Chapter 4

DATA ANALYSIS: CODING AND FREQUENCIES

Introduction

In the previous chapter four research themes were identified that provided the major areas of focus for this study. The themes were developed in turn into a set of research questions. This chapter addresses the first of these themes.

Theme 1
What natural language do children use to describe patterns associated with numbers in Years 5, 6, 7 and 8?

Question 1.1
Can children's natural language descriptions of number patterns be classified into a discrete set of categories?

Question 1.2
Can children's symbolic language descriptions of number patterns be classified into a discrete set of categories?

Question 1.3
Are the response categories stable across stimulus items that vary in complexity and context?

The discussion below is conducted in four sections. The first section reports on a number of issues with regard to the coding process, with particular emphasis on the issue of reliability. The second section presents the response categories by providing the descriptor for each category together with an example of a response for that category. The discussion in the third section reports on the validity of the categorisation. Additionally, the findings with regard to Research Questions 1.1, 1.2 and a partial discussion of 1.3 are provided. The final section of the chapter reports on the frequencies for each category and sub-category. The frequency data assists with the answer to research question 1.3 and provides the data necessary for the analysis of the relationships between the components in Chapter 6.

THE CODING PROCESS

This section consists of four parts. The first part provides a rationale for coding the responses, while the second part identifies the procedures for establishing the coding categories. The third part of the section resolves some decisions on specific issues of this coding analysis. Finally, the issue of reliability is presented in some detail, and the procedures that were followed to ensure inter and intracoder reliability are described.

Overview of the Coding Process

The survey produced 1435 responses to four stimulus items that each consisted of four components. The 22960 separate pieces of data required considerable organisation to facilitate analysis. The open-ended nature of the Components B and D did not lend itself to simple scaling procedures. To provide access to the data, coding techniques were borrowed from the technique of content analysis. Kerlinger (1986) defined coding as:

> ... an objective and quantitative method for assigning types of verbal and other data to categories.
>
> (p. 382)

Such a procedure is also known as data reduction (Cohen & Manion 1989, p. 116), and is to allow investigation and analysis of the data. It does this by allowing the data to be quantified and entered into a computer for further processing. Initially the level of measurement is nominal; however, ranking of the responses is facilitated by the analysis of the nominal data using the SOLO Taxonomy. This is discussed the next chapter.

Of considerable importance is the method of generating the coding categories. The strategies used as a basis for developing the coding classification are discussed in the next part of this section.

The Basis of Coding Classification

There are two fundamental approaches to the initial task of classifying children's responses to the stimulus items used in this study. The first is to create hypothesised categories on the basis of the theoretical framework. Cohen and Manion (1989, p. 117) called this “pre-coding” and believed it is best suited to closed questions. It was not the purpose of the survey to categorise the population into a right/wrong dichotomy but rather to look for the range of responses that children provide and hence a simple coding system based on this dichotomy would be inappropriate. The pre-coding approach assumes that the researcher has a deep understanding of the field under investigation and that the theoretical framework has a predictive ability. As noted in the research design outlined in Chapter 3, this study was explorative in nature and as such is an attempt to describe hitherto unreported responses. The theoretical framework is essentially a taxonomy for describing observed outcomes and is not detailed enough to be able to predict responses. One of its purposes is to provide a framework for ordering the responses once they are identified. To the extent that this research project is theory testing it would seem unsuitable to risk "forcing" the data into the theoretical framework. This criticism would apply equally to any predetermined set of categories. That is, they have the major disadvantage of trying to predict the range of responses prior to coding with the possibility of imposing the researchers classification system on the data. An alternative to pre-coding is to develop the classification system from the data.

Post-coding involves developing a coding frame after the data has been collected. If used flexibly it has the advantage of not imposing a structure on the data. The system of coding should be allowed to develop in such a way as to adjust for new perceptions and responses as they are discovered. This often involves the recoding of earlier work. Post-coding is a suitable strategy for open-ended questions such as Components B and D for each stimulus item of this survey. While this approach has some merit in terms of the validity of the study, it would appear to divorce the study from the theoretical framework. Miles and Huberman (1984) provided a structure to assist the resolution of this dilemma. While the structure was developed for the analysis of field notes it is of use in clarifying the analysis of the student responses of this survey.

Miles and Huberman (1984) argued that there are three levels of coding, namely:

1) descriptive codes
2) interpretive codes, and
3) explanatory codes.

Descriptive codes merely assign "the attribution of a class of phenomena to a segment of text" (p. 56) and entail no interpretation. They deal only with the response at face value. As researchers become more familiar with the responses they can begin to interpret some of the strategies that underlie a response. It is at this level that the theoretical framework begins to provide some insight. The third - explanatory - level attempts to explain the data in terms of relationships. "They typically get used later in the course of data collection, as patterns come clear" (p. 56). While it is clear that the three levels provide different depth to the analysis they also occur at different times with the simpler descriptive analysis occuring first and the deeper analysis occuring later.

This part has addressed some general matters of coding qualitative data. The next part reports the decisions made with regard to some more specific issues of this study.

Issues of Coding

Certain specific issues with relation to this study needed to be resolved prior to the commencement of the coding activity. These were the unit of analysis, the number of categories, and some procedural details on how the categories were developed.

In choosing the unit of analysis the research questions provided guidance. The research questions require a search for differences between the stimulus items and hence each of the four pattern stimulus items needed to be coded separately. In addition the research questions are looking for relationships between three of the four components (B, C, and D) for each stimulus item so each of these required a sepatate code. As a result the minimum unit of analysis had to be each component of each stimulus item. As these components are also closely related to the research questions they are a suitable unit of analysis. This resulted in 16 pieces of data in addition to school, class, and gender, that is, a total of 19 pieces of data were recorded for each student.

Two fundamental questions for the development of coding categories are: (i) how many categories should be used? and, (ii) should the categories be mutually exclusive? Firstly, while it would seem advantageous to have a large number of coding categories so as to retain as much detail in the data as possible, there are some countervailing pressures to limit the number of categories. After all, the major purpose of coding is for data reduction to facilitate an overview of 1435 sets of responses. In addition, each category needs to have a significant frequency to enable generalisations to be made. Too fine a categorisation system would result in some categories having very low frequencies. Kerlinger (1986) warned that counting should only take place when a category is representative. Further, if category membership is too small, generalisation from statistical analysis is unwarranted. Secondly, the number of categories needs to be manageable but still reflect the diversity of children's responses. In creating the categories it was decided to ensure that each response was only allocated to one category. By pursuing the mutual exclusivity of each category the range of statistical procedures that can be used is increased since the groups for each variable are independent.

Rather than provide algorithmic guidance for the number of categories, Miles and Huberman (1984) referred to the fluidity of the situation when they talked of continually revising codes (p. 60). They pointed out that some codes decay, while others flourish. Other codes emerge during the coding process. Such fluidity would be expected to be more evident in the post-coding procedures adopted in this study than if the coding frames had been established prior to the coding process. Hence, it was decided not to predetermine the number of categories but rather allow the data to specify the number of categories during an exploratory phase of coding.

This policy had implications for the structure of the categories. They were set up as major categories with the ability to divide the major category into subcategories. The major categories were identified with a number that was a multiple of ten while the subcategories were identified with a number in the units column of the multiple of ten. Such a system allowed for the creation of new subcategories as required and also allowed for the easy amalgamation of categories that proved to be of little significance.

The procedure for developing the codes was exploratory in nature. It involved selecting approximately 100 responses and allocating them to like groups. The groups were then recorded and described and a number allocated to them. This enabled each response to be coded by recording that number. The 100 responses were then recoded and the results compared with the first attempt. It was found that there were great discrepancies between the two coding attempts. This resulted in further clarification of the categories and another attempt at coding. This procedure was repeated approximately ten times before satisfactory intracoder reliability was achieved. When responses from other schools were superimposed on the coding system some further amendments were required by way of additional categories. At this point the reliability of the coding system was checked by introducing a second person. The issues of inter and intracoder reliability are discussed fully in the next part of this section.

Reliability

Of particular interest here are intercoder reliability and intracoder reliability (Miles & Huberman 1984). Intercoder reliability refers to the ability of two or more researchers to achieve the same or similar coding results from a set of responses. One measure of intercoder reliability is to express the number of agreements as a percentage of the number of coded items, i.e.,

$$\text{Intercoder reliability} = \frac{\text{Number of agreements}}{\text{Total number of coded items}}$$

Miles and Huberman (1984, p. 63) reported that, using such indices, agreement measures as high as 90% can be achieved when content analysis techniques are being used. Clearly such values would depend on the type of material being coded. Kerlinger (1986, p. 481) reported agreement coefficients of 70% and 80% as being achieved in judging the creativity of student essays. Under conditions of this study, where the unit of analysis is considerably smaller than an essay and the criteria for each category is relatively explicit it is reasonable to expect that intercoder reliability coefficients in the high range can be achieved. As a consequence an intercoder reliability coefficient of 90% was targeted. To achieve this the researcher trained two assistants in the coding procedures and in the definitions of each category. The first assistant was involved in the early coding procedure to assist in the stabilisation of the coding process discussed earlier and in the clarification of the coding categories. He was asked to code a random sample of 50 responses. These were checked against the researcher's codings and disagreements discussed. Differences at this stage arose from three sources. The first was a lack of clarity in the category definitions. The definitions were amended to enable agreement. This was done early in the coding process and involved adjustments to codings as a result of the changed definitions. The second source of difference was a result of error on behalf of the researcher. While the number of such errors was few (less than 1%) they were nevertheless seen as a potential threat to reliability, and a decision was taken to slow the rate of coding and to employ an assistant to enable time constraints to be met. In addition a measure of intracoder reliability was taken at regular intervals to monitor the accuracy of the researcher's coding. The third source of difference involved varying perceptions of the child's response between the assistant and the researcher. In cases where the differences could not be resolved because of a different perspective between the researcher and the assistant the codes of the researcher prevailed. This remained a threat to reliability. This phase of the reliability check was in the developmental phase. At the end of this process an intercoder reliability coefficient between the researcher and the first assistant of 98% was achieved. This was considered to be entirely satisfactory when compared to the criterion of 90% set previously. The second assistant provided a further check and was used in the second half of the coding procedure.

The second assistant coded approximately half of the 1435 responses. Initially an intercoder reliability measure of 90% between the researcher and the second assistant was to be achieved during the training period. Two strategies were used to ensure high reliability in the assistant's coding. The first was that she consulted with the researcher about any responses she was unsure of, hence avoiding the possibility of varying perceptions. This strategy was used less and less frequently as she developed confidence and familiarity with the procedures and responses. To continue the monitoring of the reliability of her coding the researcher recoded approximately 10% of the responses on three separate occasions and on each occasion an intercoder reliability coefficient was calculated. This coefficient remained above 98% and within the range of values considered acceptable.

As mentioned earlier, a measure of intracoder agreement was taken to monitor the stability of the researcher's coding activity. There were two potential causes of instability to be checked for. The first is error due to carelessness, that was often of a clerical nature, which was referred to above. While steps were taken to eliminate this source of instability it was felt necessary to continue to monitor its extent.

The second source of intracoder instability was the tendency for subtle changes in meaning, in the mind of the coder, of both the codes and the children's responses, to develop over time. This is what Miles and Huberman (1984) referred to as "retrospective hindsight" (p. 57). This happens as the coder became more familiar with the data and began to understand the processes respondents use in formulating their response. It seems plausible that such a phenomenon would be exacerbated by coding taking place over a long period of time. The great length of time spent developing stable codes at the beginning of the analysis phase was in part a recognition of this threat to reliability; however, two other strategies were used to address this problem. The first was the employment of a research assistant (described above) to help speed up the coding process. She also provided the opportunity for discussion of the responses and hence a shared meaning to be developed. The second was the calculating of an intracoder reliability co-efficient using the same formula as for the intercoder reliability coefficient. This was done on three occasions by randomly selecting 10% of the researcher's coded responses, recoding them and performing the necessary calculations. These calculations showed that differences occurred in fewer than 1% of cases. It appeared that the clerical errors referred to earlier had been largely eliminated. The errors that were found tended to be between subgroups of the larger categories and hence were not of importance when smaller groups were subsequently amalgamated.

This section has described the coding process in considerable detail. The next section reports the category definitions for each component.

RESPONSE CATEGORIES

A set of categories were developed for each of Components A, B, C and D and applied to the responses for each of the four stimulus items. As a result, four sets of categories are reported in this section; one for each of Components A, B, C and D.

Component A

This component was designed to determine if the children had understood the question, that is: Had they interpreted the pattern in a manner similar to the researcher? The question asked children to provide the next value of the dependent variable in the sequence. It did not necessarily require children to generate a function; however, it was clear that some children did use this strategy. Other strategies, such as, counting matches or wheels, drawing the next member of the sequence, and counting the number of items required by the question were also evident. It was presumed that other children performed some mental calculation since there was no evidence of other activity on their response sheet.

In all, there were three response categories for Component A, namely, no response, incorrect and correct. The abbreviated codes with their meaning and descriptors are reported in Table 4.1. Figure 4.1 provides example of children's responses in two of the categories for each stimulus item.

Table 4.1

Component A Response Categories

Code	Code Meaning	Descriptors for Component A
NA	No attempt	The children made no attempt to answer the question
IC	Incorrect	An attempt to answer the question had been made; however, the response did not demonstrate any understanding of the question. Responses that indicated arithmetic errors were not included in this category.
C	Correct	This group of responses were correct or indicated a clear understanding of the question but had made an arithmetic error or had counted inaccurately. An example of such an error is 9+4=12.

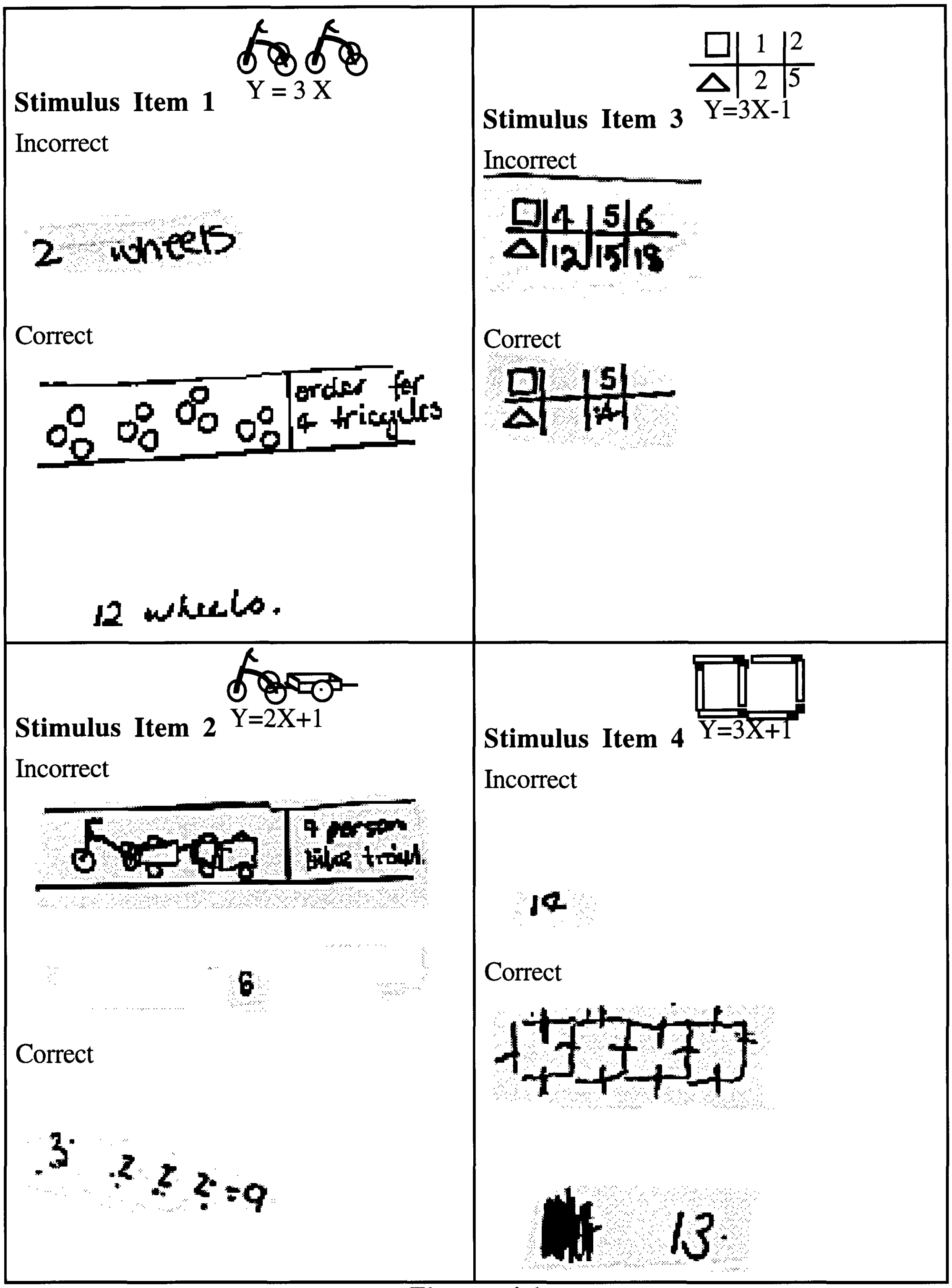


Figure 4.1

Sample of Component A Responses

Component B

Component B was the major focus of the study as it required children to write a natural language sentence to describe a pattern. The focus of the question was on the ability of each child to express generality. Since this was an open-ended question there was a greater diversity of responses than in Component A. The responses were eventually reduced to five major categories, namely, no attempt, inappropriate, one example, successive and function, with three of the categories having some subcategories. The categories and subcategories are described in Tables 4.2 to 4.5. Figures 4.2 to 4.5 present a sample of children's responses for each stimulus item in each response category.

Table 4.2

Component B Response Descriptors for the No Attempt (**NA**) and Inappropriate (**IA**) Categories

Code	Code Meaning	Descriptors
NA	No Attempt	No attempt was made to answer the question
IA	Inappropriate	This group of responses indicated a failure to understand the question. The responses sometimes included a vague statement that did not indicate a rule or procedure. The student may have asked another question such as "how many wheels do you need?"

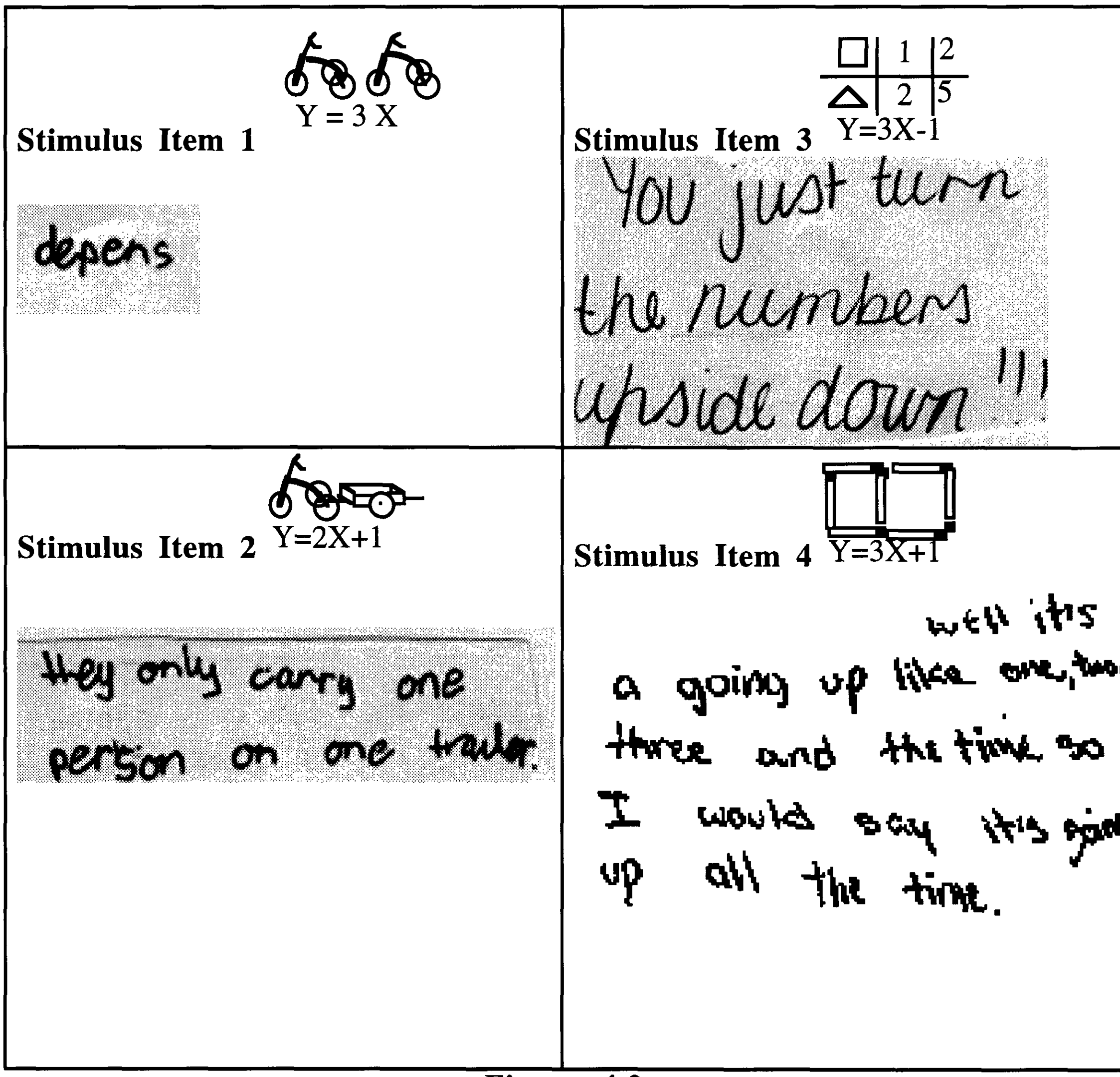


Figure 4.2

Sample of Component B Responses for Inappropriate (**IA**) Category

Table 4.3

Component B Response Descriptors for the One Example (1EG) Category

Code	Code Meaning	Descriptors
1EG	One Example	This class of response gave the value for a specific example rather than a general description, e.g., "four tricycles need twelve wheels" or "six squares need 19 matches".
	Subgroup 1	Responses in this group had incorporated an algorithm for calculating larger values of the independent variable but had used only one piece of data, e.g., "for eight squares you would need 32 matches."
	Subgroup 2	These responses successfully used the value of the dependent variable, e.g., "for eight squares you would need 25 matches."
	Subgroup 3	This was a very small group that reflected a transition phase to using a successive description, e.g., "If you had one box you would need 3 matches and it would equal two boxes."

Stimulus Item 1 Y = 3 X

Subgroup 1

three x eleven = thirtythree.

Subgroup 2

No example

Subgroup 3

No example

□	1	2
△	2	5

Stimulus Item 3 Y=3X-1

Subgroup 1

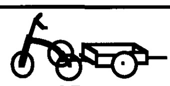
4
9

Subgroup 2

You would put 29 under 10

Subgroup 3

PLUS 3 TO 5 TO GET 8



Stimulus Item 2 Y=2X+1

Subgroup 1

$$\begin{array}{r} 2\times \\ 10 \\ \hline 20 \end{array}$$

Subgroup 2

$$\begin{array}{r} 8\times \\ 2 \\ \hline 16 \end{array} + 3 = 19$$

Subgroup 3

for each trailer, you count 2 wheels, so you add 3 wheels on for 2 people

Stimulus Item 4 Y=3X+1

Subgroup 1

to make 5 squares you need 4 ~~squares~~ matches 5x4=20

Subgroup 2

10 squares needs 31 matches

Subgroup 3

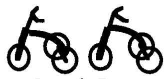
3 onto 4 boxes .. gives 5 boxes

Figure 4.3
Sample of Component B Responses for One Example (**1EG**) Category

Table 4.4

Component B Response Descriptors for the Successive (SUCC) Category

Code	Code Meaning	Descriptors
SUCC	Successive Description	This group of responses only made use of the dependent variable. The essence of the response was to describe how to calculate a term in the series given the previous term. They did not relate the independent variable with the dependent variable.
	Subgroup 1	This group of responses were restricted to the operation that allowed the sequence to grow, ignoring the starting point, e.g., "You add three every time."
	Subgroup 2	This group of response indicated a starting point to the sequence in addition to describing the change rule, e.g., "Starting from 4 you keep on adding three."
	Subgroup 3	This group of responses reflected a transition stage linking the successive group with the group describing relationships between the dependent and independent variables. The responses combined features of both categories in their description, e.g., "Add the next odd number to the top" or "You need to get the number of squares timesed by three and then add increasing counting numbers."



Y = 3 X

Stimulus Item 1

Subgroup 1

For each triangle first add three wheels

Subgroup 2

No example

Subgroup 3

No example

□	1	2
△	2	5

Y=3X-1

Stimulus Item 3

Subgroup 1

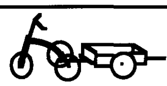
just add 3 to the top number bottom

Subgroup 2

the number goes up every 1 on the top and every 3 on the bottom.

Subgroup 3

the Computer doubles it then adds one more than it did before



Y=2X+1

Stimulus Item 2

Subgroup 1

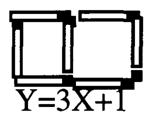
it is counting by twos

Subgroup 2

START AT 3 AND KEEP ADDING 2

Subgroup 3

the bike - 3 weeles + 2 weeles on each bike train



Y=3X+1

Stimulus Item 4

Subgroup 1

ADD 3 EVERY TIME

Subgroup 2

After one square is made you only need three matches to follow the pattern

Subgroup 3

THE FIRST SQUARE + 3 MATCHES MAKES 2 SQARES FOR HOWEVER MANY SQUARES

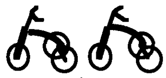
Figure 4.4

Sample of Component B Responses for Successive Description (**SUCC**) Category

Table 4.5

Component B Response Descriptors for the Function (**FUNC**) Category

Code	Code Meaning	Descriptors
FUNC	Independent/ Dependent Variable	This group of responses described a relation between the independent variable and the dependent variable. In some responses the dependent variable was implied rather than made explicit, e.g., "The number of squares times three plus one."
	Subgroup 1	This group of responses described relations that had only one operation. For the tricycle item in Stimulus Item 1 this was the best response possible. However, the response occurred for other items as a result of generalising from one element of data and ignoring the rest of the data, e.g., (For the chain of squares in question 4) "Number of matches is four times the length of the chain."
	Subgroup 2	This group of responses accurately generalised all the data provided, e.g., "The number of matches is the number of boxes times three plus 1."



Stimulus Item 1 $Y = 3X$

Subgroup 1

times the number of tricycles he orders by 3

Subgroup 2

No examples

□	1	2
△	2	5

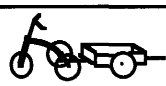
Stimulus Item 3 $Y=3X-1$

Subgroup 1

well you time that by 3 and that would give you the answer of the bottom one.

Subgroup 2

times the top number by 3 and minus one



Stimulus Item 2 $Y=2X+1$

Subgroup 1

Because you double it

Subgroup 2

You times the number of people it will carry by 2 then add

Stimulus Item 4 $Y=3X+1$

Subgroup 1

As many squares there are times 4.

Subgroup 2

multiply the number of boxes by three then add one.

Figure 4.5

Sample of Component B Responses for Function (**FUNC**) Category

Component C

This component required the children to apply the pattern to an uncountable value of the independent variable. The major purpose of this form of question in this study was to assist with the clarification of the meaning of the response to Component B. It seemed probable that only the students who provided higher-order descriptions in Component B would be able to complete this question successfully .

The responses were categorised into three groups similar to the categories in component A. The categories and their descriptors are reported in Table 4.6. Examples of children's responses are presented in Figure 4.6.

Table 4.6

Component C Response Categories

Code	Code Meaning	Descriptors
NA	No Attempt	No attempt was made to answer the question
IC	Incorrect	Answers that were incorrect due to an inappropriate strategy being used. Some students attempted to use proportions incorrectly, e.g., they would count the solution for 5 squares and multiply it by 10 for 50 squares. Other students described an iterative process rather than provide a numerical answer, e.g., "You would add three 87 times." Correct application of an incorrect rule stated in component B was not included if the rule had been consistently applied to components B, C and D. Incorrect answers that were due to simple arithmetic errors were not included.
C	Correct	These responses were correct, or the correct application of an incorrect rule. Incorrect answers due to arithmetic errors were also included.

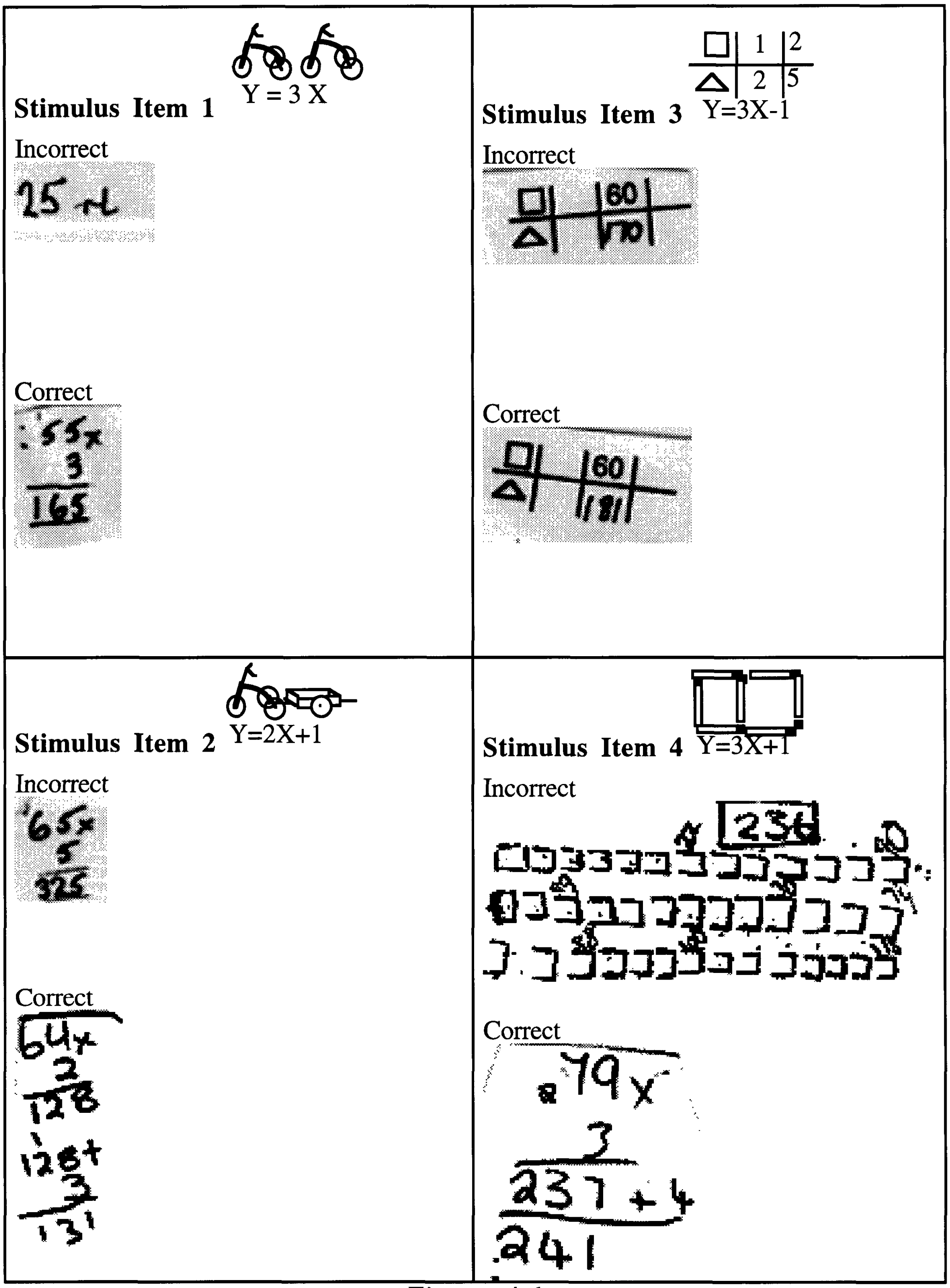


Figure 4.6

Sample of Component C Responses

Component D

This component was more problematic than the previous three due to the very disparate backgrounds of the children involved in the study. The question was designed to determine if children could express their natural language descriptions given as answers to Component B in a symbolic notation of mathematics. Year 7 and 8 children at the time of the survey had all been introduced to algebra in their mathematics classes. To encourage the best possible responses they were asked to write their description in Component B in **algebra symbols**. However, the children of Years 5 and 6 were not familiar with the word algebra and hence were asked to write their description in Component B using **maths symbols**.

The five major categories for Component D are presented in Table 4.7 to 4.10. A sample of children's responses is provided in Figures 4.7 to 4.10.

Table 4.7

Component D Response Descriptors for the No Attempt (**NA**) and Operation Symbols (**OS**) Categories

Code	Code Meaning	Descriptor
NA	No attempt	No attempt was made to answer the question
OS	Operation Symbols	Attempts at symbol use were restricted to those commonly used in arithmetic
	Subgroup 1	This group of responses used operation symbols in place of words such as plus, times, equals. All other words remained, e.g., "You need 3 × number of tricycles."
	Subgroup 2	This group of responses closely resembled the single example given in Component B, e.g., $1+(14\times 3)=43$.
	Subgroup 3	Use of a place holder which seemed to be a transition stage to the concept of a pronumeral. It was distinguished from an ideogram by the use of the same symbol to stand for both the dependent and independent variable, e.g., $\_=\_\times 3+1$, or $\square=\square\times 3+1$.

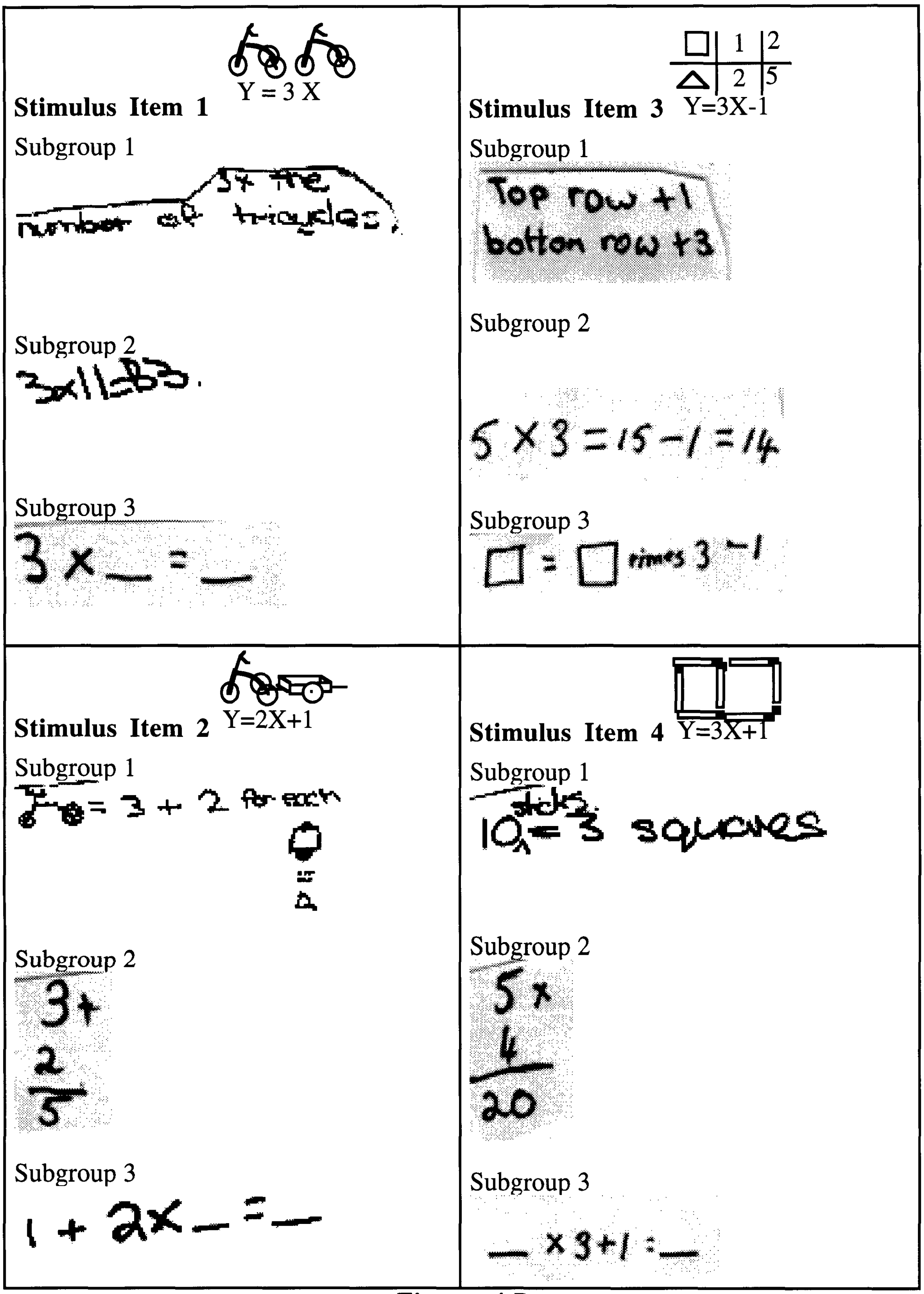


Figure 4.7

Sample of Component D Responses for Operations Symbols (**OS**) Category

Table 4.8

Component D Response Descriptors for the Arbitrary Use of Letters (AL) Category

Code	Code Meaning	Descriptor
AL	Arbitrary or inappropriate use of letters	These responses indicated a knowledge that letters were used in mathematics but reflected little knowledge about their correct use in algebra.
	Subgroup 1	This group of responses used letters for letters sake. Their use in no way related to the description in Component B, e.g., $abc =$
	Subgroup 2	These responses used letters as words, e.g., $1S + 3M = 2S$
	Subgroup 3	These responses replaced every number replaced by a letter, e.g., $a \times b = c$
	Subgroup 4	These responses allocated a specific value to the pronumeral, e.g., $x \times 3 = 15$ or $y = 43$

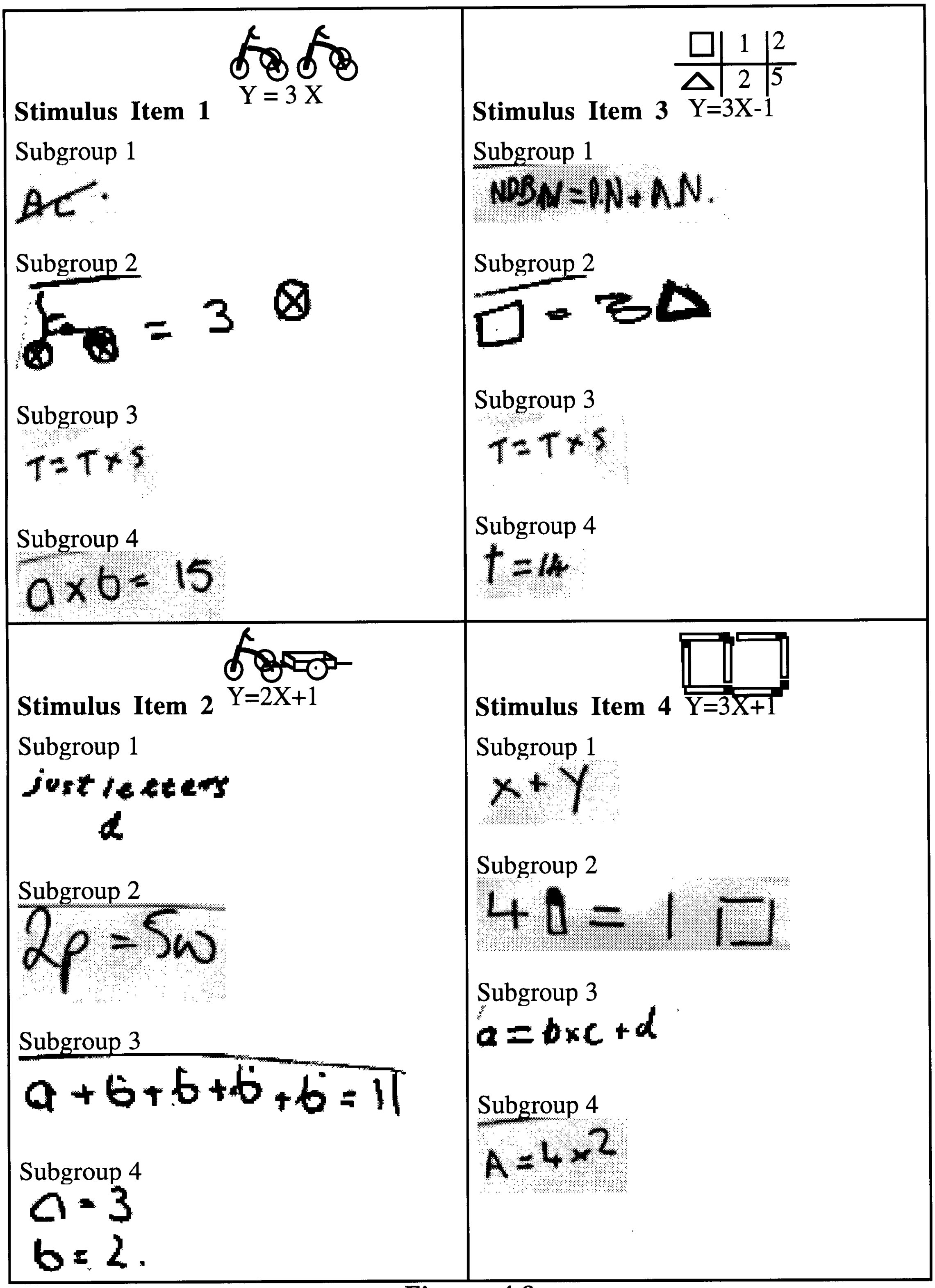


Figure 4.8

Sample of Component D Responses for Arbitrary Use of Letters (**AL**) Category

Table 4.9

Component D Response Descriptors for the Iterative Description (**REPT**) Category

Code	Code Meaning	Descriptor
REPT	Iterative description	Symbols reflect the repeated adding of a constant term
	Subgroup 1	This response group attempts to symbolise the successive description of Component B, e.g., $a+3+3+3$....., or $x+y+y+y$+...., or $4+3^{\bullet}$ (i.e., 4+3 repeater).
	Subgroup 2	This group of responses is very close to satisfactory algebra but is included here because it reflects successive thinking and, in doing so, fails to correctly reflect the value of the independent variable, e.g., $4+3x$ (the intended meaning is "start with four for the first square and add three for the rest." The correct response is $4+3(x-1)$.

$Y = 3X$

Stimulus Item 1

Subgroup 1

+3 for every tryc

Subgroup 2

No Example

□	1	2
△	2	5

$Y=3X-1$

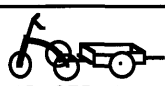
Stimulus Item 3

Subgroup 1

11+3+3+3 AND KEEP ON GOING

Subgroup 2

2n+1=△
2n+2=△
2n+3=△
etc



$Y=2X+1$

Stimulus Item 2

Subgroup 1

3+2+2+2+—

Subgroup 2

Z×2+3=x

$Y=3X+1$

Stimulus Item 4

Subgroup 1

4+3c =

Subgroup 2

4+3x

Figure 4.9

Sample of Component D Responses for Iterative Description (**REPT**) Category

Table 4.10

Component D Response Descriptors for the Algebra (**ALG**) Category

Code	**Code Meaning**	**Descriptor**
ALG	Algebra	Successful use of algebraic ideas
	Subgroup 1	This group of responses used ideograms as pronumerals, e.g., $? \times 3 + 1$
	Subgroup 2	This group of responses consisted of accurate descriptions of the wrong rule, e.g., $y = 4x$
	Subgroup 3	This group of responses consisted of attempts that involved poor syntax such as failure to close parentheses, e.g., $y = 4 + 3(x - 1$
	Subgroup 4	This group of responses reflected accurate use of standard algebra notation, e.g., $x = 4 + 3(n - 1)$ Cases where the dependent variable was not made explicit were also included, e.g., $3x + 1$.

Stimulus Item 1 Y = 3 X

Subgroup 1

$\otimes =$ ⊗ $\times 3$

Subgroup 2

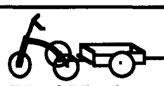
No example

Subgroup 3

No example

Subgroup 4

w = + x 3



Stimulus Item 2 Y=2X+1

Subgroup 1

? x 2 + 1 = □

Subgroup 2

w = t x 2

Subgroup 3

3 + 2 x X - 1 =

Subgroup 4

w = 3 + [(p - 1) x 2]

□	1	2
△	2	5

Stimulus Item 3 Y=3X-1

Subgroup 1

△ = □ x 3 - 1

Subgroup 2

Y = 3 X

Subgroup 3

No example

Subgroup 4

x x 2 + x - 1

Stimulus Item 4 Y=3X+1

Subgroup 1

□ x 3 + 1 = ▯

Subgroup 2

(k x 4) - 1

Subgroup 3

y = 4 + 3 (x + 1

Subgroup 4

~~x~~ x 3 + 1 = x

Figure 4.10
Sample of Component B Responses for Algebra (**ALG**) Category

Discussion

The processes that produced the above coding categories were subject to measures of inter and intracoder reliability. It is appropriate that the validity of the coding categories is also considered. It was suggested in Chapter 3 that one method of reviewing the validity of the categories was to compare the results of this survey with the findings of other researchers using similar stimulus items. Due to the qualitative nature of the categories described for Components B and D, and the variable nature of the data from other researchers, it is not possible to apply a simple statistical test to measure the extent of agreement. However, by comparing the reported responses from other studies with the categories identified in this study, some indication of the validity of the analysis is provided.

The comparison provides two indications of validity. The first is the extent to which the other researchers' stimulus items generate the same types of responses. The second asks: Which descriptions of responses from other researchers correspond to the descriptors provided in this study? However, this method of testing validity is restricted by the differences between the studies. No other researcher asked the same range of questions, nor did they ask them under the same conditions. Hence, some variation in response patterns was anticipated. The expectation was that this study would provide the widest range of responses due to four aspects of the design:

1. The wide range of age groups included in the study,
2. The four components probing children's perceptions of number patterns,
3. The open-ended nature of Components B and D, and
4. The qualitative post-coding procedures used.

To facilitate the comparison, each of the responses reported in the studies reviewed in Chapter 1 was compared with the categories of this study. Tables 4.11 to 4.13 list the authors who reported responses which could be allocated to each category. These tables support the view that the response categories developed in this study are not idiosyncratic since nearly all categories can be identified in the research literature.

Table 4.11

Comparison of Component B with Results from Other Studies

Response category	**Researchers**
No attempt (**NA**)	O'Brien (1991) MacGregor and Stacey (1993)
Inappropriate (**IA**)	
One Example (**1EG**)	Stacey (1989) O'Brien (1991)
Successive (**SUCC**)	Stacey (1989) O'Brien (1991) Ursini (1991) Meira (1990) MacGregor and Stacey (1993)
Function (**FUNC**)	Stacey (1989) (Subgroups 1 and 2) O'Brien (1991) (Subgroups 1 and 2) Meira (1990) (Subgroups 1 and 2) MacGregor and Stacey (1993)

Table 4.12

Comparison of Component C with Results from Other Studies

Response category	**Researchers**
No attempt (**NA**)	O'Brien (1991) MacGregor and Stacey (1993)
Incorrect (**IC**)	Stacey (1989) (iterative and proportions) O'Brien (1991) (proportions) MacGregor and Stacey (1993)
Correct (**C**)	Stacey (1989) O'Brien (1991) MacGregor and Stacey (1993)

Table 4.13

Comparison of Component D with Results from Other Studies

Response category	**Researchers**
No attempt (**NA**)	MacGregor and Stacey (1993)
Operation symbols (**OS**)	Stacey (1989) Ursini (1990) Meira (1990)
Arbitrary use of letters (**AL**)	Ursini (1990) MacGregor and Stacey (1993)
Iterative description (**REPT**)	Ursini (1991)
Algebra (**ALG**)	Ericksen (1988) Stacey (1989) Ursini (1990) Meira (1990) MacGregor and Stacey (1993)

There are some exceptions to this representation. The first is that no responses were reported for Component A. This was because the specific request for the next term in the sequence was not asked in any of the studies. However, Stacey (1989), and Stacey and MacGregor (1993) asked for a "near" term and these responses were allocated to the Component C group in Table 4.12. There were three categories that are not represented by multiple reports. One of these is the no attempt (**NA**) categories in Component D. While O'Brien (1991) expressed surprise at the large number of no attempts in his study, the category was mentioned as "omit" in the tables of Stacey and MacGregor (1993). However, other reports indicated that not all responses were classifiable or that some responses were incorrect. It would be reasonable to assume that no attempt (**NA**) were classified in either of these two ways. A similar suggestion is made with regard to the other two poorly represented categories. These two (inappropriate (**IA**) in Component B (see Table 4.11), and iterative description (**REPT**) in Component D (see Table 4.13)) both represent incorrect answers and hence would be treated as unclassifiable or incorrect.

Some of the response categories identified in the study have been also reported in other contexts. The symbolic responses that have been classified as "arbitrary use of letters" represent a number of subcategories that have been widely reported elsewhere in a number of environments (Küchemann 1981; Sleeman 1984; MacGregor 1990). Rojano and Sutherland (1993), for example, reported a teaching experiment involving the development of algebraic concepts in a spread-sheet environment. They reported that initially children did not think in terms of a general algebraic object but rather:

> Their thinking was initially situated on the specific example with which they were working.
>
> (p. 189)

This would appear to closely relate to the one example (**1EG**) category reported here. Elsewhere in their study Rojano and Sutherland (1993) reported a spread-sheet strategy that was remarkably similar to the successive (**SUCC**) strategy used by children to describe number patterns. These reports support the view that such responses represent a wider set of behaviour than that specifically generated by the stimulus items of this study.

As a result of this discussion the first two of the research questions of Theme 1 can be answered.

Theme 1

> The natural language that children use to describe patterns associated with numbers in Years 5, 6, 7 and 8 has been described.
>
> 1.1
> Children's natural language descriptions of number patterns can be classified into a discrete set of categories both reliably and validly.
>
> 1.2
> Children's symbolic language descriptions of number patterns can be classified into a discrete set of categories both reliably and validly.

While the categories described have been applied to all four stimulus items and the children's responses have been mapped onto those categories, the third research question of Theme 1 has not as yet been addressed, i.e.,

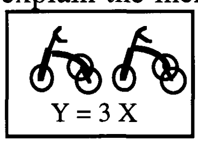
> Question 1.3
> Are the response categories stable across stimulus items that vary in complexity and context?

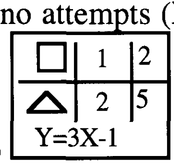
Stimulus Item 1 placed an upper limit on the complexity of the descriptions that could be provided to describe the pattern because of the simple data set (one arithmetic operation). Hence, children could not be classified in some subgroups of the categories of Component B and Component D. Thus, the item complexity of the stimulus item has some effect on the response patterns. However, to consider this matter in more detail the frequency of responses needs to be considered. This is the focus of attention in the next section.

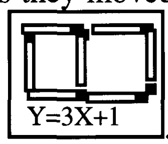
FREQUENCIES

In this section frequency tables are reported for each component of each stimulus item, together with a graph of the frequencies of each category. Finally, prior to responding to the third research question of Theme 1, some general issues are discussed .

Component A.

Table 4.14 reports the frequencies of responses to Component A of Stimulus Items 1 to 4. This information is presented graphically in Figure 4.11. Of some concern is the frequency of no attempts and incorrect answers for Stimulus Items 3 and 4. Teachers reported a degree of frustration from many of the children due to the unusual nature of the questions. This may explain the increased number of no attempts (**NA**) as they moved from Stimulus Item 1 

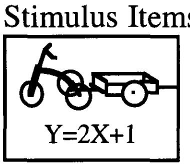
to Stimulus Item 3 

and 4 

. Teachers were asked to create a relaxed environment by removing tension from the activity. It may be that this was overemphasised in some classes and hence children tended not to persevere when the items became difficult. A third explanation for the increase in non-attempts towards the end of the items may be that the teacher stopped the class working before all students had completed the task. Indeed one Year 6 boy wrote on his script

> I think I could do this but the teacher just stopped us working.

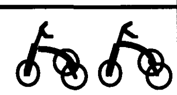
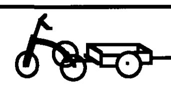
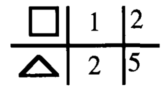
These explanations remain speculative and the relatively high frequencies of no attempts (**NA**) remain a consequence of the decision to use classroom teachers to administer the items so as to enable a large sample size.

Also of interest is the large number of incorrect (**IC**) answers to Component A for Stimulus Items 3 and 4. The rate of incorrect answers for Stimulus Items 1 (3%) and 2 

(13%) was relatively small and could largely be accounted for by simple arithmetic errors by the respondents. However, error rates for Stimulus Items 3 (22%) and 4 (35%)

Table 4.14

Frequency of Response for Component A on each Stimulus Item

Code		Stimulus Item 1	Stimulus Item 2	Stimulus Item 3	Stimulus Item 4
		 Y = 3 X	 Y=2X+1	 Y=3X-1	 Y=3X+1
NA No attempt	Total	4	12	168	112
IC Incorrect	Total	48	187	312	499
C Correct	Total	1383	1236	955	824
		n=1435	n=1435	n=1435	n=1435

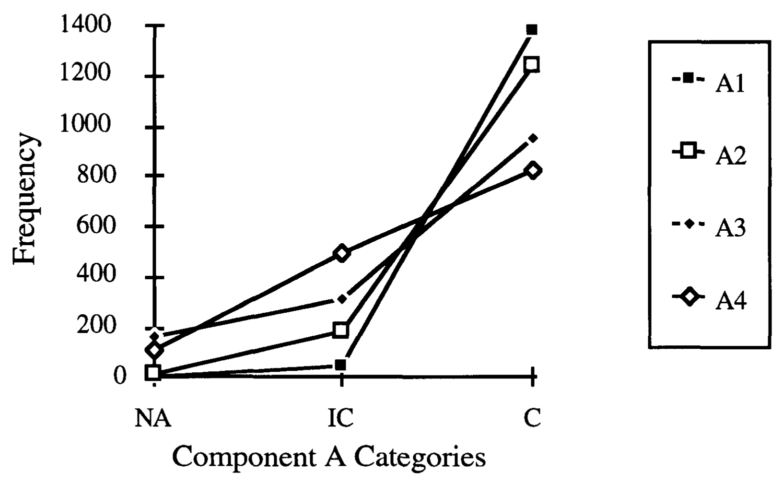


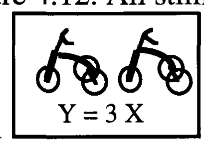
Figure 4.11

Distribution of Responses to Component A

were surprisingly large. An analysis of the responses to Item 4 showed that the most common error was to provide an answer of 14. This was calculated by counting the number of matches for three squares from the diagram and adding four. It appeared as if the image of a square with four sides overrode aspects of the stimulus item. In terms of the SOLO Taxonomy, the visual picture that the children were given in the diagram provided an inappropriate data element. This ikonic miscue did not occur for Stimulus

Items 1 and 2. For Item 1 the image of a tricycle stimulated the appropriate incremental number of three and this consistency applied in Item 2, with the trailer prompting the two as the dominant feature of the question. The errors made in Stimulus Item 3 were more varied and seemed to reflect a vagueness about the intent of the question. The only group of responses that appeared associated with ikonic imagery were those that suggested two should be placed under 5. This was arrived at by turning the diagram upside down. so that $\frac{2}{5}$ becomes $\frac{5}{2}$.

Component B

Table 4.15 presents the frequencies for Component B (pattern descriptions using natural language) for the four stimulus items. An overview of category totals is presented in Figure 4.12. All stimulus items are represented in all major categories. The first stimulus item 

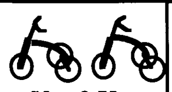
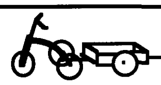
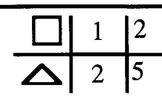
is not represented in some subgroups due to it being a single operation question.

There is a general diversity of responses within each stimulus item, which was expected given the large sample (n=1435) across four school years and 13 different schools. This variability is investigated further when the effect of school year and the performance on each item is considered in Chapter 6.

Apart from the variability within each stimulus item there is also considerable variability between the stimulus items. Figure 4.12 presents the shape of the distributions for each stimulus item and it can be seen that there is not only variability as to the modal response category but also differences in the shapes of the distributions. In particular, the distributions for Stimulus Items 1 and 3 are bi-modal in nature. This is caused by the very low frequency of the **SUCC** category for Stimulus Item 1 and a similar low frequency of the **1EG** category for Stimulus Item 3. It is tentatively suggested that this difference is due to some factors other than item complexity. That is, there are some item effects that caused the children to avoid responding in these categories.

Table 4.15

Frequency of Response for Component B on each Stimulus Item

Code	Subgroup	Stimulus Item 1	Stimulus Item 2	Stimulus Item 3	Stimulus Item 4
		 Y = 3 X	 Y=2X+1	 Y=3X-1	 Y=3X+1
NA No Attempt	Total	94	154	458	360
IA Inappropriate	Total	117	129	55	95
1 EG One Example	Subgroup 1	496	27	3	31
	Subgroup 2	0	429	22	267
	Subgroup 3	24	11	2	8
	Total	520	467	27	306
SUCC Successive Description	Subgroup 1	38	171	320	195
	Subgroup 2	0	254	5	97
	Subgroup 3	0	12	212	2
	Total	38	437	537	294
FUNC Independent/ Dependent Variable	Subgroup 1	666	59	110	85
	Subgroup 2	0	189	248	295
	Total	666	248	358	380
		n=1435	n=1435	n=1435	n=1435

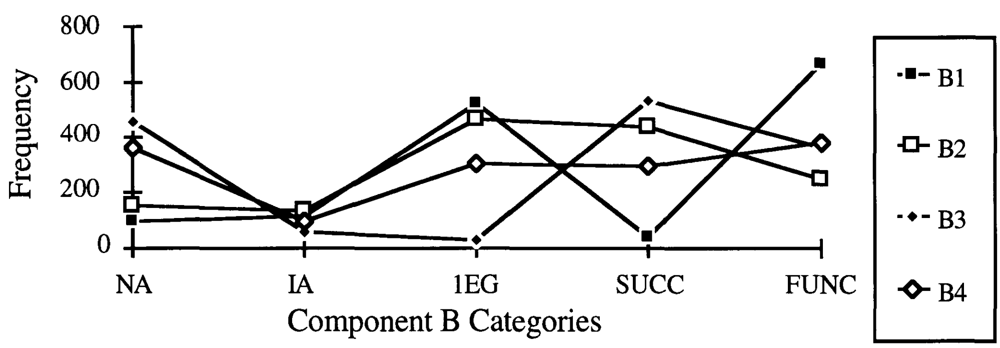


Figure 4.12

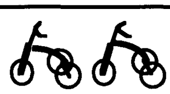
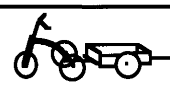
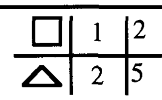
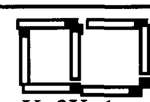
Distributions of Responses to Component B

Component C

Table 4.17 and Figure 4.13 present the frequencies for Component C (calculation of an uncountable example). One reason for this component being included in the study was to assist in the coding of Component B. It was intended to use Component C to shed additional light on any responses to Component B that were ambiguous. It would therefore seem to be reasonable to expect that a functional response to Component B would be required to facilitate a correct answer to an uncountable example. However, the ratio of correct responses to Component C to functional descriptions for Component B ranged from 1.8 to 1 for Stimulus Item 3 to 4.3 to 1 for Stimulus Item 2. These ratios seem very large and the issue of the interaction between Component B and Component C is addressed in more detail in the discussion of Theme 3 in Chapter 6.

Table 4.16

Frequency of Response for Component C on each Stimulus Item

Code		Stimulus Item 1	Stimulus Item 2	Stimulus Item 3	Stimulus Item 4
		 Y = 3 X	 Y=2X+1	 Y=3X-1	 Y=3X+1
NA No Attempt	Total	27	121	487	335
IC Incorrect	Total	80	227	298	216
C Correct	Total	1328	1087	650	884
		n=1435	n=1435	n=1435	n=1435

While there is variability in the success rates across the four stimulus items on Component C, this variability is not as great as for Component B. The differences that do exist could well reflect the item complexity rather than the other specific effects hypothesised in the discussion of Component B.

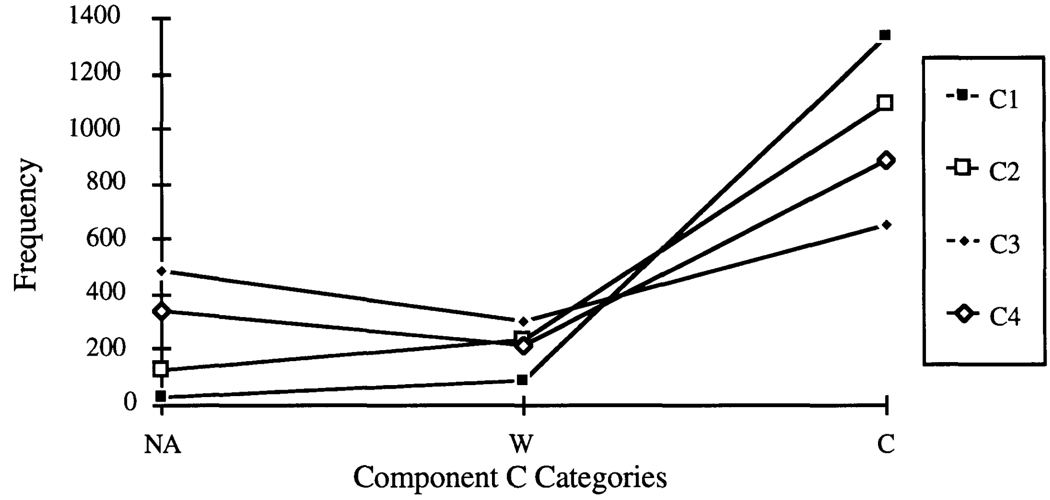


Figure 4.13
Distribution of Responses to Component C

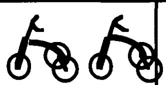
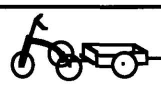
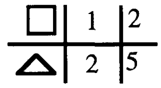
Component D

Table 4.17 presents the frequencies for Component D which investigated children's ability to use mathematics symbolism. The category totals are presented graphically in Figure 4.14. One of the significant features of the table is the very much higher proportion of no attempt (**NA**) responses than the other components. Once again this proportion increased with the later stimulus items. The low frequency for the repeat (**REPT**) category in Stimulus Item 1 raises two issues. First, it suggests an association between that category of symbolic response and the successive (**SUCC**) category of the natural language response. This and other associations are investigated in Chapter 6. Secondly, it represents the only major deviation from the pattern of responses across the four stimulus items. There is the large difference in the number of no attempts between the items but the shape of the curves for Stimulus Items 2, 3 and 4 are very similar.

It would appear from the frequencies that the simpler data set of Stimulus Item 1 has led to a much larger number of children successfully using algebraic notation to describe the number pattern.

Table 4.17

Frequency of Response for Component D on each Stimulus Item

Code	Subgroup	Stimulus Item 1	Stimulus Item 2	Stimulus Item 3	Stimulus Item 4
		 Y = 3 X	 Y=2X+1	 Y=3X-1	 Y=3X+1
NA No Attempt		242	359	677	557
OS Operation Symbols	Subgroup 1	64	73	63	47
	Subgroup 2	347	290	141	224
	Subgroup 3	18	11	9	8
	Total	429	374	213	279
AL Arbitrary Use of Letters	Subgroup 1	37	52	32	36
	Subgroup 2	144	139	58	102
	Subgroup 3	67	41	38	26
	Subgroup 4	41	28	8	18
	Total	289	260	136	182
REPT Iterative Description	Subgroup 1	14	73	150	97
	Subgroup 2	0	160	11	43
	Total	14	233	161	140
ALG Algebra	Subgroup 1	49	27	38	34
	Subgroup 2	3	38	25	66
	Subgroup 3	8	19	4	15
	Subgroup 4	401	125	181	162
	Total	461	209	248	277
		n=1435	n=1435	n=1435	n=1435

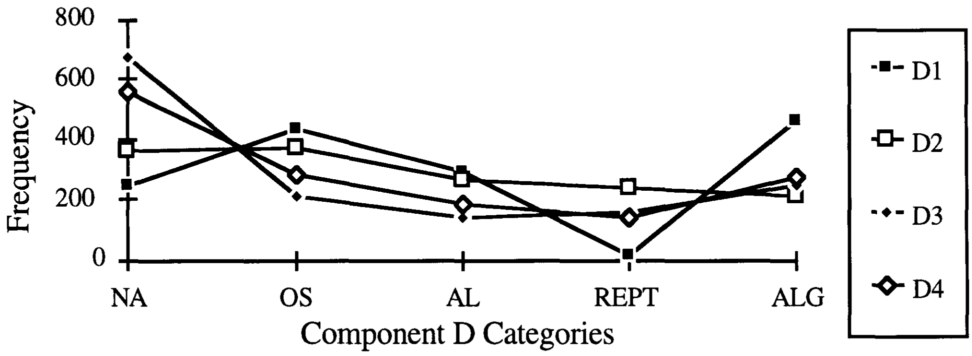
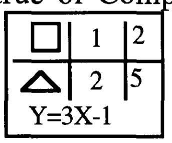
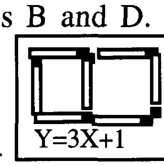


Figure 4.14
Distribution of Responses to Component D

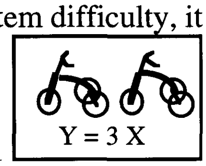
Discussion

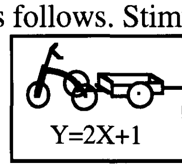
The above analysis has focused on each component individually. This part considers some general comparisons of the relative difficulty of the stimulus items.

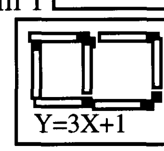
Children did not find the four items easy judging by the number of no attempts and incorrect responses. This was particularly true of Components B and D. The large number of no attempts for Stimulus Items 3 

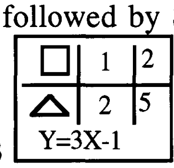
and 4 

could be indicative of a large attrition rate during the survey. It should be noted that Stimulus Item 3 has the largest number of no attempts for each of the four components. The fact that this number exceeds the no attempts for Stimulus Item 4 may say something about the relative difficulty of the items.

Some additional clues about relative difficulty of the stimulus items are given by the response frequencies for Component C. To the extent that children's ability to correctly extend a pattern to a specific, but uncountable, example is a measure of item difficulty, it can be inferred that the order of difficulty was as follows. Stimulus Item 1 

was the easiest followed by Stimulus Items 2 

and 4 

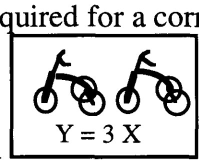
with Stimulus Item 3 

being the most difficult. Making inferences about item difficulty based on the response frequencies of Components B and D is more problematical. As an example, it is useful to compare the no attempt responses with the function responses of Component B (see Table 4.15). Stimulus Item 3 scores highest in the no attempt category and second on the function category of all the two operation items. This data would seem to be in conflict and is suggestive of a different response pattern across the four stimulus items. The issue of order of difficulty is taken up again in Chapter 5 when the Rasch model is applied in order to rank the components across the four stimulus items.

The final research question of Theme 1 can now be answered, i.e.,

> Research Question 1.3
> Are the response categories stable across stimulus items that vary in complexity and context?

The response categories identified are stable in that they can be used to represent the responses to the four stimulus items and components of this study. However, very different patterns of responses between stimulus items were noted when the category frequencies were analysed. The extent to which these differences in the frequency distributions are due to the attrition caused by the difficulty children had with the unusual nature of the items, or to the lack of appropriate encouragement given by the class teacher, is unknown. However, it should be remembered that similar difficulties were reported by both O'Brien (1991) and MacGregor and Stacey (1993).

Two issues of particular interest have been noted in the descriptive analysis of this chapter. The first is the considerable number of children who provided correct responses to Component C (uncountable example) without providing a function description in Component B. The second issue is the tendency for many children to provide descriptions in many categories that use only one arithmetic operation when more than one operation is required for a correct response. It was reasonable to expect that the one-operation question 

would result in more successful responses than the more complex two-operation items. However, it was not expected that children would use one data element in the more complex environments as a strategy. This strategy was used for all stimulus items in all response categories except the no attempt and inappropriate categories. These matters, and others, are the subjects of consideration in the next chapter.

SUMMARY

This chapter has provided a detailed account of the coding procedures used in the initial analysis of the responses of a sample of 1435 children in Years 5 to 8. This analysis has provided the first descriptive phase of three phases of the analysis hierarchy described earlier and adapted from the work of Miles and Huberman (1984).The interpretive and explanatory phases are undertaken in Chapters 5 and 6.

Detailed attention was paid to the validity and reliability of the coding process and particular measures were taken during the categorisation of responses to ensure the results are both reliable and valid. These measures include using a number of coders, calculating intercoder and intracoder reliability coefficients, using post-coding procedures to avoid data forcing, recoding until categories stabilised and comparing categories with those of other researchers.

As a result of these procedures, sets of categories were presented with descriptors and frequencies. Thus, the first two research questions of Theme 1 were answered in the affirmative. That is, the natural language and symbolic language that children use to describe number patterns can be categorised into a discrete number of categories. In addition, one set of categories was able to be applied to the responses to the four stimulus items that varied in complexity and context. However, quite different distributions for the items were noted. These differences become the focus of the analysis in the next chapter, which undertakes the interpretation of the data within the context of the SOLO Taxonomy.

Chapter 5

SOLO ANALYSIS

Introduction

The previous chapter presented a descriptive overview of the children's responses to the stimulus items and their components. This was done by outlining the coding categories, their descriptors and the frequency for each component. In this chapter the process of interpreting the data is begun. Initially, the focus is on the response categories for Component B of the study, namely, the natural language the children used to describe the number patterns, thus facilitating the consideration of the second research theme outlined in Chapter 3

Theme 2
Can the children's descriptions of number patterns be analysed within the theoretical framework of the SOLO Taxonomy?

> Question 2.1
> Do the categories of pattern description using natural language exhibit properties that would enable mapping them onto the modes and levels of the SOLO Taxonomy?
>
> Question 2.2
> Can a hierarchy of growth be postulated from the response categories of pattern description using natural language?
>
> Question 2.3
> Do the stimulus items that vary in complexity and context reveal differences in difficulty?
>
> > Are single-operation patterns easier than two-operation patterns?
> > Are physical contexts such as matchstick diagrams easier than decontextualised number pairs?

The first section of the chapter compares the response categories with the modes and levels of the SOLO Taxonomy. As a result of this discussion a two-dimensional model is suggested that integrates two hierarchies of development identified in the data. The second section tests the model by using it to develop a hypothesised order of difficulty among the 16 components. This hypothesised sequence is compared with the component difficulty delta values produced by a Rasch analysis of the data. The Q statistic is used to investigate differences between the component difficulties. The final section summarises the findings in relation to the three research questions which are the focus of the chapter.

SOLO CATEGORISATION

A detailed set of categories and descriptors for Component B was presented in Tables 4.2 to 4.5 in Chapter 4. In each of the categories in which children experienced some success (**1EG, SUCC** and **FUNC**), a set of subgroups was described that provided a more fine-grained classification of the responses. A simplified version of the categorisation is presented here in Table 5.1. These categories provide the main focus of this study and will be the initial focus of the analysis in the context of the SOLO Taxonomy.

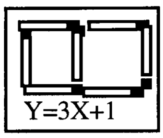
Table 5.1

Overview of Response Categories for Component B

Category	Subgroup
NA No attempt	
IA Inappropriate	
1EG One example	One data element
	Two data elements
	Transitional
SUCC Successive description	One data element
	Two data elements
	Transitional
FUNC Function	One data element
	Two data elements

In Chapter 2 the SOLO Taxonomy was identified as a suitable theoretical framework with which to interpret the data. The application of the Taxonomy to the data is a two-step process. The first step involves identifying the mode in which to place the responses and the second is identifying the appropriate level within the mode. A detailed description of the second of these steps is included as Appendix 5.3

It was argued in Chapter 2 that the mode of development is the concrete symbolic mode since the children are being asked to use a symbol system to describe a pattern that has a concrete referent. The use of a symbol system implies that the child is not operating in the ikonic mode, or at the very least the implication is that the target mode, in the terms of Biggs and Collis (1991), is the concrete symbolic mode. The children whose responses were placed in the **NA** and **IA** categories may have been able to provide ikonic responses but were unable to experience success in the target mode.

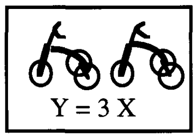
Ikonic support through multimodal interaction may have played a role in influencing the precise nature of some responses that were classified in the concrete symbolic mode. In Chapter 4 it was reported that there is evidence that some children used ikonic support in the formulation of their responses via the visual cues of the stimulus item providing inappropriate data elements which overrode other cues within the questions. An example of this is the number of children who used four as the incremental value in Stimulus Item 4 

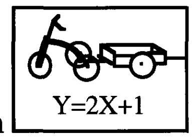
instead of the correct value of three. A little later in this chapter the distinction between responses of $3x+1$ and $4+(x-1)\times 3$ is made explicit and could be attributed to differences in the ikonic images generated among children. However, while the ikonic mode remains available to the children of the study, and may well influence the precise nature of their responses, the target mode for the content of the study, and the media of the responses, ensure that successful responses are in the concrete symbolic mode. The levels of the responses are a little more problematic.

From the overview of the categories presented in Table 5.1, it can be seen that there is a pattern of repetition in each of the categories **1EG**, **SUCC** and **FUNC**. In interpreting the results in terms of the SOLO Taxonomy there appear to be two hierarchies of development. One concerns the use of data from the question, and the other concerns the sense of an overview of the data that can be provided in the form of an expression of generality in their pattern description. It would appear that these two concepts are not necessarily developed in parallel. But a child may in fact develop in one or the other independently.

The two dimensions alluded to above are considered in more detail. The first dimension is designated the data processing dimension. It refers to the number of data elements from the stimulus item that the child can hold in memory and successfully process. The second is designated the generality dimension and refers to the child's ability to describe a general relationship that represents an overview of the information provided in the stimulus item. In the following a developmental hierarchy is postulated for each dimension and then their possible interaction is discussed.

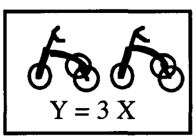
Data Processing Dimension

Children exhibited the ability to deal with a differing number of data elements. Data elements, in this context, means the number facts provided by the stimulus item. For example, in the first stimulus item 

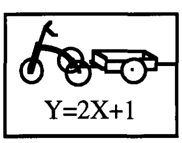
there is one element of data, namely, each tricycle has three wheels. In the second stimulus item 

there are two data elements. The precise nature of the elements of data depended on the child's perception of the question. Some children saw that three wheels were needed for the first person and then two wheels were needed for each remaining person ($y = 3 + 2(x - 1)$). Others perceived the data as two wheels are needed for each person with one extra wheel needed for the tricycle ($y = 2x + 1$).

Some children showed that they can respond at a higher level to stimulus items that had only one piece of data, but were unable to provide similar descriptions when more than one piece of data needed to be processed, e.g., a child provided a functional description for Stimulus Item 1 and a successive response for Stimulus Item 2:

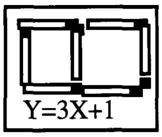


The number of wheels is the number of tricycles timesed by three.



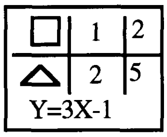
You just keep on adding 2

Other children seized on one piece of data from situations when more than one piece of data was supplied and attempted to generalise from this restricted data set, e.g.,



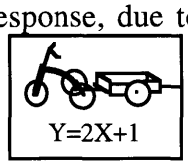
Matches equals squares times four.

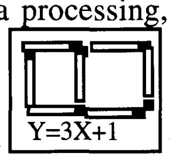
or;



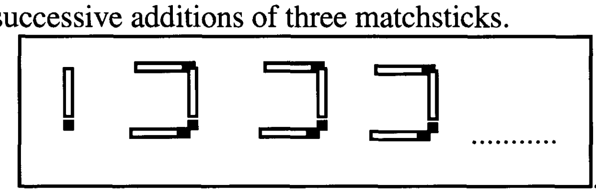
You double the top row.

The above responses are described, in SOLO terms, as lying in the concrete symbolic mode at the unistructural level (CSU). They have been classified as unistructural responses since only one piece of data is being processed at a time in offering a generalisation.

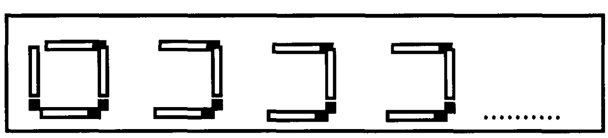
A second group of responses are classified at a higher level than CSU since they require children to successfully process more than one piece of data. Hence the responses should be classified at one of either concrete symbolic multistructural (CSM) or concrete symbolic relational (CSR). The nature of the stimulus items placed an upper limit on the quality of the response, due to data processing, that the children could produce. In Stimulus Items 2 

and 4 

the data were presented in such a way as to enable a multistructural analysis of the data for the children to be able to produce an adequate data set. As with Stimulus Item 2, the squares constructed from match sticks in Stimulus Item 4 can be perceived in two different ways. Some children see the pattern as one matchstick with successive additions of three matchsticks.



Others see the pattern as four match sticks with successive additions of three match sticks.



These views of the data are analogous with Biggs and Collis' (1991) description of multistructural as sets of unrelated elements. Children who can adequately describe these data are capable of responding at a multistructural level in the concrete symbolic mode.

This does not imply that the same children cannot respond at a higher level if required, but rather it describes the minimum level of potential required to experience success with the item. The child is not forced to describe relationships between parts of the data to identify an adequate data set.

On the other hand, in Stimulus Item 3

□	1	2
△	2	5
Y=3X-1		

such a multistructural view of the data does not lead to an adequate data set. The children who described the data as "the top line goes up by one and the bottom line goes up by three" do not have the necessary overview of the data to describe the relationship between the top line and the bottom line. The data needed are a set of arithmetical operations that had to be performed on the top line (independent variable) to give the values in the bottom line (dependent variable). Because a common set of elements that describe the relationship has to be found, the identification of an adequate data set for Stimulus Item 3 is classified as concrete symbolic relational (**CSR**).

The data processing dimension has facilitated three levels of response within the concrete symbolic mode. The unistructural level is characterised by children selecting only one data element from the stimulus item and proceeding to respond to the component questions using only this single data element. The distinction between the multistructural level and the relational level is facilitated by the complexity of the data structure within the stimulus item. Stimulus Items 2 and 4 provide data structures that are multistructural and hence place an upper limit on the level of response that is required to successfully identify the required data elements. Stimulus Item 3, on the other hand, required a relational level response to identify the data elements. The implication of this is that children who can successfully identify the appropriate data elements in Stimulus Item 3 should be able to successfully identify the appropriate data elements for Stimulus Items 2 and 4.

Children's responses on the stimulus items, however, were not solely dependent on the ability to identify the correct data elements. The second dimension identified was the expressing generality dimension.

Expressing Generality Dimension

In the discussion of coding classifications in Chapter 4 for Component B, five categories was described. The first two were called no attempt (**NA**) and inappropriate (**IA**) respectively. Both these categories reflect the child's inability or unwillingness to successfully engage the stimulus items. The term inappropriate reflects the properties of the pre-structural level of the first version of the SOLO Taxonomy (Biggs and Collis, 1982). It is plausible that these children are operating in the ikonic mode and are relying on imagery and intuition but are unable to express their thoughts in the concrete symbolic mode as required by the question. It may be feasible to investigate these responses further in an interview situation. However, on the data available it is not possible to classify them in terms of the SOLO Taxonomy.

The first group that experienced some success in responding to Component B of the stimulus items were those children who were able to give a specific example (**1EG**) rather than a general pattern description. Since they focus on one example these responses are classified as within the concrete symbolic mode at the unistructural level (**CSU**).

The next group identified have been described as successive (**SUCC**). The tendency in this group is to focus on outcomes only, and to describe how to move from one term in a sequence to the next term. This group has been classified as concrete symbolic multistructural (**CSM**). Under some circumstances there is a tendency to see successive descriptions as a relatively sophisticated iterative function. Such a function, however, implies the existence of a feedback loop to enable the calculation of the next term as in;

$$y_{n+1} = y_n + 3$$

Such strategies have relevance in many mathematical situations such as Newton's Method for approximating the roots of a function. In many computer settings similar strategies are used for incrementing variable values. Examples include the For-To-Next loop in BASIC programming, recursion in the LOGO environment, and spreadsheets. However, these iterative functions are not the focus of the successive category in this study. The predominant feature of the successive description was the "keep on adding" focus. No description that implied a sophisticated iterative process was identified.

The final group are classified as concrete symbolic, relational (**CSR**) since they have presented an overview of the data by describing a functional relationship (**FUNC**) between a dependent variable and an independent variable.

Developmental Model

In the previous part the expressing generality dimension and the data processing dimension were described separately. In this part the dimensions are integrated to form a developmental model.

Using the above classification of the children's responses into SOLO levels, a hierarchical model of development for pattern descriptions can be hypothesised. Figure 5.1 is a diagrammatic representation of the model. The model represents children who are capable of responding in the concrete symbolic mode and their earliest entry-point to the model is in the top left-hand cell (U_DU_G) which reflects the unistructural level of both the data processing dimension (U_D) and the expressing generality dimension (U_G). The addition of multistructural and relational responses in each dimension result in a 9 (3x3) cell model.

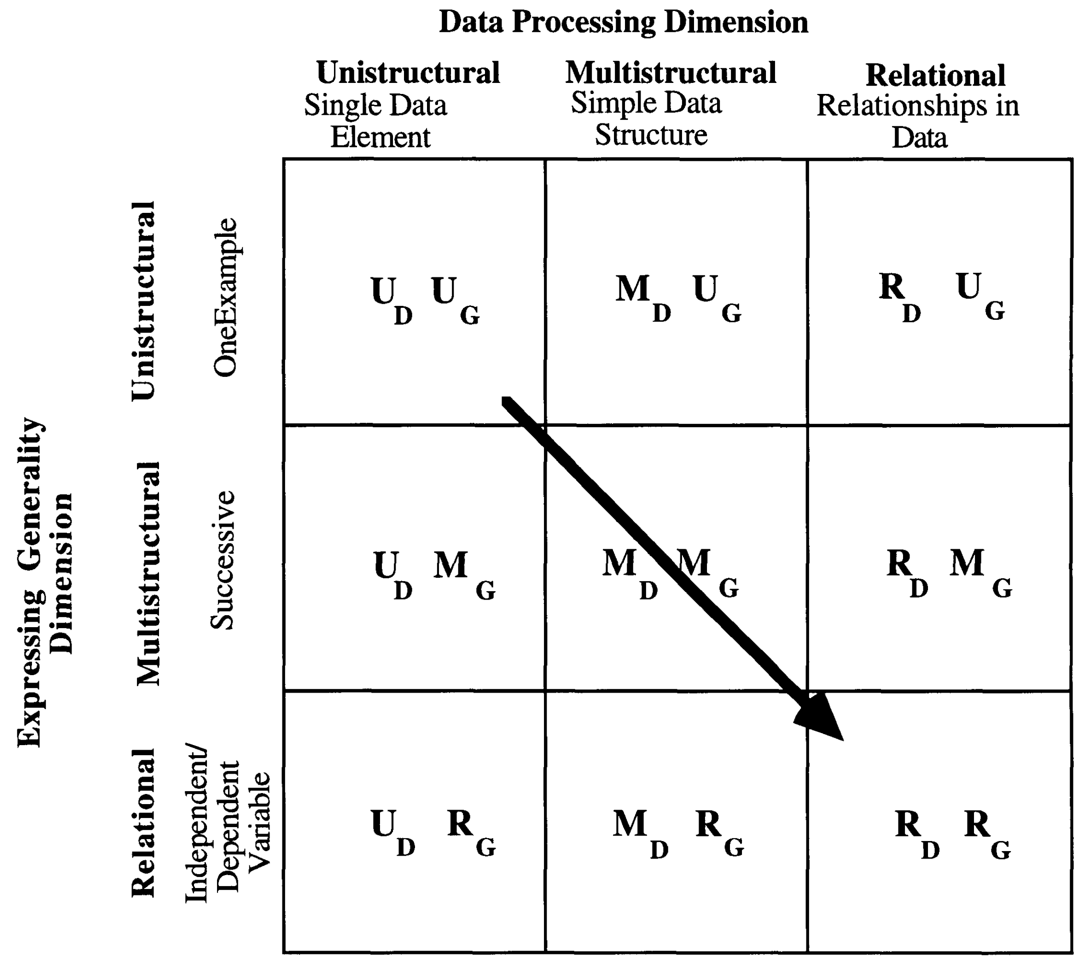


Figure 5.1
Hypothesised Development Model

The SOLO Taxonomy would predict an additional classification on each dimension to represent prestructural responses on that dimension, resulting in a model that has 16 (4x4) cells. While responses to Component B were identified that were clearly indicative of prestructural responses (these were coded as **NA** and **IA**) none were able to be classified at a higher level on one dimension and remain prestuctural on the other dimension. Hence, an extension of the model could be conceptualised as a 10 cell (3x3+1) as in Figure 5.2 (a) but not as a 16 cell (4x4) model as in Figure 5.2 (b).

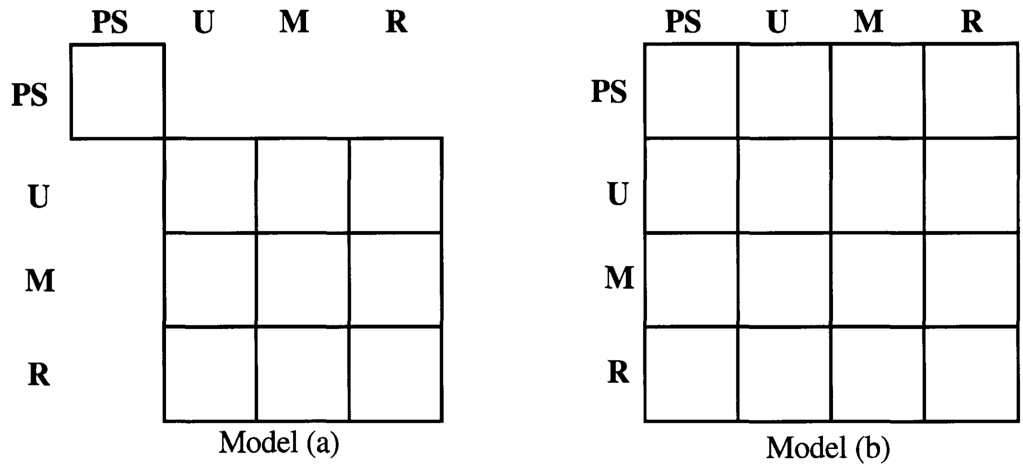
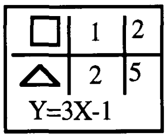


Figure 5.2
Alternative Development Models

A possible reason for the lack of prestructural responses to complete a 16 cell model is the nature of the stimulus item which place limits on the range of responses that can be classified as successful. The stimulus items did not allow for success in one dimension without some success in the other dimension. That is, generality cannot be expressed at any level without at least one data element being identified. Alternatively, once one data element has been identified, in the context of the components associated with each stimulus item, then some kind of generality is expressed according to the coding system developed in Chapter 4. A similar argument could be mounted for the absence of extended abstract responses. That is, the survey questions did not provide an opportunity for such responses to be detected, hence, the discussion is restricted to development within the 9 cell model presented in Figure 5.1.

In general, the desired development path would be represented by the diagonal line in Figure 5.1. This line represents a transition from providing unistructural responses in both dimensions (U_DU_G) to being able to respond at the relational level in both dimensions (R_DR_G). However, it is clear from the survey data that the progress exhibited by children is probably not linear, nor restricted to the diagonal cells. It can be seen that there is a possibility of a large number of paths from U_DU_G to R_DR_G. Some of the development paths that are evident in the survey data are considered in the next paragraph. It is not the researcher's intention to quantify the likelihood of a student following any of these paths, but rather to provide evidence for their existence. A little

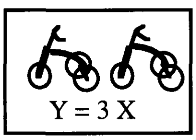
later in the discussion some possible reasons for the variability in the paths are discussed.
1. The path from U_DU_G to M_DU_G to R_DU_G. Some children demonstrated an ability to identify the appropriate data elements for the relational data set but only provided specific examples, e.g.,



Under 7 you would put 20.

While it could be argued that such a response could be provided by counting on, further evidence for the ability to identify the complex data set was provided in Component C. A small group of children were able to provide the correct answer to an uncountable example in Component C even though their pattern description was low on the generality dimension. Thus, they had correctly identified the necessary data set to provide a general description but their response to Component B indicated that the concept of expressing generality had not yet developed.

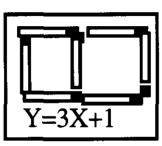
2. Path from U_DU_G to U_DM_G to U_DR_G. These children demonstrated an ability to express generality for patterns that had one data element such as in Stimulus Item 1. They were able to write succinct sentences that relate the number of wheels (dependent variable) to the number of tricycles (independent variable), e.g.,



The number of wheels equals three times the number of tricycles.

but were unable to write such relationships for the more complex data sets of Stimulus Items 2, 3 and 4.

Another set of responses that provided evidence for the existence of this path were those which use one piece of datum from a more complex data set and based all the answers for the components of that stimulus item on this restricted view of the data. An example of this type of response is the group of students who, in Stimulus Item 4, focus on the fact that a square has four sides and then go on to describe the rule as:



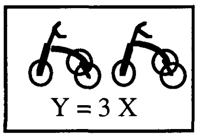
Multiply the number of squares by four to get the number of matches.

and then answered Component C with;

80x4=320

thus ignoring the fact that only three matches were required for each additional square.

3. Interaction Paths. Thus far in the discussion the two dimensions of growth have been treated as independent of each other. However, there was some evidence of an interaction effect between the two dimensions. This interaction effect was made evident by the variability in response structure across different stimulus items. One common difference was for children to provide a relational description (R_G) to Stimulus Item 1



and a successive response (M_G) to the more complex data sets of Stimulus Items 2, 3 and 4. In terms of the model in Figure 5.1 this represents a shift from U_DR_G to M_DM_G. Of course this is not a movement over time since both responses were collected on the same day. A plausible explanation is that the reduction in the level of expression of generality is a consequence of the additional cognitive load brought about by the need to process a more complex data set.

In this section the data have been categorised in terms of the SOLO Taxonomy and a model of development hypothesised that implies there are multiple pathways of development for children as they develop the ability to describe number patterns of the type used as stimulus items in this study. (A more detailed consideration of this process is included as Appendix 5.3 which includes a sample of responses fitted onto the model.) In the next section, the efficacy of the model is tested by using it to make some predictions about the data and then to test those predictions using Rasch analysis.

TESTING THE MODEL

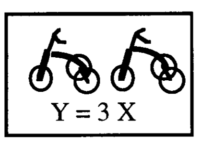
Thus far the survey data has been discussed from a qualitative perspective and a model of children's growth in the understanding of number patterns has tentatively been proposed using the SOLO Taxonomy as a theoretical framework. In this section the model is used to predict a sequence of difficulty of the stimulus items and a sequence of difficulty of the item components. This predicted sequence is then tested by a Rasch analysis of the data.

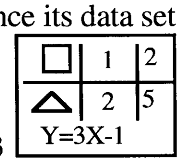
An Hypothesised Order of Component Difficulty

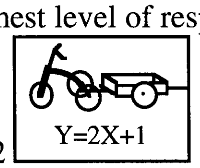
The developmental model suggests that Components A (next member of the pattern sequence) and C (uncountable member of the sequence) should be easier for children to respond to than Components B (natural language pattern description) and D (symbolic pattern description). This suggestion is made because Components A and C require development in only one of the two dimensions of the model, i.e., the data processing dimension. Further, since Component A requires the next term in the sequence it can always be calculated in these stimulus items by the addition of a single number (U_D), thus requiring a single piece of data for a correct response. Hence, it would be expected that children would be more successful in responding to Component A than Component C.

Components B and D require the development of both dimensions of the model for the child to be able to respond at the highest level ($R_D R_G$). While the model does not provide a direct indication of which of Components B or D children might find easier, the SOLO Taxonomy does. This indication comes from the mode of symbolism. Since responses to Component B are reported by children in their natural language we might expect them to experience more success in Component B than in Component D. There is an additional contributor to this expectation. The learning experiences of children in Years 5 and 6 have not included the use of algebraic notation and, hence a higher level of response to Component B than to Component D would be expected.

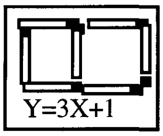
Thus far it has been argued that the order of difficulty of the components, from easiest to hardest, is Component A, Component C, Component B, and finally Component D. The model also provides some guidance as to the order of difficulty of the stimulus items.

Stimulus Item 1 

would be predicted to be the easiest item since its data set only requires analysis at the unistructural level (U_D). Stimulus Item 3 

would be the most difficult since the data set necessary for the highest level of response has been classified at the relational level (R_D). Thus Stimulus Item 2 

and 4



should lie between items 1 and 3 since their data set for the highest level response has been classified as multistructural (M_D).

In the above, a sequence of difficulty has been predicted for the stimulus items and a sequence of difficulty for the components of the stimulus items. It has not been predicted exactly how the two sequences interact. Table 5.2 tabulates a hypothesised sequence of difficulty. In this table the components are grouped into 12 categories and the rank of each group recorded, a rank of one being associated with the easiest components and a rank of 12 being associated with the components considered to be the most difficult.

Table 5.2
Hypothesised Rank Order of
Difficulty of Components

Rank	Components
1	A1
2	A2 A4
3	A3
4	C1
5	C2 C4
6	C3
7	B1
8	B2 B4
9	B3
10	D1
11	D2 D4
12	D3

To test this hypothesised rank order it was necessary to develop an index of difficulty from the children's responses. The strategy that seemed to have the least number of assumptions about the nature of the data and also provided parameter estimates of item difficulty was the Rasch Model (Masters 1982). That is discussed in detail below. It was decided to compare the hypothesised order of difficulty with the component difficulty indices of the Rasch Model using a Spearman rank correlation analysis. If the resulting correlation was sufficiently high (rho>0.8) the validation of the model could be inferred.

Rasch Analysis of Survey Data

A method of testing the predicted sequence of component difficulty is provided by the Rasch Partial Credit Model (Masters 1982; Andrich 1988) and represents an extension of the Andrich (1978) rating scale model. An implementation (referred to hereafter as Quest) of this model for polychotomous data is provided by Adams and Khoo (1993). Masters

(1982) described partial credit as a method of classifying responses that leads to a more precise estimate of a persons ability than a simple pass/fail score (p. 150). The allocated number to each category (0, 1, 2, 3,... in the Quest application) indicates only an ordering of the response categories and is not used as a category weight (p. 150). An additional feature of the partial credit model is that the response alternatives are free to vary in number and structure (p. 150).

Before proceeding, the ordinal nature of the categories for each component needs to be made explicit. For Components A and C a similar set of categories have been described and their ordinal nature is obvious. The categories in order from lowest to highest are no attempt (**NA**), incorrect (**IC**) and correct (**C**). These were allocated the numerical codings of 0, 1 and 2 respectively that are required by Quest.

The two lowest categories for Component B of no attempt (**NA**) and inappropriate (**IA**) were allocated the numerical values of 0 and 1 respectively. The SOLO Taxonomy was used to sequence the three remaining categories of one example (**1EG**), successive (**SUCC**) and function (**FUNC**) for Component B, which in turn were allocated the numerical values 2, 3 and 4.

The sequence of categories for Component D was a little more problematical to determine. Is was clear that the no response (**NA**) category should be classified as the lowest category and that the algebra (**ALG**) category should be classified as the highest category. The intermediate categories for Component D, the operating symbols (**OS**) category, the iterative (**REPT**) category and the arbitrary use of letters (**AL**) category remained to be sequenced. Using the Linchevski and Sfard model from Chapter 1 that characterised a shift from arithmetic thinking to algebraic thinking, together with the obvious close link between the **1EG** category of Component B and the **OS** category of Component D, the decision was made to place **OS** second in the sequence. This implied that **REPT** and **AL** categories were seen as transition stages between arithmetic notation and symbolic algebra. However, no theoretical support could be found for differentiating between these two categories in terms of a sequence. As a result they were given the same numerical coding value for the purpose of the Rasch analysis. In summary then, the categories for Component D were placed in the following sequence.

NA **OS** **(REPT & AL)** **ALG**

and were allocated the numerical values of 0, 1, 2 and 3, respectively.

With the categories for each component ordered, the data assumptions of the Quest application of the Rasch modelling process are consistent with the data of this study;

hence, the software (Adams & Khoo 1993) can be used to provide estimates of item difficulty together with measures of model fit. In addition to the tables of delta scores for item difficulty and the corresponding mean square values, the Quest program presents the information in the form of an item map that indicates the goodness of fit of the components to a latent trait. A variable map is presented as Appendix 5.1 that provides a sequence of component category difficulty together with a histogram of children's response frequencies for each level of difficulty.

Table 5.3 provides a summary of component estimates and fit statistics for the model that analysed all 16 components for the 1435 members of the sample. The component separation reliability indices as defined by Wright and Masters (1982) are also reported. The reliability of estimates is :

> The proportion of the observed estimate variance that is considered true.
> (Adams &Khoo 1993, p. 24)

Table 5.3

Summary of Component Estimates

Component Estimates (Thresholds)			
all on all (N = 1435 L = 16)			
Mean	-.01		
SD	1.37		
SD (adjusted)	1.35		
Reliability of estimate	.98		
Fit Statistics			
Infit Mean Square		Outfit Mean Square	
Mean	1.01	Mean	1.08
SD	.12	SD	.29
Infit t		Outfit t	
Mean	-.15	Mean	.16
SD	3.31	SD	3.36
0 items with zero scores			
0 items with perfect scores			

The fit statistics are the means and standard deviations of the infit (weighted) and outfit (unweighted) fit statistics in the mean square form. When the observed data and estimates are compatible the expected value of the mean square is 1. Adams and Khoo (1993) reported that:

> A fit mean square of (1+X) indicates (100X) per cent more variation between the observed and model-predicted response patterns than would be expected if the data and the model were compatible. Similarly, a fit mean square of (1-X) indicates less variation between the observed and model-predicted response patterns.
>
> (p. 86)

The value here of 1.01 reflects an adequate fit between the data and the model and can be interpreted as the data containing very few reversals. A reversal occurs when a child responds at a higher level on a harder item than his or her response level on an easier item.

Table 5.4 presents the 16 components, the component parameter estimates for difficulty, and the standard error for each parameter. The resulting rank order of the components

Table 5.4

Rasch Parameter Estimates and Ranks

Component	Difficulty Delta	Standard error	Infit Mean Square	Rank
A1	-3.11	0.14	1.05	1
A2	-2.17	0.08	1.12	2
A3	-0.32	0.05	1.03	6
A4	-0.41	0.05	1.11	5
B1	0.05	0.03	1.02	7
B2	0.70	0.03	1.00	10
B3	0.93	0.03	1.03	13
B4	0.87	0.03	1.08	12
C1	-1.85	0.09	1.04	3
C2	-0.67	0.05	1.07	4
C3	0.74	0.04	1.11	11
C4	0.20	0.04	1.16	8
D1	0.65	0.03	0.83	9
D2	1.33	0.04	0.79	14
D3	1.64	0.04	0.85	16
D4	1.43	0.03	0.81	15

based on the parameter estimates is reported for latter comparison with the predicted component rank.

The component infit mean square values are presented in graphical form in Figure 5.3 to assist in interpretation. The figures on the horizontal scale represent the infit mean square scale and the asterisks indicate the magnitude of the fit statistic for the component on the same line. Fit statistics that lie within the two dotted vertical lines are considered acceptable. The well fitting nature of the components to the model indicates that the components represent aspects of a latent trait.

```
Components Fit
all on all (N = 1435 L = 16)
----------------------------------------------------------------
INFIT
 MNSQ       .63      .71      .83     1.00     1.20     1.40
-------------+---------+---------+---------+---------+---------+-----
 1 A1                  .          | *           .
 2 A2                  .          |   *         .
 3 A3                  .          |*            .
 4 A4                  .          |  *          .
 5 B1                  .          |*            .
 6 B2                  .          *             .
 7 B3                  .          | *           .
 8 B4                  .          |  *          .
 9 C1                  .          | *           .
10 C2                  .          | *           .
11 C3                  .          |   *         .
12 C4                  .          |    *        .
13 D1                  .  *       |             .
14 D2                  . *        |             .
15 D3                  .  *       |             .
16 D4                  . *        |             .
```

Figure 5.3

Map of Component Fit

A comparison was made of the ranks of the hypothesised sequence and the Rasch order of difficulty, using the Spearman rank correlation coefficient. The correlation between the two sets of ranks was found to be 0.92 (see Table 5.5). This measure provided considerable support for the view that the hypothesised difficulty sequence reflected the children's response patterns and indirectly provided evidence that the two dimensional developmental model of pattern description based on the SOLO Taxonomy accurately reflected children's response patterns.

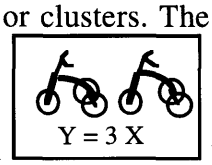
Table 5.5

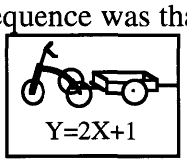
Spearman Rank Correlation for Rasch Order of Difficulty (X) and Hypothesised Sequence (Y)

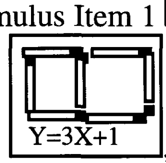
N	16	
ΣD^2	56	
Rho	.918	
Z	3.554	p = .0004
Rho corrected for ties	.917	
Z corrected for ties	3.553	p = .0004
#X tied groups: 4	#Y tied groups: 0	

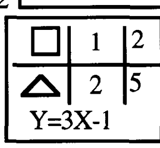
Thus far in the discussion only the sequencing of the item difficulties has been considered. In the next part of this section an analysis is conducted to investigate the significance of the differences in the component difficulties calculated in the Rasch analysis.

Tests for Homogeneity of Effect Sizes

In developing the hypothesised component difficulty sequence it will be remembered that the 16 components were allocated to 12 groups or clusters. The guiding principle in developing the sequence was that Stimulus Item 1 

would be easier than Stimulus Items 2 

and 4 

, which in turn would be easier than Stimulus Item 3 

. The rationale for this ordering was the complexity of the data sets for each stimulus item. Using the item difficulty delta values (δ) reported in Table 5.4 each stimulus item can be placed in a sequence for each component. These sequences, from easiest to hardest, are presented in Table 5.6.

Table 5.6

Components in Order of Difficulty

Component A (Next term in the sequence)	* A1	A2	A4	A3
Component B (Natural language)	* B1	B2	B4	B3
Component C (Uncountable example)	* C1	C2	C4	C3
Component D (Symbolic notation)	* D1	D2	D4	D3
* Indicates that the Q value was significant (p<0.005)				

It can be seen that the sequences of order of difficulty agreed with the predicted sequences in all components. Before discussing these sequences it was necessary to investigate the significance or otherwise of the differences between the item difficulty delta values (δ) for each of the four sequences. A test for investigating the existence or otherwise of a significant difference among the item difficulties was described by Hedges and Olkin (1985) as the "test for homogeneity of effect sizes" (p. 122). The test investigates the null hypothesis

$$H_0: \delta_1 = \delta_2 = \delta_3 = \ldots = \delta_k$$

versus the alternative hypothesis that at least one of the effect sizes, δ_i, differs from the remainder. The test statistic

$$Q = \sum_{i=1}^{k} \frac{(\delta_i - \delta_+)^2}{\sigma^2(\delta_i)}$$

is described by Hedges and Olkin (1985) as suitable for a test of homogeneity. δ_+ is the weighted estimate of effect size given by

$$\delta_+ = \sum_{i=1}^{k} \frac{\delta_i}{\sigma^2(\delta_i)} / \sum_{i=1}^{k} \frac{1}{\sigma^2(\delta_i)}$$

and $\sigma^2(\delta_i)$ is the variance of δ_i. Hedges and Olkin suggested that δ_+ is a satisfactory estimator for samples that exceed 10 and hence is suitable for this study.

> If all k studies have the same population effect size (i.e., if H_0 is true) then the test statistic Q has an asymptotic chi-square distribution with k-1 degrees of freedom.
>
> (Hedges & Olkin 1985, p. 123)

Hence if Q exceeds the required critical value of the χ^2 distribution with k-1 degrees of freedom the H_0 that all item difficulties (δ_i) are equal is rejected.

The test was initially applied to each set of four item difficulties to determine if there existed at least one significant difference. That is, the null hypothesis:

$$H_0 : \delta_1 = \delta_2 = \delta_3 = \delta_4$$

was tested for each component. As indicated in Table 5.6 there was at least one pair of delta values significantly different for each component and thus the null hypothesis was rejected. Details of Q values are reported in Appendix 5.2. Subsequently, a pair-wise investigation was conducted to identify which pairs were significantly different. The developmental model predicted that the stimulus items would be in three groups since Stimulus Item 2 and 4 were seen to be of similar complexity. Table 5.7 identifies the differences between pairs that were predicted by the model to be significant.

Table 5.7

Predicted Pair-wise Differences Between Components

Component A (Next term in the sequence)	A1	*	(A2, A4)	*	A3
Component B (Natural language)	B1	*	(B2, B4)	*	B3
Component C (Uncountable example)	C1	*	(C2, C4)	*	C3
Component D (Symbolic notation)	D1	*	(D2, D4)	*	D3
* Indicates predicted significant differences between adjacent pairs					

To avoid the risk of a high type 1 error rate a very conservative α value needed to be set. Using the relation:

$$\alpha_{FW} = C\alpha_{PW}$$

which relates the family-wise error rate to the number of comparisons and the pair-wise error rate, an alpha value of 0.005 would seem adequate to achieve a pair-wise error rate of 0.05 with 12 comparisons.

Once again Q values are reported in Appendix 5.2. Table 5.8 indicates which pairs of components are significantly different.

Table 5.8

Significant Differences Between Components

Component A (Next term in the sequence)	A1	*	A2	*	A4		A3
Component B (Natural language)	B1	*	B2	*	B4		B3
Component C (Uncountable example)	C1	*	C2	*	C4	*	C3
Component D (Symbolic notation)	D1	*	D2		D4	*	D3
* Indicates that the Q value for adjacent pair was significant ($p<0.005$)							

Only Component D followed the predictions of the developmental model exactly, with the appropriate order of stimulus items and the lack of a significant difference between Stimulus Items 2 and 4. The existence of a significant difference between A2 and A4 could be due to the attrition effect reported during the discussion on frequencies in the previous chapter. There was, of course, a significant difference between A2 and A3 as the model predicted and between B2 and B3. The analysis of Q values failed to support the existence of a significant difference between B3 and B4. This was particularly puzzling as there was significant differences between all four stimulus items on Component C, which required the application of the rule developed in Component B. Component C reflected the predicted order of difficulty; however, it included a significant difference between C2 and C4. These non-hypothesised differences could once again be attributed to attrition.

The developmental model successfully predicted the order of difficulty in all components across the four stimulus items. Two pairs of predicted difference were not supported by the data, namely (B4 and B3) and (A4 and A3). Additionally, three pairs were found to be significantly different that were not predicted by the model. These pairs were (A2 and A4), (B2 and B4) and (C2 and C4).

SUMMARY

The SOLO Taxonomy has been used in this chapter to assist in the interpretation of the response categories described in Chapter 4. A limitation of the useability of the Taxonomy was initially found to be the number of data dimensions that could be included in a single set of taxonomic categories. As a result a two-dimensional model was identified with each dimension reflecting the properties of unistructural, multistructural and relational levels of the concrete symbolic mode. However, while all the response categories could not be allocated to a unique set of SOLO levels, it was the process of reflecting the data onto the SOLO Taxonomy that facilitated the identification of a repeating pattern within categories. This in turn led to the second developmental dimension being identified.

This use of the SOLO Taxonomy in model building, described in this chapter, is different from most other applications of SOLO reported in Chapter 2. While the difference is noted, it should not be seen a contradiction of the Taxonomy as a tool of analysis. The more traditional applications of SOLO are analogous to holding one dimension of the developmental model constant and coding responses on the other dimension only. Such a procedure, while resulting in a more traditional application of the Taxonomy, would restrict the variability of response that is provided for.

The answer to research question 2.1 is that:

> The categories of pattern description using natural language do exhibit properties that would enable mapping them onto the modes and levels of the SOLO Taxonomy.

However, the mapping process was not as straightforward as was anticipated (see Appendix 5.3). It should not be inferred that there are no other dimensions influencing children's responses other than the two identified here; rather, it was that other dimensions could not be identified within the context of this study, using the coding systems described earlier. Indeed, it may be that the divergence between the predicted sequence of component difficulty and the Rasch sequence of component difficulty could be due to these other unidentified dimensions. These other dimensions could well include the attrition rate referred to above and the classroom conditions under which the instrument was administered. These matters can be better controlled for in the longitudinal study described in Chapters 7 and 8.

The response to research question 2.2 is that:

> A hierarchy of growth can be postulated from the response categories of pattern description using natural language.

In fact, the development model identified multiple pathways of growth. Since the major focus of this study is on expressing generality, it is the categories of the expressing generality dimension that are the subject of further investigation. The developmental pathway that has been identified is:

$$\boxed{\textbf{IA}} \longrightarrow \boxed{\textbf{1EG}} \longrightarrow \boxed{\textbf{SUCC}} \longrightarrow \boxed{\textbf{FUNC}}$$

The SOLO Taxonomy was used to justify this as a hierarchical sequence. Further testing of this postulated hierarchy needs to be undertaken. This process begins in the next chapter when the Component B categories are compared with the school year of the respondents. The investigation continues in the longitudinal study where the opportunity is taken to investigate how children's responses change over time.

Research question 2.3 needs to be answered in two parts. Item complexity, defined as the number of arithmetic operations in a functional description of a number pattern, does influence item difficulty. The children taking part in this study found:

> describing a single operation pattern easier than describing a two operation pattern.

The second part of research question 2.3 is concerned with the context of the stimulus item. In general:

> The children found the in-context stimulus items easier than the out-of-context stimulus item and this difference was reflected in the responses to all four Components A, B, C and D.

Two of the differences in difficulty (B4 & B3 and A4 & A3) were not found to be significant at the conservative alpha value of 0.005. However, the cause of this difference is not entirely due to children being more confident with the physical characteristics of the matchsticks and wheels of Stimulus Items 2 and 4. The two-operation questions were presented in two forms, namely, a physical context (Stimulus Items 2 and 4) and as decontextualised number pairs (Stimulus Item 3). These two question types were initially considered to have the same mathematical structure since they both could be described in the form $y=mx+b$. However, on studying the children's responses carefully, it was realised that the different contexts of the stimulus items changed the availability of the data elements. This phenomenon has been characterised, in SOLO terms, as multistructural or relational data sets. By this it is meant that for the in-context stimulus items (2 and 4) the children could develop the appropriate data set sequentially without the requirement of having to see relationships between the data. In the out-of-context stimulus item, (3) this sequential strategy did not lead to the identification of an appropriate data set. For this item the children had to develop a relational overview of the data. Hence, it cannot be said that the variability of item difficulty between contexts is solely due to children being more confident with one context than another. An additional factor is that as the contexts were changed so was the accessibility of the data elements.

The analysis in this chapter has focused on the research questions of the second research theme of using the SOLO Taxonomy to investigate the children's responses. The next chapter considers the interaction between the various components of the study together with the children's school year.

Chapter 6

RELATIONSHIPS BETWEEN COMPONENTS

Introduction

The previous chapter investigated the children's responses to the research instrument in the context of the SOLO taxonomy. The order of difficulty and item scaling were examined using correlation techniques and the Rasch scaling procedure. However, those analyses did not bear upon the research questions concerning associations between components of the stimulus items that were outlined in Chapter 3 as the third theme of this study.

Theme 3
Are the natural language responses related to the other response types such as use of symbolic notation and calculating an uncountable example, or to the child's school year?

> Question 3.1
> Is there an association between the categories of pattern description using natural language and the school year of the respondent?
>
> Question 3.2
> Is there an association between the categories of pattern description using natural language and the categories of symbolic language used?
>
> Question 3.3
> Is there an association between the categories of pattern description using natural language and being able to apply the rule to a large (uncountable) value of the independent variable?

These research questions are the focus of this chapter and are discussed in two sections. The first section provides an overview of the modelling process, the techniques of model selection, and the methods used to systematically investigate the relationships between the components of the study. The second section considers the relationship between Component B and the school year of the respondent, Component B and Component C, and finally Component B and Component D.

LOGLINEAR MODELLING

An Overview

A systematic way of investigating a set of polychotomous variables, where the data are measured on nominal or ordinal scales, is provided by multi-way frequency procedures that employ the likelihood ratio statistic, G^2, which is distributed as χ^2 (Everitt 1977). These procedures are commonly applied to contingency table data in the form of loglinear models. For each cell in these tables the natural logarithm of the ratio of the observed frequency to the expected frequency is obtained and multiplied by the observed frequency. The sum of these values is computed over all cells and then doubled to produce the likelihood ratio statistic. Hence,

$$G^2 = 2\sum f_{observed} \ln(f_{observed} / f_{expected}).$$

Likelihood ratio statistics can be derived for all main (first order) and interaction (second order and higher order) effects of interest.

In general, two kinds of loglinear analysis can be applied to contingency table data (Haberman 1978) - hierarchical and non-hierarchical. In the hierarchical form, as in multiple regression analysis, if a variable is in the model as a kth order effect it must be in the model as a (k-1) order effect. In the non-hierarchical form only those main and interaction effects of interest to the hypothesis, and for which a well fitting model can be found, are taken into the analysis. Usually, the hierarchical form is used if the research questions are concerned primarily with measuring the strength of association between categories of the variables, and the non-hierarchical form is employed when the focus of interest is on developing a predictive model. In both cases a search is made for the most parsimonious model that produces a set of expected frequencies which closely match the observed counts (Everitt 1977). A parsimonious model is one that employs the fewest main and interaction effects. In this study, the principal concern was with measuring the strength of association between pairs of components and so an hierarchical approach was adopted.

Three important assumptions of loglinear analysis are that:
1. The data is obtained from a randomly selected sample (Gilbert 1981),
2. The sample size should be sufficient to yield five times the number of subjects as there are cells in the contingency table; and
3. That no more than 20% of the cells of the effects included in the model should have an expected frequency of less than five, under an equi-probability assumption (Norusis 1990).

The issues underlying the first assumption were addressed in Chapter 3 when the sampling procedures were described. It is assumed that the procedures used were adequate to ensure the sample is representative of N.S.W. The inclusion of Components B, C, D and the school year (**Class**) in the model gives (5x3x5x4) 300 cells, thus violating both assumptions 2 and 3.

To resolve these problems the technique of collapsing categories (Kerlinger 1986) was used with Components B, C and D. The no attempt (**NA**) and inappropriate (**IA**) categories of Component B were collapsed since both categories reflect an inability or unwillingness to interact with the question. This combined group was coded as **IA**. The three categories of Component C were reduced to two by combining the no attempt (**NA**) category with the incorrect (**IC**) category. The new classification was coded as **NA/IC**. The five categories of Component D were reduced to four by combining the categories arbitrary use of letters (**AL**) and repeat notation (**REPT**) categories and were named as **REPT**. This was rationalised by perceiving both groups as transition groups between arithmetic notation and algebraic notation.

Table 6.1 presents data to demonstrate how the second of the assumptions, relating the ratio of sample size to the number of cells in the model, was satisfied after these structural decisions were made. The ratio of subjects to cells is 11.21, thus comfortably exceeding 5.

Table 6.1

Cell Frequency Analysis: Ratio of Sample Size to Number of Cells

Stimulus Items 1, 2, 3 and 4	**Number of cells (C)**	**Number of Subjects (S)**	**S/C**
B by C by D by Class	4x2x4x4=128	1435	11.21

The third assumption regarding the expected frequencies of the cells was also assisted by the collapsing of categories. Appendix 6.1 reports the expected frequencies of the cells of the terms included in the models for each stimulus item. In none of the four models was the limit of 20% of low frequency cells exceeded.

With the data assumptions satisfied the procedures for choosing the parsimonious model are discussed.

Choosing a Model

The four variables being considered have the potential to generate a model of considerable complexity. Figure 6.1 shows the fully saturated model of 15 main, two-way, three-way and four-way effects. In the process of choosing a satisfactory model to represent the data it was necessary to decide which of the 15 effects to include in the model for each stimulus item. All these terms are represented in Figure 6.1. To achieve this a search was undertaken for the model, with the fewest terms, that provides a goodness of fit statistic that is significant at the .05 level (Gilbert 1981). The goodness of fit statistic examines how closely the observed frequencies fit the expected frequencies. Using the maximum likelihood estimate for the goodness of fit (G^2) statistic the following hypothesis is tested.

> H: The model is a satisfactory approximation for the data.

Hence if $p<0.05$ the hypothesis is rejected and the model is considered to be ill fitting. Thus a G^2 statistic with a $p>0.05$ is required if the model is to be considered a satisfactory approximation of the data.

However, Gilbert warns that with large samples

> some relationships which are statistically significant may be of little practical significance.
>
> (p. 89)

The process to overcome this problem involved comparing the fit of the model including the effect in question with the fit of the model excluding the effect, and hence determining the difference in fit due to the effect. However, before applying this principle two issues had to be resolved. These were:

1. What pair of models are to be compared?
2. How is the difference in the fit to be assessed?

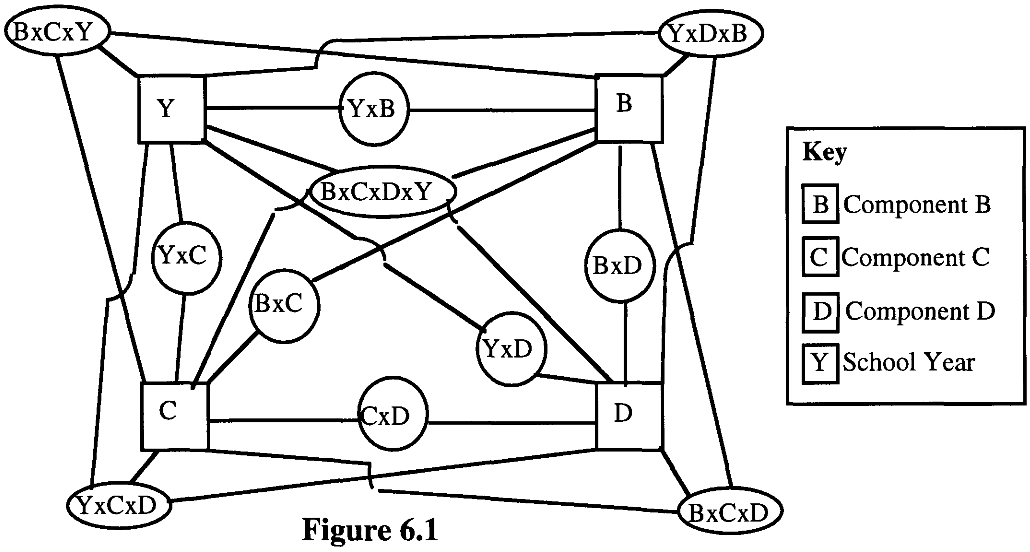


Figure 6.1
Saturated Model: All Main and Interaction Effects

To resolve the first of these issues it was decided to use a stepdown procedure (Norusis 1990) beginning with a saturated model generated by the four-way effect (BxCxDxClass) for each stimulus item. All possible effects are taken into the model initially. The four-way and three-way effects were removed first, with the partial associations for the saturated model providing the order in which the effects were taken from the model. The effects of least significance were taken from the model initially. Appendix 6.2 contains a summary of the partial association G^2 values for all main and interaction effects produced by the model. The order in which the effects were considered is indicated in Tables 6.2 to 6.5. The second question of assessing the difference in fit involved the use of the change in G^2 value and the resulting change in the degrees of freedom between successive models. These were used to test the hypothesis that the effect being investigated was zero. If the observed level of significance for the change was small, the hypothesis that the effect under investigation was zero was rejected. A significance level of $p<0.05$ was accepted as a suitable standard for this comparison.

In addition to the above procedures, a second test was applied to determine the suitability of the resulting model. This test involved ensuring that the fit was adequate. The advice of Gilbert (1981) was accepted in setting the significance level at $p>0.05$ as an indication that the data fits the model to an adequate degree. Tables 6.2 to 6.5 indicate the order in which the effects were deleted from the model and the consequent changes in G^2 values and degrees of freedom. The model that was accepted for each stimulus item can now identified.

Table 6.2

Stepdown Model Selection for Stimulus Item 1

Model	G^2	df	Prob	Change in G^2	Change in df
All effects	0	0	1.000		
BxCxDxClass	11.41	10	0.327	11.41(ns)	10
BxCxD	14.86	17	0.606	3.45(ns)	7
CxDxClass	23.38	29	0.759	8.52(ns)	12
BxCxClass	36.87	46	0.661	13.49(ns)	17
BxDxClass	86.2	81	0.268	49.3(ns)	35
CxClass	99.41	84	0.120	13.21	3
(ns) Not significant at 0.05					

Table 6.3

Stepdown Model Selection for Stimulus Item 2

Model	G^2	df	Prob	Change in G^2	Change in df
All effects	0	0	1.000		
BxCxDxClass	15.14	27	0.967	15.14(ns)	27
CxDxClass	27.12	36	0.857	11.98(ns)	9
BxCxD	39.99	45	0.684	12.87(ns)	9
BxCxClass	50.07	54	0.701	10.08(ns)	9
BxDxClass	90.81	81	0.214	40.74(ns)	27
CxClass	90.97	84	0.283	0.16(ns)	3
(ns) Not significant at 0.05					

Table 6.4

Stepdown Model Selection for Stimulus Item 3

Model	G^2	df	Prob	Change in G^2	Change in df
All effects	0	0	1.000		
BxCxDxClass	20.56	12	0.107	20.56(ns)	12
CxDxClass	31.11	21	0.103	10.55(ns)	9
BxDxClass	62.67	52	0.169	31.56(ns)	31
BxCxD	78.38	68	0.185	15.71(ns)	16
BxCxClass	93	77	0.095	14.62(ns)	9
CxClass	99.63	84	0.117	6.63(ns)	7
(ns) Not significant at 0.05					

Table 6.5

Stepdown Model Selection for Stimulus Item 4

Model	G^2	df	Prob	Change in G^2	Change in df
All effects	0	0	1.000		
BxCxDxClass	16.79	20	0.801	16.79(ns)	20
CxDxClass	24.08	32	0.842	7.29(ns)	12
BxCxClass	32.68	42	0.844	8.60(ns)	10
BxDxClass	67.67	70	0.559	34.99(ns)	28
BxCxD	81.21	81	0.473	13.54(ns)	11
CxClass	83.39	84	0.498	2.18(ns)	3
(ns) Not significant at 0.05					

Initial exploration for a well-fitting model was conducted with Stimulus Item 4. As a check on reliability of the model the initial exploratory analysis was conducted on a random sample of half of the response set. The sample was chosen using the sampling procedures of SPSS-X (Norusis, 1990). When a satisfactory model was identified it was confirmed using all 1435 responses.

From Table 6.5 it can be seen that a well-fitting model was produced for Stimulus Item 4 that excluded the 4-way effect, all the 3-way effects and the C by CLASS effect. The resulting model for Stimulus Item 4 had a goodness of fit statistic of 83.39 and was significant at the p>0.05 level. A diagrammatic view of this parsimonious model is presented in Figure 6.2.

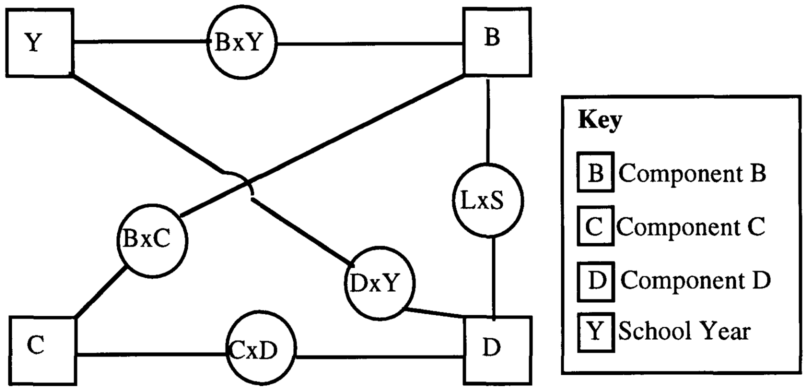


Figure 6.2

Parsimonious Model: Main Effects and Significant Interaction Effects

The use of the stepdown procedures for Stimulus Items 2 and 3 resulted in similar models being deemed well-fitting using the criteria established above. A difference between the three stimulus items was the order in which the effects were removed from the model. These differences can be determined from Tables 6.2 to 6.5. Table 6.3 shows that Stimulus Item 2 had a goodness of fit value of 90.97, significant at p>0.05, while Table 6.4 shows that the goodness of fit statistic for Stimulus Item 3 was 99.63 which was also significant at the p>0.05 level.

Attempts to apply the same stepdown procedures to Stimulus Item 1 resulted in a slightly different model being accepted. The 4-way and 3-way effects were removed; however, the attempt to remove the C by CLASS effect was unsuccessful since the change in the goodness of fit statistic brought about by its removal was found to be significantly different from zero. The partial associations for Stimulus Item 1 indicate that the B by CLASS effect should have been removed from the model prior to C by CLASS, but, it was decided not to do this for two reasons. The first was to maintain as much consistency as possible in the models across all stimulus items, and the second was that the B by CLASS effect was the subject of a research question in this study. Table 6.2 shows that Stimulus Item 1 had a goodness of fit value of 86.20, significant at p>0.05.

It will be recalled that one of the assumptions underlying the use of the parameters in loglinear analysis is that no more than 20% of cells in each effect taken into the model should have an expected frequency of less than five. The expected frequencies of all effects being considered here are reported in Appendix 6.1. Stimulus Item 1 had the highest number of cells with an expected value of less than five. However these cells still represented less than 5% of the total and hence were well within the 20% limit. Consequently all data assumptions have been satisfied.

The final part of this section discusses the methods of investigating the association between components.

Methods of Investigating Associations

From the loglinear models produced, two components of data were used to investigate the research questions. These components are partial associations and parameter estimates. Before proceeding to this investigation a brief description of these data components is given.

Partial association values were used to determine the order in which to exclude effects from the model. The values of the partial associations for the term included in the model are reported in Table 6.6.

Table 6.6

Partial Association Chi-Square Values for the Four Stimulus Items

Effect	Stimulus Item 1	Stimulus Item 2	Stimulus Item 3	Stimulus Item 14
B	815.50*	101.38*	658.99*	45.64*
C	1227.93*	399.48*	12.72*	77.98*
D	91.76*	119.23*	342.44*	136.94*
Class	38.44*	38.44*	38.44*	38.44*
BxC	42.68*	60.56*	73.14*	67.86*
BxD	461.44*	613.00*	664.42*	683.28*
BxClass	21.06*	31.52*	18.80*	54.77*
CxD	22.67*	45.86*	129.86*	98.81*
CxClass	11.00*	.04	1.51	2.19
DxClass	432.90*	297.32*	286.38*	200.53*
*Significant at p<0.05				

The likelihood ratio chi-square for the B, C, D and Class main effects indicates only that there are not an equal number of cases in each cell. The significant α values for the interaction terms mean that at least one of the set of possible interactions between the categories of the variables involved in the interaction are significant at that α value. The parameter estimates provide a method of further investigating the nature and level of significance of the interactions. (A suitable α value for the investigation of interactions between categories of variables is discussed below.)

The loglinear procedure of SPSS (Norusis 1990) produces parameter estimates (Lambda coefficients) for each cell. The parameter estimates provide a means of assessing the importance of the effects in determining cell frequency (Everitt 1977). These lambda coefficients can be expressed as standard scores (γ / se) and used to measure statistically the strength and direction of association between categories of components. For example, if a criterion of $p<0.05$ was assumed, the table of γ / se values would be scanned for $|z|$ values exceeding 1.96; such values would indicate a significant association between categories of the components represented by the rows and columns of the table. In this context, a positive association between categories means that a subject scoring high on one component tends to score high on the other component, and a negative association means that a subject scoring high on one component tends to score low on the other.

Before examining the γ / se values in more detail, it is necessary to consider an appropriate α level for the full set of tests. Because the tests are non-independent, a very conservative α level needs to be set to guard against a rapidly escalating type 1 error rate. A suitable pair-wise error rate can be calculated using the following formula which relates the family-wise error rate (α_{FW}) to the number of comparisons (C) and the pair-wise error rate (α_{PW}).

$$\alpha_{FW} = C\alpha_{PW}$$

Thus to achieve a family-wise error rate of 0.05 a pair-wise α of 0.0005 (Z=3.291) would seem adequate protection.

With the processes of loglinear analysis described and suitable confidence limits set, a systematic consideration of the associations between the components can be undertaken.

ANALYSIS OF ASSOCIATIONS BETWEEN COMPONENTS

The research questions outlined in the introduction to this chapter included several that require investigation into the nature of associations between components of the stimulus items. In particular the following research questions were asked:

> Question 3.1
> Is there an association between the categories of pattern description using natural language and the school year of the respondent?

This question required the consideration of the relationship between Component B and Class for each stimulus item.

> Question 3.2
> Is there an association between the categories of pattern description using natural language and the categories of symbolic language used?

This required a comparison of Components B and D.

> Question 3.3
> Is there an association between the categories of pattern description using natural language and being able to apply the rule to a large (uncountable) value of the independent variable?

An investigation into the association between Components B and C would help shed light on this question.

These relationships are considered in turn in the three parts to this section. To facilitate this analysis the frequency data for each pair of interest is presented as a contingency table and graphically. In addition, the parameter estimates and their associated Z scores are reported. To enhance readability, the main text includes a complete set of data for Stimulus Item 4 and a graphical representation only of the data for the other three stimulus items. The contingency table, parameter estimates and Z scores for Stimulus Items 1, 2 and 3 are included as Appendix 6.3.

Component B by CLASS

To investigate the research question:

> Is there an association between the categories of pattern description using natural language and the school year of the respondent?

the Component B by CLASS associations are considered. From the data presented in Table 6.6 it can be seen that the null hypothesis

> That Component B and Class are independent

is rejected due to the partial associations for B by CLASS being significant ($p<0.05$, $df=9$) for each of the four stimulus items. Therefore there is a significant association between Component B and the school year of the respondent. A more detailed picture of this relationship can be gained by considering the contingency table and the associated parameter estimates. This information for Stimulus Item 4 is presented as Tables 6.7 and 6.8, and Figure 6.3.

Table 6.7

Contingency Table Comparing B by CLASS for Stimulus Item 4

COMPONENT B	**CLASS**				
	YR5	YR6	YR7	YR8	Totals:
IA	113	116	110	116	455
1EG	116	92	66	32	306
SUCC	28	39	117	110	294
FUNC	38	59	117	166	380
Totals:	295	306	410	424	1435
Partial association=54.77				p<0.0005	

The frequency data of Table 6.7 is presented in graphical form in Figure 6.3. The change in modal category between school years is modelled by the saddle shape of the responses surface.

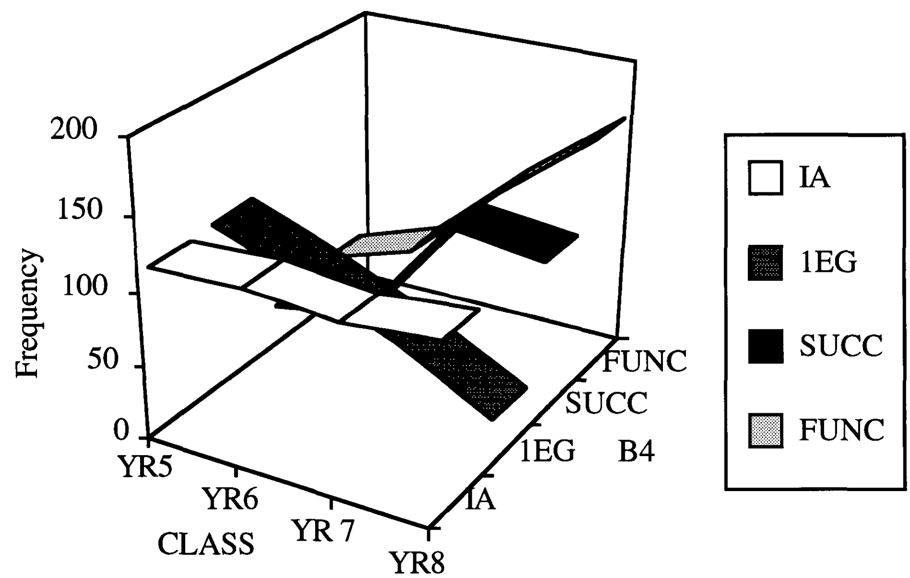


Figure 6.3

Response Surface of B by CLASS for Stimulus Item 4

The high and low points of the response surface are reflected by the positive and negative parameter estimates respectively (Table 6.8).

Table 6.8

Parameter Estimates of B by CLASS for Stimulus Item 4

COMPONENT	CLASS			
B	YR5	YR6	YR7	YR8
IA	0.196	0.054	-0.212	-0.037
1EG	0.454*	0.231	-0.058	-0.627*
SUCC	-0.424	-0.249	0.312*	0.360*
FUNC	-0.225	-0.036	-0.042	0.304*
* Significant at p<.0005				

The associated Z scores for the parameter estimates for each pair of categories are presented in Table 6.9.

Table 6.9

Z Values of B by Class for Stimulus Item 4

COMPONENT	CLASS			
B	YR5	YR6	YR7	YR8
IA	1.889	0.549	-2.237	-0.377
1EG	4.743	2.465	-0.602	-5.347
SUCC	-3.185	-2.114	3.314	3.653
FUNC	-1.786	-0.327	-0.412	5.225

Consideration of the parameter estimates indicates some general trends worth noting. There are positive associations between one example (**1EG**) and Year 5 (**YR5**) categories, with the association becoming weaker as the children move to higher classes. However, responses at this level continue to appear in the highest classes. Conversely, there is a negative association between Years 5 and 6 categories and the successive and function categories. This relationship also reverses in Year 7 and 8. These intuitive impressions are supported by evidence presented in the form of parameter estimates and their associated Z scores.

The parameter estimates for Stimulus Item 4 show that the following pairs were positively associated at a significance level of $p<0.0005$.

(1EG, YR5) (SUCC, YR7) (SUCC, YR8) (FUNC, YR8)

Only one negatively associated pair was found to be significant at the $p<.0005$ level. The pair was:

(1EG, YR8)

The response surface for Stimulus Item 2 (Figure 6.4) provided a similar pattern to that of Stimulus Item 4; however, the pattern of relationships was not as strong, and hence few were significant at the 0.0005 level.

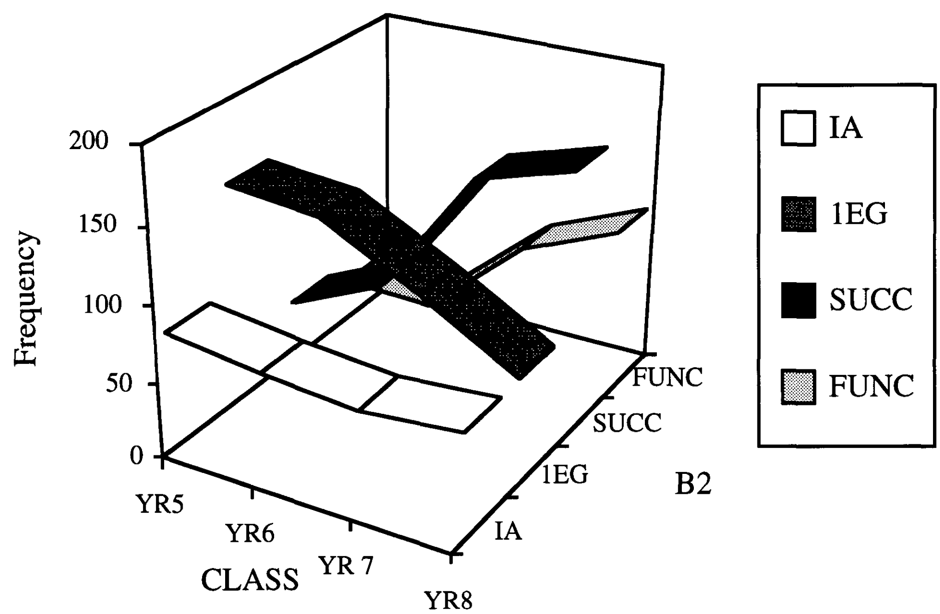


Figure 6.4
Response Surface of B by CLASS for Stimulus Item 2

The response surfaces for Stimulus Items 1 (Figure 6.5) and 3 (Figure 6.6) appear to be much flatter, indicating that there is less variability due to Class. While the partial associations indicate a significant association at the 0.05 level, the parameter estimates indicate no significant associations at the more conservative 0.0005 level. The general trends identified with Stimulus Items 2 and 4 appear, in that the lower Component B responses are positively associated with the lower school years and the higher Component B responses are positively associated with the higher school years.

However, the lack of successive (**SUCC**) responses in Stimulus Item 1 and one example (**1EG**) responses for Stimulus Item 3 make significant differences to the overall response surface. It is worth noting that the lack of responses in these categories is consistent across all school years. There appears to be an item effect that induces a changed pattern of responses.

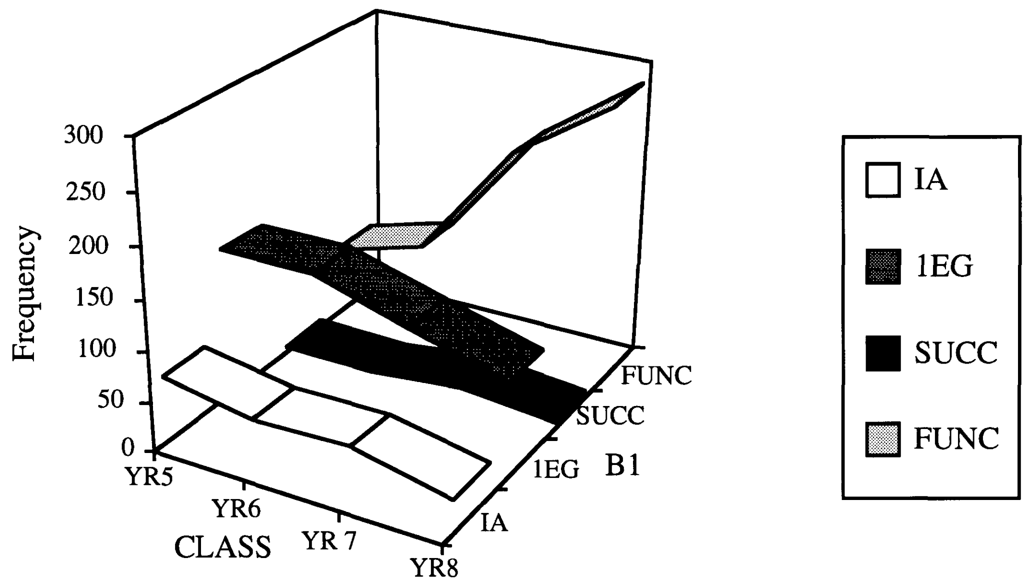


Figure 6.5
Response Surface of B by CLASS for Stimulus Item 1

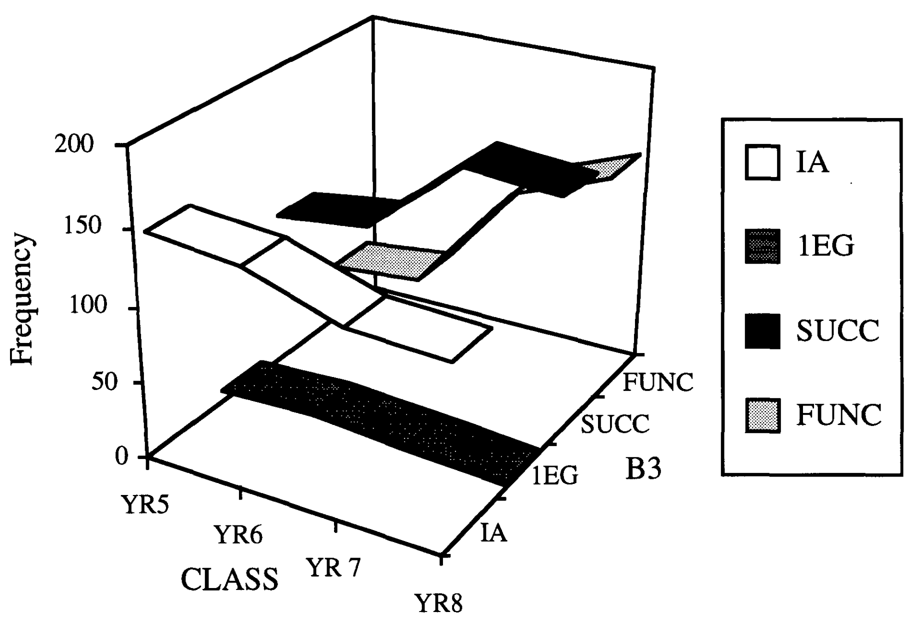


Figure 6.6
Response Surface of B by CLASS for Stimulus Item 3

In response to the research question 3.1:

> Is there an association between the categories of pattern description using natural language and the school year of the respondent?

It can be said that there is a general pattern of association between the categories of natural language descriptions and the respondents' school year, but the strength of that association is variable and seems to vary between stimulus item type.

There is evidence to suggest that there exists a development hierarchy from

$$\boxed{\text{IA}} \rightarrow \boxed{\text{1EG}} \rightarrow \boxed{\text{SUCC}} \rightarrow \boxed{\text{FUNC}}$$

with the one example (**1EG**) category being associated with younger children and the function category being associated with older children. The association between successive descriptions and Year 7 in Stimulus Item 4 suggests that the successive (**SUCC**) category might be a transition stage in expressing generality. Such a view is consistent with the earlier classification of the successive category representing the multistructural level in SOLO nomenclature. Whether or not all children pass through the stages cannot be determined from this study and the matter is investigated in the context of a longitudinal investigation discussed in Chapter 8.

Component B by Component D

To investigate the research question:

> Is there an association between the categories of pattern description using natural language and the categories of symbolic language used?

Component B by Component D associations were considered. From the data presented in Table 6.6 it can be seen that the null hypothesis

> that Component B and Component D are independent

is rejected due to the partial associations for B by D being significant ($p<0.05$, df=9) for each of the four stimulus items. Therefore there is a significant association between the responses to Component B and the Component D responses. A more detailed picture of this relationship can be gleaned by considering the contingency table and the associated parameter estimates. This information for Stimulus Item 4 is presented as Tables 6.10 and 6.11.

Table 6.10

Contingency Table Comparing B by D for Stimulus Item 4

COMPONENT B	COMPONENT D				
	NA	OS	REPT	ALG	Totals:
IA	374	41	26	14	455
1EG	93	149	56	8	306
SUCC	53	30	181	30	294
FUNC	37	59	59	225	380
Totals:	557	279	322	277	1435
Partial association=683.28					$p<0.05$

The frequency data is presented in graphical form in Figure 6.7. The pronounced saddle shape indicates the changing level of association between pairs of categories.

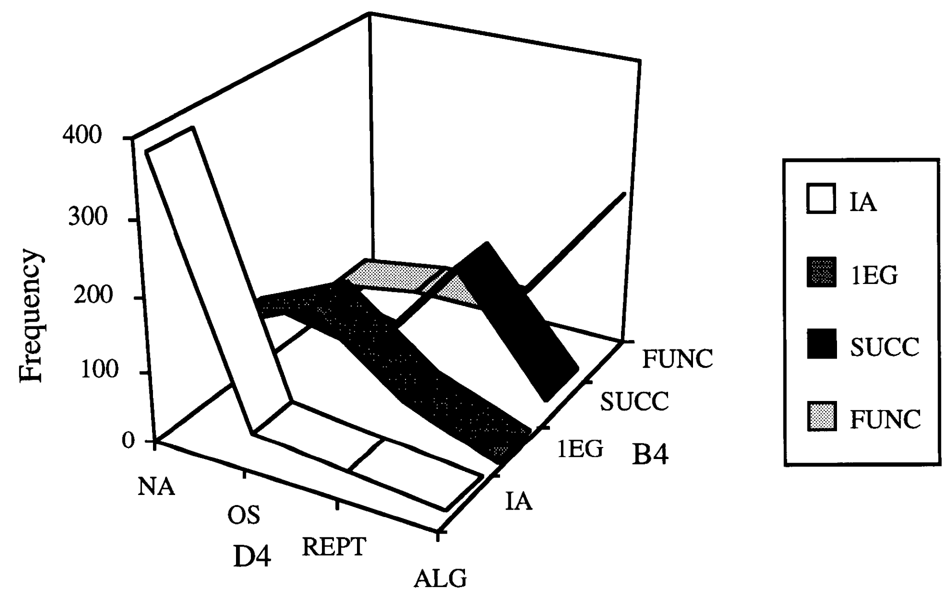


Figure 6.7

Response Surface of B by D for Stimulus Item 4

The strong positive associations indicated by the parameter estimates of Table 6.11 are represented by the high points on the ridge of the responses surface.

Table 6.11

Parameter Estimates of B by D for Stimulus Item 4

COMPONENT B	COMPONENT D			
	NA	OS	REPT	ALG
IA	1.358*	0.008	-0.416*	-0.950*
1EG	-0.287	0.829*	-0.384	-0.158
SUCC	-0.677*	0.089	0.993*	-0.405*
FUNC	-0.393	-0.926*	-0.192	1.512*
* Significant at p<.0005				

The associated Z scores for the parameter estimates are presented in Table 6.12 and indicate a number of pairs of association significant at the $p<.0005$ level.

Table 6.12

Z Values of B by D for Stimulus Item 4

COMPONENT B	COMPONENT D			
	NA	OS	REPT	ALG
PS	14.104	0.072	-3.748	-8.162
1EG	-2.193	7.322	-2.869	-1.353
SUCC	-4.952	0.744	10.555	-3.787
FUNC	-2.157	-4.257	-1.295	12.905

The high ridge on the response surface for Stimulus Item 4 reflects the significant positive associations between the following pairs of categories.

(**IA**, **NA**), (**1EG**, **OS**), (**SUCC**, **REPT**) and (**FUNC**, **ALG**).

Further, the only significant positive associations found among all four stimulus items lay on this ridge. This ridge represents the diagonal of the contingency tables going from top left to bottom right. A similar ridge is evident for Stimulus Item 2 (see Figure 6.8) reflecting a similar pattern of associations.

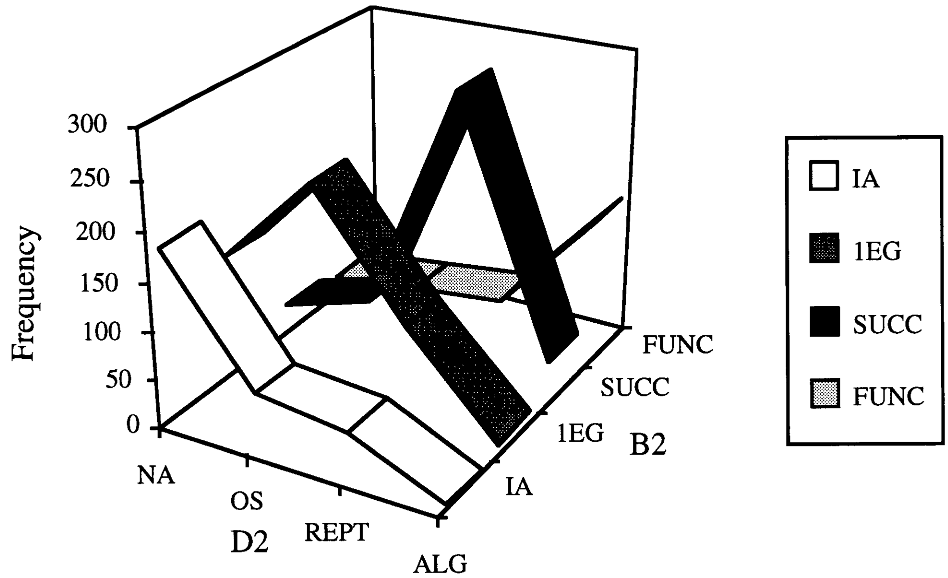


Figure 6.8
Response Surface of B by D for Stimulus Item 2

The response surfaces for Stimulus Items 1 (Figure 6.9) and 3 (Figure 6.10) differ from the saddle shape due to their low frequency counts in the successive (**SUCC**) and one example (**1EG**) categories respectively. However, with these exceptions, the pattern of significant associations remain the same.

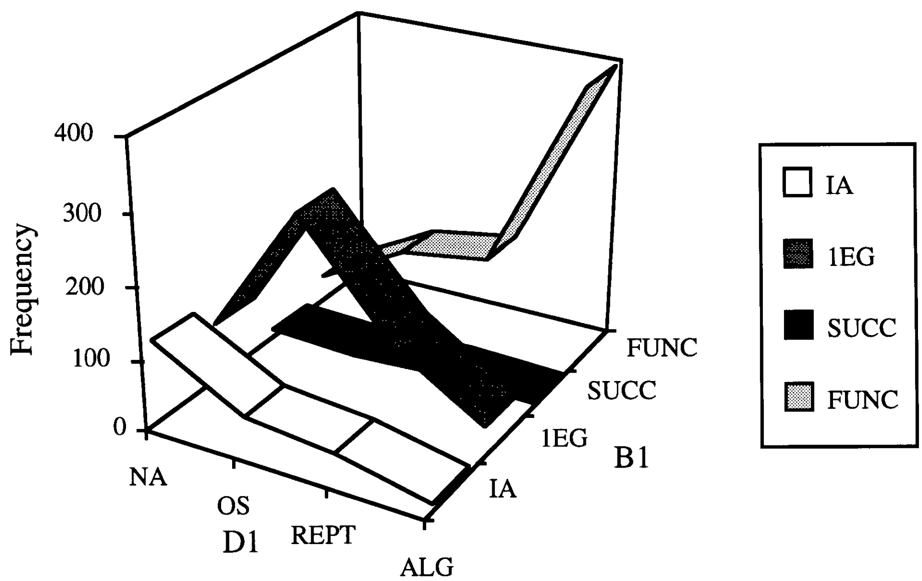


Figure 6.9
Response Surface of B by D for Stimulus Item 1

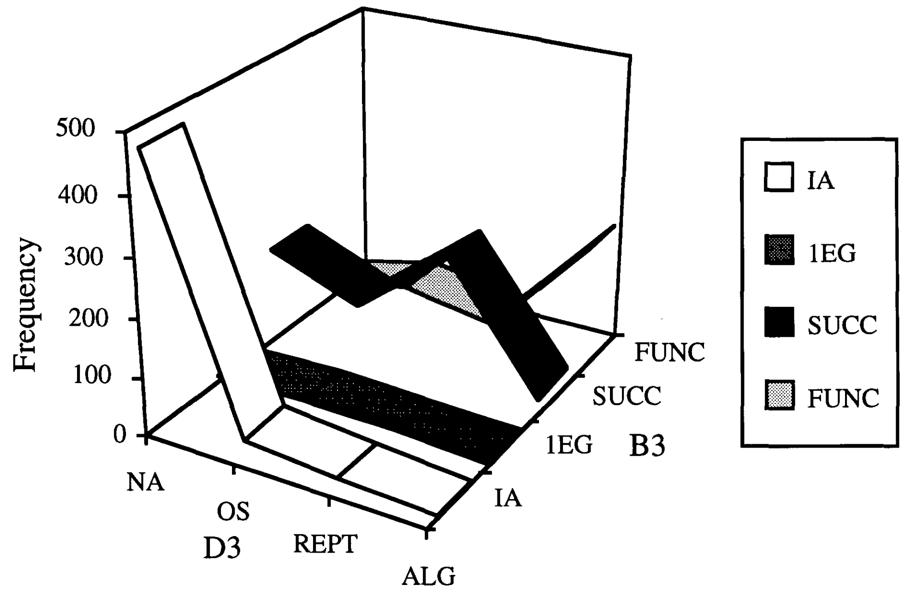


Figure 6.10
Response Surface of B by D for Stimulus Item 3

Conversely to the above positive associations, the significant negative associations all lie off the diagonal identified. In all four stimulus items the negative association between the pairs (**FUNC**, **NA**) and (**IA**, **ALG**) was significant. For Stimulus Items 1, 2 and 4 the negative associations between (**FUNC**, **OS**) were significant.

The direction and strength of the associations of categories between Component B and Component D indicate a consistent relationship between the natural language of pattern description and the symbolic notation to represent that language. However, there was a design issue that could be seen to interfere with the clarity of this relationship. It will be recalled that in designing the wording for Component D the problem of using the word 'Algebra' in the items for Years 5 and 6 was addressed. The matter was resolved by having two forms of the survey stimulus items. The word algebra was only to be used on the questions for Years 7 and 8. This was because these children had some experience of learning algebra in their normal class work. It seemed reasonable to pose the question of whether the associations identified above would be similar among just the Year 7 and 8 students who had had experience of learning algebra at school and who had been explicitly asked for an algebraic description with the instruction:

Now write your rule for part (b) in algebra symbols instead of words.

Previously, the B by D by Class effect was not included in the loglinear model, with the implication that there is a similar pattern of association for B by D among the Year 7 and 8 children as there is for the children in the whole sample. To investigate this a confirmatory loglinear analysis was performed on the data for Stimulus Item 4 that included only those respondents from Years 7 and 8. This subset of 834 respondents had a likelihood ratio statistic of 41.54 with 34 degrees of freedom with a probability value of 0.175; hence it satisfied the criteria for a well-fitting model described earlier.

The contingency table and corresponding response surface are presented in Table 6.13 and Figure 6.11. The partial chi-square value of 416.16, with 9 degrees of freedom for the contingency table, confirms the existence of a significant association between the variables ($p<0.0005$). The familiar saddle shape is once again in evidence with the exception of the one example (**1EG**) category having a low frequency among this older group of children.

Table 6.13

Contingency Table Comparing B by D for Stimulus Item 4 YR 7 and 8

COMPONENT B	**COMPONENT D**				
	NA	OS	REPT	ALG	Totals:
IA	187	6	20	13	226
1EG	38	14	38	8	98
SUCC	40	5	145	27	217
FUNC	24	22	42	195	283
Totals:	289	57	245	243	834
Partial association=416.16					p<0.0005

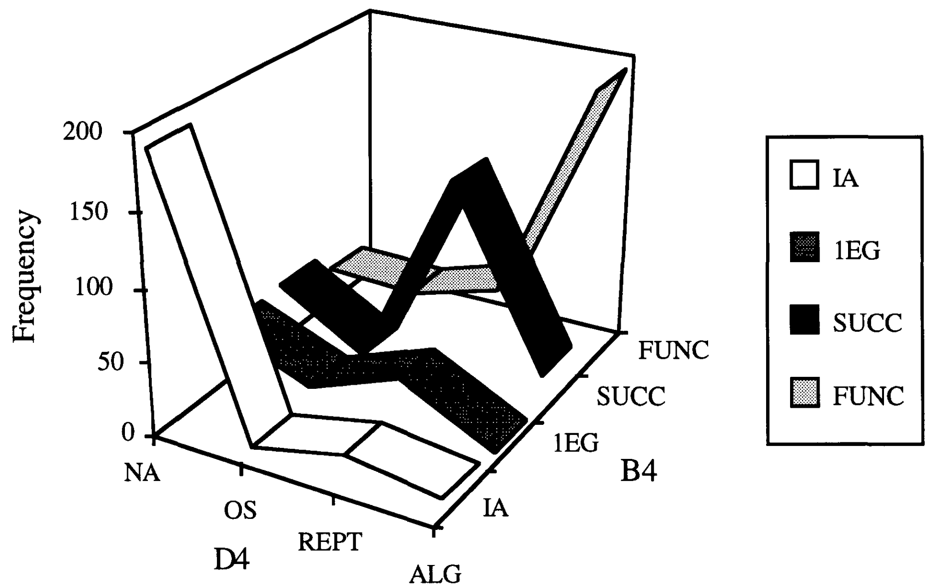


Figure 6.11
Response Surface of B by D for Stimulus Item 4 Yr 7 and 8

The parameter estimates and the corresponding Z values are reported in Tables 6.14 and 6.15. As with the discussion above, all the significant positive associations lie on the top left to bottom right diagonal. The significant pairs with positive associations are (**IA, NA**), (**SUCC, REPT**) and (**FUNC, ALG**). The (**1EG, OS**) pai,r with a parameter estimate of 0.047 and a Z value of 2.351, was significant at the p<0.01 level but not at the conservative 0.0005 level accepted as the necessary criterion of this study. There are also some significant negative associations. The pairs of categories (**IA, ALG**), (**SUCC, ALG**) and (**SUCC, NA**) were significant associations in common with the whole sample. In addition to these three the whole sample revealed significant negative associations between (**IA, REPT**) and (**FUNC, OS**). These two pairs of categories failed to achieve the 0.0005 significance level. However, they were significant at the 0.005 level.

Table 6.14
Parameter Estimates of B by D for Stimulus Item 4 YR 7 and 8

COMPONENT B	**COMPONENT D**			
	NA	OS	REPT	ALG
IA	1.349*	0.027	-0.411	-0.965*
1EG	-0.441	0.476	-0.128	0.094
SUCC	-0.567*	0.225	0.875*	-0.534*
FUNC	-0.341	-0.728	-0.335	1.405*
* Significant at p<.0005				

Table 6.15

Z Values of B by D for Stimulus Item 4 YR 7 and 8

COMPONENT	COMPONENT D			
B	NA	OS	REPT	ALG
IA	9.950	0.177	-3.059	-6.470
1EG	-1.706	2.351	-0.681	0.533
SUCC	-3.327	1.544	7.808	-5.322
FUNC	-1.685	-3.211	-2.117	10.951

It can be seen that there are very similar patterns of association between natural language and use of symbolic notation between the restricted sample of Year 7 and 8 children, who had a formal exposure to algebra in a classroom setting, and the whole sample. The differences that do exist reflect the higher level of pattern description of the older children.

The above discussion established a statistical association between natural language and symbolic language. The issue that arises from the identification of this association is whether the functional language description is a necessary precursor for the successful use of algebraic notation. More specifically, can some children use sophisticated symbolic language without using natural language of a similar level of sophistication? Within the context of this study the issue is: can children provide accurate algebraic descriptions of number patterns without functional relationships in natural language?

Across the four stimulus items approximately 80% of all algebraic responses were associated with a functional description in the natural language component. What of the other 20%? Had this group of children been forced to reconceptualise the pattern when asked to provide a symbolic response? Were they able to produce a symbolic response independently of their natural language? Is the association between natural language and algebraic symbolism a transitory one? It could be that the process is analogous to a person using a foreign language. Initially the person formulates a response to some stimulus in his or her native tongue and then translates it into the new language. As competence in the new language develops people formulate responses in the new language thus bypassing the native language. It seems plausible to suggest that beginning algebra students need to formulate descriptions of patterns and relationships in their natural language prior to translation into the symbolic language of algebra. As confidence and competence develop, the natural language can be bypassed and descriptions formulated directly into the symbolic language. The opportunity to pursue these and other issues is afforded by the more detailed investigation of the student interviews conducted in the longitudinal study.

Component B by Component C

To investigate the research question:

> Is there an association between the categories of pattern description using natural language and being able to apply the rule to a large (uncountable) value of the independent variable?

Component B by Component C associations were considered. From the data presented in Table 6.6 it can be seen that the null hypothesis

> that Component B and Component C are independent

is rejected due to the partial associations for Component B by Component C being significant ($p<0.05$, $df=3$) for each of the four stimulus items. Therefore there is a significant association between the responses to Component B and the Component C responses. A more detailed picture of this relationship can be seen by considering the contingency table and the associated parameter estimates. These frequency values for Stimulus Item 4 are reported in Table 6.16.

Table 6.16

Contingency Table Comparing B by C for Stimulus Item 4

COMPONENT B	**COMPONENT C**		
	NA/IC	CORRECT	Totals
IA	305	150	455
1EG	138	168	306
SUCC	71	223	294
FUNC	37	343	380
Totals:	551	884	1435
Partial association=67.86		p<0.05	

In addition, the frequency data is presented in graphical form in Figure 6.12.

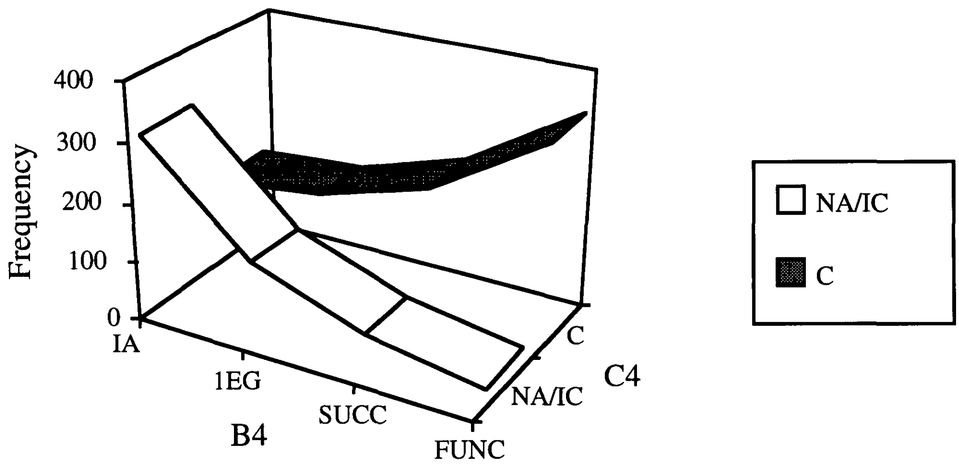


Figure 6.12
Response Surface of B by C for Stimulus Item 4

The direction and strength of the associated pairs can be determined from the parameter estimates for the pairs and their Z scores shown in Table 6.17.

Table 6.17
Parameter Estimates and Z Values B by C for Stimulus Item 4

	PARAMETERS		**Z VALUES**	
COMPONENT	**COMPONENT C**		**COMPONENT C**	
B	NA/IC	C	NA/IC	C
IA	0.393*	-0.393*	6.701	-6.701
1EG	0.222*	-0.222*	3.716	-3.716
SUCC	-0.147	0.147	-2.190	2.190
FUNC	-0.468*	0.468*	-5.913	5.913
* Significant at p<.0005				

In Stimulus Item 4 the successive (**SUCC**) and function (**FUNC**) categories of Component B were positively associated with the correct (**C**) category of Component C and negatively associated with the incorrect (**NA/IC**) category for all stimulus items. Conversely, the inappropriate (**IA**) and one example (**1EG**) categories of Component B were negatively associated with correct (**C**) responses to Component C. The Component B categories were all sequenced in the same order in terms of their association with the correct responses to Component C. The order, from weakest to strongest was:

IA → 1EG → SUCC → FUNC

Moreover, these patterns of association between the categories and indications of a hierarchical development were consistent across all stimulus items (see Appendix 6.3 for detailed information with regard to Stimulus Items 1, 2 and 3).

For all four stimulus items, the association between the categories of Component C and the inappropriate (**IA**) category of Component B was significant at the 0.0005 level. Additionally the Component C categories were significantly associated with the function (**FUNC**) category at the 0.0005 level for Stimulus Items 2, 3 and 4.

Once again variability in the associations between stimulus items was evident in some categories (see response surfaces in Figures 6.13, 6.14 and 6.15). On comparing the successive (**SUCC**) responses for Stimulus Items 2, 3 and 4 (Stimulus Item 1 was excluded due to its low number of responses in this category) it was noted that the number of correct (**C**) responses to Component C for Stimulus Items 2 and 4 (88% and 76%, respectively) was considerably higher than the number of correct responses for Stimulus Item 3 where only 51% were successful. Similar differences were noted between Stimulus Items 1, 2 and 4 for the one example (**1EG**) category of Component B. For Stimulus Item 1, 91% of the one example (**1EG**) respondents correctly answered Component C. This percentage dropped to 70% for Stimulus Item 2 and to 55 % for Stimulus Item 4. (Stimulus Item 3 was omitted from this discussion due to the low frequency of its one example (**1EG**) category). A much more even success rate for Component C was evident amongst the function (**FUNC**) respondents to Component B. This success rate ranged from 99% for Stimulus Item 1, through 94% and 90% for Stimulus Items 2 and 4, respectively, with Stimulus Item 3 having the lowest rate of 84%.

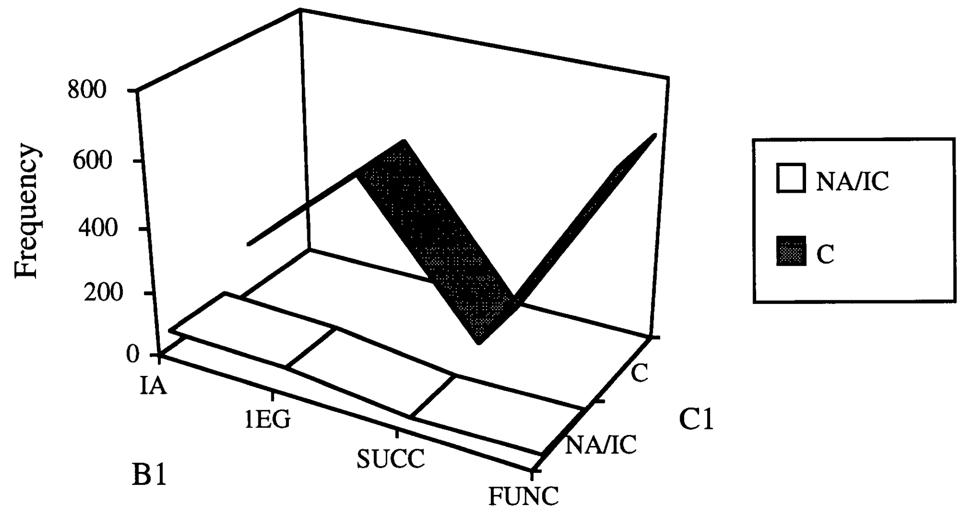


Figure 6.13
Response Surface of B by C for Stimulus Item 1

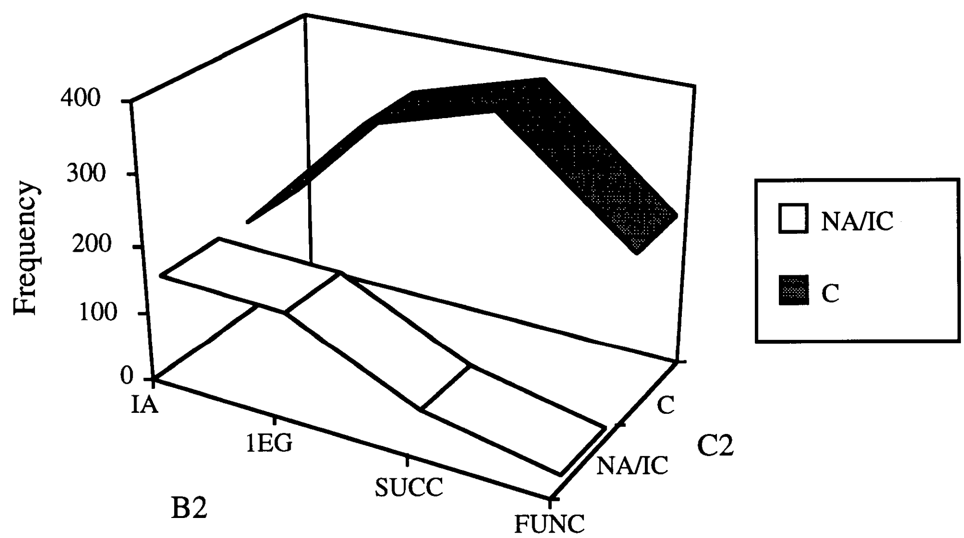


Figure 6.14
Response Surface of B by C for Stimulus Item 2

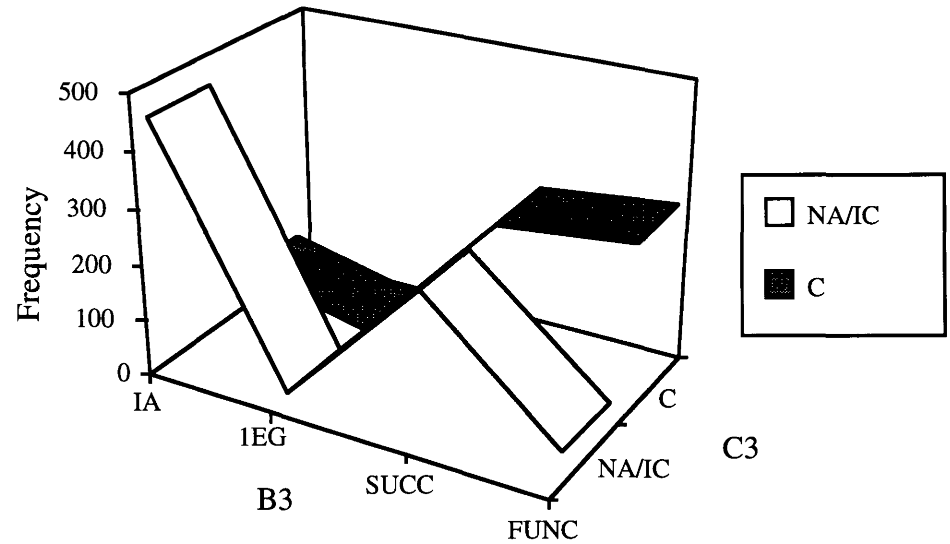


Figure 6.15
Response Surface of B by C for Stimulus Item 3

There are three conclusions that can be drawn from the data discussed in this section. The first is that children are more likely to respond correctly to an uncountable example of a number pattern if their language description is in the higher response groups of Component B. However, it is clear that a considerable number of children were able to successfully respond to Component C when they had provided a low level response to Component B.

Secondly, there is considerable variability among question types in the association between the categories of Components B and C. The reduced item complexity of Stimulus Item 1 enabled low level respondents to Component B to respond successfully to Component C. The more complex relational data set of Stimulus Item 3 could account for the lowest rate of correct responses among the successive (**SUCC**) category respondents. This variability may have implications for teachers who are trying to induce children to provide functional descriptions of number patterns by asking for uncountable examples to be evaluated. It would appear that some children can achieve the uncountable example without the functional expression of generality. However, this ability varies with question type.

Finally, the following question should be asked:

> How do children calculate the uncountable example when their language description seems inadequate?

Is it that an algorithm can exist without the language to describe it? Is it that the child is forced to reconceptualise the algorithm when faced with a question such as Component C? Does this imply that, initially, it is not sufficient to ask children to express generality on its own without generating a need for a functional description? These questions cannot be answered here but are among those investigated in the longitudinal study.

SUMMARY

The first section of this chapter described the loglinear modelling procedure used to identify the parsimonious model that adequately described the data. All four-way and three-way interaction terms were excluded from the model together with one two-way interaction term. The resulting main effects and two-way effects were used to generate the parameter estimates of association and the Z scores. Three pairs of association were particularly relevant to the research questions addressed in this chapter: Component B by CLASS, Component B by Component C, and Component B by Component D.

A significant association was identified between Component B and CLASS; however, the nature of that association varied with question type. There was evidence to support the proposition that a hierarchy of development in the Component B responses existed. The question of the necessity for children to pass through each developmental stage is further investigated in the longitudinal study.

The interaction of Component B and Component D responses provided strong evidence of an association between natural language and symbolic language. The data did not provide evidence of a cause and effect relationship between the components since a temporal difference in the appearance of the response sets was beyond the scope of the survey. However, this issue has also been identified as one appropriate for further investigation in the longitudinal study.

The final relationship discussed in this chapter was the association between Component B and Component C. A significant interaction between the two components was identified, with strong positive association between a correct calculation of an uncountable example and a functional natural language description of the number pattern. However, a significant number of children were identified who were able to correctly provide an answer to Component C after using a lower natural language description than a function. This and the other issues identified above that arise from the discussion of the relationships between the components of the study are addressed in the longitudinal study described in Chapter 8. The design of this component of the study is the subject of the next chapter.