Four levels of competence were defined by Dale (1970) in his the chemical experiment :

Level A

(a) Made the colour in two ways.

(b) knew that all of 1, 3 and 5 were necessary,

(c) knew that 2 had no effect or did not help,

(d) knew that 4 removed or prevented the colour,

where 1, 2, 3, 4 and 5 were dilute sulphuric acid, water, hydrogen peroxide sodium thiosulphate solution and potassium iodide solution respectively.

Level B

Made the colour in at least one way plus any two of (b), (c) and (d) above.

Level C

Made the colour in at least one way plus any one of (b), (c) and (d) above.

Level D

Made the colour in at least one way.

Table 4.1 shows the percentage of subjects in each age group attaining each level of success.

If *completely solve* is interpreted as achievement of level A, far fewer than 75% of the subjects in either age group, spanning the range of fifteen to seventeen years, were able to solve the problem completely. This finding is in agreement with Dale who showed that 30% of his subjects at age fifteen performed at level A.

Table 4.1 Achievement of Various Levels of Success on Piaget's First Chemical Experiment

Level of Success	Percentage of Subjects	
	Standard 8	Standard 9
A	40	49
At least B	69	86
At least C	99	100
At least D	100	100

In Table 4.2 each of the factors which contributes towards complete solution of the problem is considered separately.

The first line of Table 4.2 shows the percentage of subjects in Standards 8 and 9 who were aware that all three of the liquids 1, 3 and 5 were necessary to reproduce the yellow colour. Almost the entire sample realised that all three liquids were required and this part of the problem was the simplest for most subjects. It is of interest to compare this result with Dale's figure of 70% for his slightly younger sample of fifteen-year olds.

The next line of Table 4.2 shows the percentage of subjects who knew that liquid 2 had no effect or did not contribute towards the colour. These results not only support Dale's finding of 55% (sample of fifteen-year olds as above) but it is informative to compare the percentages of 52% and 64% for Standards 8 and 9 respectively with the proportion of subjects who had produced the colour twice i.e. used combinations of 1, 3, 5 and 1, 2, 3, 5. The percentages of such subjects from the Standards 8 and 9 groups were 58% and 70% respectively This indicates that the role of liquid 2 was readily understood by most of the subjects who had formed yellow in both ways.

The most difficult aspect of the solution appeared to be the contribution of liquid 4 to the problem. Two correct responses were possible, namely, that liquid 4 removed the yellow colour and that the presence of liquid 4 in any combination of the liquids prevented the appearance of yellow. A deeper comprehension of the role of liquid 4 is indicated by the latter response, but, seemingly illogically, the latter

Table 4.1 Achievement of Various Levels of Success on Piaget's First Chemical Experiment

Level of Success	Percentage of Subjects	
	Standard 8	Standard 9
A	40	49
At least B	69	86
At least C	99	100
At least D	100	100

In Table 4.2 each of the factors which contributes towards complete solution of the problem is considered separately.

The first line of Table 4.2 shows the percentage of subjects in Standards 8 and 9 who were aware that all three of the liquids 1, 3 and 5 were necessary to reproduce the yellow colour. Almost the entire sample realised that all three liquids were required and this part of the problem was the simplest for most subjects. It is of interest to compare this result with Dale's figure of 70% for his slightly younger sample of fifteen-year olds.

The next line of Table 4.2 shows the percentage of subjects who knew that liquid 2 had no effect or did not contribute towards the colour. These results not only support Dale's finding of 55% (sample of fifteen-year olds as above) but it is informative to compare the percentages of 52% and 64% for Standards 8 and 9 respectively with the proportion of subjects who had produced the colour twice i.e. used combinations of 1, 3, 5 and 1, 2, 3, 5. The percentages of such subjects from the Standards 8 and 9 groups were 58% and 70% respectively. This indicates that the role of liquid 2 was readily understood by most of the subjects who had formed yellow in both ways.

The most difficult aspect of the solution appeared to be the contribution of liquid 4 to the problem. Two correct responses were possible, namely, that liquid 4 removed the yellow colour and that the presence of liquid 4 in any combination of the liquids prevented the appearance of yellow. A deeper comprehension of the role of liquid 4 is indicated by the latter response, but, seemingly illogically, the latter response did not always include an understanding of the former. A considerable percentage of the subjects gave only one or the other of the answers. A full understanding of the role of liquid 4, displayed by a small percentage of the subjects, elicited both responses. The last three lines of Table 4.2 show that 41% and 44% of the Standards 8 and 9 groups respectively were able to state that liquid 4 removed the yellow colour, (as compared with 12% of Dale's sample). Further, 12% and 15% of these two groups respectively knew that the presence of liquid 4 prevented the appearance of yellow, (as compared with 14% of Dale's sample). Finally, complete understanding of the role of liquid 4 was exhibited by 17% and 24% of the two groups respectively.

Table 4.2 Comprehension of Different Facets of Piaget's First Chemical Experiment

Correct Response	Percentage of Subjects	
	Standard 8	Standard 9
135 necessary	98	99
2 has no effect or does not help	52	64
4 removes colour	41	44
4 prevents colour	12	15
4 both removes and prevents colour	17	24

Dale (1970) stated that the problem can be solved at level D by the use of trial and error methods. However, success in solving at level C or above is dependent upon

(i) *whether or not the necessary additional information was discovered.*
(ii) *the subject's ability to remember what was discovered.*
(iii) *the subject's ability to realise the significance of what had been discovered.* (Dale, 1970, p. 281).

This does not imply that ability to solve the problem is valid as the sole criterion for the ability to apply combinatorial operations to the experimental data. In the present study, the roles of liquids 2 and 4 were often established by direct and fortuitous formulation as opposed to a specific method of proof. In particular, thirteen subjects response did not always include an understanding of the former. A considerable percentage of the subjects gave only one or the other of the answers. A full understanding of the role of liquid 4, displayed by a small percentage of the subjects, elicited both responses. The last three lines of Table 4.2 show that 41% and 44% of the Standards 8 and 9 groups respectively were able to state that liquid 4 removed the yellow colour, (as compared with 12% of Dale's sample). Further, 12% and 15% of these two groups respectively knew that the presence of liquid 4 prevented the appearance of yellow, (as compared with 14% of Dale's sample). Finally, complete understanding of the role of liquid 4 was exhibited by 17% and 24% of the two groups respectively.

Table 4.2 Comprehensior fferent Facets of Piaget's First Chemical Experiment

Correct Response	Percentage of Subjects	
	Standard 8	Standard 9
135 necessary	98	99
2 has no effect or does not help	52	64
4 removes colour	41	44
4 prevents colour	12	15
4 both removes and prevents colour	17	24

Dale (1970) stated that the problem can be solved at level D by the use of trial and error methods. However, success in solving at level C or above is dependent upon

(i) *whether or not the necessary additional information was discovered.*
(ii) *the subject's ability to remember what was discovered.*
(iii) *the subject's ability to realise the significance of what had been discovered.* (Dale, 1970, p. 281).

This does not imply that ability to solve the problem is valid as the sole criterion for the ability to apply combinatorial operations to the experimental data. In the present study, the roles of liquids 2 and 4 were often established by direct and fortuitous formulation as opposed to a specific method of proof. In particular, thirteen subjects performed at level A, using empirical procedures characterised only by random association of the liquids. It is evident that solution of the problem alone is not equivalent to combinatorial reasoning.

Conversely, six subjects were able to produce all possible combinations of liquids without a full understanding of the role which each combination played among the total number of possible combinations. This shows that the ability to generate a complete factorial array is not equivalent to solution of the problem which necessarily includes comprehension of the roles of each of the liquids. This observation would suggest the inadequacy of the criterion of a systematic approach alone as an indication of formal operations, as adopted by several workers, among whom are Mecke and Mecke (1971) in their investigation of Piaget's pendulum problem.

The inference of the above observations is that, when assigning Piagetian levels of performance on the first chemical experiment to subjects, the measurement scale must have a composite nature and should focus on both the combinatorial system and its use as an instrument of conclusive deduction, as originally outlined by Inhelder and Piaget (1958).

4.4.1.2 The Use of Combinations

Inhelder and Piaget (1958) stated that children use all types of combination of liquids – twos, threes, fours and five – spontaneously after nine years (substage II–B), while they are able to use all possible combinations at ages 11/12 to 14/15 years (substage III–A).

Table 4.3 shows the percentage of subjects who used various combinations.

The upper part of Table 4.3 indicates the percentage of subjects who used each type of combination at least once. As expected, 100% of subjects tried twos and threes. A somewhat lesser percentage used combinations of four liquids, namely, 93% and 95% of Standards 8 and 9 subjects respectively. This finding is roughly comparable with Dale's figure of 90% for his slightly younger sample of fifteen-year olds. It occurred to only 70% and 66% of these two groups to use the combination of all five liquids. (These figures have been corrected for the fact that some of the subjects elected not to investigate all five liquids because they knew that formation of the yellow colour would be excluded from any combination containing liquid 4).

These results are in agreement with Dale's figure of 70% for his sample of fifteen-year olds.

The lower part of Table 4.3 shows the percentage of subjects who used all possible combinations. (The calculation of these percentages did not include the ten subjects who were assessed at substage III-B as described later in this chapter, since they employed a solution method which may be described as a generalisation of substitution and addition and which did not necessarily involve all or even most of the possible combinations). It is apparent that there is a large discrepancy between these percentages and the behaviours reported by Inhelder and Piaget. The present results resemble more closely those of Dale, who found that 50%, 36% and 0% of his sample generated all combinations of twos, threes and fours respectively. It is difficult to explain these results, particularly in the absence of a clear description of the test procedure used by Inhelder and Piaget. The demonstration of the formation of the yellow colour to each subject at the beginning of the test suggests the inclusion of liquid 5 with the colour. Despite this obvious implication, very few subjects, either in the present investigation or in Dale's investigation, chose to select liquid 5 as a possible element of the correct combination, namely, adding each liquid in turn to liquid 5 before constructing other combinations of liquids. Since the majority of subjects seemed unaware of this limitation to the number of possible combinations, the consequence was a self-imposed additional information-processing load, whereas this complication does not seem to have arisen in the study by Inhelder and Piaget.

Table 4.3 Use of Combinations in Piaget's First Chemical Experiment

	Combinations	Percentage of Subjects	
		Standard 8	Standard 9
Use of different types of combinations	twos	100	100
	threes	100	100
	fours	93	95
	fives	70	66
Use of all possible combinations	twos	43	55
	threes	12	12
	fours	23	23
	at least one of the above	53	64
	all of the above	5	3

There are several ways of organising a systematic search for a correct combination. Two main search patterns were used by the subjects, both of which coincide with the patterns most often encountered by Dale (1970) and DeLuca (1979).

The first was a comple., pattern of addition which nearly always caused difficulties after which subjects reverted to random combinations or adopted another system. An example of such a sequence, bearing in mind the initial failure of nearly all the subjects to associate the liquid used in the demonstration, namely, liquid 5, with the yellow colour, is

12 123 1234 12345 13 134 1345 etc.

The second pattern was the one which was most frequently used by subjects. It consisted of a block system which could be modified but was usually used in the following sequence

12 13 14 15 23 24 25 34 35 45
123 124 125 134 135 145 234 etc.

In effect this system was used unwittingly by some subjects who generated combinations of four liquids by deliberately omitting each of the liquids in turn from the combination. For example, combination 1234 omits liquid 5 while the combination 1235 omits liquid 4.

The subjects who were assessed at substage III–B as described later in this chapter, tended to deviate from a strict block pattern in their search for information to support effects and proofs. Perfect search patterns were more commonly used by subjects at substage III–A, who were often more interested in the organisation of combinations than in the solution to the problem. This finding supports the observations made by DeLuca (1979).

The legitimacy of keeping written records had been pointed out to each subject prior to execution of the task. Furthermore, stationery occupied a conspicuous position on the work-bench. Written records of the combinations tried and which liquids produced the yellow colour were kept by 17% of the subjects. It seems that the majority of subjects believed that they would be able to solve the problem without straining their generative or memorial capacities. Most performances were characterised by frequent repetition of combinations, that is, *'forgetting'* of information. Later, nine pupils spontaneously commented that they had not anticipated the complexity of the problem and wished that they had been both more systematic in their approaches and had kept records.

4.4.2 Allocation of Scores

The detailed report on Piaget's first chemical experiment in the previous sections of this chapter and in Appendices A and B has been directed towards a three-fold purpose, namely,

(i) to present clear particulars of a suitable experimental procedure for this task, which has been called for in the literature as discussed earlier.

(ii) to replicate Dale's investigation, enlarging the size of sample in the mid-adolescent age range, thus focusing on the high school situation.

(iii) to establish the attributes of a valid measurement scale for quantitative assessment of subjects' performances as a pretest in the intended training study.

From this point in the chapter onwards, the main line of investigation in this thesis will be adhered to and the last of the three objectives above will now be elaborated upon.

The first step towards classification of the reasoning strategies used by subjects on the chemicals task is to establish the level of measurement which may be achieved. The formation of propositional logic, marked by the establishment of a combinatorial system, as conceived by Inhelder and Piaget (1958), constitutes an ordinal scale. In intellectual development, all members of the substage III–B are more advanced than all members of the substage III–A. Substage II–B members, in turn, are more advanced than those belonging to substage II–A but the ratio between the intervals constituting substage II–B and substage II–A, for example, cannot be specified.

Although the Piagetian stage construct has been interpreted as implying discontinuous intellectual growth (Novak 1977c), Herron (1978) argues that there is nothing in the construct which suggests that the transition from one stage to another is abrupt or that development is discontinuous. The latter view seems to be more plausible when cognitive stage formulations are seen in the proper perspective as provided by Ausubel and Sullivan (1970) :

> *... developmental stages imply nothing more than identifiable sequential stages in an orderly progression of development that are qualitatively discriminable from adjacent phases and generally characteristic of most members of a broadly defined age range.*
> (Ausubel and Sullivan, 1970, p. 102).

The legitimacy of stages of cognitive development is not therefore deniable by the well-documented existence of intraindividual, interindividual and intercultural differences in the acquisition of intellectual competencies. Stage descriptions do not insist that the manner of achievement of successive stages must necessarily be abrupt.

Since it seems that cognitive development is not marked by the appearance of discontinuous new structures of reasoning, members of any one substage need not be equivalent in their cognitive abilities. For valid statistical manipulation, rank ordering of members within each substage is therefore demanded, a procedure which will moreover sensitise statistical analysis compared with merely partitioning the class of subjects into the set of Piagetian subclasses.

The results described earlier in this chapter have shown that a measurement scale must be constructed in such a way that the different aspects of the chemical experiment bear the same weighting as in the original procedure outlined by Inhelder and Piaget. In particular, judgement of the ability to use combinatorial reasoning may not be based on the sole criterion of either ability to solve the problem or the ability to generate a complete factorial array, attractive as either of these modifications would be in terms of simplification.

One of the major criteria for combinatorial reasoning is the degree of organisation displayed by a given subject. No elucidation upon the term *systematic* was given by Inhelder and Piaget (1958). As a consequence, the interpretation of this term in various Piagetian studies is necessarily subjective and has not actually been defined in the literature. Dale rated a subject as having worked systematically if he used *at least three of the possible combinations of any one group (twos, threes, fours), allowing one repetition of each choice, and permitting the combinations concerned to be in a broken sequence, for example,*

15 25 125 135 35. (Dale, 1970, p. 282).

While Dale's study involved subjects ranging from six to fifteen years, his criterion for a systematic approach seems too lenient for the present purpose where the measurement scale must endeavour to differentiate between late-concrete and formal intellectual modes of tackling the chemicals task, particularly when one realises that Dale's criterion would permit assessment of a subject as having worked systematically if he had tried, for example, combinations of the liquids in twos but had in fact omitted seven of the ten possible combinations.

In the present investigation, the term *systematic* is arbitrarily defined as adopting a search pattern which is successful to the extent that at least 80% of the possible combinations of the liquids are generated, namely, twenty-one of the twenty-six combinations, allowing both repetition and broken sequences.

At substage III–B, subjects spontaneously organise proof of the experiment, that is, they employ a method of generalisation of substitution and addition, obviating as a redundancy the need to examine most of the possible combinations. Since the present definition of the term *systematic* means that the number of combinations produced is one of the attributes of the intended measurement scale, it is apparent that the assignment of ranks to substage III–B subjects within that substage must be based on a method of mapping which is different from that applied to the rest of the sample.

According to Inhelder and Piaget (1958), spontaneity in searching for the solution to the problem of colour formation, as opposed to experimenter-provided prompting, is also observed in subjects at substage III–A, although to a lesser extent than that exhibited by substage III–B subjects. In the present investigation, only fifteen subjects of the total sample displayed any degree of spontaneity, of whom ten were assessed at substage III–B. It is evident that very few subjects below substage III–B searched for relationships beyond those requested. This finding is in agreement with those of Lovell (1961) and Dale (1970). However, suggestion by the researcher elicited from most subjects unused combinations and ready experimentation, indicating that further anticipatory strategies were in fact available. Provided the other requirements of substage III–A were satisfied as described later in this chapter, such subjects would be assessed at substage III–A.

Inhelder and Piaget (1958) stated that subjects at substage II–A are in possession of systematic association of liquid 5 with each of the others but without any other combination. Subjects at substage II–B, as at substage II–A, commence by combining each liquid in turn with liquid 5 or by taking them all at once, but the innovation of this substage is that subjects then spontaneously use combinations of three liquids or four liquids. In the present study, among the subjects classified as at the concrete operational stage on the basis of a haphazard approach to the problem, association of the liquids two at a time did not precede the use at random of combinations of three, four or all five liquids. These subjects thus exhibited characteristics of substage II–B without the serial ordering of combinations reported at this substage by Inhelder and Piaget. Since no differentiation between substages II–A and II–B was observed, a single composite stage of concrete operations would be incorporated into the measurement scale, with its upper limit defined by the fact that combinations at this stage are not systematic.

The definition of the term *systematic*, as adopted in this investigation, implies the existence of search patterns intermediate between random and systematic, namely, those which do not succeed in generating at least 80% of the possible combinations. This would suggest a stage on the measurement scale which is transitional between the concrete and formal operational stages. This modification of the Piagetian categories is further suggested by the fact that many subjects exhibited characteristics of different stages simultaneously. Inhelder and Piaget (1958) stated that subjects at substages II–A and II–B do not attribute formation of the yellow colour to the combination of several elements as such. In the present investigation, although a considerable proportion of subjects was rated at the concrete operational stage, 98% and 99% of Standards 8 and 9 subjects respectively understood that the colour was due to the combination, an innovation supposed to appear at substage III–A. Further, Inhelder and Piaget state that subjects at substages II–A and II–B discover the negative effect of liquid 4 by direct and fortuitous formulation. However, many of the present subjects, apparently at the concrete operational stage, understood and could explain the reciprocal exclusion of liquid 4 with the yellow colour, an innovation associated with substage III–A. These observations support the view of Herron (1978) that Piaget's stage theory does not imply discontinuous intellectual growth. With specific reference to a transitional stage, these observations concur with studies by Lovell (1961), Case and Collinson (1962), Randall (1967), Rowell and Hoffman (1975), Shayer, Adey and Wylam (1981) and others which indicate that many adolescents are in a transitional state between the concrete and formal operational stages.

In summary, therefore, the construction of the measurement scale was defined by an analysis of the literature and by the examination of the subjects' responses described in the preceding subsection of this chapter. The number assigned to each subject was composite and consisted of a base assessment relating to substage classification, to which was added a secondary score involving points allocated within that substage. The scaling for the base assessment is given in Table 4.4 while that for the secondary assessment is detailed in Table 4.5. If, for example, a subject utilised a search pattern in the experiment, which generated at least 80% of the possible combinations of the liquids, his mode of approach to the task would be classed as early formal and given a rating of 200 (Table 4.4). To this basic score would be added the sum of points allocated in terms of the schedule in Table 4.5, say, 85 points. His composite score on the task would then be 285 points. (This score would next be converted to the corresponding score on the continuous measurement scale which was constructed as described below, resulting in a final score of 210 points). However, for reasons which have been discussed earlier in this chapter, Table 4.5 does not apply to subjects assessed at the late formal operational substage. The method which was used to assign secondary scores to these subjects is shown in Table 4.6.

Table 4.4 Base Assessment of Characteristic Behaviours on Piaget's First Chemical Experiment

Main Trait	Operational Substage	Base Score
Random search for solution to the problem	Concrete	0
Use of a search pattern evident in at least six consecutive combinations of the liquids	Transitional	100
Use of a search pattern which succeeds in producing at least 80% of the possible combinations	Early formal	200
Use of the combinatorial system for conclusive deduction, that is, integration of spontaneously organised discovery and proof (Inhelder and Piaget, 1958)	Late formal	300

Table 4.5 Secondary Assessment of Solution to the Chemicals Problem by Concrete, Transitional and Early Formal Modes of Approach

Aspect of Problem	Details of Score Allocation	Maximum Score
Comprehension that all of 1, 3 and 5 are necessary for colour formation	Comprehension : 5 points Partial comprehension : 3 points No comprehension : 0 points	5
Comprehension of the roles of liquids 2 and 4	Both 2 and 4 understood : 15 points One understood, the other partially : 13 points Either 2 or 4 understood : 10 points Both partially understood : 8 points Only one partially understood : 5 points	15
Use of combinations	Twos : 10 points plus 1 point for each combination Threes : 10 points plus 1 point for each combination Fours : 10 points plus 1 point for each combination Five : 10 points	65
Use of redundant combinations	No repeated combinations : 15 points One repeated combination : 10 points Two repeated combinations : 5 points Three or more repeated combinations : 0 points	15
	Maximum combined secondary score :	100

Table 4.6 Secondary Assessment of Attributes of the Late Formal Mode of Approach to the Chemicals Task

Attribute*	Score Allocation
Subject hypothesises	1, 2 or 3
Subject investigates spontaneously	1, 2 or 3
Subject generalises	1, 2 or 3
Validity of logic	1, 2 or 3
Efficiency of approach	1, 2 or 3
Non-redundancy of combinations**	1, 2 or 3
Maximum combined secondary score :	18

*Subjects obviously manifested these attributes to have their modes of approach classed as late formal but this scaling achieves rank ordering of the subjects within the substage. 1 = weak 2 = moderate 3 = good

**The score for this attribute is inversely related to the number of repeated combinations.

The measurement scale which is detailed in Tables 4.4, 4.5 and 4.6, ranged from zero at its lower limit and stopped at 318, (although theoretically there should be no upper limit to the late formal substage). It is apparent that the criteria adopted to define the scale disallowed scores of between 0 and 11, between 100 and 116, between 200 and 250 and between 300 and 306. For instance, considering the extreme case at the lower limit of the concrete operational substage, the base assessment would be zero to which would be added eleven points for producing a single combination of two liquids, giving a minimum composite score on the task of eleven points. Although the assumption that the scores are drawn from an underlying continuous distribution need not necessarily hold for valid statistical analysis, as discussed in Chapter 3, the scale was translated into a continuous scale which had the same intervals as the former. Specifically, the ranges 11–100, 116–200, 250–300 and 306–318 on the original scale corresponded to the ranges 0–89, 90–174, 175–225 and 226–238 on the new scale. This may sound unnecessarily involved to readers, but in practice the latter scale was neither difficult nor tedious to use and aided interpretation of results in view of its continuity. This scale was then used to assess the performances of subjects on the chemicals task and is delineated in Figure 4.1. It should be borne in mind that the scores may appear to be more precise than ranks but they will not meet the requirements of a higher level of measurement and the scale must be properly viewed as ordinal.

The writer scored the written protocols of the task for all the subjects. There was no difficulty in the assignment of base scores, namely, identification of Piagetian operational behaviours on the task, for 71% of the total sample. (Allocation of secondary scores then followed easily.) All uncertainties regarding the remaining 29% of the sample involved the concrete/transitional border where it was sometimes not clear whether a particular observed approach to the task was random (concrete) or in fact exhibited a small degree of system and should be rated as transitional. An auxiliary judge with whom the criteria had been discussed in detail, scored ten protocols of the task selected as being among the most obscure of those in this 29%. The two judges were in agreement on nine of these protocols, differing by one substage on the tenth one. Furthermore, since the secondary score of the subjects at the late formal substage was based on several abstract attributes, the auxiliary judge assessed the protocols for the ten late formal operational subjects. The two judges concurred on five of these and differed on the rest by one point out of the possible score of eighteen.

Typical responses for each of the Piagetian behaviours on the task as defined earlier, are reproduced in Appendix C.

The histogram in Figure 4.2 shows the distribution of the scores obtained by the subjects on the chemicals task.

The percentage of subjects exhibiting each type of strategy on the task, is as follows :

Concrete stage	:	27%
Transitional substage	:	41%
Early formal substage	:	26%
Late formal substage	:	6%

Chiappetta (1976) has reviewed a number of Piagetian studies that assessed cognitive levels of secondary students and concludes that more than 50% of the

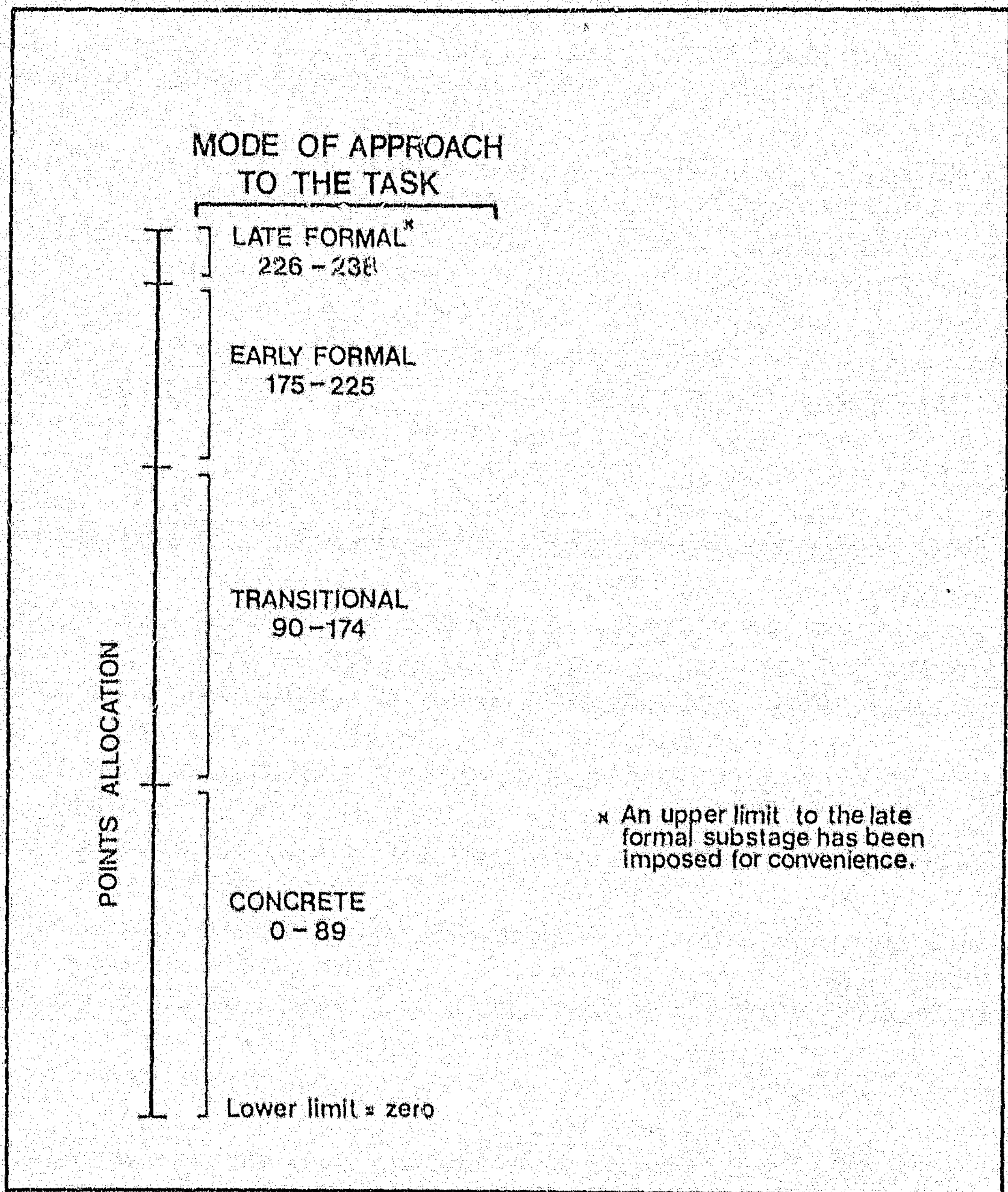

Figure 4.1 The Structure of the Measurement Scale Used for Piaget's First Chemical Experiment

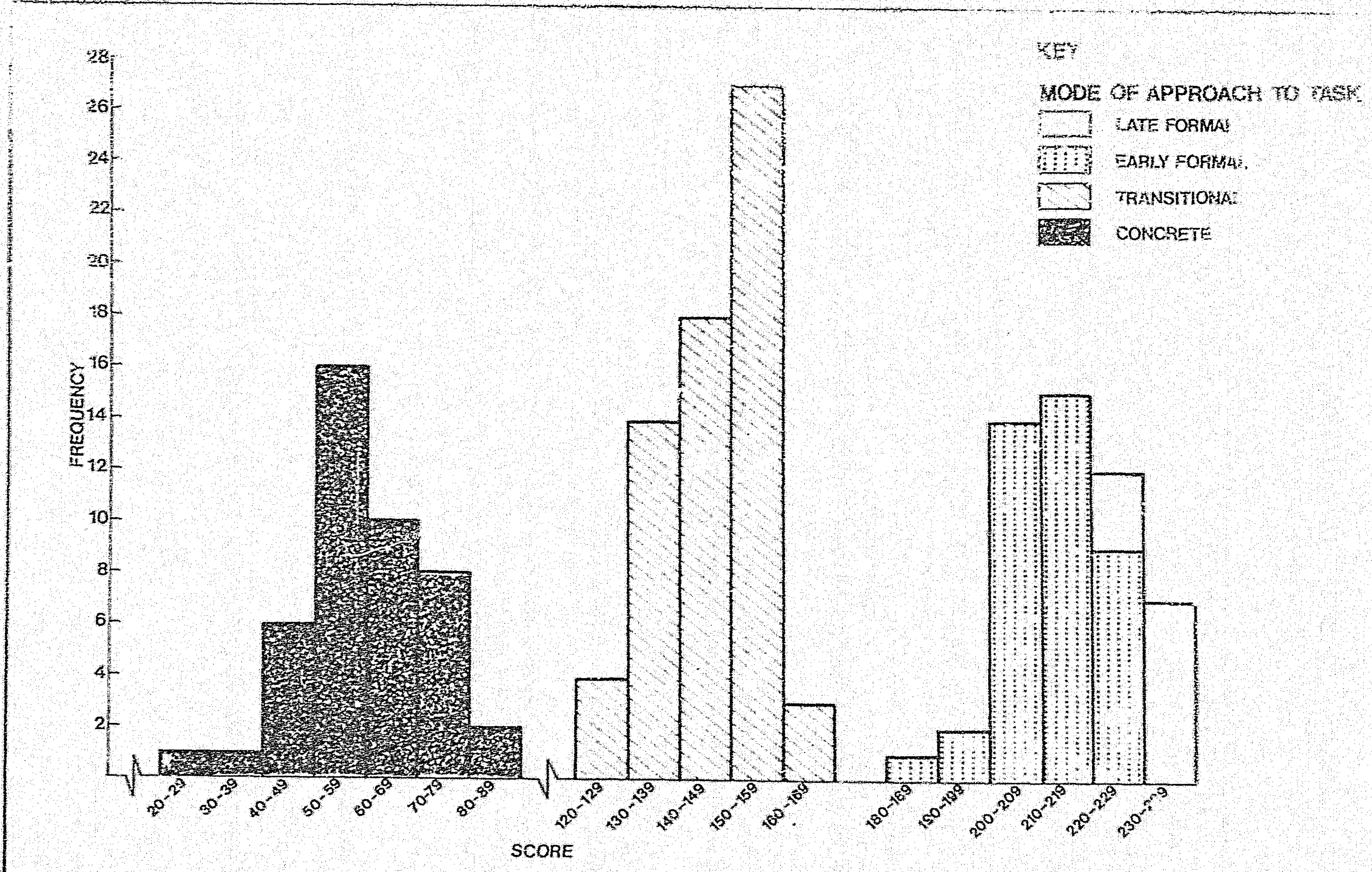

Figure 4.2 Distribution of the Scores Obtained by the Total Sample (N = 161 Pupils) on Piaget's First Chemical Experiment

American population of age 16 and over usually functions at the concrete operational level. This trend seems to be confirmed by the present results, although quantitative comparison with figures reported in the literature is not possible. Apart from the major considerations of task content and context and of sample size and distribution, there is a lack of consistency among studies as to experimental procedures and success criteria where these are reported at all.

4.4.3 Statistical Decisions

The data yielding information on the acceptability of null hypotheses 4. [illegible] 4.4 are summarised in Tables 4.7, 4.8, 4.9 and 4.10.

Table 4.7 shows that there is no significant difference in the performances of the experimental subjects and of the control subjects comprising any one block, except for the block of lower IQ Standard 8 boys. In the case of the latter it seems that the experimental subjects probably scored better than the control subjects. However, the value of z lies only just within the region of rejection of H_0 and may even be accounted for by statistical variation. This means that it will very likely still be possible to judge the effect of training on this block, relying on the relative probabilities obtained in the pre- and posttests. Unfortunately, scoring of the pretest prior to continuation of the investigation was inadmissible on theoretical grounds owing to its time-consuming nature. (Had this been permitted, subjects in this block could have been re-distributed within the experimental and control groups constituting the block.) As already mentioned, any delay in execution of the experimental sequence could have meant possible intellectual development of the subjects due to maturation which would have invalidated the study.

The rationale for use of the one-tailed test to determine the effect of age (Table 4.8) was the Piagetian tenet of intellectual development as a function of maturation (Inhelder and Piaget, 1958). Table 4.8 shows that there is a significant difference in the scores of the lower IQ subjects of both sexes, where the older subjects scored better. However, closer examination of the results shows that, in every block, the sum of the ranks of the older subjects is higher. Furthermore, the value of z for the block of higher IQ boys lies quite close to the region of rejection of H_0. The general conclusion is that Standard 9 subjects achieved higher scores in the chemical experiment, but significantly so only in the case of lower IQ subjects. This result suggests that division into age groups by school standard, a procedure

Table 4.7 Rank Ordering of Scores on Piaget's First Chemical Experiment as Pretest : Null Hypothesis 4.1 under Test

Mann-Whitney U Test, two-tailed, $\alpha = 0,05$

The test statistic was corrected for ties in borderline cases only, using the normal approximation (Siegel, 1956).

Block	Sample Size		Sum of Ranks		Value of U		z	p	Decision re H_0
	Experimental Group	Control Group	Experimental Group	Control Group	Observed	Critical, $\alpha = 0,05$			
Boys									
High IQ Std. 9	14	14	217	189	84	55	–	–	accept
Low IQ Std. 9	14	14	219,5	186,5	81,5	55		–	accept
High IQ Std. 8	14	15	211	221	101	59	–	–	accept
Low IQ Std. 8	14	14	246,5	159,5	54,5	55	–2,00	0,0456	reject
Girls									
High IQ Std. 9	6	6	48,5	29,5	8,5		–	> 0,132	accept
Low IQ Std. 9	6	6	36	42	15	–	–	0,700	accept
High IQ Std. 8	6	6	40	38	17	–	–	0,938	accept
Low IQ Std. 8	6	6	49,5	28,5	7,5	–		> 0,094	accept

which accepts some degree of overlap, is not sufficiently fine to achieve the statistical sensitivity required to detect significant performance differences due to age within the comparatively small range of fifteen to seventeen years. Sample size is one of the considerations that limits the number of variables controlled within the homogeneous blocks since there must be sufficient subjects within each block for statistical purposes. Under the inevitable restrictions upon the size of the present sample, the merits of alternative block designs must be weighed : division of the sample into exact age groups or into school standards which correspond to commensurate levels of experience in physical science but to only rough age classification. Although the ages of the subjects were readily available for the construction of a larger number of blocks, the former type of subdivision would give inappropriately few pupils per block. The latter subdivision of the sample – into school standards – was thus retained. Blocks based on academic levels might be viewed as a compromise by protagonists of maturation, in contrast to task content and context, as the main predictor of attainment on the pretest. However, these blocks may in fact be advantageous in the latter half of the thesis which is aimed at pupil achievement in science instruction. It has been suggested by Butts (1981) in his review of research in science education that, within the comparatively narrow band of age differences in the high school, age *per se* does not seem to be a reliable indicator of achievement outcome. Instead, the most significant contributor to success in science learning is the past experience of the pupils, especially when such experience is allied to the instructional context (Boulanger, 1981; Butts, 1981).

Table 4.8 Rank Ordering of Scores on Piaget's First Chemical Experiment as Pretest : Null Hypothesis 4.2 under Test

Mann-Whitney U Test, one-tailed, $\alpha = 0,05$

Whenever the number of cases in the larger group was greater than twenty for a given block, the normal approximation was used to determine the probability associated with the observed value of U (Siegel, 1956).

The test statistic was corrected for ties in borderline cases (Siegel, 1956).

Block	Sample Size		Sum of Ranks		Value of U		z	p	Decision re H_0
	Std. 9	Std. 8	Std. 9	Std. 8	Observed	Critical, $\alpha = 0,05$			
Boys									
High IQ	28	29	909	744	309	–	-1,55	0,0606	accept
Low IQ	28	28	898,5	697,5	291,5		-1,65	0,0495	reject
Girls									
High IQ	12	12	169,5	130,5	52,5	42	–	–	accept
Low IQ	12	12	200	100	22	42	–	–	reject

It seems reasonable to use the one-tailed test to determine the effect of IQ, although this must not be taken to imply that Piagetian measures of intelligence are regarded as coincident with IQ standardised instruments. Table 4.9 shows that the sum of the ranks of the high IQ subjects in every block is higher. However, the effect of IQ is not significant, except in the case of the Standard 9 boys. It is noteworthy that the observed value of z for this block of boys lies only just within the region of rejection of

H_0 while the z values for both the Standard 8 blocks lie very close to the rejection region. This absence of a clear differentiation in general between higher IQ and lower IQ subjects does not support the finding of Mealings (1963) and Keating (1975), that students' formal reasoning abilities as defined by Inhelder and Piaget, are strongly IQ-dependent. Kuhn (1976), on the other hand, reported that, for her sample of subjects with superior mental abilities, although a high correlation was found between IQ and progression towards Piaget's stage of concrete operations, correlation was nonsignificant between IQ and progression towards Piaget's stage of formal operations. Staver and Harty (1979) and Marek (1981) found modest correlations between IQ and performance on Piaget's chemicals task in the high school situation.

Table 4.9 Rank Ordering of Scores on Piagets' First Chemical Experiment as Pretest : Null Hypothesis 4.3 under Test

Mann-Whitney U Test, one-tailed, $\alpha = 0{,}05$

Whenever the number of cases in the larger group was greater than twenty for a given block, the normal approximation was used to determine the probability as[illegible]d with the observed value of U (Siegel, 1956).
The test statistic was corrected for ties in borderline cases (Siegel, 1956).

Block	Sample Size		Sum of Ranks		Value of U		z	p	Decision re H_0
	High IQ	Low IQ	High IQ	Low IQ	Observed	Critical, $\alpha = 0{,}05$			
Boys									
Standard 9	28	28	899,5	696,5	290,5		-1,66	0,0485	reject
Standard 8	29	28	940,5	712,5	306,5		-1,59	0,0559	accept
Girls									
Standard 9	12	12	163	137	59	42			accept
Standard 8	12	12	178	122	44	42	-1,62	0,0526	accept

Table 4.10 shows that there is no significant difference in the scores achieved by the boys and by the girls. Moreover, the probabilities associated with the observed values of the test statistic are fairly high. However, the Mann-Whitney statistics for each block indicate a tendency for the girls to perform better than the boys in the Standard 9 blocks and for the boys to perform better than the girls in the Standard 8 blocks.

Inhelder and Piaget (1958) did not mention sex differences but Dale (1970) found that boys scored significantly better than girls. DeLuca (1979), using an electronic equivalent of the first chemical experiment, did not confirm the significant sex effect reported by Dale. However, caution is necessary in comparing findings from studies requiring different activities of the subjects,viz. manipulation of chemicals in the former and manipulation of switches in the latter.

It is of doubtful validity to extend findings from studies focusing on different age groups and tasks to combinatorial operations as investigated in this thesis. Nevertheless, the results reported in the literature seem to suggest inconsistency in general as to bias due to sex. Elkind (1961) and Bridgham (1967) found significant sex effects. Bridgham inferred from his research that separate pedagogical techniques may be necessary for boys and girls at the concrete operational stage. Graybill (1975) reported that girls lagged behind boys in her qualitative study of three pairs of boys and girls in each of four age groups, using four Piagetian tasks of which one was the first chemical experiment. Sayre and Ball (1975), with their sample of four hundred and nineteen junior and high school students, found that boys did not differ significantly from girls in their overall performances on five formal operational tasks which included the first chemical experiment. Farmer *et al* (1982) reported that there was no significant difference in achievement between males and females on the Longeot tasks (which were developed as a Piagetian-based battery). Ehindero (1982) found no significant sex-related differences in *content-free* tasks. In contrast, several studies have reported that males performed better than females on tasks requiring logical operations (Brendzel, 1978; Doddy, 1980; Fogelman, 1972; Howe and Shayer, 1981; Lawson, 1975; Shayer and Wylam, 1978).

Table 4.10 Rank Ordering of Scores on Piaget's First Chemical Experiment as Pretest : Null Hypothesis 4.4 under Test

Mann-Whitney U Test, two-tailed, $\alpha = 0,05$

Since the number of cases in the larger group was greater than twenty for each block, the normal approximation was used to determine the probability associated with the observed value of U (Siegel, 1956).

Block	Sample Size		Mann-Whitney Statistic		Observed Value of U	z	p	Decision re H_0
	Boys	Girls	Boys	Girls				
Standard 9								
High IQ	28	12	179	157	157	−0,32	0,7490	accept
Low IQ	28	12	183,5	152,5	152,5	−0,46	0,6456	accept
Standard 8								
High IQ	29	12	166	182	166	−0,23	0,8180	accept
Low IQ	28	12	136	200	136	−0,94	0,3472	accept

4.4.4 Summary of Statistical Results*

1. There was no significant difference in the performances of the experimental subjects and the control subjects, except for the block of lower IQ Standard 8 boys where $p\ (z \leq -2,00) = 0,0456$.

2. Standard 9 subjects achieved higher scores than Standard 8 subjects, but significantly so only in the case of the lower IQ subjects of both sexes.

3. Higher IQ subjects achieved higher scores but significantly so only in the case of the Standard 9 boys.

*A summary of statistical decisions is provided at the end of every chapter where hypotheses have been tested. All qualitative findings are discussed in detail in the body of each chapter. This is so that the reader may have an uninterrupted view of the statistical analysis of the work in the thesis as a whole.

4. There was no significant difference in the performances of the boys and the girls. There was a tendency for girls to perform better at Standard 9 level and for boys to perform better at Standard 8 level but these tendencies were very weak as the probabilities associated with the observed values of the Mann-Whitney test statistic were high.