In view of the fact that the train task is in early stages of development, there need not be a significant correlation between the performances of subjects on this task and on the traditional task. For the present purpose, however, the former task is an appropriate measure of the effects of training since it demands the same logical operations as the training endeavoured to promote. The train task may thus be used for the assessment of selected aspects of intellectual advancement as described in the next paragraph, but Piagetian thought categories may not be allocated on the basis of scores assigned on the train task.

Since any measurement scale that may be constructed for the train task will not therefore necessarily be valid as an indicator of the presence or absence of the combinatorial scheme in pupils, the scale need not strive to embody both the Piagetian criteria, namely, that of systematic procedure and that of the formulation and testing of hypotheses in order to reach general conclusions. The present scale was constructed by the selection of only the criterion which satisfied the intentions behind the training. The scope of the thesis was limited to direction of the training towards a single major aspect of combinatorial reasoning, namely, the generation of a complete factorial array by means of a systematic procedure. In order to accept as valid and to sensitise assessment of training effects, observations on the train task explored acquisition of this particular skill only.

6.6.2 The Assignment of Scores to Observations

The set of task scores lay on an ordinal scale ranging from 0 to 100 with an underlying continuous distribution. Each score consisted of the sum of two components with arbitrary relative weighting. One component was assigned on the basis of the type of approach used in execution of the task and the other component reflected the number of repeated combinations.

The former component classified each subject into one of five levels on a scale which ranged from 0 to 80 as follows :

Type of Approach to Task	Points Allocation
Random	0
Systematic to a minor degree (evidenced by a pattern containing at least five successive combinations)	20

Systematic to a marked degree (evidenced by a pattern or patterns containing at least ten combinations, not necessarily consecutive)	40
Systematic in all respects with partial success (evidenced by the omission of combinations)	60
Systematic in all respects with attainment of objective (evidenced by a complete factorial array, although some combinations might be repeated)	80

The researcher could usually identify the type of approach utilised by a particular subject after the execution of a moderate number of trials. In a few instances, subjects were detained at the end of the task and asked to explain the procedure which had been used in finding the solution.

The second component of the total score reflected a penalty for repeated combinations. It was considered likely that the more combinations a subject had been permitted to test on the task, the more repeated combinations would have been produced. To avoid discrimination on the basis of the number of trials which had been allowed, the measurement scale was constructed in such a way that a reasonable range of repeated combinations would result in the same penalty. The number of points allocated to individual subjects bore an inverse relationship to the percentage of redundant combinations and was calculated on a scale of 0 to 20 according to the formula :

$$\text{Number of points} = \frac{100 - (\text{percentage of redundant combinations})}{5}$$

The expression, *percentage of redundant combinations*, has no single established meaning or rule of usage. The above formula generates scores which are dependent upon what the expression is taken to mean. If, for example, *percentage of redundant combinations* in the formula were to be defined as

$$\frac{\text{number of repeated combinations}}{\text{number of different combinations}} \times 100$$

then the possibility of negative scores exists. In such an event, the final (composite) score would reduce the performance of the individual into the next lower thought category and thereby also affect ranking within the set of scores achieved by the group of subjects. The generation of such spurious results must be avoided by defining the expression, *percentage of redundant combinations*, in the sense required in the special context of the formula. *Percentage of redundant combinations* was therefore defined as

$$\frac{\text{number of repeated combinations}}{\text{total number of combinations attempted}} \times 100$$

This definition describes a scale which tends to zero at its lower limit if the number of redundant combinations is taken as the number of times combinations are repeated, including in this total all the times a specific combination is repeated. The points therefore range on a continuous scale of 0 to 20. It was thought likely, however, that subjects would tend to achieve fairly well as a preponderance of repeated trials by a subject would suggest an unusual level of incompetence. For example, ten points or fewer corresponds to fifty per cent of redundant combinations or more.

6.7 Results

6.7.1 The Effectiveness of the Training

The histograms in Figures 6.2, 6.3, 6.4 and 6.5 show the scores which were obtained. Although the histograms appear to be discontinuous, the measurement scale which was employed to generate them, was actually continuous as has been discussed above. The apparent discontinuities in the histograms are due to the fact that no subject repeated trials often enough to achieve less than ten points in that component of the score which reflected a penalty for repeated combinations.

Of the eighty experimental subjects, seventy-two achieved the maximum score of 100. The lowest score achieved by an experimental subject was 76.

The median score for the control group was 37, with the highest score being 80 and the lowest score being 12. Among the control subjects, nineteen pupils gave up completely on the task, on trial numbers ranging from 16 to 67, mean trial number 41. Further, fourteen other control subjects were struggling to produce combinations

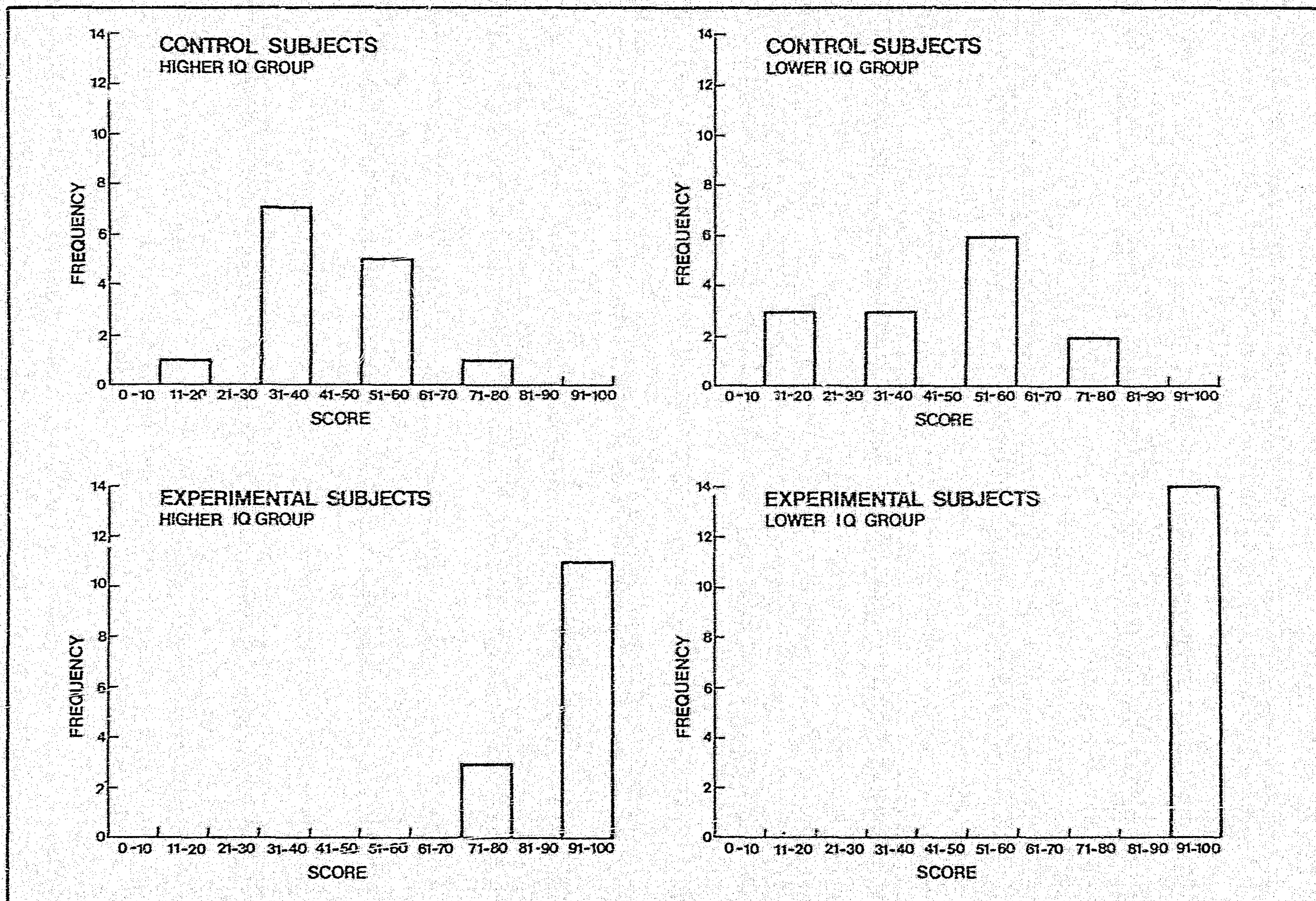

Figure 6.2 Scores Achieved on the Train Task by Standard Nine Boys

Note : Although these histograms appear discontinuous, the measurement scale used to generate them is continuous (see text).

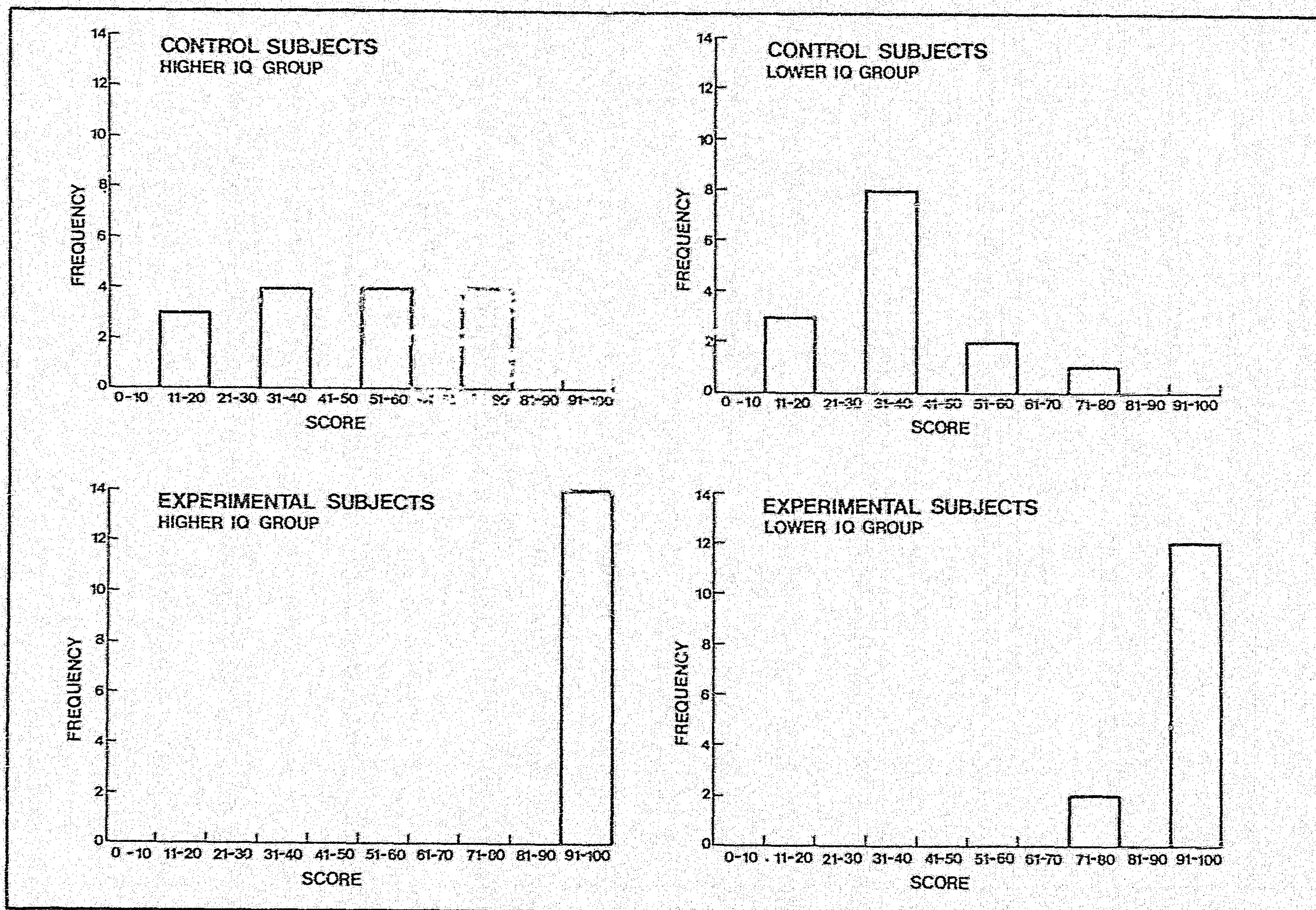

Figure 6.3 Scores Achieved on the Train Task by Standard Eight Boys

Note : Although these histograms appear discontinuous, the measurement scale used to generate them is continuous (see text).

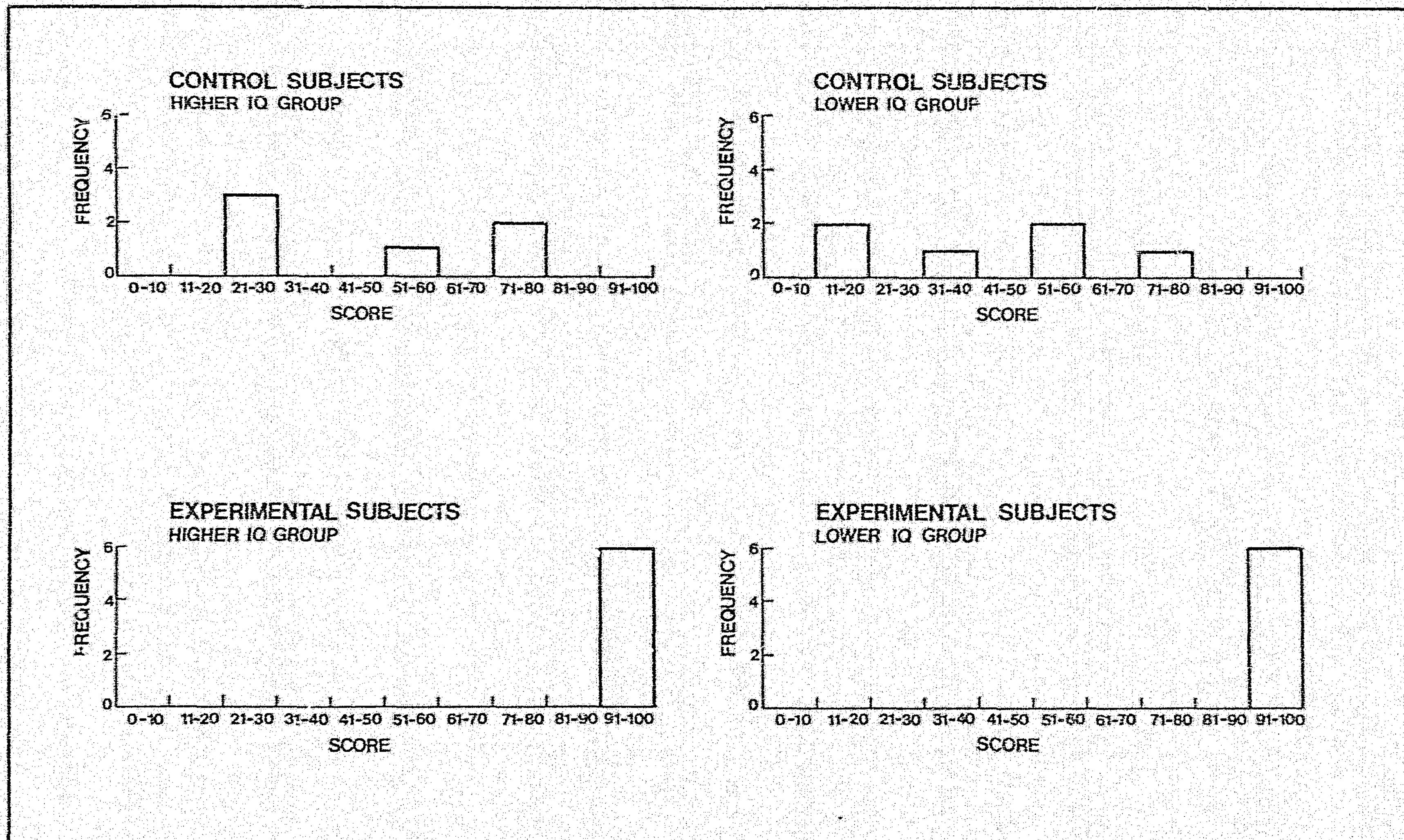

Figure 6.4 Scores Achieved on the Train Task by Standard Nine Girls

Note : Although these histograms appear discontinuous, the measurement scale used to generate them is continuous (see text).

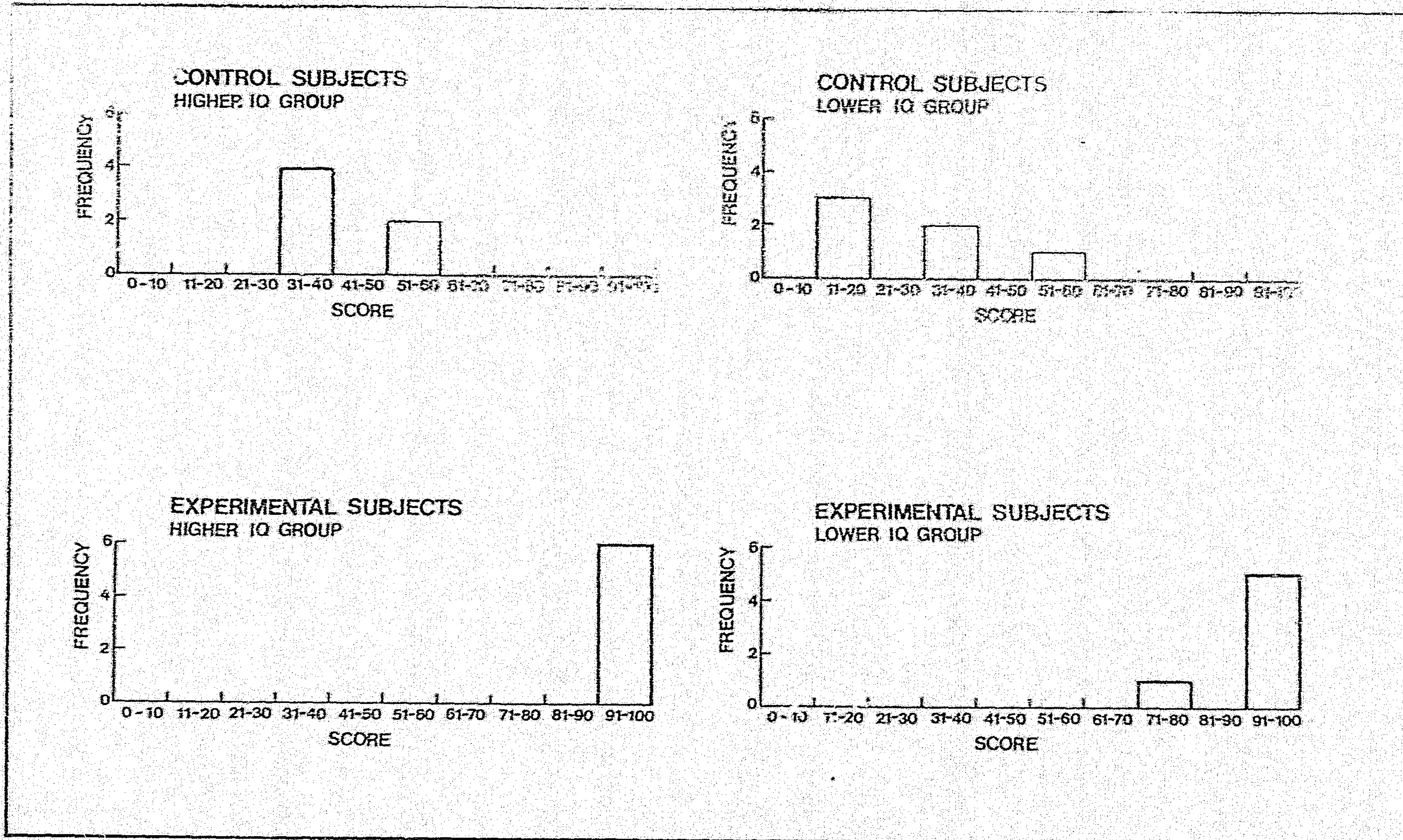

Figure 6.5 Scores Achieved on the Train Task by Standard Eight Girls

Note : Although these histograms appear discontinuous. the measurement scale used to generate them is continuous (see text).

so ineffectually by the time they had reached trial numbers ranging from 28 to 68 mean trial number 43, that it was apparent that they also would have withdrawn from the task, despite encouragement to continue with their efforts, at which stage the train was opportunely made to run.

It is clear without statistical analysis that the training was beneficial to the experimental group. The statistical significance of the training is given in Table 6.1, showing that null hypothesis 6.1 must be rejected in favour of the research hypothesis in every block. The extreme values of the Mann Whitney test statistic are noteworthy. The value of U is particularly important in the case of the lower IQ Standard 8 boys (U = O) since the experimental subjects in this block performed better in the pretest than the control subjects ($U = 54.5$, $p = 0.0456$). Comparison of the two values of U indicates that the training was indeed beneficial to this group of subjects. The question of whether any particular group of pupils based on division by age, IQ or sex benefited more from the training than other groups, will be discussed later.

6.7.2 The Effects of Age, IQ and Sex on the Task

The data yielding information on the acceptability of null hypotheses 6.2 to 6.4 are summarised in Tables 6.2 to 6.4.

Table 6.2 examines age effects and shows the same results as those reported for Piaget's first chemical experiment in Chapter 4, namely, that the older subjects tended to score better but that the rough age classification of pupils into Standards 8 and 9 does not permit the analysis to reveal any significant differences in performance which could be attributed to age.

Table 6.3 shows that the effect of IQ is not significant, except in the block of Standard 8 boys. However, one should view the latter result with caution as the test statistic borders on the region of rejection such that the Mann-Whitney test statistic accepts the null hypothesis and the normal approximation, whether a ties correction is incorporated or not, rejects the null hypothesis. In this somewhat awkward case, the preferred choice is use of the normal approximation with adjustment for non-continuity. It should be borne in mind however that the specified level of significance, compared with a more stringent level, has increased probability of committing the Type I error, that is, rejecting the null hypothesis when in fact it is true.

Table 6.1 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.1 under Test

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

Block	Sample Size		Sum of Ranks		Value of U		Decision re H_0
	Experimental Group	Control Group	Experimental Group	Control Group	Observed	Critical, $\alpha = 0{,}05$	
Boys							
High IQ Std. 9	14	14	300	106	1	61	reject
Low IQ Std. 9	14	14	301	105	0	61	reject
High IQ Std. 8	14	15	315	120	0	66	reject
Low IQ Std. 8	14	14	301	105	0	61	reject
Girls							
High IQ Std. 9	6	6	57	21	0	7	reject
Low IQ Std. 9	6	6	57	21	0	7	reject
High IQ Std. 8	6	6	57	21	0	7	reject
Low IQ Std. 8	6	6	57	21	0	7	reject

Decision : Reject H_0 in favour of H_1 : the experimental subjects performed significantly better on the evaluation task than the control subjects.

Table 6.2 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.2 under Test (Control Group)

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

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

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	14	15	200,5	234,5	95,5	66		–	accept
Low IQ	14	14	236	170	65	61	-1,52	0,0643	accept
Girls									
High IQ	6	6	46,5	31,5	10,5	–		> 0,120	accept
Low IQ	6	6	45	33	12	–	–	0,197	accept

In Table 6.3 it is seen that the sum of the ranks of the higher IQ subjects is higher in every block except for the Standard 9 boys. The latter block leads one to ponder on the findings of DeLuca (1979). DeLuca reported that, in accordance with Inhelder and Piaget (1958), the general trend in cognitive development was for both the organisation of combinations and the use of proofs to appear in more systematic fashion as age increased. However, in some cases, there was a retrogression in the degree of organisation with intellectual growth, with maximum organisation taking place at substage III–A. Relevant to the present consideration of Table 6.3, seven of the higher IQ pupils in the Standard 9 block of male subjects had been assessed at substage III–A on the pretest. This number constituted the highest percentage in any control section of the eight blocks. In the lower IQ section of this block, only three of the fourteen subjects had been assessed at substage III–A. As has previously been shown, scores on the train task were based mainly on the criterion of organisation of combinations, the cognitive design skill which is claimed by DeLuca as reaching its peak of development at substage III–A. It could have been expected, therefore, that the higher IQ Standard 9 boys would achieve better on this task than their lower IQ peers. Their failure to do so, as seen in Table 6.3, might, in the

Table 6.3 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.3 under Test (Control Group)

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

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

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	14	14	193	213	88	61	–	–	accept
Standard 8	15	14	263,5	171,5	66,6	66	–1,68	0,0465	reject
Girls									
Standard 9	6	6	44	34	13		–	0,242	accept
Standard 8	6	6	45	33	12	–	–	0,197	accept

In Table 6.3 it is seen that the sum of the ranks of the higher IQ subjects is higher in every block except for the Standard 9 boys. The latter block leads one to ponder on the findings of DeLuca (1979). DeLuca reported that, in accordance with Inhelder and Piaget (1958), the general trend in cognitive development was for both the organisation of combinations and the use of proofs to appear in more systematic fashion as age increased. However, in some cases, there was a retrogression in the degree of organisation with intellectual growth, with maximum organisation taking place at substage III–A. Relevant to the present consideration of Table 6.3, seven of the higher IQ pupils in the Standard 9 block of male subjects had been assessed at substage III–A on the pretest. This number constituted the highest percentage in any control section of the eight blocks. In the lower IQ section of this block, only three of the fourteen subjects had been assessed at substage III–A. As has previously been shown, scores on the train task were based mainly on the criterion of organisation of combinations, the cognitive design skill which is claimed by DeLuca as reaching its peak of development at substage III–A. It could have been expected, therefore, that the higher IQ Standard 9 boys would achieve better on this task than their lower IQ peers. Their failure to do so, as seen in Table 6.3, might, in the

Table 6.3 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.3 under Test (Control Group)

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

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

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	14	14	193	213	88	61	–	–	accept
Standard 8	15	14	263,5	171,5	66,6	66	–1,68	0,0465	reject
Girls									
Standard 9	6	6	44	34	13			0,242	accept
Standard 8	6	6	45	33	12		–	0,197	accept

absence of any other evidence, have been attributed to statistical fluctuations but a parallel deficiency was observed in the performances on this task of the experimental subjects. Assuming that DeLuca's findings are applicable to the present sample, the two sets of observations suggest that the train task involves some factor or factors which tended to inhibit optimum functioning of the higher IQ Standard 9 boys. Qualitative results concerning formal stage subjects, which seem to throw some light on the nature of the problem, are discussed later in this chapter.

Table 6.4 examines sex effects on the task and indicates that there is no significant difference in the scores achieved by the boys and by the girls. This finding is in agreement with the work of Siegler and Liebert (1975) on their train task and of DeLuca (1979) on his electronic equivalent of Piaget's first chemical experiment. Although differences are not significant, the values of the Mann-Whitney test statistic in Table 6.4 indicate that the boys achieved consistently better than the girls except for the Standard 9 higher IQ block. These results roughly correspond to the sex effects on the first chemical experiment reported in Chapter 4, where the girls outstripped the boys (total sample of subjects) at the Standard 9 level only, but not significantly so. Possibly the slight differences in performance might reflect differential attitude rather than different cognitive capacities as it seemed that the older girls exhibited a more careful approach to the task than the older boys.

6.7.3 Differential Responsivity to Instruction

6.7.3.1 Different Classificatory Groups of Subjects

The statistical decisions discussed above are that there are no differences in performance among the Standards 8 and 9 pupils in the control group, which can be attributed to age, IQ or sex. These results not only give information intrinsic in the task but also form base-lines in the assessment of whether any particular age, IQ or sex group among the experimental pupils benefited more from training than other groups. This question is analysed in the statistical consideration of null hypotheses 6.5, 6.6 and 6.7, which are summarised in Tables 6.5, 6.6 and 6.7.

Table 6.5 shows that there are no significant age effects in the performances of the experimental subjects, although, like the control subjects (Table 6.2), the older pupils tended to score better. The exception is the block of higher IQ boys where the Standard 8 pupils achieved significantly better scores. Comparing this

Table 6.4 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.4 under Test (Control Group)

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

Block	Sample Size		Mann-Whitney Statistic		Value of U		Decision re H_0
	Boys	Girls	Boys	Girls	Observed	Critical, $\alpha = 0{,}05$	
Standard 9							
High IQ	14	6	54,5	29,5	29,5	17	accept
Low IQ	14	6	39,5	44,5	39,5	17	accept
Standard 8							
High IQ	15	6	36	54	36	19	accept
Low IQ	14	6	35	49	35	17	accept

Table 6.5 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.5 under Test (Experimental Group)

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

Since the proportion of ties in each case was quite large, the test statistic was corrected for ties using the normal approximation (Siegel, 1956).

	Sample Size		Sum of Ranks		Observed Value of U	z	p	Decision re H_0
	Std. 9	Std. 8	Std. 9	Std. 8				
Boys								
High IQ	14	14	182	224	77	-1,80	0,0359	reject
Low IQ	14	14	217	189	84	-1,44	0,0749	accept
Girls								
High IQ	6	6	39	39	18	0	0,5000	accept
Low IQ	6	6	42	36	15	-1,00	0,1587	accept

result with the corresponding result for the control group (Table 6.2) which showed no significant difference in performance for the block of higher IQ boys, the conclusion is that, among boys of higher IQ, Standard 9 pupils benefited less from training than Standard 8 pupils. This is discussed further in the next paragraph in conjunction with null hypothesis 6.6.

The probabilities in Table 6.6 are similar to those in Table 6.5. This arises from the very high proportion of tied scores of 100% achieved by experimental subjects. Table 6.6 indicates that the effect of IQ is not significant, although in general, like the control group (Table 6.3), the higher IQ subjects tended to achieve higher scores. The exception is the block of Standard 9 boys where the lower IQ pupils were rated significantly better. The corresponding block in the control group (Table 6.3) did not manifest any significant differences in performance. The inference is that higher IQ Standard 9 boys benefited less from training than their lower IQ counterparts. This finding, coupled with the finding reported in the previous paragraph, namely, that higher IQ Standard 9 boys benefited less from training than their Standard 8 counterparts, leads to the suggestion that training should be presented earlier than Standard 9 for all pupils to derive maximum benefit.

Table 6.7 shows that there are no significant sex effects in the performances of the experimental subjects. Since the corresponding results for the control subjects (Table 6.4) also exhibit no significant sex effects, it is evident that the training did not discriminate against subjects on the basis of sex.

It seems from Table 6.7 that the block which most nearly approaches significance is the one which consists of higher IQ Standard 9 subjects. The probability of occurrence under H_0 of $z \leqslant -1{,}20$ is $p = 0{,}2302$ compared with probabilities of 1,0000, 1,0000 and 0,8966 for the other three blocks respectively. Comparison of $p = 0{,}2302$ with the probability of like magnitude which is associated with $U = 29{,}5$ for the corresponding control block in Table 6.4 ($z = -1{,}04$ and $p = 0{,}2984$), does not reveal any tendency for higher IQ Standard 9 boys to benefit differently from training than their female peers.

This analysis of the differential benefits of training with reference to groups divided by age, IQ and sex, should be interpreted with caution in terms of general applicability. The construction of the present task served to attain the main objective of being able to detect any differences between the performances of the trained and untrained groups of subjects. It also served to supply information, from the

Table 6.6 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.6 under Test (Experimental Group)

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

Since the proportion of ties in each case was quite large, the test statistic was corrected for ties using the normal approximation (Siegel, 1956).

Block	Sample Size		Sum of Ranks		Observed Value of U	z	p	Decision re H_0
	High IQ	Low IQ	High IQ	Low IQ				
Boys								
Standard 9	14	14	182	224	77	−1,80	0,0359	reject
Standard 8	14	14	217	189	84	−1,44	0,0749	accept
Girls								
Standard 9	6	6	39	39	18	0	0,5000	accept
Standard 8	6	6	42	36	15	−1,00	0,1587	accept

Table 6.7 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.7 under Test (Experimental Group)

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

Since the proportion of ties in each case was quite large, the test statistic was corrected for ties using the normal approximation (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	14	6	51	33	33	−1,20	0,2302	accept
Low IQ	14	6	42	42	42	0	1,0000	accept
Standard 8								
High IQ	14	6	42	42	42	0	1,0000	accept
Low IQ	14	6	41	43	41	−0,13	0,8966	accept

performances of the members of the control group, differentiable from one another within this group, on the question of any age-, IQ- and sex-bias intrinsic in the task. However, the trained subjects, in contrast totheir untrained counterparts, found the task a relatively easy problem. Thus most of the scores achieved by the experimental subjects were tied observations on the upper limit of the measurement scale and were therefore not differentiable. Without extension of this research and alteration to the existing structure of the task, differentiation among experimental subjects is not feasible. If the task were merely increased in complexity to allow this differentiation, then any differences between the experimental group and the control groups may be obscured ow to loss of motivation in control subjects on being confronted with a task requiring manoeuvres beyond their powers of organisation. In summary, the resulting tied scores among the experimental group mean that no inference may be made that the differential benefits of training will be the same when measured by performances on another combinatorial task which, apart from the usual content differences from one task to another, may allow better discrimination among the experimental subjects.

The foregoing paragraphs have examined the differential benefits of training with reference to different classificatory groups of subjects. The discussion will now be extended to individuals at different stages of intellectual growth.

6.7.3.2 Individuals at the Formal Stage

All nine experimental subjects who had been assessed at substage III–B in the Piagetian pretest displayed retrogressive approaches towards the train problem. Confronted with this task, they abandoned their former strategies involving the combinatorial system as a means of conclusive deduction. They each constructed a tree diagram and almost mechanically executed one combination of switches after the other without pursuit of tests to produce information to support effects and proofs. It was apparent that the previous spontaneity in problem-solving had vanished.

Similar approaches towards the problem were observed in the control subjects at the formal stage. There was no evidence of organisation of the experiment with an eye to proof. However, both the experimental group and the control group claimed to like the train task as much as the chemicals task (see Appendix H). This factor is probably linked to, but not on its own accountable for, performance. Since the experimental and control groups had received different treatments, the explanations for listless behaviour could differ for subjects from the two groups.

It appears that the untrained subjects viewed the chemical experiment and the train task as markedly different in content. Few control subjects attempted to solve the train problem with the same approach that they had used on the chemicals task (see Section 6.7.5). In Chapter 2, there is reference to studies which suggest that

> *... it is the contextual aspect of a task for a pupil rather than its logical structure which determines performance.* (Driver, 1981, p. 9).

By the time the pupils participating in the present investigation reached senior high school, they had had regular laboratory experience. Although their school experience was not directly applicable to the chemicals task, it is possible that the control subjects viewed a task which required handling of test-tubes and mixing of solutions as a reasonably familiar assignment. On the other hand, the train task was unlikely to have been affected by a similar process of educational diffusion and therefore may have constituted for the pupils a novel problem. If this were so, then, in the case of the control group, there would seem to be justification for the observed behaviour. Ausubel (1968) has stated,

> *Generally mature students tend to function at a relatively concrete or intuitive level when confronted with a particularly new subject-matter area in which they are totally unsophisticated.*

However, the experimental subjects could hardly be described as unsophisticated in the principles behind the train task when each of them had produced a tree diagram which indicated all possible combinations of switches prior to the actual execution of the experiment. Possibly the training which they had received in the construction of tree diagrams, had suggested to them that the generation of a complete factorial array was an efficient method of problem-solving in contrast to their usual hypothetico-deductive methods. Nevertheless, lack of resistance to counter-suggestion seems far less likely than the explanation that these pupils were simply responding within the confines of what they considered to be the stereotype of performance expected by the teacher, in accordance with their well-established habit of *'switching off'* in answer to lack of stimulation in the usual classroom situation where teaching is necessarily directed mostly towards the pupils of average ability. In particular, the performances of three boys from the higher IQ Standard 9 experimental group were noteworthy. The boy with the highest IQ, who had previously been assessed at substage III–B, produced a tree diagram but carried out the experiment without consulting the diagram, mentally following the block pattern indicated by the tree. He omitted 22% of the combinations, later ascribing his omission to the fact that he did not have to think with such a mechanical method and would prefer a more demanding pattern of combinations.

Another boy from the group, also at substage III–B, exhibited an unexpected degree of hesitation on the task. He explained that this was due to confusion between the binary counting pattern which he often used and the block [illegible] specified by the tree diagram. The boy with the third highest IQ also po[illegible] an established system of generating combinations. He had been assessed at the upper limit of substage III–A, that is exhibiting maximum organisation of combinations. However, he became confused on the task by the tree diagram to the extent that he could execute neither pattern satisfactorily.

The adverse effect of training on late formal subjects leads to the suggestion that such subjects should be screened before the application of a training programme in the regular school situation. In Chapter 2, the problems associated with intertask reliability have been discussed. Several combinatorial tasks, therefore, should be employed to give an overall impression of each individual's stage classification with respect to the combinatorial scheme.

6.7.4 The Effect of Record-Keeping

Of the experimental group, 98% kept a written record. Each of these records consisted of an advance scheme of all possible combinations (with no redundant combinations) which guaranteed solution of the problem. This type of record is superior to those merely listing combinations actually tried.

Each record was in the form of a tree diagram, although this had not been requested or directly suggested. The possibility of interpupil communication concerning the use of a tree diagram to facilitate execution of the train factorial experiment, has to be acknowledged. In this case, pupils would have had to establish which treatment other individuals had received before suggestions as to technique were passed on. No control subject produced a tree diagram.

Four pupils commenced the drawing of an incorrect tree diagram but of these, three spontaneously corrected themselves. Minor errors, due to carelessness, were made by a further four pupils, which were not corrected.

A difficulty experienced by many pupils was the spatial organisation of the diagram with respect to the size of the paper provided, as a result of which many pupils requested extra paper on which to redraw the diagrams using a smaller scale. This applied to 51% of the pupils who drew tree diagrams and 10% of these (i.e. 5% of those presenting tree diagrams) had to draw three diagrams.

Of the group which produced trees, 97% correctly interpreted the alternative courses of action indicated by their diagrams. Most of the group (92%) extended their methodical approach to marking off each combination on the diagram as it was experimentally executed.

Of the control group, 62% noted in writing one or more of their responses. It is of interest to compare this percentage with the corresponding figure of 12% on the pretest. This seemed to indicate within the control group a growing awareness, arising from experience, of the desirability of keeping records. In spite of this, there was no control subject who avoided redundant combinations. The mean number of redundant combinations was 21.

The effect of record-keeping on experimental design proficiency was examined statistically by dividing the control group into record keepers and non-record keepers as shown in Table 6.8. It had been expected that record-keeping should result in higher scores as it ought to prevent trials of redundant combinations and should assist in the generation of further non-redundant combinations. This led to application of the one-tailed Mann-Whitney U test also contained in Table 6.8. Examination of the Mann-Whitney test statistic indicated that in four of the seven blocks, record keepers achieved higher scores than non-record keepers, while in the other blocks the reverse occurred. In all cases, there was no significant difference between record keepers and those who did not keep records.

The above results indicate that it is not the mere keeping of records which is significant for the efficient generation of combinations but that the kind of record kept is of prime importance. To illustrate, many record keepers tried each successive combination without any reference to previous trials on their lists. A systematic approach was essential to achieve a high score on the task but record-keeping did not exclude random trials by pupils. Moreover, only one control subject prepared a written advance scheme of combinations.

Table 6.8 Rank Ordering of Scores on the Evaluation Task : Null Hypothesis 6.8 under Test

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

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

Block	Sample Size		Mann-Whitney Statistic		Value of U		z	p	Decision re H_0
	Record Keepers	Others	Record Keepers	Others	Observed	Critical, $\alpha = 0{,}05$			
Boys									
High IQ Std. 9	6	8	12	36	12	–	-1.56	0,0594*	accept
Low IQ Std. 9	9	5	27,5	17,5	17,5	9	–	–	accept
High IQ Std. 8	9	6	18	36	18	12	–	–	accept
Low IQ Std. 8	8	6	29,5	18,5	18,5	–	–	> 0,245	accept
Girls									
High IQ Std. 9	6	0	–	–	–	–	–	–	–
Low IQ Std. 9	4	2	2,5	5,5	2,5	–	–	> 0,267	accept
High IQ Std. 8	4	2	3,5	4,5	3,5	–	–	> 0,400	accept
High IQ Std. 8	4	2	5,5	2,5	2,5	–	–	> 0,267	accept

6.7.5 Classification of Search Patterns

In cases where control subjects systematised their approaches to the task (to a greater or a lesser degree), in general the search patterns used tended to be less complex than those used in the first chemical experiment, possibly occasioned by the slightly greater information-processing load of the former task. Table 6.9 summarises the main search patterns which were identified. The simplest patterns were symmetrical or progressive in form (columns 1 and 2 of Table 6.9). Both these patterns necessarily terminated after a short sequence of combinations and usually left subjects confused over the manner in which to continue the task. The more successful patterns were various block systems. The block pattern shown in column 3 was the most frequently used but inherently omits many combinations. It appeared that only a minority of control subjects recognised the train task as essentially the same problem as the first chemical experiment. This estimate formed 12% of the control group. Such subjects

Table 6.9 Search Patterns Used By Control Subjects

Key u = switch in up position m = switch in mid position d = switch in down position

	Type of Search Pattern					
	Symmetrical	Progressive	Block	Block	Block	Block
Switch numbers	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Switch positions	u u u u u	u u u u u	u u u u u	u u u u u	u u u u u	u u u u u
	d d d d d	d u u u u	d u u u u	d d u u u	d u u u u	u u u ...
	d u d u d	d d u u u	u d u u u	d u d u u	d d u u u	+ all possible
	u d u d u	d d d u u	u u d u u	d u u d u	d d m u u	combinations
	u u d u u	d d d d u	u u u d u	d u u u d	d d d u u	of switches
	d d u d d	d d d d d	u u u u d	u d d u u	d d u m u	4 and 5 (xy)
	*	*	d d u u u	u d u d u	d d u d u	
			u d d u u	u d u u d	d d u u d	u u m ...
			u u d d u	u u d d u	d d m m u	+ x y
			u u u d d	u u d u d	d d d m u	
			d d d u u	u u u d d	d d d d u	u u d ...
			u d d d u	d d d u u	d d d u d	+ x y
			u u d d d	d d u d u	d d d m u	
			d d d d u	⋮ ⋮	d d d d u	Repeat scheme
			u d d d d	d d d d u	d d d d d	for
			d d d d d	d d d u d	Proceed to	u d...
			*	d d u d d	blocks	d u ...
				d u d d d	commencing	d d ...
				u d d d d	d u ...	
				d d d d d	u d ...	
				*	u u ...	

*The trials following this stage were usually combinations involving midway settings of switches 3 and 4.

were identified as those who transferred the block pattern which had been the most commonly used in the chemicals task, to the train task (column 4). Identification was assisted by the fact that many of these pupils spontaneously commented that the same procedure could be used for both tasks. An ... r block pattern which was more efficient than those previously described but whic... uld not possess the ability to produce a complete factorial array, is exemplified in column 5. Only exceptional control subjects attempted to use a block pattern which essentially corresponded to a tree diagram, resulting in the generation of all possible combinations if correctly evolved.

These observations confirm the work of DeLuca (1979) on his electronic equivalent of Piaget's first chemical experiment in which he reports that the search patterns most frequently utilised by subjects were block systems. The choice of block patterns can be interpreted in terms of Miller's (1956) concept of chunking or encoding information, cited by DeLuca in his research. This involves a reduction procedure which simplifies a complex task to well within the immediate memory span by encoding relevant pieces of information into chunks. In the present task, a block pattern apparently represents the most efficient encoding of information in relation to memory load and the number of chunks required. Each chunk then consists of a set of combinations which contain a fixed number of switches at a specific setting. This is tantamount to holding a given number of variables constant in turn until all combinations have been generated.

6.7.6 Summary of Statistical Results*

1. Trained subjects achieved significantly higher scores than untrained subjects. Even in the case of the lower IQ Standard 8 boys (where experimental subjects had performed better in the pretest than control subjects, $U = 54{,}5$, $p = 0{,}0456$), the extreme value of zero observed for the Mann-Whitney test statistic indicated that training was beneficial to this group of subjects also.

2. The train task itself did not inherently discriminate to a significant extent among groups of subjects divided by standard (age), IQ and sex.

---

*See the footnote at the end of Chapter Four.

3. Groups of subjects divided as above did not differ significantly in their responsivity to training with the exception of the higher IQ Standard 9 boys who did not benefit from instruction to the same degree as the other groups.

4. Untrained subjects who kept written records did not score significantly better than those who did not do so.