

https://dspace.org/

The acquisition of formal scientific reasoning by physical science pupils in standards eight and nine

Chandler, Heather Anne

2014-12-12
http://hdl.handle.net/10539/15967

Downloaded from a DSpace repository. https://dspace.org/

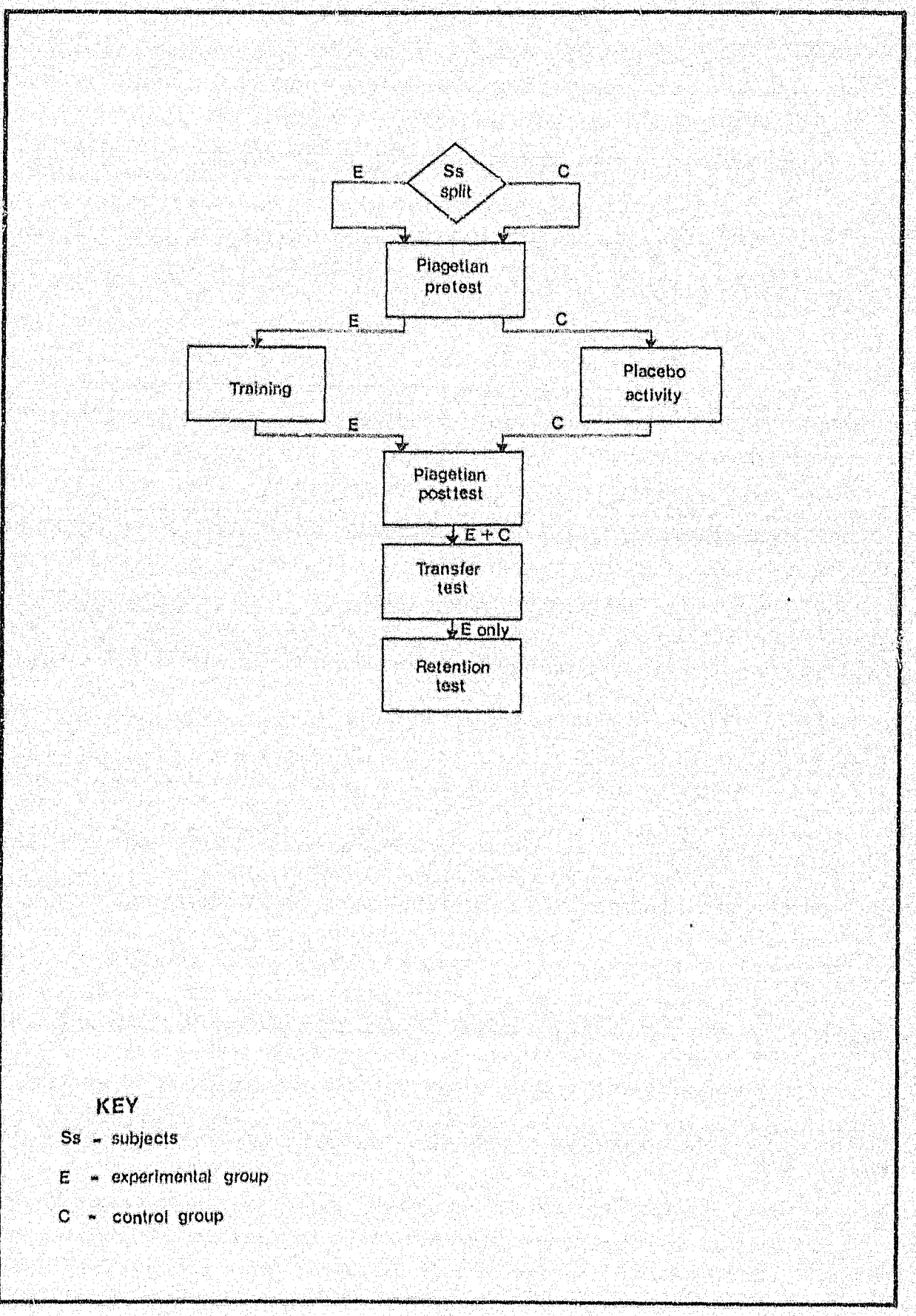


Figure 3.1 Flow Diagram Showing the Principle of the Piagetian Approach to a Training Study Designed to Enhance Intellectual Growth

The general attributes and experimental sequence of a study utilising Ausubel's principle of progressive differentiation to promote learning, are shown in Figure 3.2. The investigation commences with the administration of a pretest to all the subjects on the intended academic material. Relevant subsuming concepts are next introduced to the experimental subjects. As was discussed in Chapter 2, the meaning of an advance organizer has not been operationally defined by Ausubel and there is not yet any apparent consensus concerning such a definition. The instructional material to be learned is then presented, followed by a posttest which usually corresponds to the pretest. Finally, in the more valid studies, retention effects and transfer of learning are examined.

If one wishes to synthesise the approaches of Piaget and Ausubel, then in principle the sequence of the experimental procedure is given in the flow diagram shown in Figure 3.3. On cursory inspection, the learning of academic material, designated as the seventh operation, seems to have been solely a replication from the Ausubelian flow diagram in Figure 3.2. Closer inspection should reveal that this operation is the connecting link between the Piagetian and Ausubelian models. Although the Piagetian posttest (operation 6) which is an indicator of the effectiveness of training, precedes the learning of the academic material, the ultimate aim of the training in formal reasoning is to enable better understanding of abstract concepts presented in the instructional material, since the basic idea behind a synthesis of the two paradigms is the fostering of optimum learning of the same novel academic matter. Since the Piagetian and Ausubelian facets of the investigation are not independent of one another, the ninth and tenth operations in the flow diagram, which refer to transfer and retention, now take on a slightly different meaning and involve measures relating to both the training and the learning of the academic material.

The order of procedure, as indicated by the flow diagram, must not be regarded as totally inflexible. Theoretically, the pretest on academic material (operation 2) may precede the Piagetian pretest (operation 1). As a further example, the Piagetian posttest (operation 6) could be administered prior to the introduction of subsuming concepts (operation 4). On the other hand, the subsuming concepts could be introduced before the training (operation 3) takes place or could even be presented concurrently with the training and Piagetian posttest. Similarly, the Piagetian aspects of the transfer test (operation 9) and the retention test (operation 10) may precede the learning of academic material (operation 7) and the posttest on the academic

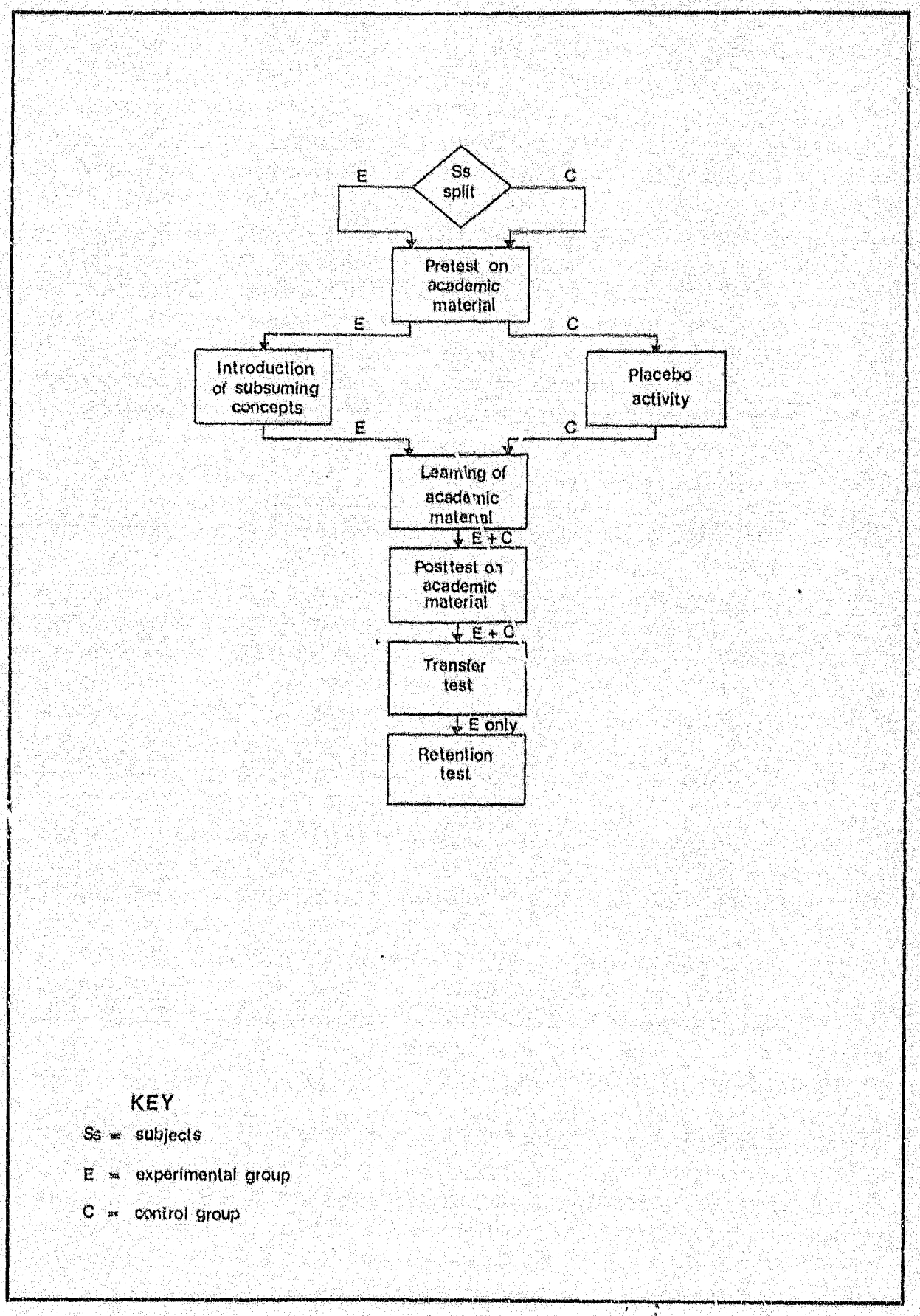


Figure 3.2 Flow Diagram Showing the Principle of the Ausubelian Approach to a Study Designed to Promote Learning

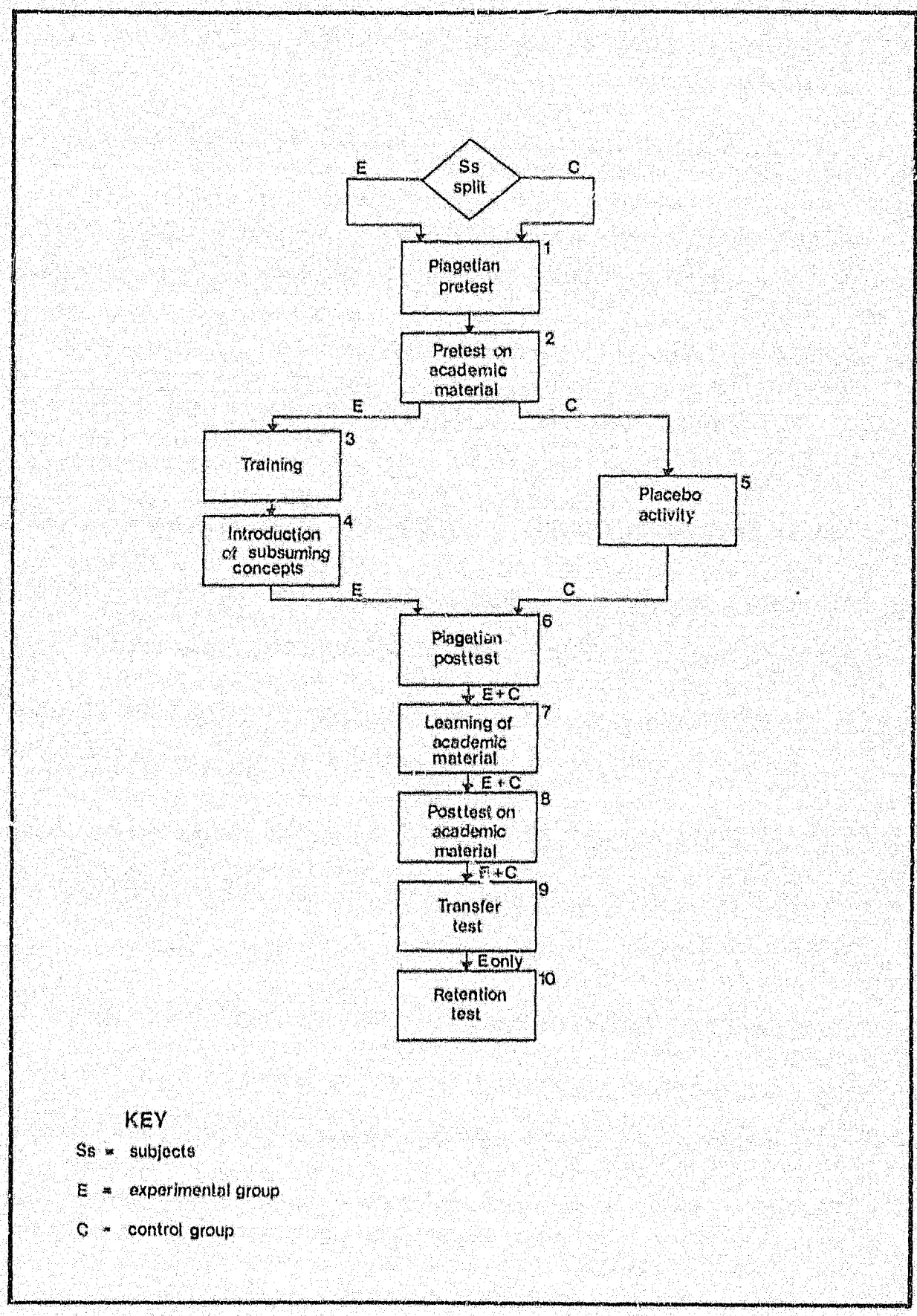


Figure 3.3 Flow Diagram Showing the Principle of the Synthesis of Piagetian and Ausubelian Approaches

material (operation 8). Any changes in sequence, however, should endeavour to maintain a methodical order of procedure to avoid confusion. The exact order of procedure, as adopted in practice, was determined by several considerations which will be discussed in the following paragraphs.

The sequence of the experimental procedure as performed is given in the flow diagram shown in Figure 3.4. The most salient feature of the modifications to the arrangement depicted in Figure 3.3, is the concomitance of the training and the introduction of advance organizers. The second operation in the flow diagram given in Figure 3.4 is thus bifunctional : the training procedure as designed and presented in this thesis was identical with the provision of advance organizers. It was hoped that limiting the number of different experimenter-provided strategies would assist the learning process by avoiding the mental confusion which might arise among the subjects from a multiplicity of contrived teaching aids. Discussion of this dual function is clearly necessary and will be found in Chapter 5 which details the training method. For the moment, let it be said that the training met the conditions for advance organizers on two counts. Firstly, the training consisted of material relevant to both the Piagetian pretest and the learning task which was based on the school physical science syllabus. Further, the training presented this material to the subjects at a higher level of generality and inclusiveness than both these tasks. The extent of deviation from Ausubelian definitions is discussed in Chapter 5.

Any Piagetian study provokes the question as to whether a posttest in the form of a single task can be an adequate and reliable measure of the scheme involved, as was discussed in Chapter 2. The limitations imposed on the present study did not allow extensive investigation into this aspect of training. However, this study is not concerned with precise allocation of Piagetian stages as it is focused on assessment of the effectiveness of training. Nevertheless, the Piagetian posttest would be put on a more valid scientific footing by the introduction of an additional task requiring the same logical operations. With the prime objective of increasing the value of any conclusions concerning training benefits, an addition to the basic scheme as shown in Figure 3.3, was the Piagetian-like task which is designated as the fourth operation in Figure 3.4 and whose exact nature is discussed in Chapter 6. This task was presented before the posttest which, like the pretest, consisted of a traditional Inhelder task.

The Piagetian and Ausubelian aspects of this study are closely interwoven by means of the common links, the single operation of the training and introduction of advance organizers (operation 2) and the learning of academic material (operation 9).

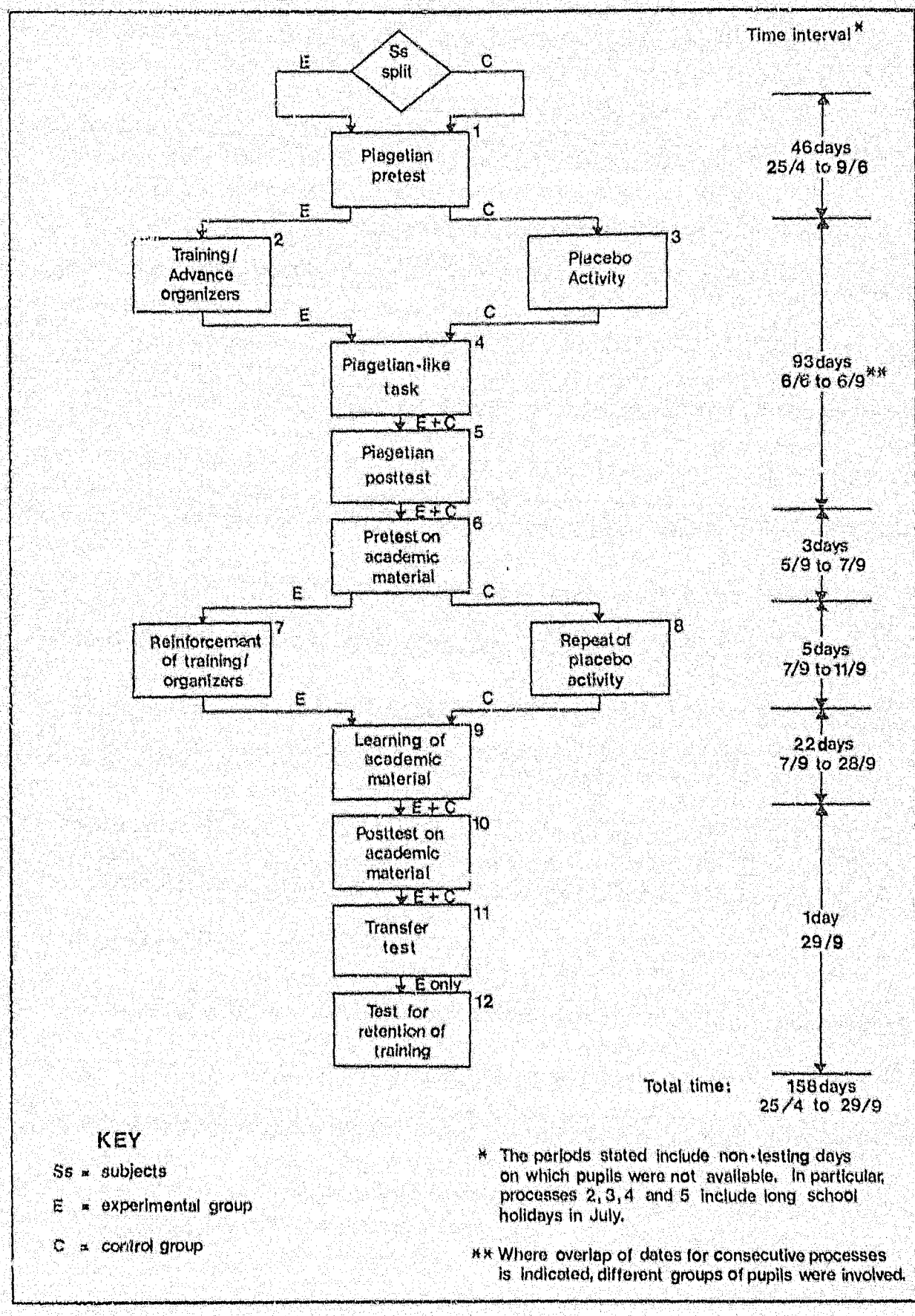


Figure 3.4 Flow Diagram Showing the Synthesis of Piagetian and Ausubelian Approaches as Performed

Nevertheless, the re-allocation of the pretest on the academic material to the sixth position in the flow diagram of Figure 3.4 served to divide the investigation into two roughly distinguishable sections, the first half with a Piagetian focus, comprising operations one to five and the second half with an Ausubelian focus, comprising the remaining operations except for the posttest, transfer and retention tests which involved both approaches and were placed at the end of the study. This division seems natural and methodical. Further, the pretest as the sixth operation served a purpose in addition to its fundamental capacity : it also functioned as a test of whether participation in the earlier portion of the study had affected the knowledge or experience of the experimental and/or control groups in any way which might influence subsequent assimilation of the instructional material. This information can be furnished by including, in the pretest only, a third group of subjects who have been non-participants in the study as a whole.

The right-hand side of Figure 3.4 gives a time-scale against which the various phases of the investigation may be measured. Testing was conducted seven days a week in order to minimise the duration of the training aspects of the study which was necessary to preclude, as far as possible, intellectual maturation resulting from natural encounters as opposed to the contrived experiences of the laboratory setting. Inevitably, however, there were days when subjects were not available for testing, particularly during the long school holidays in July which occurred during operations two to five. The completion of operations two to five took ninety-three days, as shown in Figure 3.4. This period clearly applies to the subjects as a group, not singly. Each subject underwent individual training or placebo treatment and completed the Piagetian-like task and the Piagetian posttest, both individually administered, all within the space of three days, as described later in this chapter.

The large sample of subjects necessary for meaningful statistical analysis of the study, meant that considerable time elapsed between the operation of training/ introduction of advance organizers to the first experimental subjects and administration of the pretest on the academic material. Reinforcement of training/ advance organizers was thus considered important and was inserted as the seventh operation in Figure 3.4, taking place in small groups just prior to presentation of the instructional material. A corresponding activity, the eighth operation, was organised for the control subjects, consisting of a repetition of the placebo treatment in the third operation.

The subjects were taught the academic material under conditions simulating the classroom situation, namely, in classes (of which there were nine, all of non-random composition as explained in Chapter 9). The time interval between instruction and administration of the posttest needed to be the same for the nine classes. Since all the classes had attended the course of tuition within the space of the same seven working days, (not including drill and revision lessons), the subjects would have to be present on approximately the same day for the administration of the posttest. A similar situation clearly applied to the experimental subjects and the requirements for their test on the retention of training. These considerations dictated group assessment by means of written test instruments. The test of transfer could conveniently be included in the same testing schedule. In practice, the posttest, transfer and retention tests, shown in Figure 3.4 as operations ten, eleven and twelve respectively, were typed on the same question-paper and sat as a formal examination on a single occasion by the subjects.

The time-scale in Figure 3.4 seems to indicate that the posttest on the learning material was administered on the day following the completion of instruction. However, as the more detailed commentary in Chapter 9 points out, instruction on subject-matter that was actually unfamiliar, was given for the first seven working days of the ninth operation in Figure 3.4, after which drill and revision lessons were presented. The posttest was therefore a kind of delayed posttest administered fourteen days after presentation of the last novel academic material. Finally, the dates accompanying the time-scale show that the test for the retention of training took place between eighteen and twenty-two days after group reinforcement of the training, depending on the group to which a particular subject had been allocated, and anything up to one hundred and twelve days from the time that subject had undergone individual training.

This section has introduced the basic research design, further particulars of which will be supplied at appropriate stages in the thesis.

3.2 Choice of the Statistical Test

3.2.1 Statistical Requirements of the Research Design

Each of the hypotheses featured in the research design necessitates testing whether two (or more) samples were drawn from populations which differ in location (central tendency).

3.2.2 Rationale for Choosing among Alternative Statistical Tests

Since alternative tests are available, choice of a test must take the power of the test into consideration. Further, the most appropriate test will be the one whose statistical model most closely approximates the conditions of the research in terms of the assumptions which qualify use of the test and whose measurement requirement is met by the measurement scale achieved in the research. Should two or more tests satisfy these criteria, final choice must depend on power-efficiency.

3.2.2.1 Choice between Parametric and Nonparametric Statistical Techniques

There is considerable controversy in the literature regarding the choice of statistical method. The following paragraphs present arguments for the position adopted in the present research while the counter-arguments will be considered at the end of this section.

Measurement is the process of assigning numbers to observations. The level of measurement achieved by a researcher is a function of the rules under which the numbers or scores are assigned. These rules define and limit admissible operations on the scores, which must preserve the formal properties of the measurement scale. The manipulations and operations permissible in handling the scores must be those of the numerical structure to which the measurement is isomorphic. If two systems are isomorphic, their formal properties are the same in the relations and operations they admit (Siegel, 1956).

An ordinal scale incorporates equivalence and the relationship *greater than,* but, unlike the interval scale, the ratio of any two intervals cannot be specified. In an interval scale, the differences in the scale are isomorphic to the structure of arithmetic. Parametric statistical tests use means and standard deviations, i.e. require the operations of arithmetic on the original scores and may only be used with data measured in at least an interval scale. Nearly all of the measurements made by behavioural scientists constitute ordinal scales. When parametric techniques of statistical inference are used with such data, any decisions about hypotheses are doubtful (Siegel, p. 26).

Many behavioural scientists employ parametric statistical tests. Usually apparent success is a consequence of the untested assumptions which the scalemaker

is willing to make, for example, that the population has a normal distribution. Under these circumstances, it is difficult, if not impossible, to estimate the extent to which a probability statement about the hypothesis under test is meaningful. In addition to the condition that the variables involved must have been measured in at least an interval scale, the models of the parametric tests specify a variety of strong assumptions about the parameters of the population from which the research sample was drawn. Conditions associated with the parametric statistical model of, say, the two-sample t test, which must be satisfied before any confidence can be placed in any probability statement obtained by the use of the test, are at least these :

(i) The observations must be independent.

(ii) The observations must be drawn from normally distributed populations.

(iii) These populations must have the same variance.

In the case of more than two samples, an analysis of variance (the F test) is performed. Another condition is then added to those listed above :

(iv) The means of these normal and homoscedastic populations must be linear combinations of effects due to columns and/or rows i.e. the effects must be additive.

If these conditions are in fact met by the data under consideration, a parametric statistical test will be the most powerful and should certainly be chosen for analysis of the data. However, ordinarily these conditions are not tested before application of the t test or the F test. The possible exception is the assumption of homoscedasticity (equal variances), but the behavioural scientist who can very seldom achieve more than an ordinal scale cannot validly examine this point. In the behavioural sciences, knowledge about the parameters of the population in a research desig. is often lacking. Deviations in meeting the assumptions underlying a parametric statistical test obviously determine the meaningfulness of the probability statement arrived at by the test (Siegel, p. 19).

Nonparametric tests do not specify conditions about the parameters of the population from which the sample was drawn (Siegel, p. 31). Certain assumptions are associated with most nonparametric tests :

(i) The observations must be independent.

(ii) The variable under study has an underlying continuous distribution.

Nevertheless these assumptions are fewer and much weaker than those associated with parametric tests. Moreoever, corrections for ties are available, so that assumption (ii) need not hold.

It is believed by some researchers that utilisation of nonparametric methods results in a marked waste of data. However, if the measurement is weaker than that of an interval scale, by using parametric tests a researcher would *'add information'* and thereby create distortions which may be larger than those arising from the *'throwing away of information'* when scores are converted to ranks (Siegel, p. 32). The degree of wastefulness of a given nonparametric test is indicated by its power-efficiency but the power can in most cases be increased by simply enlarging the sample size slightly (Siegel, p. 20). In fact certain nonparametric tests approach the power of the comparable parametric procedures, examples of which are given in Section 3.2.2.2.

Many research workers have employed parametric methods for the analysis of their data but a number of these have been severely criticised for using such methods rather than nonparametric statistics. Examples are Hale's (1979) criticism of DeLuca (1978); Phillips' (1977) criticism of Blake, Lawson and Nordland (1976); Treagust's (1979) criticism of Lawson, Karplus and Adi (1978). In countering these attacks, it was generally conceded that nonparametric methods are strictly correct but contended that, in most cases, parametric methods give similar results. The most frequently quoted source in defence of parametric statistics is Gardner (1975) :

> *If Siegel is correct, parametric procedures, such as t-tests, product-moment correlations, factor analysis and analysis of variance simply have no place in the statistical armory of any researcher whose data possess less than interval strength. This would effectively prevent most researchers in the behavioural sciences from using parametric statistics.* (Gardner, 1975, p. 45).

Gardner concludes that parametric procedures are, in any case, robust and yield valid conclusions even when mildly distorted data are fed into them. This view results principally from a consideration of works by Gaito (1959), Heerman and Braskamp

(1970) and McNemar (1969). The situation has been succinctly summarised by Stevens (1968) in his consideration of the question of the extent to which an inappropriate statistic may lead to a deviant conclusion :

> *By spelling out the costs, we may convert the issue from a seeming proscription to a calculated risk.*
> (Stevens, 1968, p. 852).

In view of this risk, three specialist statisticians were independently asked their opinions regarding this controversy and each was emphatic that nonparametric methods were the only valid choices. Their views were independently similar to that of Treagust (1979) who stated that the use of complex parametric tests of statistical significance, while lending an air of scientific objectivity, do not necessarily support the validity of the conclusions.

3.2.2.2 Choice of Appropriate Nonparametric Tests

Two-sample tests are usually required as, in most cases, the researcher wishes to establish whether the experimental and control treatments are significantly different.

The tests available for two related samples that have high power-efficiency are not permissible in this study as the strongest level of measurement attainable in the present research is ordinality. Both the Walsh and randomisation tests demand measurement in at least an interval scale (Siegel, pp. 83 ff.) The otherwise suitable Wilcoxon matched-pairs signed ranks test requires measurement in an ordered metric scale which is intermediate in strength between an interval scale and an ordinal scale (Siegel, p. 76).

The tests available for two independent samples are more or less sensitive to different kinds of differences between samples. In the present study, the researcher wishes to determine whether two samples represent populations which differ in location. The level of measurement achieved precludes use of the most powerful test of location, the randomisation test (Siegel, p. 152), but use of this test in statistical analysis is hardly ever considered owing to the tediousness of its application. The appropriate test to choose is the Mann-Whitney U test (Conover, 1971; Lehmann, 1975; Siegel, 1956).

Compared with its most powerful parametric counterpart, the t test, when applied to data under conditions where the assumptions and requirements underlying use of the t test are met (*inter alia*, normality), the Mann-Whitney U test has asymptotic power-efficiency of 95,5%. Its power-efficiency is close to 95% even for moderate-sized samples.

The function of the two pretests in this investigation is to ensure that prior knowledge and experience do not differ significantly between the experimental and control groups. The variables most likely to cause differences in the performances of the two groups, apart from the different treatments, are thus age, sex and IQ. In order to maximise sensitivity of the Mann-Whitney U test, subjects should be divided into groups defined by these variables. The sampling procedure which is discussed fully in the next section, produces eight homogeneous blocks. This statistical design does not contemplate an over-all decision about the total sample by the application of a single test but involves computation of the value of the U test for each block. It must be emphasised here that one may not proceed from the decisions governing the individual blocks to an over-all decision concerning the total sample (Siegel, p. 159). The present research design means that the total sample may be discussed validly in terms of trends only.

The significance of the observed value of U in the Mann-Whitney test is determined by reference to different statistical tables whose selection depends on the number of cases in each of the two groups under consideration, on the proximity of U to the rejection region and on the presence or absence of a correction for ties. This explains the slight variations in statistical manipulations, presented in tabular form in this thesis, which result in blank spaces for particular cases in certain columns of the tables. These spaces are indicated with a dash (–).

The Mann-Whitney U test is the most appropriate test for analysis of the scores achieved by the experimental and control groups, but a third group of pupils is involved in the pretest prior to the instruction unit. In this instance, the researcher wishes to decide whether the three samples were drawn from the same population. Using the same criteria which led to the selection of the Mann-Whitney U test for two samples, the Kruskal-Wallis one-way analysis of variance by ranks is used for three samples. The Kruskal-Wallis test is simply the extension of the U test to test for differences among three or more samples.

If the Kruskal-Wallis test is applied to data which might properly be analysed by its most powerful parametric counterpart, the F test, the Kruskal-Wallis test has asymptotic power-efficiency of 95,5%.

3.3 The Sampling Procedure

The research was conducted at an English-medium co-educational Johannesburg high school situated in a middle class residential area. A rough analysis of parental employment defines the term, *middle class*, as used in the present context. The occupations of the fathers of 151 out of the 161 pupils selected for the research as described below, were available from the school records. The positions held could be categorised as follows :

38%	managerial and/or directorial
38%	professional
13%	sales representative
6%	technical
5%	diverse
0%	artisans

None of the pupils attending this school was taught by the author at the time of the investigation.

As was discussed in Chapter 2, Piaget (1964a) emphasised that instructional strategies designed to enhance intellectual development should be based, for meaningful learning, on the more elementary structures which the learners already possess. On this ground, many researchers who have attempted to induce logical operations in young children have been censured by Piaget and by Herron (1975). Since appropriate levels of maturation and experience are important Piagetian criteria in the selection of subjects for research involving formal operations, Standard 8 was circumspectly chosen as the lower academic limit for the subjects in the present study.

If results are to be meaningful, a training study related to logical operations should be conducted within a reasonably short period of time to exclude the factor of natural maturation. This condition was satisfied by the selection of Standards 8 and 9, since the pupils in these forms, unlike the matriculation pupils, still had a fair amount of free time to devote to the research project.

It was envisaged that the sample would be divided in terms of age, sex and IQ to maximise sensitivity of statistical procedures. In Standard 8 there were 189 pupils (105 boys and 84 girls) while in Standard 9 there were 171 pupils (93 boys and 78 girls). Of these, only pupils who were studying physical science were considered in order to eliminate, as far as possible, the extraneous variable of inconsistent scientific backgrounds among subjects, particularly as concepts in physical science would feature prominently in the training study. Furthermore,this limitation to pupils who had elected to take the option of physical science tended to achieve a certain degree of homogeneity in attitude towards this discipline, which was desirable for sensitivity in statistical analysis. This restriction reduced the number of pupils available for selection to 97 boys and 32 girls in Standard 8 and to 84 boys and 45 girls in Standard 9. It is evident that, compared with the proportion of the boys who opt to study physical science, the proportion of the girls is much lower. As a consequence the sample, unless it is made undesirably small, cannot comprise equal numbers of both sexes, nor can division of the sample by age, sex and IQ result in blocks of the same size for statistical analysis.

IQ was now the final criterion to be considered before drawing the sample. Here IQ is defined as average non-verbal IQ. It was felt that non-verbal IQ would be most suited to research involving logical operations. The specification of average IQ lessened the risk of inaccuracies in IQ values. This required the rejection of pupils for whom only one IQ value was available. Furthermore, pupils for whom only two IQ values were available and where these two values differed by 20 points or more, were also rejected. These exclusions decreased the number of pupils available for selection to 77 boys and 28 girls in Standard 8 and to 60 boys and 39 girls in Standard 9.

The histograms in Figures 3.5 and 3.6 show the IQ distributions for these pupils. The central tendencies of these distributions governed the division of the pupils into lower and higher IQ groups. Pupils with IQ below the median were classified as belonging to the lower IQ group while those with IQ above the median were allocated to the higher IQ group.

The sample consisted of pupils divided according to age and sex, drawn at random in equal numbers from the lower and upper IQ ranges of each histogram. The size of the sample was limited by the number of pupils which one researcher could handle but was large enough to accommodate possible losses owing to

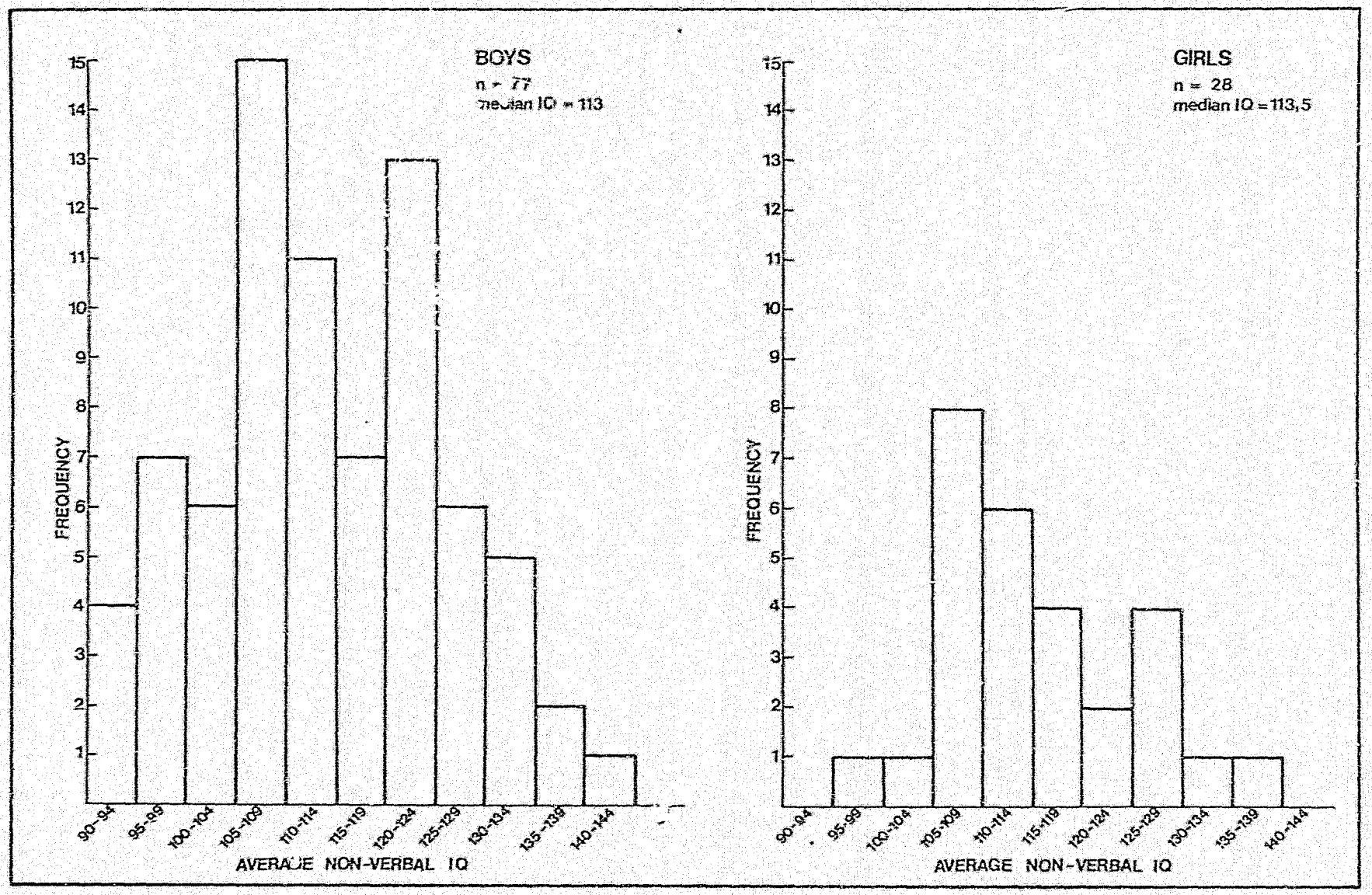


Figure 3.5 IQ Distributions for Standard Eight Physical Science Pupils

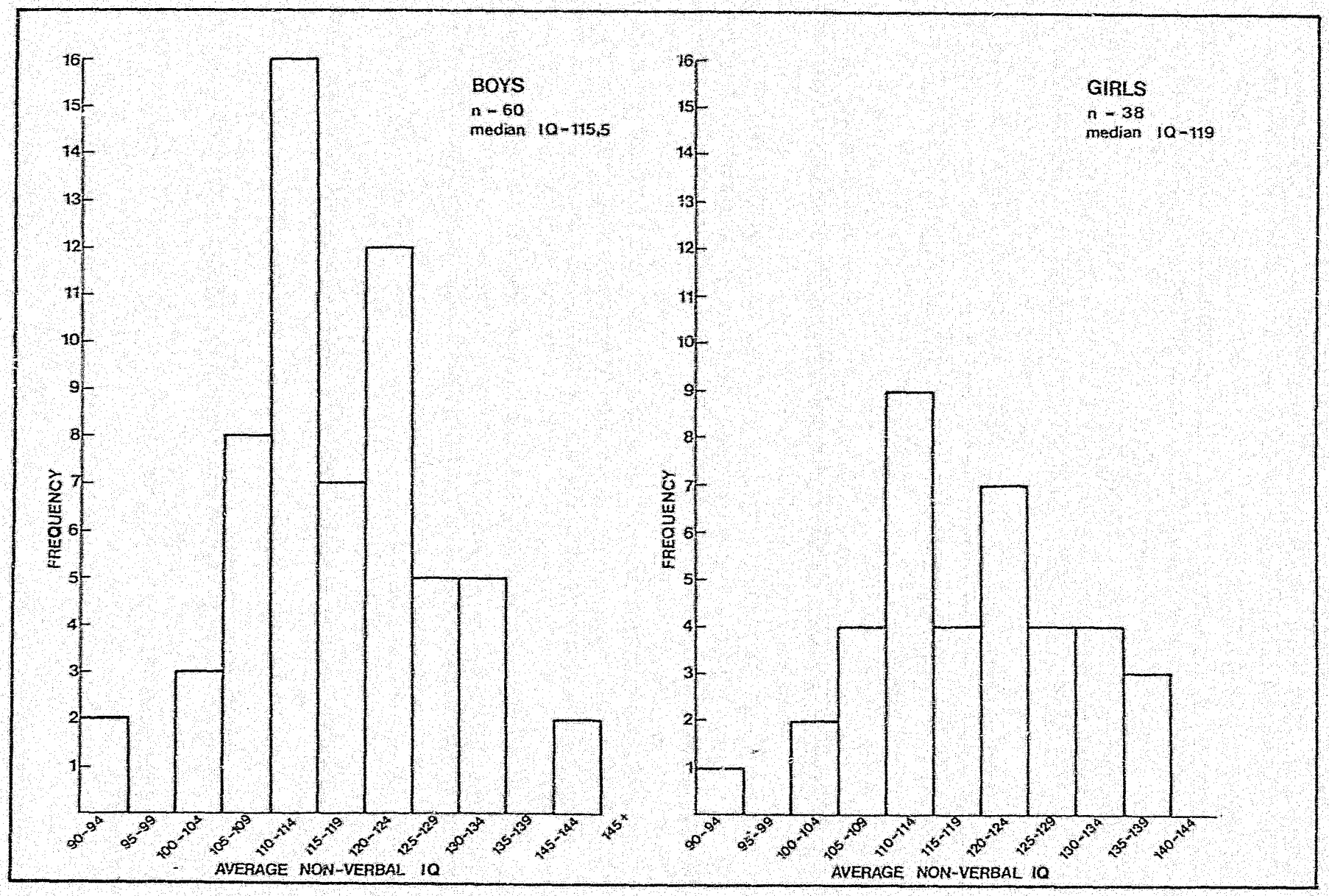


Figure 3.6 IQ Distributions for Standard Nine Physical Science Pupils

absentees and school-leavers. The total number of subjects in the sample was intended to be 160, but an extra pupil joined the research project through an administrative error. The 161 pupils in the sample were subdivided into eight blocks as follows :

Standard 8:
- 28 lower IQ boys
- 29 higher IQ boys
- 12 lower IQ girls
- 12 higher IQ girls

Standard 9:
- 28 lower IQ boys
- 28 higher IQ boys
- 12 lower IQ girls
- 12 higher IQ girls

The experimental treatment was assigned at random to half of the pupils in each block while the control treatment was assigned to the rest of the pupils in the block.

The sixteen groups of subjects obtained by this procedure are indicated in Table 3.1

Table 3.1 Division of the Sample of 161 Pupils into Sixteen Groups

	Experimental Group				Control Group			
	Standard 8		Standard 9		Standard 8		Standard 9	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Lower IQ	14	6	14	6	14	6	14	6
Higher IQ	14	6	14	6	15	6	14	6

3.4 Administrative Aspects of the Research

Permission to conduct the research was obtained from the Transvaal Education Department, the principal of the school, the pupils and the parents of the pupils.

The Transvaal Education Department and the school principal stipulated certain times only which would be allowed for the investigation. Each pupil provided the researcher with his or her individual timetable. In addition, each pupil filled in a form prepared by the researcher, indicating his or her preferences as to days and times for attending the investigation. All four timetables had to be correlated and the final schedule compiled. This schedule had to be flexible, however, allowing modification to accommodate absentees. Particular care had to be taken in planning that section of the research which dealt with the training or placebo activity, the Piagetian-like evaluation task and the Piagetian posttest. Here the schedule had to be arranged in such a way that each pupil would complete this trio of tasks within the space of three days.

To avoid disruption of the normal routine of the school, pupils were always summoned individually by notes delivered at early morning registration. During weekends and holidays, a telephone reminder system was added to the notes delivered on the last appropriate school day. Transport to and from the school was provided when necessary.

At the end of the experimental stage of the project, a report on the research was distributed to all the participants (Appendix I). As well as appreciation to the pupils for their co-operation, the report contained pertinent but harmless information, observations and *'findings'*, since the actual results were related to IQ and could not be divulged.

CHAPTER 4

THE PRETEST : REPLICATION OF PIAGET'S FIRST CHEMICAL EXPERIMENT

4.1 Background to the Pretest

Inhelder and Piaget (1958) exemplify the use of the combinatorial schema with the problem of combination of colourless chemicals. Two chemical experiments are reported, of which the first receives more mention in the subsequent protocols, particularly in those for the later stages. Their intention is the diagnosis of mental form rather than study of the content of thought, accounting for the use of the interview as a flexible diagnostic probe evidently independent of experimental method. Dale (1970) has replicated in detail Piaget's first chemical experiment for purposes of clarification, stating that much of the criticism which has been levelled at the published work from Piaget and his team, is deserved :

> *... experimental procedures have been sketchily outlined; sample size and distribution, data obtained and procedures used in the analysis of data often have been omitted completely; inferences drawn are often not justified by the evidence given; theoretical constructs often appear to be developed imaginatively without adequate experimental evidence. These factors have led to an initial lack of confidence in Piaget's findings and explanations and a desire to produce sound evidence to validate or refute them.* (Dale, 1970, pp. 277--278).

In his thorough replication study, Dale's findings support the basic structure of Piaget's theory of development of logical thinking, but show that use of this task should be accompanied by clear statements of test procedure and the criteria for success. In studies by subsequent investigators, experimental procedures are often ill-defined and success criteria vary so that direct comparison of data is not possible.

Among the many researchers who have used Piaget's first chemical experiment to test for combinatorial reasoning, are Graybill (1975), Joyce (1977), Kolodiy (1977), Neimark (1970), Pallrand (1977), Piburn (1977), Renner (1977) and Sayre and Ball (1975). Although the diversity of studies replicating Piaget's work is a tribute to his research, by the same token, criticism is implied. This is the opinion of Levine and Linn (1977), who remark on Piaget's imprecision, particularly with regard to standardisation of categories.

Several alternative tasks to the first chemical experiment have been devised. These will be examined in Chapter 6 which discusses the selection and implementation of a task in addition to the combination of chemicals for evaluation of the effectiveness of training. The disadvantages inherent in many of the tasks allied to the chemical experiment, motivate the development of a definite experimental procedure for the latter. The case for traditional Inhelder tasks has been presented in Chapter 2 and the first chemical experiment forms the pretest for the training study in the present thesis. However, in view of the uncertain and controversial background to this experiment, the thesis is also intended to establish and analyse a clear experimental procedure for administration of this task, which might help to provide guidance for disheartened investigators like Joyce (1977) who commented :

> *The chemicals posed special problems for the interviewer e.g. contamination, obtaining proper concentrations, acid burns and generally inconstant behaviour. While acceptable in textbook form, this task was troublesome to administer. It could conceivably be replaced by another task on combinatorial logic.* (Joyce, 1977, p. 157).

4.2 Administration of the Pretest

4.2.1 Pilot Study

Appendix A describes the technical procedure for trouble-free administration of Piaget's first chemical experiment, including details of the equipment, preparation and use of solutions and the prevention of colour formation irrelevant to the task.

It was envisaged that test presentation would approximate the design of Dale (1970), whose chemical experiment had been structured to follow the procedure outlined by Inhelder and Piaget (1958) as closely as possible. However, before presenting subjects with the task, two trial runs were conducted with the school science staff as subjects. This pilot study enabled the streamlining of questioning procedures, of efficient data recording and planning for each day's testing schedule.

4.2.2 Presentation of the Chemical Experiment

Each subject was tested individually. Before the subject were four similar volumetric flasks containing colourless solutions and labelled from one to four. The liquids used were :

(i) dilute sulphuric acid,

(ii) distilled water,

(iii) hydrogen peroxide solution,

(iv) sodium thiosulphate solution.

A smaller volumetric flask contained potassium iodide solution and was labelled five.

Two test-tubes, each containing a similar volume of unidentified colourless liquid, were placed in front of the subject. In fact, one test-tube contained sulphuric acid and hydrogen peroxide while the other simply contained water. While the subject watched, the potassium iodide solution was selected from the row of five volumetric flasks and several drops added to each test-tube. The researcher indicated that the contents of the latter test-tube remained the same while the contents of the former yielded a yellow colour. The subject was then asked to reproduce the yellow colour, using all or any of the five flasks as he wished.

For each subject, a written record of the combinations of solutions used was kept, together with key points in the conversation between subject and researcher. A standard list of questions was used (Dale, 1978), limiting questions to those on the list except when others were considered essential for elucidation upon a subject's mode of thought. This list is reproduced in Appendix B. In addition, for the first ten subjects, a tape recording of each interview was made. This precautionary measure was discontinued only when the researcher was satisfied that her approach towards each subject was consistent and that tapes would yield no further information concerning subjects' performances compared with written records.

The duration of testing for each subject varied considerably. The mean time was thirty minutes with a standard deviation of ten minutes.

There was intense interest in the investigation, both among participants and non-participants. Accordingly, provision was made to overcome the anticipated problem of inter-subject communication. The *'correct'* combinations of solutions were in fact altered several times daily by changing the numbered labels on the volumetric flasks. Subjects were told (falsely) that the chemicals in the flasks were changed regularly. It was stressed to each subject that obtaining the yellow colour would not ensure success in the test but that the idiosyncratic manner in which each individual tackled the task would be of prime importance. The subjects believed

the statements and these measures effectively prevented mere repetition of specific combinations, sequences or procedures which had been tried or considered appropriate by previously tested pupils. Subjects who were tested at a later stage in the course of the investigation did not appear to perform any better than those who had been tested earlier but it became apparent that the precautions were justified since it was obvious that considerable discussion on the project was taking place among the pupils.

As mentioned in Chapter 3, the Transvaal Education Department and the school principal had stipulated specific times only which would be permitted for the research. With this limitation, testing of the 161 pupils took six-and-a-half weeks.

4.3 Hypotheses to be Tested

As mentioned earlier, the aims of the work described in this chapter are, firstly, the analysis of the structure of Piaget's first chemical experiment and, secondly, the use of the experiment as the pretest for the training study which is the main line of investigation in this thesis. The first null hypothesis, formulated below, relates to the latter aim. The subsequent three hypotheses at first glance appear to relate to the former aim only. However, it will be remembered that the statistical design, described in Chapter 3, requires division of the sample by age, IQ and sex and therefore hypotheses 4.2, 4.3 and 4.4 are directly relevant to both of the aims above.

4.3.1 Null Hypothesis 4.1

H_o: There is no significant difference in the performances of the experimental subjects and the control subjects.

The interpretation of performances by the assigning of scores to observations can be specified to only some extent prior to an examination of subject responses. The precise construction of the measurement scale will necessarily depend on empirical analysis of the pretest. Score allocation will therefore be discussed in the results section of this chapter.

4.3.2 Null Hypothesis 4.2

H_o: There is no significant difference in the performances of the Standard Eight subjects and the Standard Nine subjects.

4.3.3 Null Hypothesis 4.3

H_0: There is no significant difference in the performances of the lower IQ subjects and the higher IQ subjects.

4.3.4 Null Hypothesis 4.4

H_0: There is no significant difference in the performances of the boys and the girls.

4.4 Results

4.4.1 Analysis of the Structure of Piaget's First Chemical Experiment

A considerable portion of this analysis replicates the study of the growth of systematic thinking by Dale (1970). In the years since Dale's publication, other workers, referred to earlier in this chapter, have used Piaget's first chemical experiment to test for combinatorial reasoning, but none has replicated the detailed analysis of Dale, although several have emphasised the need for consistent standards in evaluating formal thought. The present investigation focuses on the high school situation with an adolescent sample larger than that of Dale whose vertical study involved two hundred subjects ranging from six to fifteen years. The rationale for the replication was based on the thought that, although Dale's results were illuminating, data for a larger and different sample of adolescent subjects might be needed before Dale's conclusions could be better evaluated in their implications for high school science teaching. It must be borne in mind, however, that the present study was restricted to a single high school so that generalisation of results to other secondary school populations is not strictly possible.

4.4.1.1 Solution of the Problem

Inhelder and Piaget (1958) reported that children at substage III–A (11/12 to 14/15 years) are able to solve the problem completely i.e. they comprehend and can state the role of each of the five liquids. However, the criteria for success are not specified, although the translator (p.x.) stated that *a test is not considered passed until 75% of the children tested succeed.*

Author Chandler H A
Name of thesis The acquisition of formal scientific reasoning by physical science pupils in standard eight and nine 1984

PUBLISHER:
University of the Witwatersrand, Johannesburg
©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.