THE TRANSITION FROM CONCRETE
TO FORMAL THINKING

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

by

Susan Clare Page (née Somerville)

Australian National University
December, 1970.

Copy No. 1
CHAPTER 5


5.1 Measures of Ability used in Selection of the Sample

Performance on a number of other tests and school examinations could be expected to relate to performance on the Four Group Task. It was not a major aim of this investigation to study such effects in their own right. Rather performance on these other tasks was seen as an appropriate basis for the selection of a representative sample. Furthermore, it is quite possible that there would be effects due to the particular teacher a subject has had and even to the particular school he attends. Again, the investigation was not specifically designed to study these latter effects, although some post hoc comparisons may be made.

Assessed I.Q. scores, based on A.C.E.R. verbal tests Junior B and Intermediate D (A.C.E.R. 1948-49 and 1939-53 respectively) administered at the ages of about 9 and 11 respectively, were available from school record cards. These tests are group tests administered by School Counsellors. However, since performance on the Four Group Task would be more likely to relate to a nonverbal ability score, the Standard Progressive Matrices (A.C.E.R. 1966) was administered, as a group test, to all classes of pupils in 4th, 5th and 6th classes at two Primary Schools and in 1st, 2nd and 3rd forms at two Secondary Schools. In addition the most recent mark in a school mathematics examination was recorded for each of the
pupils in these classes. This data, together with considerations of equal representation of sexes and representation of the schools in proportion to the size of their enrolments, formed the basis of selection of the sample for testing. Selection procedures will be discussed in detail in Section 5.2.

5.2 Selection of the Sample

5.2.1 Determination of five age groups, and the representation of different schools.

In approaching schools for permission to do the testing, no attempt was made to find a sample representative of the Canberra population as a whole. The only concern was to ensure that the schools providing the samples of primary school attenders drew on the same areas of Canberra and in the same proportions as did those providing the samples of secondary school attenders. Two secondary schools were approached initially and gave permission for the study to be conducted using their pupils. Then primary schools whose pupils pass on to these secondary schools were approached and permission obtained from two out of the three concerned. Thus for one secondary school, only half of the primary school children who would feed into it were available for testing and it was decided therefore to take less than half of the total sample from schools in this area. The other available primary school was one with very large enrolment, providing almost all of the intake to the other secondary school. Thus in each of the
age groups in the sample, subjects were drawn from two schools approximately in the ratio 3 : 2.

A decision was made to test five different age groups, rather than a sample varying continuously in age, for a number of reasons. The methodological reasons for wanting to correlate performance on tasks within age groups have already been discussed in Chapter 2. For this to be possible, a reasonably large number of subjects, falling within a rather narrow age range, is required, at each age level. Secondly, if different age groups are to be compared, a separation of each group from the next (by about 6 months) is likely to make the differences between them more pronounced. Thirdly, it is administratively easier to sample, on the basis of two stratifying variables, from groups in which age can be regarded as constant, than it is while at the same time trying to achieve a good distribution over a continuous age range.

From a study of the age distributions within the classes pre-tested on the Standard Progressive Matrices, the following five age groups were selected:-

(i) 9 yrs. 11 months to 10 yrs. 5 months inclusive - "10 year-olds"
(ii) 10 yrs. 11 months to 11 yrs. 5 months inclusive - "11 year-olds"
(iii) 11 yrs. 11 months to 12 yrs. 5 months inclusive - "12 year-olds"
(iv) 12 yrs. 11 months to 13 yrs. 5 months inclusive - "13 year-olds"
Groups (i), (ii) and (iii) were all Primary School pupils and Groups (iv) and (v) were Secondary School pupils. The age ranges are the ages at time of testing on the Four Group and Pendulum Tasks. The testing on Standard Progressive Matrices occurred from two to six months prior to the individual testing. In the individual testing the groups were tested in the order (iii) (iv) (v) (ii) (i) for a variety of practical reasons concerned with school examinations and the transfer of 12 year olds from primary to secondary school.

The detailed considerations entering into the selection of the age groups, described above, will now be presented, and in the discussion the schools involved will be referred to as follows:-

School A(S) - the secondary school contributing approximately 60% of the 13 and 14 year-old samples.

School A(P) - the primary school contributing approximately 60% of the 10, 11 and 12 year-old samples, drawing pupils from the same area as School A(S).

School B(S) - the secondary school contributing approximately 40% of the 13 and 14 year-old samples.

School B(P) - the primary school contributing approximately 40% of the 10, 11 and 12 year-old samples, drawing pupils from the same area as School B(S).

A number of considerations entered into selection of the sample for testing at each age group. Firstly, there was no alternative to
complete confounding of age with educational experience, since fairly
strict age limits are placed on entry to particular grades in school.
As a result, children whose age is markedly above or below that of the
majority in their class are likely to be atypical. It would seem a
mistake, therefore, to include such children in an age sample drawn
primarily from the classes one grade above or below their own. Thus the
first decision made was to select each age sample from a single grade
in the schools. The distributions of ages in school grades were inspect-
ed to determine the smallest possible age range which would give a reason-
ably large number of subjects from which to select the sample for testing.
These distributions are shown in Table 5 below and the sections from which
the sample was drawn are indicated. As mentioned previously the age
range of those eligible for selection is always from \((x-1) \) yrs. 11 mths.
to \(x\) yrs. 5 mths. \(\text{(where } x \text{ is the subject's "nominal age"})\). Thus, for the
first column of the table, showing frequency and cumulative frequency dis-
tributions over age of all children in 5th classes at Primary school, the
figures are to be interpreted in terms of a "nominal age" of 10 years.
Specifically, the children eligible for selection for the "10 year-old"
group are 107 children (56 at School A(P) and 51 at School B(P) in the
age range 9 yrs. 11 mths. to 10 yrs. 5 mths. As the table indicates,
of these 107 a total of 48 (27 in School A(P) and 21 in School B(P))
were actually selected and tested individually. The table also indi-
cates that of the 48 tested, 24 were male (14 in School A(P) and 10 in
School B(P) and 24 were female (13 in School A(P) and 11 in School
B(P)). The table is to be interpreted similarly for the other four
<table>
<thead>
<tr>
<th>Age at Time of Individual Testing</th>
<th>Class Group Tested</th>
<th>Primary Schools</th>
<th>Secondary Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Relative to Nominal Age, x)</td>
<td>5th Class 1960</td>
<td>(Nominal age, x-10)</td>
<td>(Relative to Nominal Age, x)</td>
</tr>
<tr>
<td>(Nominal age, x+11)</td>
<td>School</td>
<td>School</td>
<td>Both</td>
</tr>
<tr>
<td>(Relative to 5th Class)</td>
<td>A(P)</td>
<td>A(P)</td>
<td>B(P)</td>
</tr>
<tr>
<td>(Relative to 6th Class)</td>
<td>Schools f.c.</td>
<td>f.c.</td>
<td>f.e.f.</td>
</tr>
<tr>
<td>(Relative to 1st Form)</td>
<td>Schools f.c.</td>
<td>f.c.</td>
<td>f.e.f.</td>
</tr>
<tr>
<td>(Relative to 2nd Form)</td>
<td>Schools f.c.</td>
<td>f.c.</td>
<td>f.e.f.</td>
</tr>
<tr>
<td>(x-1) : 5</td>
<td>1 1 1 1 2 2 0 0 0 0 0 0</td>
<td>1 1 7 7 8 8</td>
<td>7 7 7 7 14 14</td>
</tr>
<tr>
<td>(x-1) : 6</td>
<td>1 2 0 1 1 1 0 0 1 1 7 1</td>
<td>2 3 5 12 7 15 0 13 3 10 9 23</td>
<td>7 7 7 7 14 14</td>
</tr>
<tr>
<td>(x-1) : 7</td>
<td>2 4 0 1 2 2 0 0 2 4 4</td>
<td>5 8 4 10 9 24</td>
<td>15 18 5 15 10 23</td>
</tr>
<tr>
<td>(x-1) : 8</td>
<td>0 3 0 1 0 3 0 1 3 2 3</td>
<td>1 1 13 10 6 30</td>
<td>13 30 7 22 21 53</td>
</tr>
<tr>
<td>(x-1) : 9</td>
<td>1 5 4 5 2 10 5 4 1 2 3 1 3 4 6</td>
<td>1 1 15 12 16 29</td>
<td>9 3 4 12 7 24 0 14 13 65</td>
</tr>
<tr>
<td>(x-1) : 10</td>
<td>15 10 7 12 12 22 15 10 12 12 14 15 13 12 23</td>
<td>5 3 21 26 6 45</td>
<td>7 17 22 30 21 53</td>
</tr>
<tr>
<td>(x-1) : 11</td>
<td>17 12 15 8 30 17 12 9 5 5 10 13 7 14 23 11</td>
<td>11 34 2 26 15 60</td>
<td>5 37 11 51 16 108</td>
</tr>
<tr>
<td>x = 0</td>
<td>11 5 5 23 11 41 1 24 0 27 13 50</td>
<td>1 3 9 4 30 0 60</td>
<td>11 68 8 50 10 127</td>
</tr>
<tr>
<td>x = 1</td>
<td>11 32 1 30 18 50 9 32 4 31 13 63</td>
<td>8 47 8 36 10 85</td>
<td>11 68 7 50 10 127</td>
</tr>
<tr>
<td>x = 2</td>
<td>11 33 8 40 18 70 3 35 4 35 7 70</td>
<td>12 50 5 43 17 102</td>
<td>10 50 14 54 26 170</td>
</tr>
<tr>
<td>x = 3</td>
<td>10 33 8 47 17 91 8 43 6 41 14 84</td>
<td>8 67 3 40 11 113</td>
<td>10 50 14 54 26 170</td>
</tr>
<tr>
<td>x = 4</td>
<td>12 57 7 55 21 112 9 52 11 52 20 104</td>
<td>7 74 5 51 12 125</td>
<td>11 106 7 101 18 207</td>
</tr>
<tr>
<td>x = 5</td>
<td>0 62 8 63 10 128 10 62 4 56 16 118</td>
<td>10 84 2 53 12 127</td>
<td>10 115 13 143 22 220</td>
</tr>
<tr>
<td>x = 6</td>
<td>12 89 3 66 18 186</td>
<td>12 74 0 55 21 130</td>
<td>3 87 1 54 4 141</td>
</tr>
<tr>
<td>x = 7</td>
<td>6 89 6 72 12 158</td>
<td>7 81 2 68 10 160</td>
<td>2 80 3 57 5 140</td>
</tr>
<tr>
<td>x = 8</td>
<td>5 93 1 73 6 166</td>
<td>7 88 7 75 14 163</td>
<td>2 91 1 58 3 140</td>
</tr>
<tr>
<td>x = 9</td>
<td>7 98 3 70 10 178</td>
<td>4 92 0 75 4 167</td>
<td>1 92 0 58 1 130</td>
</tr>
<tr>
<td>x = 10</td>
<td>5 103 1 77 6 180</td>
<td>5 97 2 77 7 174</td>
<td>2 94 0 58 2 152</td>
</tr>
<tr>
<td>x = 11</td>
<td>10 107 2 70 6 186</td>
<td>5 103 3 56 8 180</td>
<td>3 97 1 50 4 155</td>
</tr>
</tbody>
</table>

| No. eligible for Selection (i.e. in range (x-1): (x+1)) | 55 | 50 | 90 | 92 | 83 | 94 |
| No. Selected: TOTAL | 27 | 27 | 43 | 44 | 40 | 48 |
| MALE | 14 | 14 | 10 | 10 | 23 | 23 |
| FEMALE | 13 | 13 | 11 | 11 | 25 | 25 |

Table 5: Frequency and cumulative frequency distributions of subjects pretested on the Standard Progressive Matrices over age. These distributions define the populations available for individual testing. The ages shown are relative to the "nominal age" of the class group. The age ranges from which subjects were actually chosen are those given by (x-1) yrs. to (x+1) yrs. in each of the nominal age groups. The total number of children (in the specified age range) available for testing, in each of the schools, and in each of the nominal age groups, is shown below each distribution. The number of those subjects actually tested in the sample (with a breakdown into sexes) is then shown for comparison.
age groups. The other considerations which entered into the selection of those for testing from those in the eligible age range will now be discussed.

5.2.2 Use of the variables discussed in Section 5.1 in selection of the sample.

In broad terms, the study aimed to investigate the performance of the typical 10, 11, 12, 13 and 14 year-old in the population of these schools. Rather than relying on a random samples (from those in the eligible age ranges) to provide typical groups for study, it was decided to stratify the groups of those available on variables likely to be relevant to performance, and then sample in such a way as to make the sample representative of the population on these variables. The two variables chosen as most relevant were nonverbal I.Q. score and performance in mathematics at school. As mentioned in Section 5.1, these were measured by the Standard Progressive Matrices and by the most appropriate mark in a recent school mathematics examination, respectively.

Such a procedure enables an evaluation of the relationship of these variables to performance on the tasks in a way which might not be possible with a randomly chosen sample. From the discussion in Chapter 2 it is clear that, until recently, very few studies arising from Piaget's work have concerned themselves with other ability variables, and that, particularly with older subjects, the relationship of performance on Piagetian tasks to that on other tests is not clear. It was therefore thought important, in this study, to collect such correlational data as was possible, as an indication of the relationships which
exist for subjects of these ages. It was also considered advisable to select samples of subjects in which the correlation between the two relevant variables found in the population was preserved. A method of stratifying the populations of subjects available on the two measures, and then sampling proportionately to the number in each cell, seemed appropriate. However three additional requirements were placed on the sample. The first, already mentioned, was that the representation of the two schools concerned for each age group should be roughly 3:2, School A to School B; the second was that this ratio should also hold as far as possible within each cell of the stratified table; and the third that the number of male and female subjects should be equal overall, and equal also, as far as possible, within each school.

Tables 6, 7, 8, 9, 10 below are provided as a summary description of the samples tested in each of the five age ranges. In each case, the distribution was split at the median, for both schools combined, of raw scores on the Standard Progressive Matrices, giving categories of High and Low scores. The divisions into High, Medium and Low on mathematics mark were made at the thirty three and sixty seven percentile points for each school separately, except in the case of the twelve year olds where the distributions were sufficiently similar to be combined into one. The tables show, for each age group, the number of subjects eligible for testing in each cell (tables (a) ) the number actually tested (tables (b) ) and the ratio of number tested to number eligible (tables (c) ). Similar tables with, in addition, a break-down within each cell into schools and sexes are provided in Appendix III. They
### TABLE 6: 10 YEAR-OLDS

<table>
<thead>
<tr>
<th></th>
<th>Mathematics Mark</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw L</td>
<td>11</td>
<td>17</td>
<td>27</td>
<td>55</td>
</tr>
<tr>
<td>Raw M</td>
<td>20</td>
<td>20</td>
<td>12</td>
<td>52</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>37</td>
<td>39</td>
<td>107</td>
</tr>
</tbody>
</table>

### TABLE 7: 11 YEAR-OLDS

<table>
<thead>
<tr>
<th></th>
<th>Mathematics Mark</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw L</td>
<td>6</td>
<td>13</td>
<td>25</td>
<td>44</td>
</tr>
<tr>
<td>Raw M</td>
<td>23</td>
<td>13</td>
<td>8</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>31</td>
<td>33</td>
<td>93</td>
</tr>
</tbody>
</table>

### TABLE 8: 12 YEAR-OLDS

<table>
<thead>
<tr>
<th></th>
<th>Mathematics Mark</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw L</td>
<td>9</td>
<td>13</td>
<td>25</td>
<td>47</td>
</tr>
<tr>
<td>Raw M</td>
<td>24</td>
<td>13</td>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>26</td>
<td>33</td>
<td>92</td>
</tr>
</tbody>
</table>
### TABLE 9: 13 YEAR-OLDS

(a) Number of subjects eligible

<table>
<thead>
<tr>
<th>Mathematics Mark</th>
<th>L</th>
<th>M</th>
<th>H</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Score</td>
<td>0</td>
<td>20</td>
<td>36</td>
<td>65</td>
</tr>
<tr>
<td>S.P.M.</td>
<td>37</td>
<td>26</td>
<td>9</td>
<td>72</td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td>46</td>
<td>137</td>
<td></td>
</tr>
</tbody>
</table>

(b) Number of subjects tested

<table>
<thead>
<tr>
<th>Mathematics Mark</th>
<th>L</th>
<th>M</th>
<th>H</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Score</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>S.P.M.</td>
<td>12</td>
<td>8</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>16</td>
<td>48</td>
<td></td>
</tr>
</tbody>
</table>

(c) Ratio: Number tested/Number eligible

<table>
<thead>
<tr>
<th>Mathematics Mark</th>
<th>L</th>
<th>M</th>
<th>H</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Score</td>
<td>.50</td>
<td>.33</td>
<td>.37</td>
<td>.37</td>
</tr>
<tr>
<td>S.P.M.</td>
<td>.31</td>
<td>.31</td>
<td>.40</td>
<td>.40</td>
</tr>
<tr>
<td>Total</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
</tr>
</tbody>
</table>

Tables 6, 7, 8, 9, 10: Samples of subjects tested in each of the five age groups.

The tables show, for each age group separately, the number of subjects in the eligible age range (x to x+1 years 11 months to x years 5 months, where “x” is the nominal age in each case), the number of subjects selected for individual testing, and, finally the ratio of the number tested to the number eligible. These figures are shown for three levels of mathematics mark, and two levels of raw score on the Standard Progressive Matrices, in each age group, as well as for marginal and overall totals. A more detailed breakdown of these tables, by sex and school attended, together with details of the mathematics marks used, can be found in Tables 6A, 7A, 8A, 9A and 10A of Appendix III. Tables 6B, 7B, 8B, 9B and 10B of the same appendix provide means, standard deviations and correlations between the stratifying variables (and assessed I.Q.) for the whole school grade concerned, those in the eligible age range, and those actually tested. These data confirm that the samples tested were typical of the population in the schools from which they were drawn, in terms of these three variables at least.
are labelled Tables 6A, 7A, 8A, 9A and 10A to correspond with the condensed tables presented here.

Details of the mathematics marks used and of the mechanics of selection are also to be found in Appendix III. As confirmation of the fact that this method yielded samples typical of the population from which they were drawn (in terms of these variables) the means, standard deviations and correlations are shown in the same Appendix for the following three groups:- the sample actually tested; those in the eligible age range from which the sample was drawn; the entire population of the relevant school grade (including those in and out of the eligible age range). Since the schools' assessed I.Q. ratings were available (although not used in the sample selection procedure), figures for them accompany those for the other variables. Slight discrepancies in the number of observations contributing to the statistics are due to the fact that, for some children, an assessed I.Q. rating was not available, but these children were not discarded when plotting the bivariate distributions of raw scores on Standard Progressive Matrices and mathematics marks. These data are in Tables 6B, 7B, 8B, 9B and 10B of Appendix III.

A number of points about the samples tested, arising from the more detailed tables in Appendix III, should be noted. Firstly, the sample tested did not always reflect the sex distribution of those in the eligible age range. For example, the eligible twelve year-olds comprised 56 boys and 36 girls, while the sample tested has 22 of each. Secondly, with respect to the distribution of raw scores on the Standard Progressive Matrices, it will be noted that pupils attending schools
B(P) and B(S) score consistently higher than those attending A(P) and A(S). While some attempt was made to correct for this in selection of the sample, there are several instances where all or nearly all eligible children in one cell of the table were pupils of the same school. In such cases it was impossible to draw from the schools in the adopted ratio, 3 : 2, (School A : School B) in those particular cells. However, an attempt was still made to have the ratio for the whole table close to 3 : 2, by compensating in other cells. Finally, it was decided to keep the minimum number tested in any cell at four, with a view to using Analysis of Variance techniques to investigate the effects of intelligence and mathematics ability on such measures as the efficiency of learning the task. This consideration affected the number tested in only one cell of one of the tables. Of the six eligible eleven year-olds in the low-low category, four were tested, a much higher proportion than in the other cells of the table.

For each individual subject tested, the value of the Assessed I.Q., the raw score on the Standard Progressive Matrices and the Mathematics Mark are shown in the tables of raw data in Appendix VII.

5.3 Measure of Operational Level (Piaget)

To enable validation of the Four Group Task in a Piagetian framework, it was necessary to include one of Inhelder and Piaget's (1958) tests to locate children, by their techniques, within the stages of Concrete and Formal thinking. While it might have been desirable to do this as part of the preliminary testing, and use the information in
the selection of the sample, it was not practicable to contemplate individual testing of such a large number of children. Testing of performance on this task was therefore done together with testing on the Four Group Task. An unfortunate consequence of this was that only one of Inhelder and Piaget's (1958) tasks could be included, because of time considerations. A notional 45 minutes was allowed for testing on the Four Group Task, plus 30 minutes for one of Inhelder and Piaget's (1958) problems.

The Piagetian test chosen was the Pendulum Problem (see Chapter 2) in which the subject is asked to experiment with different weights, different lengths of string and different dropping points to find out what determines the frequency of oscillation of a pendulum. The reasons for choosing this test were as follows. It has been pointed out, in Chapter 3, that the Four Group Task cannot readily be conceptualised in terms of specific members of the set of Sixteen Binary Operations, but that it can be said to necessitate formal operations for an evaluation in terms of a "role" for each of the lights. In order to abstract such "roles" a subject must consider the effect of each light in combination with each other light and with itself. He must hold one light constant, vary the others, and deduce a principle from the results. Inhelder and Piaget's (1958) Pendulum Problem is one in which a method of holding all variables but one constant is essential to an exclusion of the irrelevant variables (such as the size of the weight). While this is not identical to the procedure required for the discovery of the "roles" of elements in the Four Group Task, it is sufficiently similar to suggest that the method of holding all but one variable constant on the Pendulum Problem
may be a necessary (prerequisite) skill for a formal operational approach to the other task. It is suggested that the method of investigation required for a formal solution of the Four Group Task is more difficult than that required for the exclusion of factors in the Pendulum Problem for the following reasons.

In the Four Group Task, the "effects" or "roles" of the lights are abstracted by the subject from the combinations he sees. In the Pendulum Problem the effects of variables are directly observable. Thus, in the Four Group Task, once the subject has conceived the idea that each light may have a formal "role", he has to realise that he may deduce it by making a generalisation about its effect on every other light (and on itself). He must ignore the "role" of other lights while he looks at that of any one. The methodological requirements are similar to those of the Pendulum Problem, in that, on the latter task, the subject must realise that if he varies other factors while looking at the effects of any one, the results which he observes may not be due to the factor which it was his intention to investigate. Thus the difference between the two tasks is that the variables involved in the Four Group Task are abstracted from observations, and "holding one constant" amounts to concentrating, mentally, only on its effects, ignoring those of others which are still, by the nature of the task, there at the same time. In the Pendulum Problem on the other hand it is possible physically, to hold all other factors constant, in order to study the effects of any one, and in addition the effects will be observable immediately, not requiring mental separation from those of other factors. In short, on the Four Group Task
a formal operational subject must realise that he needs the kind of experimental control relevant to the Pendulum Problem, but that this can be achieved not on a practical, but only on an abstract level, by means of "mental" concentration on the role of just one element at a time. It thus seems reasonable to regard the Four Group Task as more difficult, methodologically, than the Pendulum Problem; and an exclusion of irrelevant variables on the latter task seems a meaningful prerequisite to a formal operational evaluation of the former.

One other reason for choosing the Pendulum Problem rather than others from Inhelder and Piaget (1958) was that formal operational approaches to it were found in comparatively young children by the original authors and by others (Lovell 1961; Jackson 1965). It therefore seemed possible to regard it as a task whose content was such as to make it soluble early in the transition from concrete to formal thinking. Such a task was desirable in view of the methodological considerations discussed in Chapter 2. Specifically, performance on it could be expected to relate better, in a transitional phase such as Flavell and Wohlwill's (1969) Phase 3, to performance on other tasks, than could performance on a more difficult formal operational problem. In Flavell and Wohlwill's (1969) terms, the Pendulum Problem was chosen in the hope that the value of $P_b$ for it was a relatively low one, compared with the value for other formal operational tasks described by Inhelder and Piaget (1958). It seemed to be the only "easy" task with a definite emphasis on the method of holding all but one variable constant, shown to be of importance in the discussion above.
Apparatus was built to resemble that used by Inhelder and Piaget (1958) as closely as possible (see Figure 2.). The method of questioning also followed Genevan clinical lines. Inhelder and Piaget's (1958) analysis of the problem and a detailed description of the apparatus and instructions used are given in Sections 5.2.1 to 5.2.5 below.

5.3.1 Apparatus.

The apparatus consisted of a peg board in front of which weights could be suspended on a string. The weights were brass (with a hook on top) and comprised one each of weights 20gm, 50gm, 100gm, 200gm, 500gm. The length of the hook varied with the size of the weight so that when any one was attached to a given length of string, the total length was always the same. Three lengths of string were provided, to fit over a rod at the top, and onto which weights could be hooked at the bottom. The ratio of lengths was 3 : 2 : 1. The weights could be swung backwards and forwards in front of the board and two pairs of lines were drawn on the cardboard backing to give an indication of the limits of the path travelled. A clock was available to time the oscillations.

5.3.2 Problem.

The subject was asked to find out what makes the pendulum swing more quickly. (i.e. the factor determining the frequency of the oscillations). The variables which on seeing the apparatus, might seem to be relevant, according to Inhelder and Piaget (1958), are:-
Figure 2: The Pendulum Apparatus. The diagram (drawn to scale) shows the apparatus used to seat subjects on the Pendulum Problem (from Inhelder and Piaget, 1958). Weights were made of brass and the three pendulum lengths (string length plus length of attached weight) were 17 in, 21 in, and 25 in (long). The diagram shows the 500 g weight on the long string. Pairs of red and green lines (labeled B and G respectively in the diagram above) were drawn to mark two amplitudes of oscillation. Seats were grouped as the weight passed the center black line.
(i) the length of the string
(ii) the weight of the object hooked onto the string.
(iii) the height of the dropping point (amplitude of oscillation).
(iv) the force of the push given by the subject.

Only factor (i) does have an effect on the frequency of the oscillation.

Thus the subject has to exclude factors (ii), (iii) and (iv) in order to isolate factor (i) and establish its causal role. The logical model for the situation, in detail, is as below:-

If $p$ is a statement that there is modification in the length of the string, and $\bar{p}$ that there is no such modification;

If $q$ . . . . . . modification in weight ($\bar{q}$ similarly)
$r$ . . . . . . modification in height of drop ($\bar{r}$ similarly)
$s$ . . . . . . modification in impetus ($\bar{s}$ similarly)
$x$ . . . . . . modification of result (i.e. of frequency of oscillations) and $\bar{x}$ no such modification; then:

the subject has $2^5$ (i.e. 32) basic possible combinations of events to observe, only half (16) of which are actually true. These can be shown as in Table 11, below.

To solve the pendulum problem, then, a subject must make appropriate observations and deduce from them that $p$ (a change in length of string) implies and is implied by $x$ (a change in frequency of oscillations), independently of change or no change in any other variable or
The tree diagrams show the 32 possible combinations of events in the Pendulum Problem, and the truth values show that the actual relationships between variables can be expressed in the logical statement:

\[(p \lor x) \land (q \land x) \land (r \land x) = p(q \lor r \lor x)\]

in words, \(p\) implies and is implied by \(x\), while \(q\), \(r\), and \(s\) are irrelevant to \(x\).
combination of variables.

If the subject varies, for example, both the weight and the length of string systematically, the observed events will be:

\[(p \cdot q \cdot x) \lor (p \cdot \neg q \cdot x) \lor (\neg p \cdot q \cdot x) \lor (\neg p \cdot \neg q \cdot x)\]

from which the following two truth tables (a) and (b), for the roles of p and q can be constructed.

Taking p (p) and x (x) only, as the basis of classification of events, Table (a) shows the truth values -

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>(\neg p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
</tbody>
</table>

Logical statement of relationship of p and x.

\[(p \equiv x)\] or \[(p = x)\]

In words, p implies and is implied by x; or p is equivalent to x.

Table (a) Truth table for the relationship of p and x.

Taking any other factor, say q(q), and x (x) as the basis of classification, Table (b) shows the truth values.

<table>
<thead>
<tr>
<th></th>
<th>q</th>
<th>(\neg q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

Logical statement of relationship of q and x.

\[(q \equiv x)\]

In words, q and x are independent.

Table (b) Truth table for the relationship of q and x.

The relationship of the height of the dropping point r(r), and that of the impetus s(s), to x(x) are identical to that of q(q) to x(x). Thus the pendulum problem involves the subject in the logical deduction of only two of the sixteen binary operations - an equivalence of p and x and independence of x from all other factors.
5.3.3 Instructions to subjects

"I want you to find out something about these weights and strings, which can be used to make a pendulum swinging from side to side in front of the board. Let's put a string on the rod like this, and hook one of the weights to the bottom of the string. We can pull the weight over here (or here, or here) let it go, and it will swing from side to side. What would be a good way to see how long it takes to swing from where it starts to where it goes on the other side?" (An attempt was made to have the subject himself suggest counting the swings 1-2-1-2- to get an idea of the rhythm. Occasionally, with older subjects, accurate timing (using a clock) was permitted to find either the number of swings per ten second period, or the time taken to do ten swings).

The subject was then told: "I want you to see if you can find out what makes the pendulum take a shorter time to swing from side to side. This is NOT a test, where I would want to know how quickly you could get the right answer. I am interested in what you think to yourself while you are trying things out, so I would like you to talk out loud while you work. You can have as much time as you like and do whatever you want with the weights, strings and so on. Please talk about what you are doing as much as you can; and I will ask you questions if I'm not sure why you are doing something, because this is what I want to find out. I am going to tape record what we say so I can remember it afterwards. Also I am going to write down the things you try out because the tape recorder can't see which ones you pick up." (for a sample record sheet, see Appendix I).
The general types of questioning used from then on were as follows:

1. If no reason was offered for action or no comment on the result, the subject was prompted by "What are you finding out now?" "What did you see?" etc.

2. If a role was claimed for some factor (with or without specification) e.g. "It's going faster" the experimenter probed for a principle by saying "And how could you make it go faster still?"

3. If the role of a factor was asserted as in 2, the experimenter went on to check on other factors by asking "That's all?" "Is anything else important?", if necessary asking "What about weight (e.g.)?"

4. If the subject had discovered the correct relationship, he was asked "How would you show (prove) that it's the length of the string to somebody who didn’t know?" "How do you know?" etc.

Experimentation and questioning of the subject continued until the experimenter was satisfied that all of the relevant variables had been considered by the subject. At the conclusion of this, the subject was asked to imagine that a friend of his was present and was asking him to give a demonstration of all the things he had found out about the pendulum; the subject was then asked to tell and show the experimenter what the best demonstration he could think of would be. Some prompting was given if the subject omitted to mention some variable (s) in his demonstration. The "demonstration" was seen as a way of checking on the level of approach
shown in the previous experimentation.

5.4 Scoring of Performance.

Details of the scoring procedure used for the pendulum problem are given in Chapter 6. No scoring took place during testing, but was done subsequently, by listening to the tape recorded interview in conjunction with the record of the materials used by the subject (see sample Record Sheet Appendix I). Detailed individual protocols are not provided since the task and scoring procedures were taken as directly as possible from Inhelder and Piaget (1958).
CHAPTER 6

SCORING OF PERFORMANCE ON THE PENDULUM PROBLEM.

6.1 The Stages Described by Inhelder and Piaget (1958).

Inhelder and Piaget (1958) describe four discernible substages of operational thought revealed by performance on the Pendulum Problem. Two are within the stage of Concrete Operations and two within that of Formal Operations. They also devote some time to a discussion of the performance of children at the Preoperational Stage. Children at this stage (4 to 7 years approximately) will approach the problem in a completely unsystematic way and be unable either to report events accurately or to make valid conclusions about the effects of variables. The preoperational child does not separate physical causes from his own expectations or from the effects of his own actions and may decide, for example, that the pendulum did not go as fast because he did not want it to. The force of "push" he gives the pendulum is typically thought to determine the frequency.

The beginning of the stage of Concrete Operations is marked by a substage, labelled Stage IIA, which represents advances in several respects on the performance of the preoperational child. In this substage, although a child may have difficulty with the serial ordering of weights, other variables, namely the length of string, heights of dropping points, and speeds of ticking, should be ordered and reported accurately. However, the child's behaviour tends to indicate that he still does not conceive of the situation in terms of variables whose effect he is trying to assess, but is merely content to play with combinations of strings and weights which
interest or look "right" to him. One major advance is that the impetus given by the child himself is separated out from physical variables such as the length of the string. Another advance, according to Inhelder and Piaget, is that since the child reports the observed events accurately he is able to find correspondences between sets of events. For example, the child is now able to report that three pendulums which differ in the length of the string (from shortest to longest) also differ in the speed of ticking (from fastest to slowest) in a corresponding manner. It is in this sense that he discovers the "role" of the length of the string. At the level of concrete operations the relationship between the two variables is not a formal one, that is, not stated in sentences which could be formalised into propositions. The child understands the relationship merely as a correspondence between certain physical events.

With regard to factors other than the length of the string, it is typical for a child at Stage IIA to fail to understand their relationship to the speed of ticking. This is because the child's method of investigation does not ensure that only one factor is varied at a time. It may thus happen that, while looking at the role of some other factor, the child will change the length of the string as well, but conclude that it is the other factor which has caused the change in frequency of oscillation. In the present investigation it was found that subjects typically did not regard the push and the dropping point as separate variables, but saw them as alternative ways to change the same thing. Comments were made such as "Well if you push it more it just seems to make it go higher up". It seems plausible, therefore, to treat these two variables as one and the same, for
children of the ages involved in the present study. It may be true that, for the preoperational child, the push he gives is something quite different, expressing his "intentions" or "expectations", rather than contributing to the physical situation in a lawful way.

One further advance is made within the stage of Concrete Operations, and this marks what Inhelder and Piaget have called Stage IIB. In this stage, while still unable to separate out variables in their investigation, the children are better able to order the effects of weight. At Stage IIA the effects attributed to the size of the weight are haphazard and include, for example, directly contradictory notions (on one occasion that the 500 gm weight is the fastest because it has more power, on another occasion that it is the slowest because it is so hard to carry), or notions that there is a "right" weight for every string (e.g. 20 gm for the short string, 100 gm for the medium string and 500 gm for the long string). In Stage IIB the effects attributed to the size of the weight show a correct ordering. This does not mean, however, that the weight is understood to have no effect on the frequency. Typically it is thought to have an effect, because of the same error made by Stage IIA subjects, namely changing both the weight and the length of string simultaneously. The error in reasoning involved here, "put into propositional language, amounts to concluding from \((p \land r \land s) \rightarrow x\) that \((p \rightarrow x)(q \rightarrow x)(r \rightarrow x)\) without suspecting the existence of other possible combinations" (Inhelder and Piaget, 1958, p. 72). A reciprocal error, described by Inhelder and Piaget as common in Stage IIB, but not found very frequently in the present investigation, is that of concluding that the one factor which has been held constant
over two or more situations cannot be effective if no effect is observed. 
Inhelder and Piaget quote as an example "Anyway, it's not the string 
because it's the same string. In other words, from \( p \lor q \lor r \lor s \lor x \) 
\( p \lor q \lor r \lor s \lor (x \lor \bar{x}) \) he concludes \( \overline{p \lor \bar{x}} \)." (Ibid., p. 72).

In summary, the concrete operational child is limited to 
obseving empirical correspondences, which technique enables him to 
discover the inverse relationship of the length of the string to the 
frequency of oscillation. When considering other variables, however, 
he fails to experiment in a way which would separate their effect from 
that of the length of the string. In fact he cannot see the need to do 
this, because he does not yet think in terms of a variety of possible 
relationships (in abstract) between which he must decide, using appro­
piate evidence. To observe what happens will be enough.

At the beginning of the stage of Formal Operations the combina­
torial system of possible relationships begins to be understood. In a 
first substage (Stage IIIA) however, the child does not know how to 
produce the empirical comparisons he needs to provide himself with valid 
evidence. According to Inhelder and Piaget (1958) he will be able to 
reason correctly if given a properly controlled demonstration of the 
effects of the variables, but he could not devise such a demonstration 
himself. In this substage, the child is aware of the need to hold other 
variables constant and to vary only one thing at a time in order to assess 
its effect, but he is susceptible to several kinds of errors when attempting 
to put this into practice.

Firstly he may make the comparisons unnecessarily complicated by
studying all possible combinations of two different variables, forcing himself to remember and compare many different frequencies of oscillation at once. For example, he may look at the 20 gm weight on the short string, then the 100 gm weight on the medium string and then say "now I have to change over the weights". It is also quite common for the child to find himself somehow holding constant the variable he wants to assess, and varying something else. Such errors cause considerable stress for children at this stage. Frequently their comments reveal the struggle involved in making their empirical tests adequate for the possible conclusions between which they want to decide. Inhelder and Piaget (1958) remark that "these subjects feel that they manage either actually to transform the factor which they want to leave unchanged . . . or to vary all factors by turns without knowing how to focus their analysis on the point being analysed" (Ibid., p. 75).

In summary, Inhelder and Piaget (1958) see the children at Stage IIIA as having a "formal logic in the process of formation . . . superior to their experimental capacity". They hold that the subjects' logic has "not yet adequately structured their method of proof" (Ibid., p. 75). Because of this, they say that the child at Stage IIIA is still able to discover only the easiest relationship, that of the length of the string to the frequency of oscillation of the pendulum. He is likely to fail to extract the relationship of exclusion which requires a more difficult combination of facts. Inhelder and Piaget assert that it is easier to manipulate the operations . . . "which state that which is and establish true implications . . . (than) those which exclude that which
is not and deny the false implications" (Ibid., p. 75).

It is only upon reaching the final substage of Formal Operations, Stage IIIB, that a child becomes capable of using a method of experimentation which changes the appropriate variable, holding all others constant, and enables him to discover its effect. In addition to being able to separate the variables, the child is capable of inferring the more difficult relationship of exclusion for the first time. He sees the situation with respect to each pair of variables in terms of the sixteen possible binary operations and knows that the equivalence relationship, which holds between the length of the string and the frequency of oscillation, is only one of these. The logical model used by Inhelder and Piaget (1958) to describe what is required for solution of the problem at a formal level has been set out in Section 5.2.2 and will not be restated here. It should be noted, however, that in addition to the ability to reason within this structure, a child at Stage IIIB performs very much neater experiments to test his ideas than does the child at Stage IIIA. For example, he is more able to generalise findings about the weights on one string to strings of other length and in general shows much less need to test out such inferences. He makes his statements using terms such as "the size of the weight" rather than "the big one" and "the small one".

6.2 Modifications to the Stage Criteria Made in the Present Study.

In the present investigation, one difficulty has arisen in using the descriptions given by Inhelder and Piaget (1958) to locate children at a developmental stage. This is the fact that the three aspects of perform-
control in experimentation, accuracy of reported observations and
nature of inferences made; do not dovetail together in quite the way they
should. If anything, rather than the formal logic being ahead of the
child's available methods of proof at Stage IIIA, it seems to be the
reverse. Methods of experimentation show meticulous control, but the
subject seems to distort the observed events in order to report what he
expects to find. This latter type of behaviour is described by Inhelder
and Piaget (1958) as characteristic of the much earlier Preoperational
Stage. They say that a child at this early stage who expects that one
particular combination will tick faster than another, when in fact they
are the same, will distort the speed at which he counts aloud to fit in
with his expectation. To do this he must ignore the compelling physical
contradiction of the rhythm which he is imposing on the pendulum. During
the present investigation one subject under such pressure disclosed "I
can definitely see it is going faster, but only when I close my eyes".
Inhelder and Piaget claim that by Substage IIA the observation of
empirical facts and correspondences is accurate. They maintain that a
child who has reached this stage will always report the speed at which
the pendulum is going accurately, and thus point out that if a role is
mistakenly attributed to weight, for example, it must be because the
weight and length of string have been varied simultaneously and not
separated.

In the present investigation it has been found that the two
aspects (conclusions drawn and method of experimentation) do not vary so
closely together. Even amongst 13 and 14 year-olds it is common to find
a child who holds every factor except the weight constant and still "perceives" or persuades himself that there are changes in speed due to changes in weight alone. It seems that a subject's expectations about the role of factors have a stronger effect on his reporting of events than has been suggested previously. Perhaps this is an effect peculiar to the pendulum task, where it is somewhat difficult to assess the frequency of oscillations exactly. Subjects who are "prepared for" the idea that weight may not change the frequency seem to be quite happy that they can see that it does not; however those who seem to expect that everything (including weight) will have an effect become completely preoccupied with judging the "differences" made by the weights (which they may study very systematically). For the latter type of child, differences made by the length of the string and the height of the dropping point may also become "hard" to judge and the child eventually may arrive at an extremely complex set of conclusions: for example that the heavier the weight on the longer string, the faster it will go; that on the medium string all weights are the same; and that on the short string the heavier the weight, the slower it will go. In coming to such a set of conclusions, the subject is certainly not reporting the empirical facts with accuracy. Thus, according to Inhelder and Piaget, he could not be said even to have reached Stage IIA. However his method of investigation may show clearly that he understands the need to separate variables and to vary only one at a time, and this would tend to place him in Stage III rather than Stage II.

The inconsistency described above makes it necessary for the
present study to modify the descriptions of substages given by Inhelder and Piaget (1958) to some extent. The point of view adopted is that the accurate observation of empirical events is a less reliable indication of attainment of concrete operations, and more subject to disruption by expectations, than has been suggested by the original authors. If, as they claim, a child at Stage II has the necessary operational structures to discover that the length of string is inversely related to the frequency of oscillation, then it could be argued that he will anticipate such a relationship between other variables (the size of the weight and the height of the dropping point) and the frequency of oscillation. Since he has not the combinatorial structure of the sixteen binary operations, he is not aware that the relationship between the frequency of oscillation and any of the variables may be one of a variety, including "no effect", or complete affirmation.

Thus for the child who is in transition from Stage II to Stage III it is possible to suggest that equivalence, or reciprocal implication, is the only one of the sixteen binary operations which it occurs to him to consider, perhaps because it is easiest to comprehend, or because it is the one most like the type of relationship he was able to discover, by correspondences, using concrete operations. If this expectation is strong, it is quite conceivable that he will distort or misperceive empirical events to fit in with the kind of interpretation he expects to make. In a sense he is unlucky if his methods of experimentation are too advanced. If he understands the need to vary only one factor at a time, and yet is not prepared to conclude that a factor has no effect (for whatever reason), he
will be in an uncomfortable situation when he observes the results of his experiments. If, on the other hand, he is lucky enough to make the mistake of varying two or three things together, no observable events will occur to upset the interpretation he wants to make.

In the present study of this problem with children transitional between concrete and formal thinking, the situation seemed very much as described above. Quite often a child who had made a meticulously correct investigation of the variables, methodologically speaking, but been unable to "tolerate" the idea that the size of the weight or height of the dropping point had no effect, would, at the end of the experiment, suggest that the best way to show someone about it would be to vary all three factors at once, so that the effect of each of them could be seen. In such cases a strong impression was created that the impossibility of a "no effect" relationship (in these children's minds) had led them firstly to report differences in speed which were not there and secondly, because they found this very uncomfortable, to abandon the methods of investigation which they felt were correct.

It is interesting to ask why or how these children's methods of investigation could have advanced beyond the mental structures available to them in interpreting results. It is possible that principles such as "changing only one thing at a time" are learned by rote, with little or no understanding of the need for them. On the whole this seems unlikely and it is more interesting to entertain the idea that some of the relationships defined by the sixteen binary operations will be easier to master than others. This may mean that children can discover the easier relation-
ships at an earlier age than they discover the others, and that, when confronted with more difficult relationships, they distort facts to force them into the easier interpretive mould. Inhelder and Piaget (1958) themselves suggest that the relationship of equivalence is easier to understand than that of "no relationship" or complete affirmation. In fact the achievement of an understanding of the second marks the advance from Stage IIIA to Stage IIIB in their description of performances on the Pendulum Problem. What they do not find, apparently, is any tendency for the methods of investigation employed to be in advance of the formal logic; rather the reverse.

The difficulty outlined above does not necessitate very extensive modifications to the method of allocation to stages outlined by Inhelder and Piaget (1958). As will be clear from the discussion, the only real discrepancy is in the significance attributed to inaccurate observation of events and the stage affected most is Stage IIIA. The next section presents a detailed description of the criteria used in the present study to locate children at a stage of operational development, on the Pendulum Problem.

6.3 Details of the Criteria Used to Locate Subjects in Stages in the Present Study.

Having available a taperecording of the complete interview with each child about the pendulum problem, and a record of each combination of weight, string length, etc., tried out in the course of his experiment, it was possible to develop a detailed set of criteria for the attainment of different stages. After studying a number of the protocols, it was decided
to specify several aspects of performance on which every child tested could be rated as successful, successful sometimes (or with difficulty), or unsuccessful (denoted by +, 0, - respectively). These aspects of performance were divided into those concerned with the experimental method used and those concerned with the content of the conclusions reached. It is recognised that such an approach automatically excludes idiosyncratic aspects of performance which may, very often, be particularly informative. The approach was adopted, however, in the interests of reliability of the allocation to stages.

Because the ratings of performance on the Pendulum Problem were made by the author, who also conducted a majority of the interviews; and because they were made using full details of every interview (in the form of a tape recording plus a record sheet); an "objective" check on their reliability, employing an independent rater, was not attempted. Had the records of performance been more concise, and had the experimenter not been involved in the ratings at all, an inter-rater reliability assessment would have been appropriate. The only indication of the reliability of ratings made is given by a re-rating of all subjects, by the same experimenter, after an interval of six months. In this re-rating, the original tape-recordings were not heard again, but only used in the form of annotations which had been made, on the record sheet, at the time of making the original ratings. The stage allocation of only 15 (3, 14 year-olds; 4, 13 year-olds; 3, 12 year-olds; 3, 11 year-olds; 2, 10 year-olds) out of the 236 subjects (6.3%) was changed, because the re-rating differed from the original rating. In no case was the changed rating more than two categories away from the
original. While this indication of reliability must be viewed with extreme caution, because of the circumstances described above, it is possible to be reasonably confident that the aspects of performance used to make the ratings were sufficiently objective to be assessed without serious error. These aspects are described in detail below. As mentioned above, they are divided into those concerned with the method of investigation used by the subject, and those concerned with the content of the conclusions he reached; the sections will be described in that order.

6.3.1 Rated aspects of performance on the Pendulum Problem concerning the method of experimentation.

(i) Some attempt to hold other variables constant.

+ A subject's performance was rated + if his work showed overall awareness that it was important to control the effects of other variables when looking at the effects of one (or two if he thought he could) in particular.

0 A rating of 0 indicated such awareness only rarely.

- A rating of - indicated that the child paid no attention to other variables at all when considering the one (or ones) in which he was interested.
(ii) **Varies only one**

to test an effect.

+ A subject's performance was rated + if, in his experimentation, he kept all except one of the variables constant in consecutive "tests".

0 A rating of 0 indicated that the above procedure was used sometimes and not other times, so that it was not really seen as necessary.

- A rating of - indicated that the above procedure was almost never followed, changing two or three things at a time being standard.

(iii) **Varies correctly**

to make the tests desired.

+ A subject's performance was rated + if the variable (or variables) he changed from one occasion to the next were those he had stated a wish to investigate.

0 A rating of 0 indicated some confusion about which one(s) to change.

- A rating of - indicated such behaviour as changing the weights to find out about differences between the lengths of string.

(iv) **Infers about the correct variable(s)**

after testing.

+ A subject's performance was rated + if, subsequent to a manipulation of the variable(s) under consideration, the conclusion stated pertained to the correct variable(s).

(Note this does not mean that the conclusion was correct).
A rating of 0 indicated some difficulty in keeping the conclusions appropriate to the tests made (e.g., changing the strings using the same weight and concluding that the weight has no effect; or changing the length of string and then adding to the conclusion "the strings are different because the long one has to go out further").

A rating of - indicated that the conclusions drawn were seldom related to the tests made at all. This was rare amongst children of the ages tested.

A subject's performance was rated + if he was able to generalise findings to other situations and did not insist on testing them all out to make sure. For example, a subject finding no difference between a 20 gm, a 50 gm and a 100 gm weight on the small string may go no further in testing weights. He infers that the other weights will be the same as the three he has tried on that string and also that on the other strings all weights will go the same speed as each other. Such inference is often expressed by "Well it isn't the weight" or "Well changing the weights isn't any good".
Less frequently the same kind of inference is shown by a subject who concludes (wrongly) that the weights do change the speed. He nevertheless does not test them all out because he knows "how they will compare with the others".

A rating of 0 indicated that the subject could predict results in different situations from the ones he had tested, but showed a need to try them out, just to make sure. Such subjects were very hard to discourage from doing the "checking up".

A rating of - indicated that the subject was quite unwilling to make a statement about any set of conditions which he had not seen. He would be insistent that he could not know how the clock would go until he had tried it. If asked afterwards "Should you have known that?" he would probably say "No, it might have been different, you would have to try and see".

A subject's performance was rated + if his testing procedure was extremely efficient, that is he made the maximum number of inferences from the minimum number of
An example would be a subject who starts with the medium length string and the 100 gm weight. After comparing it when started at the red and the green lines he may say "Well the distance it goes doesn't affect it". Then he may change the weight to another and (starting it from anywhere) find "It isn't the weight either .... so it must be the string". One test of another string with any weight on it, started at any distance, then enables him to state the relationship of the length of the string to the frequency of oscillation. Such cases did occur but were extremely rare.

A rating of 0 indicated that the testing procedure was moderately efficient and that, at least in the demonstration outlined at the end, the subject realised that it was not necessary to show every possible combination of conditions.

A rating of - indicated no tendency to eliminate unnecessary tests, particularly not from the demonstration. The ideal demonstration would be seen as one in which every possible combination was shown so that the
person watching would really know everything thoroughly.

While the last two aspects of the method of experimentation are very similar to one another, the first is intended to give a rather qualitative evaluation of whether or not the subject was able to make inferences from one situation to another at all; the second, while dependent on the first (in the sense that a rating of + on the first is essential for a rating of + on the second to be possible), is more a quantitative indication of the degree of efficiency which he was able to introduce.

6.3.2 Rated aspects of performance on the Pendulum Problem concerning the content of conclusions reached.

Secondly, ratings were made of aspects of performance showing the content of the conclusions drawn from experimentation. Only two aspects of the content were sufficiently general to enable a rating of each subject's performance on them. The first of these was whether or not a consistent ordering of the effects of each variable was achieved (for example if the medium string was said to tick more slowly than the short string, the long string should be said to be slower still). The second aspect was whether or not the effect ascribed to each variable was the correct one. As with the aspects of the method employed, a rating of +, 0 or - was made (for each variable) on each of the two aspects of content. A rating of + indicated, on the first aspect, successful ordering and, on the second, successful correct statement, of the effects of the variables; a rating of 0
indicated the same types of success, but with a good deal of wavering or confusion in the process; a rating of - indicated failure firstly to order, and secondly to arrive at the correct effects of, the variables.

Finally, to complete description of the scoring of performance on the pendulum task, an indication of the most common patterns of ratings on the aspects just described will be given for each stage. Whereas it was not possible to treat these aspects of performance as "items" forming some scale, such as a Guttman Scale, it was found that there were very typical patterns of performance for each of the stages of thought. In fact, using the ratings described above it was possible to add several transitional categories to the stages described by Inhelder and Piaget, so that the initial classification of performances was into nine different categories. This compares with the large number of categories used to describe performance on a number of tasks, including the Pendulum Problem, by Lovell (1961). A general description of each of the nine categories, with the patterns of ratings typical of each, will now be given. Every obtained pattern of ratings is included in the account which follows. The patterns of ratings for individual subjects are in the tables of raw data in Appendix VII.

6.4 Patterns of Performance Ratings Defining Nine Developmental Stages on the Pendulum Problem.

The patterns of ratings are shown on the following three pages.
### Patterns of Performance Ratings Defining Nine Developmental Stages of the Pendulum Problem

<table>
<thead>
<tr>
<th>Stage and Description</th>
<th>Content</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ordering</td>
<td>Correct Effect</td>
</tr>
<tr>
<td>Stage IIB: Finds correct effect of each factor (possibly with difficulty over one exclusion) - see (a) and is sound on all aspects of method, may not infer from minimum testing, however must show some untested inference.</td>
<td>(a) * * + + + +</td>
<td>(b) * * + + + +</td>
</tr>
<tr>
<td>Stage IIB: Some failure in one aspect of performance - perhaps fails to include one factor, but method is good and includes some untested inference (a), perhaps some faults in method but succeeds in finding all the correct effects of factors (b), some untested inference is required unless (c) the performance is otherwise faultless, (b).</td>
<td>(a) * * + + + +</td>
<td>(b) * * + + + +</td>
</tr>
<tr>
<td>Stage IIB: Finds exclusion very difficult and method, although mostly sound, shows no untested inferences or (a) although failing completely to exclude, has excellent method of investigation,</td>
<td>(a) * * + + + +</td>
<td>(b) * * + + + +</td>
</tr>
<tr>
<td>Stage IIC (a) Fails to exclude weight and height (perhaps because of uncertain ordering) but method has few faults and may show some untested inference, or (b) excludes one of weight and height with difficulty and has some shaky aspect to method and no untested inference, or (c) excludes both weight and height with difficulty, but method is shaky throughout, although may show some untested inferences.</td>
<td>(a) * * + + + +</td>
<td>(b) * * + + + +</td>
</tr>
</tbody>
</table>
### Patterns of Performance Ratings Defining Nine Developmental Stages on the Pendulum Problem

<table>
<thead>
<tr>
<th>Stage and Description</th>
<th>Content</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ordering</strong></td>
<td>Correct Effect</td>
<td>Some attempt to hold constant</td>
</tr>
<tr>
<td>ST.</td>
<td>Wt.</td>
<td>Ht.</td>
</tr>
<tr>
<td>IIIA/IIIB</td>
<td>(a) Fails to exclude weight and height (perhaps because of uncertain ordering) and method has one shaky point, and no untested inferences, or (b) excludes only one of weight and height with difficulty and method is shaky throughout with no untested inferences, or (c) succeeds in finding correct effect of each factor, but only with difficulty for each one, and method also shows one shaky point, or (d) fails to exclude weight and height and finds effect of string only with difficulty, but method has no shaky aspects although there is no untested inference.</td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>+</td>
<td>or</td>
</tr>
<tr>
<td>(b)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(c)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(d)</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

IIIB/IIIA | (a) Fails exclude weight and height (perhaps because of uncertain ordering) and method is weak throughout, but shows at least 0 on “varies only one to make a test”, or (b) excludes only one of weight and height with difficulty and method is shaky showing either 0 or - on “varies only one to make a test”, or (c) although one of weight and height is excluded, method is completely unacceptable throughout. |
| (a) | + | or | or | + | or | + | or | 0 | 0 | - | - |
| (b) | + | + | + | or | or | + | or | 0 | 0 | - | - |
| (c) | + | + | + | or | or | + | or | 0 | 0 | - | - |
### PATTERNS OF PERFORMANCE RATINGS DEFINING NINE DEVELOPMENTAL STAGES ON THE PENDULUM PROBLEM

<table>
<thead>
<tr>
<th>STAGE AND DESCRIPTION</th>
<th>CONTENT</th>
<th>METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ordering</td>
<td>Correct Effect</td>
</tr>
<tr>
<td>IIIB (a) Fails to exclude both weight and height (perhaps because of uncertain ordering) and fails in method at least to vary only one to make a test and possibly to vary and/or infer correctly as well, or (b) fails to find correct effect of string, although is able to exclude one of weight and height and method shows failure on one of the two aspects &quot;varies only one...&quot; and &quot;varies and infers correctly&quot;.</td>
<td>(a) + or or 0 0 - - + - or or 0 0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(a) + + 0 0 - - + - or or 0 0</td>
<td>-</td>
</tr>
<tr>
<td>IIIBPA (a) Fails exclude both weight and height (possibly because of uncertain ordering) and also has difficulty finding correct role for string; method also fails either to vary only one, or to vary and infer correctly (b) fails to find correct role for all three variables (and may also fail to order one or more) but method, although shaky, shows some ability to vary only one and vary and infer correctly,</td>
<td>(a) + + 0 0 - - + - or or 0 0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(a) + or or 0 0 - - + - or or 0 0</td>
<td>-</td>
</tr>
<tr>
<td>IIAYIV (a) Fails exclude weight and height (perhaps because of uncertain ordering) and either fails or has difficulty finding correct effect of string - method also fails (or has 0) to vary only one and to vary and infer correctly, or (b) fails to find correct effect of any variable and may be unable to order the effects as well - method in addition is extremely weak.</td>
<td>(a) + + 0 0 - - + - or or 0 0</td>
<td>-</td>
</tr>
</tbody>
</table>
6.5 Summary of Stages on the Pendulum Problem.

The preceding description of stages is too detailed to give an adequate indication of the qualitative differences which exist between them. The important differences on which the classifications are based may be summarised as follows:

STAGE IIIB (Including IIIB and IIIB?A). The child has the necessary operations to understand an equivalence relationship (between the length of string and the frequency of oscillation) and the exclusion of a factor (either the size of the weight, or the height of the dropping point, or both) and also investigates these relationships in a rigorous manner (varying only one at a time, and mostly doing so appropriately). He is also able to generalise findings in one set of conditions to others which are appropriately similar.

STAGE IIIA (Including IIIA?B, IIIA and IIIA?IIB). The child has serious difficulty either with the understanding of the exclusion of a factor (the size of the weight, or the height of the dropping point, or both) or with the appropriate experimental manipulation of variables. Weakness (but not complete failure) in both aspects may also characterise this stage. The achievement of this stage is an ability to find the equivalence relationship between the length of the string and the frequency of oscillation, with some other indication that this achievement is no longer by the concrete operational method of establishing correspondences between ordered relationships. This indication comes either from the method of investigation used, or from the fact that the exclusion of a factor is understood. In either case, it appears that the child now understands that the equivalence relationship is not the only one possible and that
he needs to envisage or to investigate others as well. He has only the beginnings of the structure required for this, however, and is liable to find one or other aspect of the task too difficult.

STAGE IIB (Including IIIB?IIIA, IIB and IIB?A). The child at this stage understands that there are a number of variables involved in the situation. He is able to order the values of each and expects that the effects will correspond to this ordering. He may not be completely sure of these orderings of effects, particularly in the case of the size of the weight, and may decide, for instance, that the 100 gm weight is "just right" while those smaller and larger than it all go too slowly. It does not seem, however, that this failure to order the effects correctly is a failure which would class the child as preoperational. It seems more likely that he has an expected order for the effects of weight, and that when this is not found, he distorts his observations either in the direction of his expectation, or toward some other compromise solution such as there being a "right" weight for each string. It is not a failure to seriate as such, which would be implied by a classification of the performance as "preoperational".

The relationships discovered by children at this stage, and also the kind of difficulties they encounter, arise from the fact that they are capable only of ordering variables and their effects, and then looking for correspondences between them. Thus it is quite common, although not guaranteed, that the correct relationship between length of string and frequency will be described. It is not understood as a formal, logical relationship, however, but "observed" as an empirical correspondence. The
fact that there is a need to decide between a number of possible relationships is not known to the subject. He therefore does not carry out appropriate tests in a systematic manner and may not even know what the experiments he has performed can tell him. He simply wants to "look and see", hoping that a relationship will emerge. Typically he orders the weights by size (sometimes in a neat row) and when he finds no corresponding order of effects, is embarrassed because he must have "done it wrongly". He apparently has no way, either of being sure that he is going about things in the correct way, or of understanding how a set of ordered weights can correspond to a set of equal frequencies of oscillation. Sometimes a child at this stage will say that the weight does not seem to make any difference, but the impression is always that he is sure that some sort of an effect is there and that there is something wrong with his experimentation. He does not consider "no effect" as one of the possible relationships and may, as a result, turn to a method of investigation which ensures that the fact that a variable (such as the size of the weight) has "no effect" is hidden by the effect due to changing the length of string. This seems to be why, to a child at this stage, the best demonstration of how the pendulum works is to show first the short string with the little weight dropped from the red line, and then the long string with the big weight dropped from the green line, so that "you can really see how they all make a difference".

**STAGE IIA.** In the present investigation, no clear cases of IIA performance were observed, and only four children were classified as IIA?B. These were children who appeared not to see the situation as involving a number of variables and their interrelationships at all. They achieved almost no
success in discovering the effects of variables, or in their methods of investigation, and those successes which did occur seemed to be accidental and very easily destroyed by later events. They did not, however, confuse the effects of their own actions or wishes with the effects of physical variables and were therefore not considered to be at a preoperational level. The fact that they did not have difficulty in ordering the effects of weight, in particular, prevented their being classified definitely as in Stage IIA. However, there were so few children concerned that a clear picture of the typical behaviour and thought structures of this stage did not emerge.