routledge and Kegan Paul, 1958.

Kirk, S.A., McCarthy, J. and Kirk, K. *Illinois test of psycholinguistic abilities.* Illinois: University of Illinois, 1968.

Kohlberg, L. and Gilligan, C. The adolescent as a philosopher: the discovery of the self in a postconventional world. *Dædalus,* 1971, 100, 1051-1086.

Laughlin, P.R. Focusing strategy in concept attainment as a function of instructions and task complexity. *Journal of Experimental Psychology,* 1973, 98, 320-327.

Lawson, A.E. Developing formal thought through biology teaching. *American Biology Teacher,* 1975a, 37, 411-419 & 429.

Lawson, A.E. Sex difference in concrete and formal reasoning ability as measured by manipulative tasks and written tasks. *Science Education,* 1975b, 59, 397-405.

Lawson, A.E. *Biology reasoning test.* 1977 Duplicated.

Lawson, A.E. and Blake, A.J.D. Concrete and formal thinking abilities in high school biology students as measured by three separate instruments. *Journal of Research in Science Teaching,* 1976, 13, 227-235.

Lawson, A.E., Blake, A.J.D. and Nordland, F.H. Piagetian tasks clarified: the use of metal cylinders. *American Biology Teacher,* 1974, 209-211.

Lawson, A.E., Blake, A.J.D. and Nordland, F.H. Training effects and generalisation of the ability to control variables in high school biology students. *Science Education,* 1975, 59, 387-396.

Lawson, A.E. and Nordland, F.H. The factor structure of some Piagetian tasks. *Journal of Research in Science Teaching*, 1976, 13, 461-466.

Lawson, A.E., Nordland, F.H. and De Vito, A. Piagetian formal operational tasks: a crossover study of learning effect and reliability. *Science Education*, 1974, 58, 267-276.

Lawson, A.E., Nordland, F.H. and De Vito, A. Relationship of formal reasoning to achievement, aptitudes and attitudes in preservice teachers. *Journal of Research in Science Teaching*, 1975, 12, 423-431.

Lawson, A.E. and Renner, J.W. A quantitative analysis of responses to Piagetian tasks and its implications for curricula. *Science Education*, 1974, 58, 545-559.

Lawson, A.E. and Renner, J.W. Relationship of science subject matter and developmental level of learners. *Journal of Research in Science Teaching*, 1975, 12, 347-358.

Lawson, A.E. and Wollman, W.T. Encouraging the transition from concrete to formal cognitive functioning - an experiment. *Journal of Research in Science Teaching*, 1976, 13, 413-430.

Lengel, R.A. and Buell, R.R. Exclusion of irrelevant factors (The pendulum problem). *Science Education*, 1972, 56, 65-70.

Levine, D.I. and Linn, M.C. Scientific reasoning ability in adolescence: theoretical viewpoints and educational implications. *Journal of Research in Science Teaching*, 1977, 14, 371-384.

Levine, M. Hypothesis behavior in humans during discrimination learning. *Journal of Experimental Psychology*, 1966, 71, 331-338.

Linke, R.D. Replicative studies in hierarchial learning of graphical interpretation skill. *British Journal of Educational Psychology*, 1975, 45, 39-46.

Longeot, F. Un essai d'application de la psychologie génétique a la psychologie différentielle. *Bulletin de l'Institut National D'Etude*, 1962, 18, 153-162.

Longeot, F. Analyse statistique de trois tests génétiques collectifs. *Bulletin de l'Institut National D'Etude*, 1965, 20, 219-237.

Lovell, K. A follow-up study of Inhelder and Piaget's "The growth of logical thinking". *British Journal of Psychology*, 1961, 52, 143-153.

Lovell, K. Some problems associated with formal thought and its assessment. In D.R. Green, M.C. Ford and G.B. Flamer (Eds.), *Measurement and Piaget*. New York: McGraw Hill, 1971.

Lovell, K. and Butterworth, I.B. Abilities underlying the understanding of proportionality. *Mathematics Teaching*, 1966, 37, 5-9.

Lovell, K. and Shields, J.B. Some aspects of a study of the gifted child. *British Journal of Educational Psychology*, 1967, 37, 201-208.

Lucas, A.M. ASEP - a national curriculum development project in Australia. *Science Education*, 1972, 56, 443-451.

Mann, L. and Phillips, W.A. Fractional practices in special education: a critique. *Exceptional Children*, 1967, 33, 311-317.

Marek, E.A. and Renner, J.W. Intellectual development, IQ, achievement and teaching methodology. *American Biology Teacher*, 1979, 41, 145-150.

Mays, W. Logic for juniors. *Teaching Arithmetic*, 1965, 3, 3-10.

Mealings, R.J. Problem-solving in science teaching. *Educational Review*, 1963, XV, 194-207.

McKinney, J.D. Developmental study of the acquisition and utilization of conceptual strategies. *Journal of Educational Psychology*, 1972, 63, 22-31.

McKinnon, J.W. and Renner, J.W. Are colleges concerned with intellectual development? *American Journal of Physics*, 1971, 39, 1047-1052.

McNally, D.W. *Piaget, education and teaching*. Lewes: New Educational Press, 1974.

Modgil, S. Piagetian research: a handbook of recent studies. *NFER*. Slough, Bucks: 1974.

Moerdyk, A.P. The complexity of cognitive structure in relation to scholastic achievement. *Unpublished masters dissertation*. University of Natal, 1973.

Nay, M.A. and Associates. A process approach to teaching science. *Science Education*, 1971, 55, 197-207.

Neale, M.D. Patterns of dyslexia, predictive studies and early intervention. In B.N. Harrison and F. Hooper (Eds.), *Dyslexia Symposium: 1968*. Clayton, Victoria: Australia College of Speech Therapists, 1968.

Neimark, E.D. Intellectual development during adolescence. In Horowitz (F.D.) *Review of child development research*. Chicago: University of Chicago, 1975.

Neimark, E.D. and Santa, J.L. Thinking and concept attainment. *Annual Review of Psychology*, 1975, 26, 173-205.

Neisser, U. and Weene, P. Hierarchies in concept attainment. *Journal of Experimental Psychology*, 1962, 64, 640-645.

Nordland, F.H., Lawson, A.E. and Kahle, J.B. A study of levels of concrete and formal reasoning ability in disadvantaged junior and senior high school science students. *Science Education*, 1974, 58, 569-575.

Novak, J.D. Editorial comments: "Lawson, A.E. Nordland, F.H. and De Vito, A. Piagetian formal operational tasks: a crossover study of learning effect and reliability". *Science Education*, 1974, 58, 269.

Novak, J.D. Editorial comments: "Lawson, A.E., Blake, A.J.D. and Nordland, F.H. Training effects and generalisation of the ability to control variables in high school biology students". *Science Education*, 1975, 59, 431-432.

Nunnally, J.C. *Psychometric theory*. New York: McGraw Hill, 1967.

O'Brien, T.C. and Shapiro, B.J. The development of logical thinking in children, *American Educational Research Journal*, 1968, 5, 531-542.

Paris, S.G. *Propositional logical thinking and comprehension of language connectives*. The Hague: Mouton, 1975.

Parsons, A. Translators introduction: a guide for psychologists. In B. Inhelder and J. Piaget, *The growth of logical thinking from childhood to adolescence*. London: Routledge and Kegan Paul, 1958.

Parsons, C. Inhelder and Piaget's:"The growth of logical thinking" II. A logician's viewpoint. *British Journal of Psychology*, 1960, 51, 75-84.

Philp, H. and Kelly, M. Product and process in cognitive development: some comparative data on the performance of school age children in different

Piaget, J. *The psychology of intelligence.* London: Routledge and Kegan Paul, 1950.

Piaget, J., Inhelder, B. and Szeminska, A. *The child's conception of geometry.* London: Routledge and Kegan Paul, 1960.

Piaget, J. Development and learning. *Journal of Research in Science Teaching,* 1964, 2, 176-186.

Piaget, J. Intellectual evolution from adolescence to adulthood. *Human Development,* 1972, 15, 1-12.

Piaget, J. Genetic epistemology. In R.I. Evans (Ed.), *Jean Piaget: the man and his ideas.* New York: Dutton, 1973.

Piaget, J. and Inhelder, B. *The child's construction of quantities.* London: Routledge, 1974.

Raven, R.J. Programming Piaget's logical operations for science inquiry and concept attainment. *Journal of Research in Science Teaching,* 1974, 11, 251-261.

Rock, I. and Harris, C.S. Vision and touch. *Scientific American,* 1967, 216, 96-104.

Sayre, S. and Ball, D.W. Piagetian cognitive development and achievement in science. *Journal of Research in Science Teaching,* 1975, 12, 165-174.

Shapiro, B.J. and O'Brien, T.C. Logical thinking in children ages six through thirteen. *Child Development,* 1970, 41, 823-829.

Shayer, M., Küchmann, D.E. and Wylam, H. The distrubution of Piagetian stages of thinking in British middle and secondary school children. *British Journal of Educational Psychology*, 1976, 46, 164-173.

Shayer, M. and Wylam, H. The distribution of Piagetian stages of thinking in British middle and secondary school's children.II - 14-to 16-year-olds and sex differentials. *British Journal of Educational Psychology*, 1978, 48, 62-70.

Sheehan, D.J. *The Longeot examination.* 1971. Duplicated.

Siegler, R.S. and Altas, M. Acquisition of formal scientific reasoning by 10- and 13-year-olds: detecting interactive patterns in data. *Journal of Educational Psychology*, 1976, 360-370.

Siegler, R.S. and Liebert, R.M. Acquisition of formal scientific reasoning by 10- and 13-year-olds: designing a factorial experiment. *Developmental Psychology*, 1975, 11, 401-402.

Siegler, R.S., Liebert, D.E. and Liebert, R.M. Inhelder and Piaget's pendulum problem: teaching preadolescents to act as scientists. *Developmental Psychology*, 1973, 9, 97-101.

Sime, M. *A child's eye view.* London: Thames and Hudson, 1973.

Smalley, N.S. Modes of extracting information as a function of selection vs reception paradigms, *Journal of Experimental Psychology,* 1974, 102, 56-61.

Somerville, S.C. The pendulum problem: patterns of performance defining developmental stages. *British Journal of Educational Psychology,* 1974, 44, 266-281.

Stones, E. *An introduction to educational psychology.* London: Methuen, 1966.

Sund, R.B. *Piaget for educators.* Columbus, Ohio: Merril, 1976.

Tisher, R.P. A Piagetian questionnaire applied to pupils in a secondary school. *Child Development,* 1971, 42, 1633-1636.

Tisher, R.P. and Dale, L.G. *Manual for Understanding in science test.* Hawthorn, Victoria: Australian Council for Educational Research, 1975a.

Tisher, R.P. and Dale, L.G. *Understanding in science test.* Hawthorn, Victoria: Australian Council for Educational Research, 1975b.

Towler, J.O. and Wheatley, G. Conservation concepts in college students: a replication and critique. *The Journal of Genetic Psychology,* 1971, 118, 265-270.

Tuddenham, R.D. Theoretical regularities and individual idiosyncrasies. In D.R. Green, M.B. Ford and G.B. Flamer (Eds.), *Measurement and Piaget.* New York: McGraw Hill, 1971.

Vernon, P.E. Intelligence testing and the nature/nurture debate, 1928-1978: What next? *British Journal of Educational Psychology,* 1979, 49, 1-14.

Warburton, F.W. The British intelligence scale. In B. Dockrell (Ed.), *On intelligence.* Toronto, Ontario: Institute for Studies in Education, 1970.

Wason, P.C."On the failure to eliminate hypotheses.."a second look. In P.C. Wason and P.N. Johnson-Laird, *Thinking and reasoning.* Harmondsworth: Penguin, 1968.

Weitz, L.J., Bynum, T.W., Thomas, J.A. and Steger, J.A. Piaget's system of 16 binary operations: an empirical investigation. *Journal of Genetic Psychology,* 1973, 123, 279-284.

Wells, H. and Watson, D. Strategy training and practice in disjunctive concept attainment. *Psychological Reports,* 1965, 17, 925-926.

Wetherick, N.E. Bruner's concept of strategy: an experiment and a critique. *Journal of General Psychology,* 1969, 81, 53-58.

Whimbey, A.E. Something better than Binet. *Saturday Review/World,* 1974, 1, 50-53.

Whimbey, A.E. *Intelligence can be taught.* New York: Dutton, 1975.

Wollman, W. Controlling variables: a Neo-Piagetian developmental sequence. *Science Education,* 61, 385-391.

Wollman, W.T. and Lawson, A.E. Teaching the procedure of controlled experimentation: a Piagetian approach. *Science Education*, 1977, 61, 57-70.

Yudin, L.W. Formal thought in adolescence as a function of intelligence. *Child Development*, 1966, 37, 697-708.

Yudin, L.W. and Kates, S.L. Concept attainment and adolescent development. *Journal of Educational Psychology*, 1963, 54, 177-182.

X

APPENDIX A

THE TEST

Name:

Class:

Date of Birth:

These questions are being used to find out how people of your age think. You must try to answer each question. Once you have put in an answer you may not change it.

SECTION A

1. Watch the demonstration.
   Underline the correct answer.

   The sphere is heavier than the pancake.

   The pancake is heavier than the sphere.

   The pancake and the sphere are equally heavy.

   Give a reason for your answer starting with the following words:

   I think this is the answer because

2. Watch the demonstration.
   Underline the correct statement.

   The sphere will push the water level up more.

   Both objects will push the water level up the same amount.

   The sausage will push the water level up more.

   Give a reason for your answer starting with the following words:

   I think this is the answer because

3. Watch the demonstration.

   The red block is four times as large as the blue block.
   The red block is twice as heavy as the blue block.

   Underline the correct statement.

   Both blocks have an equal chance of sinking.

   The red block is more likely to sink.

   The blue block is more likely to sink.

   Give a reason for your answer starting with the following words:
   I think this is the answer because

4. Watch the demonstration.
Underline the correct statement.

The heavier cylinder will push the water level up more.

The heavier cylinder will push the water level up less.

Both cylinders will push the water level up the same amount.

Give a reason for your answer starting with the following words:
I think this is the answer because

5. Watch the demonstration.
Explain as fully as you can why these objects sink in water.

SECTION B

Four groups of pupils carry out an experiment with a pendulum.
Watch the demonstration of Group 1's experiment.

1. By making certain changes it is possible to vary the speed at which a pendulum swings to and fro.
What THREE things could be changed in an attempt to vary the speed at which this pendulum swings to and fro?

i)

ii)

iii)

Group 1's pendulum had a long string, a large weight, a low drop and swung to and fro slowly. (approximately 16 times in 30 seconds).

Watch the demonstration of Group 2's experiment. Group 2's experiment had

a short string, a small weight, a high drop and swung to and fro rapidly. (approximately 30 times in 30 seconds)

So far we have done two experiments.
You have seen that: The string can be long or short
The weight can be large or small
The drop can be high or low.

2. Suppose you think that the length of the string affects the speed at which the pendulum swings to and fro, which ONE of the following combinations would you use in an experiment to find out if the length of the string does affect the speed at which the pendulum swings to and fro?
Using a ruler, underline the combination you would use.

long string	large weight	low drop
long string	large weight	high drop
long string	small weight	low drop
long string	small weight	high drop
short string	large weight	low drop
short string	large weight	high drop
short string	small weight	low drop
short string	small weight	high drop

3. Suppose you think that the size of the weight affects the speed at which the pendulum swings to and fro. Which ONE of the following combinations would you use in an experiment to find out if the size of the weight does affect the speed at which the pendulum swings to and fro?
Using a ruler, underline the combination you would use.

long string	large weight	low drop
long string	large weight	high drop
long string	small weight	low drop
long string	small weight	high drop
short string	large weight	low drop
short string	large weight	high drop
short string	small weight	low drop
short string	small weight	high drop

4. Suppose you think that the height of the drop affects the speed at which the pendulum swings to and fro. Which ONE of the following combinations would you use in an experiment to find out if the height of the drop does affect the speed at which the pendulum swings to and fro?
Using a ruler, underline the combination you would use.

long string	large weight	low drop
long string	large weight	high drop
long string	small weight	low drop
long string	small weight	high drop
short string	large weight	low drop
short string	large weight	high drop
short string	small weight	low drop
short string	small weight	high drop

Next, Group 3 and Group 4 try to find out what affects the speed at which the pendulum swings to and fro.

Study the summary of results below and answer the questions which follow.

	LENGTH OF STRING	SIZE OF WEIGHT	HEIGHT OF DROP	SWINGS TO & FRO
Group 1	long	large	low	slowly
Group 2	short	small	high	rapidly
Group 3	short	large	low	rapidly
Group 4	long	small	low	slowly

5. Consider the length of the string.
Which of the following statements is correct? Underline your choice.

The length of the string has no effect on the speed at which the pendulum swings to and fro.

The longer the string the faster the pendulum swings to and fro.

The longer the string the slower the pendulum swings to and fro.

All three factors have an equal effect on the speed at which the pendulum swings to and fro.

There is not enough information to be sure of an answer.

6. Consider the size of the weight.
Which of the following statements is correct? Underline your choice.

The size of the weight has no effect on the speed at which the pendulum swings to and fro.

The larger the weight the faster the pendulum swings to and fro.

The larger the weight the slower the pendulum swings to and fro.

All the factors have an equal effect on the speed at which the pendulum swings to and fro.

There is not enough information to be sure of the answer.

7. Consider the height of the drop.
Which of the following statements is correct? Underline your choice.

The height of the drop has no effect on the speed at which the pendulum swings to and fro.

The lower the drop the faster the pendulum swings to and fro.

The lower the drop the slower the pendulum swings to and fro.

All three factors have an equal effect on the speed at which

8. Which of the following methods is best for finding out the effect of the length of the string? Underline your choice.

Compare experiments done by Group 1 and Group 3.

Compare experiments done by Group 1 and Group 4.

Compare experiments done by Group 2 and Group 3.

Compare experiments done by Group 2 and Group 4.

Compare experiments done by Group 3 and Croup 4.

One must use a combination of the above methods.

9. Which of the following methods is best for finding out the effect of the size of the weight? Underline your choice.

Compare experiments done by Group 1 and Group 3.

Compare experiments done by Group 1 and Group 4.

Compare experiments done by Group 2 and Group 3.

Compare experiments done by Group 2 and Group 4.

Compare experiments done by Group 3 and Group 4.

One must use a combination of the above methods.

10. Which of the following methods is best for finding out the effect of the height of the drop? Underline your choice.

Compare experiments done by Group 1 and Group 3.

Compare experiments done by Group 1 and Group 4.

Compare experiments done by Group 2 and Group 3.

Compare experiments done by Group 2 and Group 4.

Compare experiments done by Group 3 and Group 4.

One must use a combination of the above methods.

----- o0o -----

XI

APPENDIX B

TESTER'S COMMENTARY AND MARK SCHEME

This appendix gives details on the method of administration and the mark-scheme used when evaluating test results. The test sections and spoken instruction are arranged in the sequence in which the test was administered. When the pupils were seated the following instructions were given:

> Today we are going to do some problems. Some are easy; some are difficult. This test will not in any way affect your school marks - so you must work entirely independently.
>
> The questions will take about 40 minutes to answer.
>
> I will go through each question with you. If you do not understand what you have to do put up your hand and I will come to you and try to help you.

The papers were then given out. The tester read through the test with the pupils who waited after each item until the whole class was ready to proceed.

Name:

Class:

Date of Birth:

These questions are being used to find out how people of your age think. You must try to answer each question. Once you have put in an answer you may not change it. 1

---

1 The last statement was modified by instructions given during the test. Changes were allowed until a page was turned.

Please fill in your name,

class e.g. 4B,

full date of birth (not 1978).

We will now work through the questions together.

SECTION A

1. Watch the demonstration.

Here are two spheres of plasticine. (The spheres are given to a pupil.) To the best of your judgement are they equally heavy? If not please adjust the quantity of plasticine and make them equally heavy. (The spheres are given back.) I now deform the one sphere into a pancake.

Correct multichoice answers in A1 - A4 are underlined.

Underline the correct answer.

The sphere is heavier than the pancake.

The pancake is heavier than the sphere.

<u>The pancake and the sphere are equally heavy</u>.

Give a reason for your answer starting with the following words:

I think this is the answer because

Mark scheme for A1

The pancake and the sphere are equally heavy 1

i.e. multichoice correct.

They have the same amount of matter / mass / quantity 1

This is an affirmation of equivalence of mass -

The subject argues the conservation of weight with reference to the conservation of mass. An elaborated argument could be:

> The mass must be the same; therefore the weight must be the same as they are directly proportional to each other at any one point on earth.

A mark was not given for saying "They are the same" or "They are equally heavy" as this was deemed to be repitition of the multichoice statement.

The two pieces are the same weight. 1

A mark was given here because it indicated the correct understanding of the concept of heavy; i.e. that heaviness was a property associated with weight.

They were the same before (They could be made the same as before) 1

Argument by reversibility

No plasticine was taken away 1

No plasticine was added 1

Both the above are part of the argument by identity

The deformation was of no consequence 1

The spatial form is irrelevant

The pancake was wide but thin 1

Argument by compensation

Total 8

2. Watch the demonstration.

> Here are two more spheres of plasticine.
>
> Do these two spheres - as far as you can judge - take up the same amount of space or room? (The spheres are handed to a pupil.)
>
> If not adjust the plasticine until they take up the same amount of space or room.
>
> Are these two beakers - as far as you can judge - filled to the same level? (The beakers are placed in front of another pupil.)

If not adjust the water level until they are. (The spheres and beakers are collected.)

The one sphere is rolled into a sausage. Suppose I submerge the sphere in this beaker and the sausage in this beaker.

Underline the correct statement.
The sphere will push the water level up more.
Both objects will push the water level up the same amount.
The sausage will push the water level up more.
Give a reason for your answer starting with the following words:
I think this is the answer because

Mark scheme for A2

Both objects will push the water level up the same amount	1
i.e. multichoice item correct	
They have the same volume / space / room / size	1
This is an affirmation of the equivalence of volume.	
They were the same before (They could be made the same as before)	1
Argument by reversibility	
No plasticine was taken away	1
No plasticine was added	1
Both of the above are part of the argument by identity	
The deformation was of no consequence	1
The spatial form is irrelevant	
Because the objects were of the same volume they cause the same quantity of displacement	1
Considering objects in relation to a continuous medium acting on the objects from all directions simultaneously.	
Total	7

3. Watch the demonstration.
The red block is four times as large as the blue block.
The red block is twice as heavy as the blue block.

Here are two blocks of two different unknown substances. I want you to decide which block is more likely to sink - in other words you must decide which block has the greater chance of sinking.

Underline the correct statement.
Both blocks have an equal chance of sinking.
The red block is more likely to sink.
The blue block is more likely to sink.
Give a reason for your answer starting with the following words:
I think this is the answer because

Before you turn over make sure you have completed all questions on this page. Once you have turned over you will not be allowed to turn back and make any changes to these answers.

Mark scheme for A3

The blue block is more likely to sink	1
i.e. multichoice item correct	
The blue block is more dense than the red block	1
The subject may have limited understanding of density so marks are given for explicit reasons e.g.	
The mass volume ratio is greater in the case of the blue block	1
than in the case of the red block	1
Red is four times larger but only two times heavier	1
Total	5

Here are two metal cylinders - one copper and one aluminium.

(The cylinders are given to a pupil.)

Are both of equal width and height? Yes. Which cylinder is heavier? - Yes, the copper cylinder.

Are the water levels in the two beaders equal? If they are not, adjust the levels until they are. (The cylinders and beakers are collected.)

Suppose the cylinders were each submerged into a beaker of water.

-2-

4. Watch the demonstration.

Mark scheme for A4

Both cylinders will push the water level up the same amount i.e. multichoice item correct	1
The cylinders are of the same volume or size or take up the same space or room.	1
Because they are equal in volume or size they displace the same quantity of water.	1
The mass of the cylinders does not affect the displacement of the water	1
Total	4

Underline the correct statement.

The heavier cylinder will push the water level up more.

The heavier cylinder will push the water level up less.

Both cylinders will push the water level up the same amount.

Give a reason for your answer starting with the following words:

I think this is the answer because

5. Watch the demonstration.

I will drop in a number of objects:

a pin

a curtain ring

a marble

a drawing pin.

Explain <u>as fully as you can</u> why these objects sink in water.

Mark scheme for A5

The objects have greater density than water	*1*
Surface tension is not strong enough to play a part	*1*
The volume mass ratio leads to sinking	*1*
The packing of particles is closer in substances which are more dense	*1*
1cm^3 of the substances is heavier than 1cm^3 of water	
OR the object is heavier than an equal volume of water	
OR the upward force is greater than the downward force	*1*
Total	5

<u>SECTION B</u>

Four groups of pupils carry out an experiment with a pendulum. Watch the demonstration of Group 1's experiment.

This is a pendulum. Notice how it swings.

1. By making certain changes it is possible to vary the speed at which a pendulum swings to and fro.
What THREE things could be changed in an attempt to vary the speed at which this pendulum swings to and fro?

i)

ii)

iii)

Before you turn over make sure you have completed all questions on this page. Once you have turned over you will not be allowed to turn back and make any changes to these answers.

Mark scheme for B1i, ii, iii

Three of:

1.	*Length of string*	1
2.	*Size of weight*	1
3.	*Height of drop*	1
4.	*Degree of push*	1

and any other valid suggestions.

Group 1's pendulum had a long string, a large weight, a low drop and swung to and fro slowly. (approximately 16 times in 30 seconds).

Watch the demonstration of Group 2's experiment.

A second pendulum is allowed to swing.

Group 2's

experiment had

a short string, a small weight, a high drop and swung to and fro rapidly. (approximately 30 times in 30 seconds)

So far we have done two experiments.
You have seen that: The string can be long or short
The weight can be large or small
The drop can be high or low.

The correct answers for B2 - B10 are indicated by means of the correct answers being underlined.

2. Suppose you think that the length of the string affects the speed at which the pendulum swings to and fro, which ONE of the following combinations would you use in an experiment to find out if the length of the string does affect the speed at which the pendulum swings to and fro?

You have seen two experiments. You are now required to design a third experiment.

Using a ruler, underline the combination you would use.

long string	large weight	low drop
long string	large weight	high drop
long string	small weight	low drop
long string	small weight	high drop
<u>short string</u>	<u>large weight</u>	<u>low drop</u>
short string	large weight	high drop
short string	small weight	low drop
short string	small weight	high drop

3. Suppose you think that the size of the weight affects the speed at which the pendulum swings to and fro. Which ONE of the following combinations would you use in an experiment to find out if the size of the weight does affect the speed at which the pendulum swings to and fro?

You have seen two experiments. You are now required to design a third experiment.

Using a ruler, underline the combination you would use.

long string	large weight	low drop
long string	large weight	high drop
long string	small weight	low drop
long string	small weight	high drop
short string	large weight	low drop
short string	large weight	high drop
short string	small weight	low drop
short string	small weight	high drop

4. Suppose you think that the height of the drop affects the speed at which the pendulum swings to and fro. Which ONE of the following combinations would you use in an experiment to find out if the height of the drop does affect the speed at which the pendulum swings to and fro?

You have seen two experiments. You are now required to design a third experiment.

Using a ruler, underline the combination you would use.

long string	large weight	low drop
long string	large weight	high drop
long string	small weight	low drop
long string	small weight	high drop
short string	large weight	low drop
short string	large weight	high drop
short string	small weight	low drop
short string	small weight	high drop

Before you turn over make sure you have completed all questions on this page. Once you have turned over you will not be allowed to turn back and make any changes to these answers.

Next, Group 3 and Group 4 try to find out what affects the speed at which the pendulum swings to and fro.

Study the summary of results below and answer the questions which follow.

As each pendulum was described a demonstration was given.

Next, Group 3 and Group 4 try to find out what affects the speed at which the pendulum swings to and fro.

Study the summary of results below and answer the questions which follow.

	LENGTH OF STRING	SIZE OF WEIGHT	HEIGHT OF DROP	SWINGS TO & FRO
Group 1	long	large	low	slowly
Group 2	short	small	high	rapidly
Group 3	short	large	low	rapidly
Group 4	long	small	low	slowly

5. Consider the length of the string.
Which of the following statements is correct? Underline your choice.

   The length of the string has no effect on the speed at which the pendulum swings to and fro.

   The longer the string the faster the pendulum swings to and fro.

   The longer the string the slower the pendulum swings to and fro.

   All three factors have an equal effect on the speed at which the pendulum swings to and fro.

   There is not enough information to be sure of an answer.

6. Consider the size of the weight.
Which of the following statements is correct? Underline your choice.

   The size of the weight has no effect on the speed at which the pendulum swings to and fro.

   The larger the weight the faster the pendulum swings to and fro.

   The larger the weight the slower the pendulum swings to and fro.

   All the factors have an equal effect on the speed at which the pendulum swings to and fro.

   There is not enough information to be sure of the answer.

7. Consider the height of the drop.
Which of the following statements is correct? Underline your choice.

   The height of the drop has no effect on the speed at which the pendulum swings to and fro.

   The lower the drop the faster the pendulum swings to and fro.

   The lower the drop the slower the pendulum swings to and fro.

   All three factors have an equal effect on the speed at which the pendulum swings to and fro.

8. Which of the following methods is best for finding out the effect of the length of the string? Underline your choice.

   Compare experiments done by Group 1 and Group 3.

   Compare experiments done by Group 1 and Group 4.

   Compare experiments done by Group 2 and Group 3.

   Compare experiments done by Group 2 and Group 4.

   Compare experiments done by Group 3 and Group 4.

   One must use a combination of the above methods.

9. Which of the following methods is best for finding out the effect of the size of the weight? Underline your choice.

   Compare experiments done by Group 1 and Group 3.

   Compare experiments done by Group 1 and Group 4.

   Compare experiments done by Group 2 and Group 3.

   Compare experiments done by Group 2 and Group 4.

   Compare experiments done by Group 3 and Group 4.

   One must use a combination of the above methods.

10. Which of the following methods is best for finding out the effect of the height of the drop? Underline your choice.

    Compare experiments done by Group 1 and Group 3.

    Compare experiments done by Group 1 and Group 4.

    Compare experiments done by Group 2 and Group 3.

    Compare experiments done by Group 2 and Group 4.

    Compare experiments done by Group 3 and Group 4.

    One must use a combination of the above methods.

Before you close your booklets make sure you have completed all the questions on both these pages.

APPENDIX C

PLATES

The following plates show the apparatus and critical features of the presentation of the two sections of the test.

SECTION A

PLATE 1

ITEM A1: THE TWO PLASTICINE SPHERES

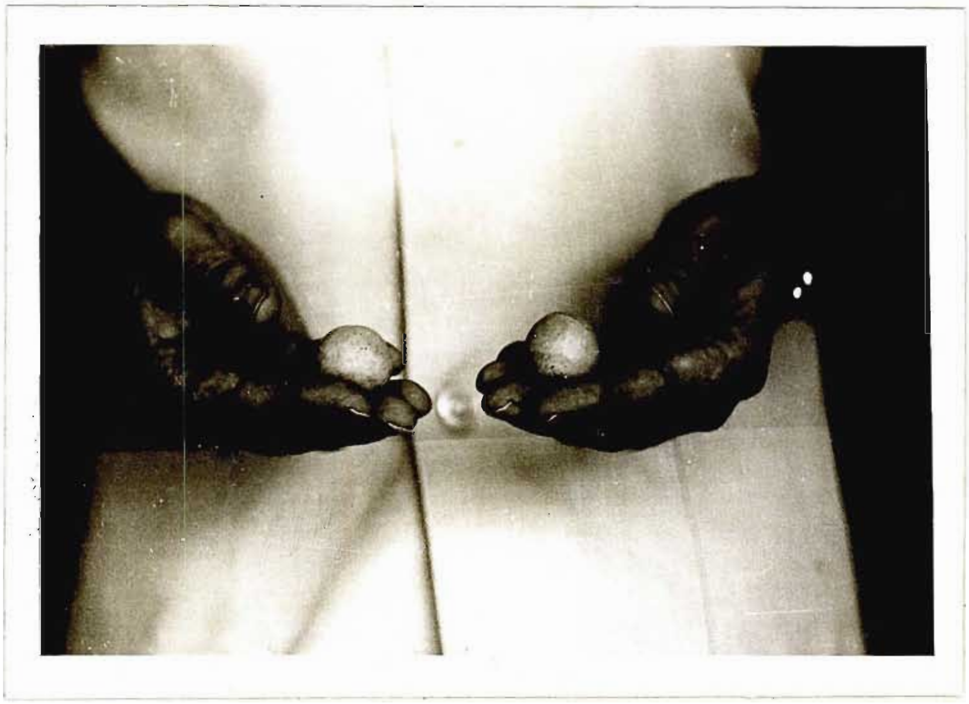


PLATE 2

ITEM A1: THE ONE SPHERE HAS BEEN DEFORMED

PLATE 3

ITEM A2: THE TWO PLASTICINE SPHERES

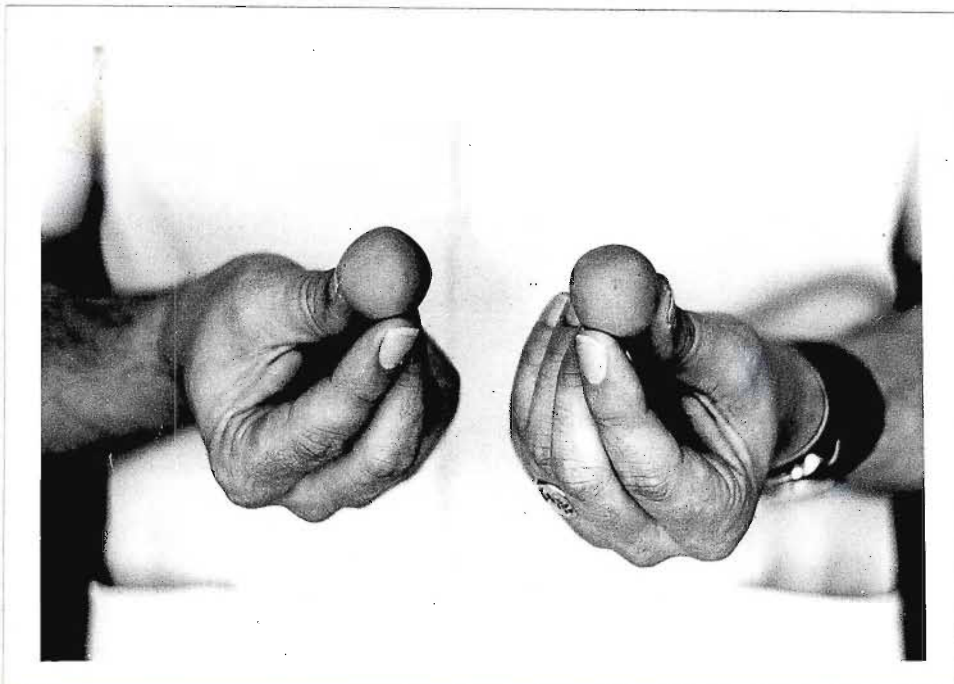


PLATE 4

ITEM A2: THE ONE SPHERE HAS BEEN DEFORMED

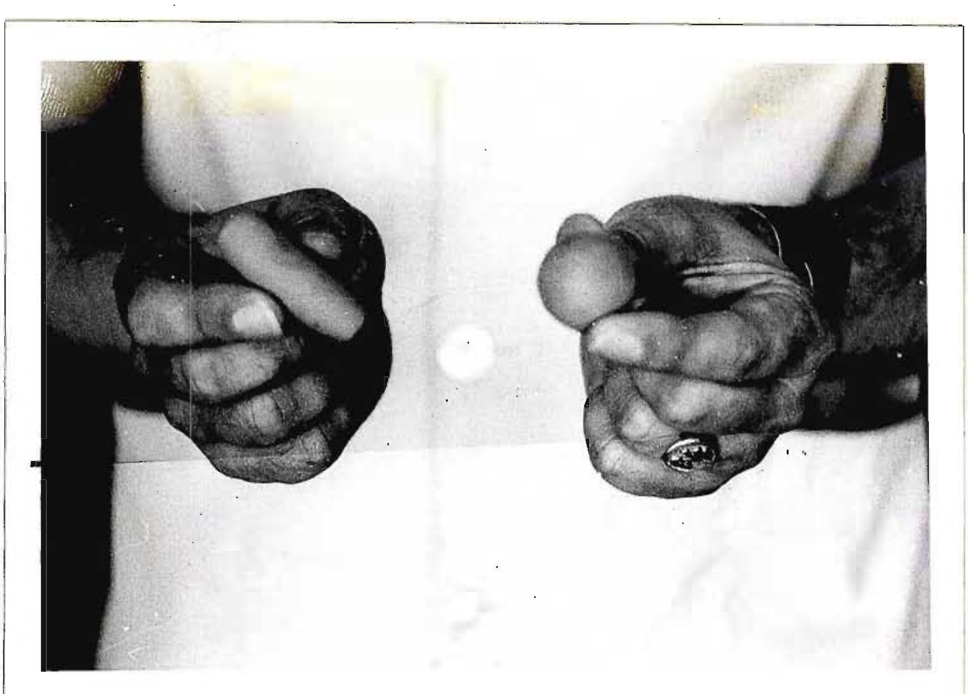


PLATE 5

ITEM A2: THE TWO OBJECTS AND THE TWO BEAKERS OF WATER

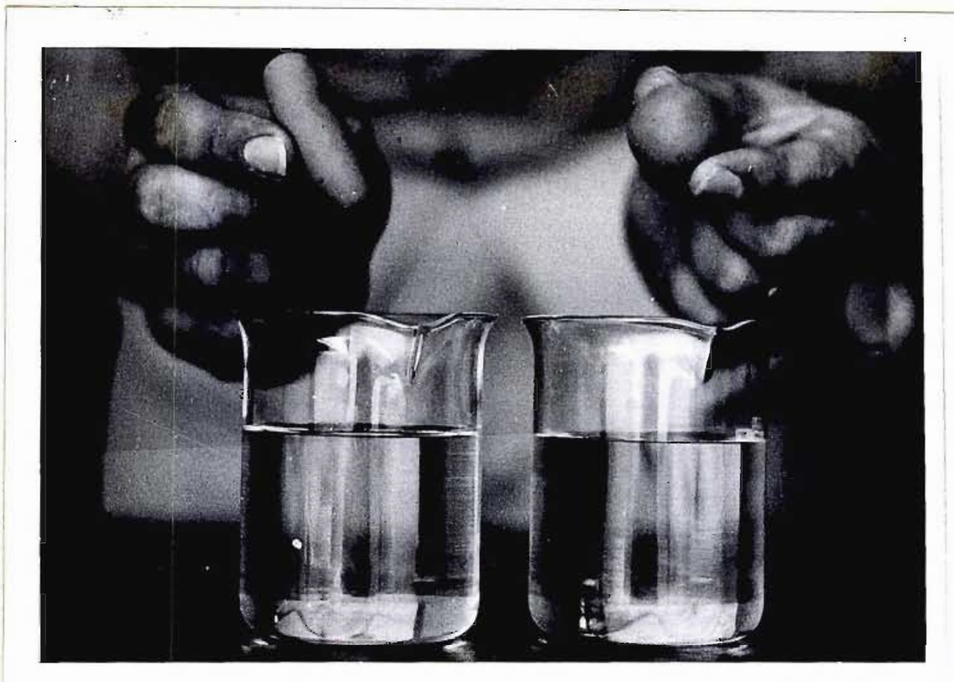


PLATE 6

ITEM A3: THE TWO BLOCKS

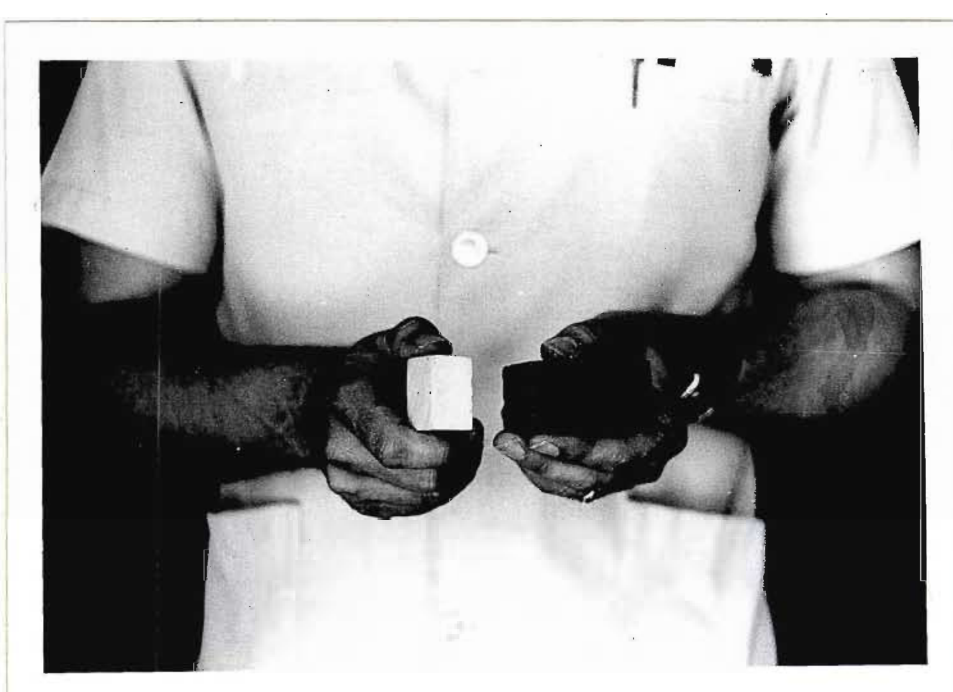


PLATE 7

ITEM A4: THE WIDTH OF THE CYLINDERS IS EQUAL

PLATE 8

ITEM A4: THE HEIGHT OF THE CYLINDERS IS EQUAL

PLATE 9

ITEM A4: THE TWO CYLINDERS AND THE TWO BEAKERS OF WATER

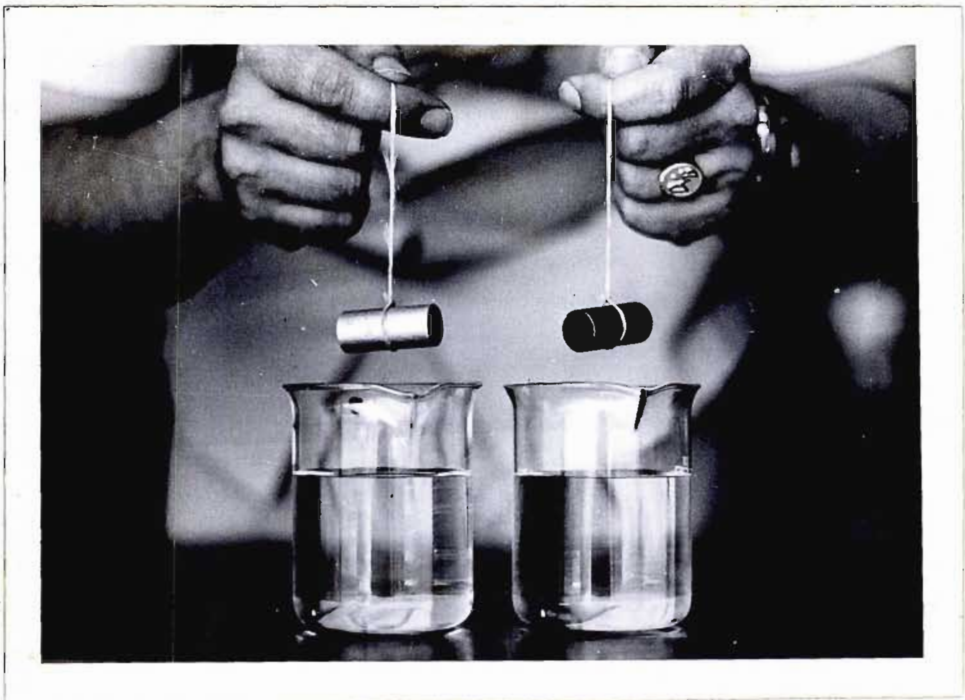
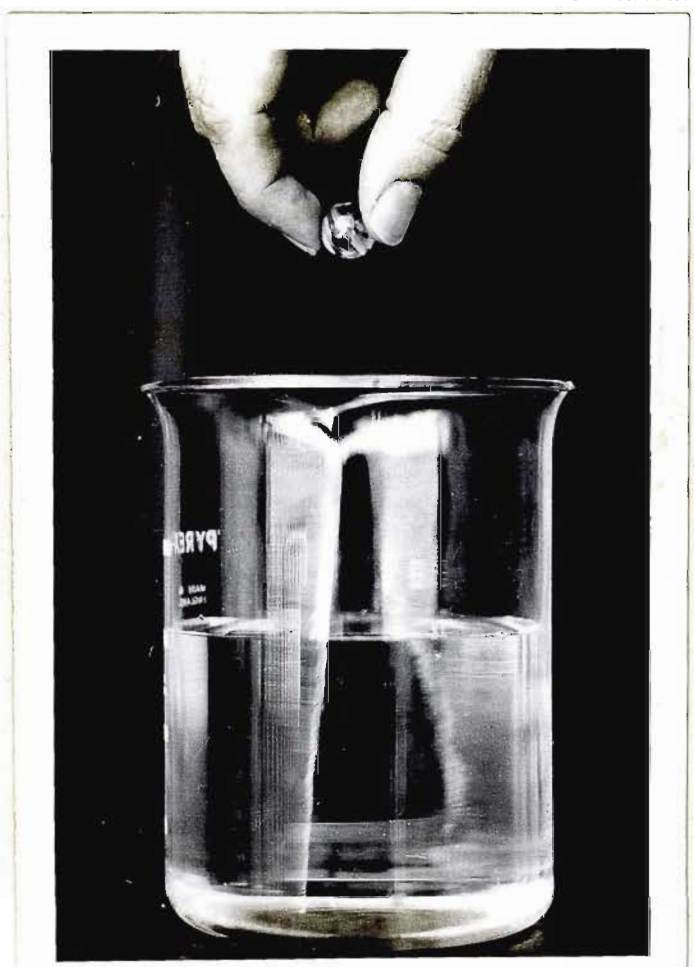


PLATE 10

ITEM A5: THE MARBLE AND THE BEAKER OF WATER



SECTION B[1]

PLATE 11

PENDULUM WITH LONG STRING,

LARGE WEIGHT AND LOW DROP

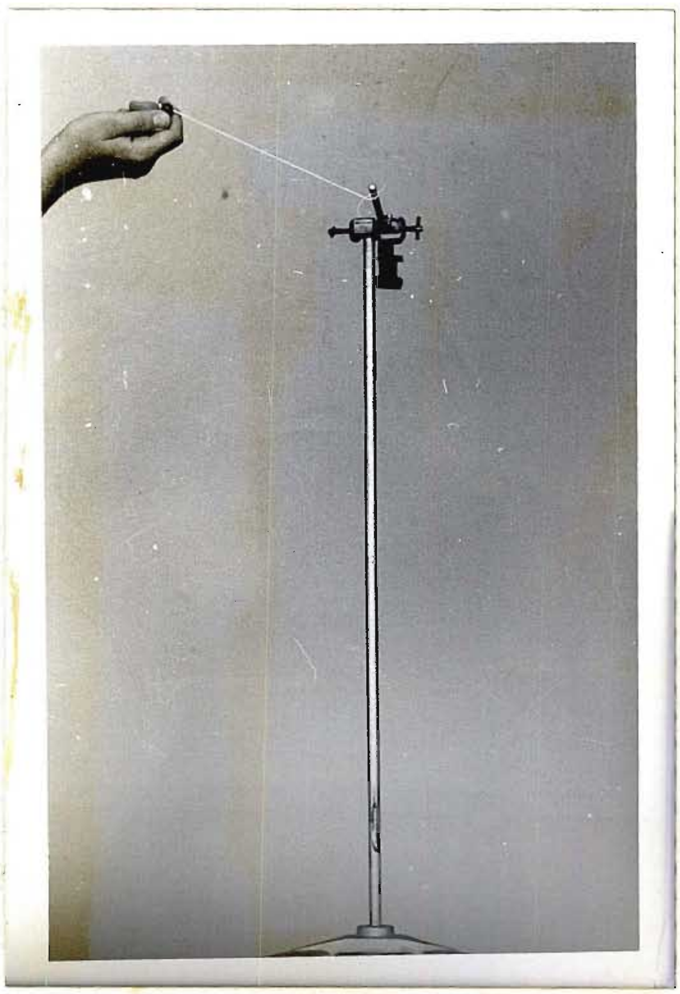


PLATE 12

PENDULUM WITH SHORT STRING,

SMALL WEIGHT AND HIGH DROP

1 It is inappropriate to assign each plate to a particular item as in

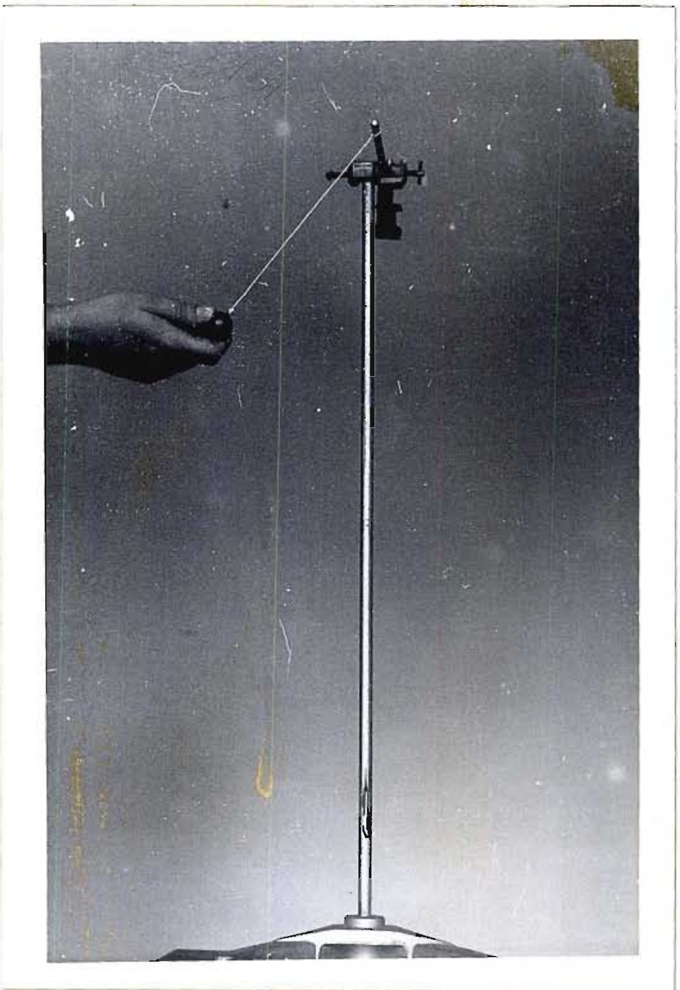


PLATE 13

PENDULUM WITH SHORT STRING, LARGE WEIGHT AND LOW DROP

PLATE 14

PENDULUM WITH LONG STRING, SMALL WEIGHT AND LOW DROP