USING NEMP TO INFORM THE TEACHING OF SCIENCE SKILLS

A REPORT TO THE NEMP BOARD

ROSEMARY HIPKINS AND NATASHA KENNEALLY

NEW ZEALAND COUNCIL FOR EDUCATIONAL RESEARCH
TE RŪNANGA O AOTEAROA MŌ TE RANGAHAU I TE MĀTAURANGA

NEMP
National Education Monitoring Project

MINISTRY OF EDUCATION
Te Tāhuhu o te Mātauranga

WELLINGTON
2003

New Zealand Council for Educational Research
P O Box 3237
Wellington
New Zealand

ISBN 1–877293–31–8

Distributed by NZCER Distribution Services
P O Box 3237
Wellington
New Zealand

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

APPENDICES

ACKNOWLEDGMENTS

A substantial proportion of this project was funded by the Unit for Studies of Educational Assessment, Education Department (USEE) at Canterbury University, via Ministry of Education funds set aside for researching aspects of the National Education Monitoring Project (NEMP) data. We are grateful to the NEMP staff in Dunedin who helped us obtain the sample of tapes and loaned us the specialist equipment with which to view them. Alison Gilmore, of the USEE team, provided valuable advice as the project was being shaped.

In this project we have attempted to draw together empirical data from several diverse fields. Expert advice is crucial in such a process. We are indebted to Associate Professor Miles Barker of The University of Waikato who challenged us to rethink aspects of the developmental framework we have proposed. As Prof. Barker noted, such an exercise is as much about saying what cannot be included/assumed as it is about shaping and organising findings that are relevant to the exercise. Dr. David Symington provided a valuable international perspective. He challenged us to address moving beyond fair testing, and the drawing of theory/evidence links, more definitively than we had done at the stage when he read the report. Section Seven substantially owes its existence to this challenge. Robyn Baker acted as a sounding board at various critical stages of the project, like Prof. Barker and Dr. Symington drawing on her deep expertise in science education. Rachel Bolstad provided insightful critique on the first draft and assisted with the focus group stage of the research.

Kristina Louis was our super-sleuth in the process of searching for relevant literature. Her skills saved us much time and effort. Edith Hodgen and her team provided data support for collating and using the empirical data generated by the NEMP tape observations. Lia Mapa helped with observations of the *Ball Bounce* task.

Finally, we thank the teachers in the three focus groups, who gave up after-school time to work with us, and generously shared their professional expertise. Two of the groups met in school staff rooms, and we thank those host schools for their hospitality.

EXECUTIVE SUMMARY

This research builds on an earlier NEMP probe study which found that many teachers do not appear to actively teach students the key objectives of the “Developing Scientific Skills and Attitudes” strand of *Science in the New Zealand Curriculum* (Gilmore, 2001). This project sought to develop rich descriptions of children’s investigative actions, and to analyse these findings in the light of research literature that describes children’s actual and potential investigative skills development. The analyses generated were then used to inform the design of teaching strategies that could help children to actively develop the skills specified in *SNZC.*

Two hundred video-taped episodes of Year 4 or Year 8 groups of children carrying out 3 different NEMP science investigation tasks were closely observed and subsequently analysed with reference to the review of the research literature. (The tasks were *Truck Track*, *Ball Bounce*, and *Emptying Rate*.) The strategies identified as a result of the analysis for the active teaching of investigative skills were tried out and critiqued by 24 primary teachers during after-school focus group sessions in 3 different areas.

We found that children’s actions are influenced in subtle ways by teachers’ instructions and dialogue. However there is a range of relatively simple measures that NEMP assessors, and indeed all primary teachers, can undertake to help children actively learn the skills of investigating “scientifically”. Key areas for making such modifications are summarised below.

The literature suggested that children can recognise “fair tests” before they are able to produce these independently. The NEMP observations showed that when children are asked to carry out pre-devised “investigations” this recognition may take the form of intuitive actions, carried out silently with no discussion during any stage of the investigation. Providing younger children with opportunities to select fair tests during assessment might allow them to better display their developing skills.

Children need to have many rich exploratory experiences from which to build a library of causal mechanisms. Only then can they draw on these to shape their own investigative questions, and/or explanations of the phenomena they explore. When presented with a prescribed task, children may perceive little meaning beyond task completion in the actions they carry out. Unless the context is familiar, children may struggle to recognise variables that need to be controlled, or to develop a considered causal theory that gives a sense of science meaning to their investigation. Providing opportunities to display planning knowledge at the end of an assessed investigation rather than at the beginning could help to overcome this challenge. If NEMP assessors are to draw out more insights into children’s causal reasoning, this will probably need to be scripted into their formal talk, because few do so spontaneously.

Year 8 children recognise and acknowledge more features of fair tests than Year 4 children. They are more likely to control at least some variables, although they do not usually display any other types of development in their approach to/understanding of fair testing. The understanding that there can be interactions between different variables is identified in the literature as an important developmental step and is an essential aspect of mature scientific reasoning. However this seems to be a neglected component of school science investigative tasks.

Children find measuring laborious and the context of a task can greatly influence the measuring skills demanded of them. Unfamiliar measuring tools distract from the main focus, vertical scales introduce errors of parallax, and so on. The act of measuring, followed by written recording, seems to partition sequential tests into distinct episodes so that they are not immediately seen as parts of a whole, coherent test design. Simplification of measuring may be an important strategy that frees children to pay more attention to the overall patterns and purposes of the tests they are carrying out. Collection of categoric rather than continuous data is one such strategy.

Even when they have planned a series of tests, children may "lose their way" and deviate from their intended plan part way through an investigation. Children's more limited memory capacity has been linked to their investigative ability and so perhaps exacerbates this effect. Visual strategies bring more of the overall investigation structure and/or results into view simultaneously, helping to transcend memory demands. Such strategies support children's ability to identify fair tests, and/or to see meaningful data patterns from single repeated tests, or from sequences of tests.

Children typically ignore experimental error, apart from occasional single instances of repetition when a result diverges too widely from what they expected. However the literature suggests that they do understand that, even though individual results vary, main effects are robust. With encouragement to explore patterns of data variability, the process of test repetition might be made more meaningful.

Teachers are attracted to strategies for teaching science investigative skills that have a strong visual element, especially simple, visual, data recording strategies that make strong links to *Mathematics in the New Zealand Curriculum.* They have a concern to use science investigations to stimulate both science learning and language development, especially for mixed-ability classes and for ESOL children.

SECTION ONE: INTRODUCTION

This report presents the findings of a three-stage research project that was conceived with an intention to inform the active teaching of investigative skills in primary science. The first stage involved the observation of 200 taped episodes of Year 4 or Year 8 groups of children carrying out NEMP science tasks. The analysis of the collated observations was informed by a review of the literature on the development of children's investigative skills. The insights generated by these 2 stages were then used to devise some strategies for the active teaching of investigative skills at the primary school level. In the third stage of the project, these activities were tested and commented on by 3 focus groups of primary teachers.

BACKGROUND TO THE RESEARCH

A previous NEMP probe study in the area of science education highlighted the challenge of helping teachers to "build bridges" between existing NEMP tasks and their actual classroom pedagogy (Gilmore, 2001). Gilmore noted that the "Developing Scientific Skills and Attitudes" strand of *Science in the New Zealand Curriculum* seems to pose particular challenges in that many teachers do not appear to actively teach students the key objectives of this strand. This finding is supported by the literature review of effective pedagogy for raising student achievement in science that was jointly carried out by NZCER and the CSTER team at Waikato University (Hipkins, Bolstad, Baker, Jones, Barker, Bell, Coll, Cooper, Forret, France, Haigh, Harlow, and Taylor, 2002). Hipkins et al. found that teachers tend to focus on "fair testing", ignoring other types of scientific investigations. This narrow focus provides a very limited basis for actively teaching investigative skills. This situation is exacerbated by the outdated focus on technology, at the expense of developing explicit nature of science (NOS) understandings, in the first strand of *Science in the New Zealand Curriculum*.

It seemed to us that a detailed analysis of carefully selected NEMP tasks in which children have been recorded on video in the process of carrying out actual scientific investigations could provide useful data to inform the active teaching process. The proposal was initially designed at a time when the team working on the Ministry of Education-funded Science Exemplars project had been charged with the additional and partially retrospective responsibility[1] of developing a matrix that included scientific investigation skills. The matrix thus devised[2] specifies the nature of student achievement for the "Developing Scientific Skills and Attitudes" strand of *SNZC* in considerable detail. In this context, we also wondered if the planned systematic comparisons of Year 4 and Year 8 children's investigative actions might support teachers' interpretation and any potential revision of the Exemplars Matrix by providing a somewhat more streamlined guide for evaluating children's learning progress in this area.

---

[1] A 'matrix' for each subject area was not part of the initial Exemplars project but was added as a requirement part way through the project.
Poskitt, J. (2002). National consultation on exemplars: What difference does it make for teachers? ***set**: Research Information for Teachers, 3* 13–16.

[2] The current version is available in the Exemplars section of the Te Kete ipurangi website – www.tki.org.nz

THE RESEARCH QUESTIONS

As they were originally conceived, the research questions were:

1. How does NEMP data for actual student achievement in developing scientific skills correlate with the relevant achievement criteria specified on the draft Exemplars Matrix for science?
2. How do primary teachers conceptualise opportunities and challenges for actively teaching investigative skills in science:
    - generally?
    - when using NEMP tasks in particular?
3. Do patterns in achievement data reflect identified strengths and weaknesses of teachers' pedagogical content knowledge in the science skills area, and if so, in what ways?

OVERVIEW OF THE TASK ANALYSIS

It was our original intention to select 3 widely different NEMP tasks, so that we could cover all of the investigative skills categories on the Exemplars Matrix. To identify suitable tasks, we developed a grid to match components of all NEMP science tasks to key criteria from the matrix. This process is described more fully in Section Two. We planned to analyse 100 instances of each selected task. In the event, the 3 tasks chosen were more similar than we had anticipated because few of the other tasks showed the same potential to cover a number of the categories of the Exemplars Matrix. (This comparative analysis of potential matches between the NEMP tasks and the Exemplars Matrix is presented in Section Two.) A consequence of the specific selection process employed has been that we too have focused on "fair testing", despite our awareness of the need to widen the types of science investigations that are carried out in schools.

For reasons that are outlined in Section Two of the report, the planned 100 observations were made for only one of the 3 tasks (*Emptying Rate*). Fifty examples of each of the other 2 selected tasks (*Truck Track* and *Ball Bounce*) were analysed as follows:

Table 1

NEMP tasks selected for this research

Name of task/year of administration	No. of Year 4 groups observed	No. of Year 8 groups observed
Emptying Rate (95 and 99)	50	50
Truck Track (95 and 99)	50	NR
Ball Bounce (95 and 99)	NR	50

All 3 tasks were "trend tasks", that is the same task was used and data reported and compared for both the 1995 and 1999 science rounds (Crooks and Flockton, 2000), although the actual tapes observed all originated from the 99 round of tests. A brief summary of each task follows.

Truck Track

A flat ramp, supported by 1, 2, 3 or 4 corks was used to investigate the travel distance of a small toy truck. As well as altering the slope of the ramp by varying the number of corks supporting it, Year 4 children could investigate the travel patterns of trucks as they ran either forwards or backwards down the ramp. Trucks ran across a provided square of material at the end of the ramp. A three-fold builder's rule was provided for measuring distance.

Ball Bounce

Year 8 students were given a selection of balls of varying diameters, weights, and material composition. Using a long, folding ruler positioned on a desk-top they were asked to determine which ball was the bounciest.

Emptying Rate

Children were provided with an improvised funnel made from a clear PEP beverage bottle with the bottom end cut off and a hole drilled in the lid. The bottle was inverted and marked at 3, 6, and 9 cm up the sides away from the lid. Using a stop-watch children were asked to determine the time it took water to drain when the funnel was filled to each of these 3 marks, and then to do the same for detergent. Groups of Year 4 children and other groups of Year 8 children completed this same task.

INSIGHTS FROM OTHER LITERATURE SOURCES

The impetus for this research was generated in part by the concern that observational data *per se* are not a sufficient basis on which to describe progressions in children's skills. Of course that is as true of our own observations of children working on NEMP tasks as of observations made by the team of science advisers who worked with primary teachers to develop the science exemplars. While our analysis of the NEMP tasks was being carried out, we also sought insights on the development of children's investigative skills from the considerable literature on this subject. The literature in cognitive psychology, in particular, has a concern to explain *how* and/or *why* skills develop in the actual sequences that children display. It seemed to us that this explanatory dimension would add to our research. While we cannot claim any direct links between this literature and our own observations, some of the findings reported do seem to us to provide useful potential explanations for the patterns we observed. Accordingly, findings from the literature are threaded through the next 3 sections of the report as follows.

In Section Three, the results of the analyses of the 3 NEMP tasks are reported. The main focus is on those stages of the investigative process at which it is possible to make detailed observations of, and comments about, children's investigative skills – that is, the actual stage of carrying out the investigation. While the tasks appear to test children's investigative skills at all stages of the investigative process, critique from the literature suggests that they actually represent an inadequate basis on which to determine children's planning skills, or to assess their ability to draw conclusions from data. This critique is explored in the context of all 3 NEMP tasks.

Section Four reviews the research literature to describe findings that illuminate the manner in which children's investigative skills may develop over time and/or with appropriate learning experiences. This research is linked to the findings from the analysis of the NEMP tapes where appropriate. In Section Five we then draw these various literature threads together to suggest some broad patterns in the development of children’s investigative skills. These are presented within a framework of 5 clusters of meta-level attributes, with descriptions of the associated investigative actions that children might display. Each cluster is accompanied by suggestions of areas of attention for the active teaching of science investigation skills.

Section Six draws the collected findings together to describe some simple strategies for the active teaching of investigative skills to primary school children. Teachers’ opinions about these activities, and their comments about their current teaching of investigative skills, are also outlined. Finally, Section Seven outlines some areas for further exploration and debate.

SECTION TWO: METHODOLOGY

TASK SELECTION

The proposal developed for this research specified an initial comparison of NEMP tasks with the Exemplars Matrix. The intention was to select 3 quite different tasks that would allow for an exploration of as many cells on the matrix, in its then most recent version, as could be covered.

Preliminary analysis of NEMP tasks

To carry out this preliminary analysis, all published NEMP tasks from both the 1995 (Crooks and Flockton, 1996) and 1999 (Crooks and Flockton, 2000) rounds of NEMP testing were considered. All those tasks that appeared to have the potential to reveal some aspect of children's thinking about science investigations were initially considered in relation to the Exemplars Matrix.

The Exemplars Matrix is actually presented as 4 separate matrices, which are variously called:

- Investigating in Science;
- Thinking as a Scientist;
- Developing and Communicating Science Understanding; and
- Developing Interest and Relating Science Learning to the Wider World.

To identify suitable NEMP tasks with a focus on investigative skills, we used the "key aspects" identified at the left-hand end of each row of matrix cells for almost the entire *Investigating in Science* matrix.[3, 4] We also used the "overall progression" summaries for the *Thinking as a Scientist* and *Developing and Communicating Science Understanding* matrices. We have labelled these aspects as follows:

"Key aspect" of matrix	**Our phrase in Tables 2–4**
Whole investigations	Whole investigations
Asking questions	Asking own questions
Relating ideas to investigations/making predictions	Making predictions
Trialing and testing	Trialing/testing
Observing and systematically recording observations	Observing
Processing and interpreting	Processing/interpreting
Evaluating the investigative process	Evaluating
Reporting	Reporting
Thinking like a scientist	Thinking like a scientist
Developing and communicating Science understanding	Communicating understanding

---

[3] The current version of the Exemplars Matrix can be located at www.tki.org.nz/r/assessment/exemplars/sci. This is changed in some detail from the version we used but not in overall layout.

[4] We omitted "Exploring a situation' from this matrix because, in the context of NEMP testing tasks, time for exploration is not typically provided.

Two researchers separately rated links between the identified NEMP tasks and the Exemplars Matrix headings. At this stage, we did not concern ourselves with levels of progression, as specified on the Exemplars Matrix. Our concern was solely with potential coverage at any level. We also noted the contextual strand(s) of *Science in the New Zealand Curriculum* (Ministry of Education, 1993) that provided the overall context for each task we analysed. We then met and compared notes to develop a consensus agreement on the broad areas that were potentially covered by the identified tasks. This process was not as straightforward as it might at first appear. The matrices are detailed and the categories overlap. Arguably this is to be expected since the process of any one investigation is ultimately a unified whole. The difficulty does however serve to highlight the challenge of laying out any matrix, which must, since it is committed to two-dimensional paper, be developed in some linear order.

Results of preliminary analysis

The potential Exemplars Matrix/NEMP task matches that were identified are summarised in Tables 2-4 on the next 4 pages. Our headings matched to the selected content of the matrix (as described above) are on the left-hand side of each table. The names and overview details of each NEMP task we considered are listed across the top of each table.

Table 2 summarises matches found for tasks whose content was linked to the contextual strands *Making Sense of the Living World* and *Making Sense of Planet Earth and Beyond*. The many empty cells on this table highlight the lack of emphasis on investigative skills in NEMP tasks that cover knowledge from these two strands. For our purposes, any tasks from these 2 strands were ruled out for possible selection at this early stage.

Table 3 summarises matches found for tasks whose content was linked to the contextual strand *Making Sense of the Physical World.* Three tasks stand out here because their relatively full columns reveal a number of potential matches to different facets of children's investigative skills. These tasks are *Truck Track*, *Ball Bounce*, and *Parachutes*. All 3 are group tasks in which small groups of children actually carry out an investigation, although this happens within pre-set parameters, rather than in response to a question of their own. NEMP staff recommended against the selection of *Parachutes*, because it had caused some practical difficulties for children as it was implemented.

Table 4 identifies *Emptying Rate* as a similarly suitable group task set in a context linked to the *Making Sense of the Material World* strand. Two individual tasks, *Powders* and *Sugar Solutions*, could potentially be used to reveal some aspects of thinking of individual students working on a one-to-one basis with the teacher.

There were fewer suitable tasks from which to choose than we had originally anticipated. At this stage the decision was made to use all 3 suitable group tasks – that is, *Ball Bounce*, *Truck Track*, and *Emptying Rate* – so that task *type* did not become an additional variable in the overall research.

Table 2

Investigative skills in NEMP tasks with links to the Living World and Planet Earth and Beyond strands of SNZC

Skill from Exemplars Matrix	***Finger games*** **N95/99** **Year 8** **One-to-one**	***Planting seeds*** **N95** **Year 4/8** **One-to-one**	***What eats what*** **N95** **Year 8** **Team**	***Rocks*** **N99** **Year 4/8** **One-to-one**
Whole investigation				
Asking own questions				
Making predictions			Use of own ideas	
Trialing/testing				
Observing	Observes measures			Describes properties
Processing /interpreting	Trends conclusions			
Evaluating				
Reporting				
Thinking like a scientist	Conflicting evidence	Identifies variables		
Communicating understanding			Warrants for explanations	Links properties/uses

Table 3

Investigative skills in the NEMP tasks with links to the Physical World *strand of SNZC*

Skill from Exemplars Matrix	***Drop coins*** **N95/99** **Year 4/8** **Station**	***Mystery wire*** **N95/99** **Year 4/8** **Station**	***Water mix*** **N95/99** **Year 4** **One-to-one**	***Truck track*** **N95/99** **Year 4** **Team**	***Ball bounce*** **N95/99** **Year 8** **Team**	***Electricity*** **N99** **Year 4/8** **One-to-one**	***String ping*** **N99** **Year 4/8** **One-to-one**	***Marbles/water*** **N99** **Year 4/8** **One-to-one**	***Travelling trucks*** **N99** **Year 4** **One-to-one**
Whole investigation				Ideas for testing	Ideas for testing				
Asking own questions									
Making predictions	Only as a guess		Only as a guess	Based on experiences	Based on experiences	POE	PO(E)	PO(E)	PO(E)
Trialing/testing				Potentially – not signalled	Potentially – not signalled				
Observing	Observation link to product		Using equip. accurately	Measuring recording	Measuring recording	Recording observations			
Processing /interpreting		Drawing conclusion		Recognise patterns/trends	Recognise patterns/trends		Recognise patterns/trends		
Evaluating				Critical comment	Critical comment				
Reporting				How data handled	How data handled				
Thinking like a scientist	Evidence refutes explanation			Evidence/ explanation links	Evidence/ explanation links				
Communicating understanding				Nature of explanation	Nature of explanation	Use of science ideas	Nature of explanation		

Table Three contin…

Skill from Exemplars Matrix	*Wheel race* N99 Year 8 Team	*Mighty magnets* N95 Year 4/8 Station	*Sound chimes* N95 Year 4/8 Station	*Reflections* N95 Year 4/8 One-to-one	*Estimation* N95 Year 4/8 One-to-one	*Circuits* N95 Year 4/8 One-to-one	*Parachutes* N95 Year 4/8 Team
Whole investigation	Ideas for testing						Ideas for testing (x2)
Asking own questions							
Making predictions	As part of planning		POE		As an estimate		As part of planning (x2)
Trialing/testing	Changes plan						Use first part as trial
Observing	Measuring recording	Measuring recording	Compare patterns	Describe observations	Range of measurements		Timing recording
Processing /interpreting	Identify patterns	Identify patterns				Draw pattern as diagram	Identify patterns
Evaluating							Link first part to second
Reporting	Justify conclusions	Identifies all components					Explaining patterns
Thinking like a scientist	Cause/effect links						Theory/observe Best argument
Communicating understanding	Science ideas in explanation	Links results to science ideas		Use of vocabulary			Links between ideas/events

Table 4

Investigative skills in the NEMP tasks with links to the Material World *strand of SNZC*

Skill from Exemplars Matrix	***Emptying rate*** **N95/99** **Year 4/8** **Team**	***Rods*** **N95/99** **Year 4/8** **One-to-one**	***Sugar solutions*** **N99** **Year 4/8** **One-to-one**	***Powders*** **N99** **Year 4** **Team**	***Separating mixtures*** **N95** **Year 4/8** **Team**	***Salt solutions*** **N95** **Year 4/8** **One-to-one**	***Raincoats*** **N95** **Year 4/8** **Station**	***Acids in home*** **N95** **Year 8** **One-to-one**
Whole investigation	Ideas for accuracy		Individual ideas	Ideas for possible tests and observations	Ideas for actions	Fair test plan		
Asking own questions								
Making predictions	From actual results	As a guess	Predict – Explain			Predict – Explain		
Trialing/testing	Changes in action			Before chart design	Modifies for equip range			
Observing	Measuring recording	Detecting trends	Timing recording	Systematic recording	Systematic recording	Carrying out plan	Test – no plan	Use of litmus
Processing /interpreting	Identify patterns		Drawing conclusion	Identify relevant patterns		Identifies patterns	Justifies conclusion	Significance of colour change
Evaluating			How to make more fair		Identify problems and how to modify			
Reporting	Explain patterns							
Thinking like a scientist	Use of evidence and/ or science ideas for predictions		Features of fair test	Links to previous expt.				
Communicating understanding			Link science ideas/events	Link science ideas/events		Link science ideas/events	Link science ideas/events	

CREATING OBSERVATION SCHEDULES

Once the tasks had been selected, the analysis began with the creation of an observation schedule for each of the 3 tasks. To create each task schedule, the two researchers watched 3 to 5 tapes together and we discussed the dynamics displayed by each group as they worked through the task. Variables relevant to the task were identified, and task-specific features likely to be of interest were debated and incorporated in the schedule as appropriate. The 3 observation schedules were re-checked and adjusted if necessary after 10 observations for each task had been completed. A typical task schedule is included as Appendix 1.

NOTES ON THE OBSERVATION PROCESS

Emptying Rate was investigated first, with the same researcher watching all 101 tapes (48 at Year 4 and 53 at Year 8). During this process it became evident that very little new behaviour was seen after the first 20 or so episodes at either age level. For this reason, only 50 tapes were analysed for each of the other two tasks – 50 Year 4 groups for the *Truck Track* task, and 50 Year 8 groups for the *Ball Bounce* task. A third researcher assisted in the analysis of the final task *Ball Bounce* and watched 24 of the tapes. At the "data cleaning" stage the main observer went through a random number of the tapes that had been viewed by the third researcher to check consistency of the data collected.

All tapes, whether at Year 4, or Year 8, began with the *Emptying Rate* task. In most cases the other task we had selected followed next on the tapes. Thus the same groups of children were observed carrying out both tasks at their year level. The time taken to complete tasks varied quite considerably. For example, the Year 4 children generally took considerably longer to complete the *Emptying Rate* task than the Year 8 children. The younger children needed more time to get organised, to decide on roles, and to carry out the tasks. They also needed more help and encouragement from the teacher to keep them focused. Behavioural problems occasionally caused longer duration sessions in both year groups and across all tasks. At times it appeared that the facilitating teacher needed to hurry children to the end of the task:

> *Come on, we are running late and we have another activity to get through.*
>
> *We have not got time to re-do it, you have to keep going.*[5]

When time was short there was usually a very superficial discussion with children at the end of the task.

---

[5] Throughout the report italicised statements on separate lines are actual quotes taken from the NEMP tapes.

SECTION THREE: THE TASK ANALYSES

This section reports the initial findings from the analysis of the NEMP tapes. The analysis is divided into 3 main sub-sections. The first reports on children's actions as they were introduced to the tasks and asked to plan their investigation. The second sub-section discusses events at the "carrying out" stage and the third reports on the reflection/discussion stage that completed each task.

PLANNING STAGE

In all 3 tasks, children were asked to plan what they would do before they began the practical part of their investigation. The observation schedules were designed to capture comments about all the key variables that would need to be managed in each task. Those that were actually mentioned by one or more groups of children are reported next.

Truck Track

Fifty-one groups of Year 4 students were observed as they carried out the *Truck Track* task. It is very evident from the results in Table 5 that there was very little discussion of the management of variables at the planning stage of this task, although children had been asked to "think about things you will need to keep the same". Just 4 variables were identified at this stage and one group mentioned 2 of these. Those variables that were most commonly discussed were concrete features of the context – things that these younger children could readily observe and anticipate:

> *Keep the rug straight and the lines lined up.*
>
> *One person takes the bumps out of the mat before each turn, and then hold it flat so surface is smoother.*
>
> *Measure from the mat edge to where the truck stops.*

Table 5

Management of variables/task discussed by groups of Year 4 children when planning for the Truck Track *task*

Item	Frequency of mentions (N = 51)
Consistent ramp and/or mat set up	1
Accurate measurement of truck travel	2
Consistent set up and release of the truck	1
Total mentions	**4**

Ball Bounce

Fifty-two groups of Year 8 students completed this task. As Table 6 shows, there were many more instances of explicit discussion of variables amongst these older children than we saw in the *Truck Track* discussions amongst the Year 4 children. Thirty-three groups were able to identify at least one variable that should be controlled, and one group identified 4 variables. Consistency in release of a ball (31 mentions) is a comparable factor to consistency in release of a truck (1 mention) in the *Truck Track* task. In this respect, at least, more than half of the Year 8 groups showed an awareness of the need to manage at least one variable in “fair testing”.

Table 6

Management of variables/task discussed by groups of Year 8 children when planning for the Ball Bounce *task*

Item	Frequency of mentions (N = 52)
Consistent height from which to drop balls	31
Ruler held straight vertically	5
Strategy for determining the top point of bounce	5
Other strategies for anticipating/accurately measuring bounce	1
Ruler oriented the same way each time (there are different scales on either side)	1
Checking consistency by replicating	1
Total mentions	**44**

Emptying Rate

Forty-eight Year 4 groups and 53 Year 8 groups were observed carrying out this task. It was the first task that any one group completed and so may have been intended as a “warm up” for the second task that followed. This second task was *Truck Track* for Year 4 children, and *Ball Bounce* for Year 8.

Perhaps because the NEMP designers anticipated that this task would be less familiar to the children than either the *Ball Bounce* or *Truck Track* tasks, each group was initially shown a video of the task set-up, and then they were asked to plan their own investigation. They were not specifically asked to think about things to keep the same. Indeed there was little that they could decide in this respect since the task was already tightly defined. Nearly half the groups discussed the requirement to vary the volume of liquid to be tested, although this was more often in the nature of repeating/clarifying task instructions than planning for accuracy by taking specific care with the actual measuring process.

Table 7

Management of variables/task discussed by groups of Year 4 and Year 8 children when planning for the Emptying Rate *task*

Item	Frequency of mentions (N = 101)
Aspects of measuring to the unit marks on the bottle	22
Aspects of using stopwatch, measuring time	4
Keeping bottle level	3
Avoiding cross-contamination of water/detergent	1
Totals	**30**

Nine Year 4 groups identified 1 of these variables and one group identified 2. Fifteen Year 8 groups identified at least 1 variable to be controlled, with 3 of these groups identifying 2 or more variables. Thus at both year levels, the groups who did identify and discuss variables at the planning stage were in the minority.

Children's seeming lack of planning skills

Children's planning skills appear to be very context sensitive. As we watched the tapes it seemed to us that 4 factors constrained children's ability to demonstrate such "fair testing" planning skills as they may actually have had.

1. The type of instructions given

At the beginning of each activity, for all 3 tasks, teachers read a set script of instructions to the students. This is the *Emptying Rate* script, read out after children watched the video that modelled the inquiry procedure:

> *…plan how you will do your experiment. Do this now and when you are ready I will ask you to tell me your plan. Remember you need to plan your activity so everyone has a job to do.*

This is the *Ball Bounce* script, read out after the context was introduced, and before students were given the folding ruler with which to measure:

> *You are to work as a team, and try to make sure that everyone helps. First you should plan how you will do the experiment. Think about what things you will need to keep the same. Think about what you will need to measure. Think about how you will use numbers to say how bouncy each ball is. Sort out who is going to do the measurements and who will do the other jobs. Everyone should have a job.*

The introduction to the *Truck Track* task was very similar. There is an important similarity, and an important difference in the emphasis given by each of these sets of instructions. The difference is that the *Ball Bounce* and *Truck Track* tasks gave direct guidance about the fair testing aspects. "What to keep the same" introduces an important aspect of fair testing — the control of variables. Although both age groups exhibited little unprompted awareness of the need to "keep things the same" when talking about the *Emptying Rate* task, many of the Year 8 groups were able to anticipate at least one way to do so for the *Ball Bounce* task that they subsequently completed. In contrast, the Year 4 children were able to generate more ideas about variables for *Emptying Rate* than for *Truck Track* – perhaps because the *Emptying Rate* video instructions allowed them to anticipate more of the actions they were about to undertake.

Table 8

The contrast in types of discussion topics raised during the planning conversations of Year 4 and Year 8 children

Year Level and Task	Social (roles)	Management of variables/task
Emptying Rate (Yr. 4/N=48)	47	11
Truck Track (Yr. 4/N=51)	45	4
Emptying Rate (Yr. 8/N=53)	50	19
Ball Bounce (Yr. 8/N=52)	50	47

All 3 scripts emphasised that each student should have a part to play. Consequently, ideas about "what to keep the same" were often displaced by conversations in which children determined the roles they would play:

> *I will hold the ruler.*
>
> *Can I do the stopwatch?*

In the light of the scripted emphasis on role allocation, this is perhaps not surprising. As shown in Table 8, the younger children focused on this aspect of their planning almost to the exclusion of actual fair testing planning. Many of these Year 4 children occupied their planning time in playing games such as "Rock, Scissors, Paper" to determine the allocation of their roles.

2. Private and public planning conversations

We gathered some evidence that aspects of the assessment context could contribute to an under-reporting of children's ability to plan for managing variables in an investigation. During our observations of the planning stage, we recorded a number of instances where children did not report back to the teacher all that they had actually discussed amongst themselves. Typically, the allocation of roles was what children reported at the end of the planning time. Where students had talked amongst themselves about *keeping the bottle straight* for *Emptying Rate*, or *holding the ruler straight* for *Ball Bounce*, the teacher was likely to be told something along the following lines: *John will hold the ruler, Sarah will drop the balls, and Amy will record the results.* The children were most likely to report role decisions even when their private discussions had included decision-making about aspects of control of variables.

After detecting this pattern in the first 2 tasks observed, we decided to record instances of private/public ideas for *Ball Bounce*, the final task to be analysed. In this task, not all students spoke within their group and in some instances it was hard to hear what students said to each other as they spoke quietly or the microphone was placed too far away. Nevertheless, themes of the conversations we heard encompassed all those aspects reported in Table 6 – 47 ideas in all. Teachers were only told about 25 of these 47 ideas, not least because 27 of the 52 teachers we observed in the *Ball Bounce* task did not ask students to discuss their planning before commencing the task itself.

3. The absence of a meaningful sense of purpose?

One important aspect of the investigative context is the sense of meaning or purpose that children bring to their investigations. Children are typically encouraged to begin their investigations from questions they have generated either individually or as a group. This approach is clearly elaborated in curriculum support materials such as the *Making Better Sense of…* primary science series produced by New Zealand's Ministry of Education to support *Science in the New Zealand Curriculum.* However within a national assessment context such an approach would create issues of comparability and validity. Children "investigate" a task of which they have had no previous ownership. Can they develop a sense of purpose in such contexts? If not, how might this lack of purpose impact on their ability to demonstrate planning skills?

The script for the *Emptying Rate* task instructed the students to *"Do this now* [the planning] *and when you are ready I will ask you to tell me your plan."* What sense did the children make of the purposes for doing this task, beyond an imperative to do as they had been told? The implicit purpose – to compare the emptying rates of liquids of differing viscosity/density – did not emerge until the "discussion" stage at the end of the experiment.

The task script for both the *Truck Track* and the *Ball Bounce* tasks requested children to *"Plan your experiment now, and tell me when you have finished your planning."* With this small shift in emphasis, a number of the teachers, especially for *Ball Bounce,* did not ask the students to report back about their planning before moving on to the carrying out stage. While it may seem that the purposes of the *Truck Track* and *Ball Bounce* tasks were self-evident, their theoretical underpinnings were never discussed (and indeed are not easy to determine for *Ball Bounce* since the balls appeared to vary in a number of their material features). We wondered whether and how children would address this aspect at the planning stage. We found that just one group discussed the purpose of the *Truck Track* task, 3 groups discussed this for *Emptying Rate* and 8 groups discussed purposes for *Ball Bounce*. However the sense of purpose expressed was task-orientated rather than related to a conceptual science idea/question:

> *We need to measure bounce* [of the balls].

In *Ball Bounce* 3 groups also expressed an opinion about what they thought may happen:

> *The orange ball* [the smallest] *will bounce the highest.*
>
> *The smallest* [ball] *will bounce the highest.*

These are seemingly guesses although it is possible that children think smallness is the property that confers bounce. Because their causal theorising was not probed we cannot know for certain. Whether and how this absence of conceptual links impeded planning is something about which we can only speculate. Indeed the literature reported in Section Four would suggest that in the absence of such links children are not really planning in a scientific sense at all.

4. The necessity to become familiar with the task at hand

We have already noted that children’s task-related planning for *Emptying Rate* was weighted towards familiarisation with the requirements of the task rather than management of variables for fair testing. While there was more discussion of variables for *Ball Bounce*, the lack of ownership of purpose, combined with the presentation of a set of sometimes unfamiliar equipment, meant that practicalities often dominated the so-called “planning” time.

After the children had completed the task, ideas related to its management and its purpose did emerge in the group discussions. We wonder if “planning” would be more appropriately assessed if placed at the end of an initial familiarisation task. Since children were asked to make predictions for new situations at the end of all 3 tasks, this would be relatively easy to do, although it would make the task time longer.

Being systematic about sequencing experimental tests

The literature that we read in conjunction with our observations of the NEMP tapes has led us to reflect on why (Section Four) and how (Section Six) teachers should and can improve children’s investigative skills by helping them to visualise the whole “experimental space”. While such visualisation should accompany a coherent sense of purpose, which we have established seemed to be lacking, we nevertheless watched to see whether and how children organised and sequenced the separate test episodes in a pre-planned way.

The *Emptying Rate* demonstration video and the accompanying instruction card both stated that students must time all 3 water levels before those of the detergent. We wondered if the children would discuss this sequencing issue prior to carrying out the task. They were also requested to pour the water back into the bottle between tests, doubtless for purely practical reasons. The *Truck Track* task required a series of tests, with 4 ramp positions and 2 truck orientations to be individually tested. Would the Year 4 students discuss how they would systematically keep track of all 8 possible tests? *Ball Bounce* featured a series of balls ranging in size, with no inherently obvious sequencing criteria. Would students even consider the order in which they should test these and, if so, how would they decide? Table 9 reports the differences between instances of sequencing discussions for the 3 tasks.

Table 9

Instances of sharing planning ideas on how to sequence separate test events within a task

Task/Level	Sequencing mentions	Comment
Emptying Rate Y.4/8	12 Year 4 (N=48) 15 Year 8 (N=53)	Sequencing comments largely a reiteration of task instructions.
Truck Track Y.4	3 (N=51)	Mainly comments to do with starting with a different number of corks.
Ball Bounce Y.8	1 (N=52)	One group decided to *“do small balls first”*. Most groups followed sequence on task sheet

Did the more detailed instructions (video/directive to "do water first") divert children from other possible planning topics when talking about how to carry out the *Emptying Rate* task? Or was it that this task had a greater number of unfamiliar contextual details to which children felt they needed to attend? Or should we reflect that children moved into the tasks, mostly ignoring sequencing implications, because in the absence of a sense of ownership of the initial question they did not visualise the experimental space beyond "one step at a time"? The literature reported in Section Four suggests that these are not trivial questions. In that section we draw the conclusion that supporting children to visualise the whole planning space may be a very important aspect of actively teaching science investigation skills.

THE CARRYING OUT STAGE

In this part of the report, we discuss the observations we made as children carried out each of the 3 tasks in the manner they had planned. During the preliminary observations it became very evident that children often *do* more than they *say*. To capture this tendency we created observation schedules that could distinguish between intuitive and explicit actions. "Intuitive" actions were deemed to be those where children simply carried out an action without any comment being made by anyone in the group (for example straightening up the improvised PEP bottle funnel for the *Emptying Rate* task). "Explicit" actions were those that were accompanied by a specific explanatory comment:

> *We need to keep this bottle straight.*

As the following results illustrate, we found that actions to do with management of variables were more likely to be intuitive, while those that concerned measurement processes were more typically explicit. These 2 aspects of each task are reported next.

Truck Track **task**

Management of variables

Table 10 records the numbers of Year 4 groups who took action to manage variables as they carried out their successive *Truck Track* tests. There was some inconsistency between actions taken in the "truck forward" and "truck backwards" series of tests in a few groups. This pattern has been reported by listing the number of truck forward actions, followed in the same cell by the matching number of truck backward actions. (Usually, students tested the trucks in a forward-facing position first.) While there were 51 groups in total, some groups did not carry out both series of tests, or the camera was positioned such that not all actions could be observed, and so not all data sets are complete. The overall tendency, as noted, was for children to do much more than they can say.

Table 10

Truck Track *variables attended to by Year 4 children when testing trucks facing forward and then backwards*

Truck forwards/truck backwards	Intuitive	Explicit	Ignored
Consistent positioning of ramp and mat	31 / 30	4 / 3	16 / 17
Ramp straight onto mat	33 / 27	0 / 3	17 / 19
Consistent set up and release of the truck	29 / 25	10 / 11	10 / 12
Total instances	**93 / 82**	**14 / 17**	**43 / 48**

Trucks were intended to roll onto a standard type of mat at the end of the ramp. These mats typically had bumps along their set folds that were hard for the children to remove. While many children intuitively or explicitly devised ways of dealing with this, others chose to ignore the bumps and to continue on with their task. Over half the teams had trucks that ran off the mat onto one or more other types of surface. In one instance the teacher had used masking tape to secure the edge of the mat, so that in a short distance the truck travelled over 3 different surfaces. Groups who attempted to manage this complexity typically positioned the ramp as close to one edge of the mat as possible, to lengthen the run room.

Aspects of measuring

While actions to manage the positioning of the ramp and mat, and the consistent release of the truck, were often intuitively kept the same for both the forward and backward tests of the truck, the students tended to explicitly comment on their actions when measuring was involved. In part, this may be because of the novelty of the measurement tool. Many children had clearly not seen a three-fold carpenter's rule before and there was much "playing around" with this at the start of the task. As a result some groups did look closely and discovered that there were separate scales for centimetres and millimetres on the opposite edges of the same side of the rule. However, for some groups, this made the ruler difficult to use consistently.

Table 11

Aspects of measuring attended to by Year 4 children during the Truck Track *testing stage*

Truck forwards/truck backwards	Intuitive	Explicit	Ignored
Accurate measurement of truck travel	5 / 6	31/ 28	11/ 11
Awareness of measuring units e.g., cm/ mm	3 / 3	22 / 22	24 / 23
Total instances	**8 / 9**	**53 / 50**	**35 / 34**

Several aspects of the context made it easier for children to carry out and record measurements in the *Truck Track* task than in the other 2 tasks. The ruler was positioned horizontally on the ground and the truck had stopped moving before the students measured its travelling distance. In comparison, *Ball Bounce* required the students to hold the ruler erect and then ascertain the top point of the bounce while the ball was still moving. However the measurement aspect of *Truck Track* was not without its management challenges. Children needed to decide where to fix their measure of the truck’s position. In most cases they measured from the back of the truck, and if not, they tended to be consistent in the alternative measurement position they had determined. However some groups remained unaware of the necessity to control this aspect of the task, measuring from a random selection of parts of the truck.

The collective travel trajectories of the individual truck runs made for some interesting patterns. Trucks sometimes slowed down when they hit the side of the ramp. On 2 occasions trucks stopped completely at a bump on the mat. They frequently careered sideways, and 21 groups set the ruler up in line with the ramp to attempt to keep the truck on a straight line or path of travel. This again slowed the truck when it hit the ruler. In these cases children typically released the truck again, and moved the ruler sideways slightly in an attempt to prevent further collisions. Some groups improvised various strategies to use the inflexible ruler to measure curved travel trajectories. Some used the lines on the mat to help them align the truck with the ruler, others moved the truck across to the ruler in line with its resting position, and some used paper or fingers as alternative measures. A few groups positioned the ruler in line with the corks at the back of the ramp. This meant that trucks usually outran the length of ruler available and the children had to devise a method for estimating. One team denoted the area after the ruler as *“past the mark”*, and they recorded this on their result sheet

Twenty-nine groups repeated test runs so that they could replace “faulty” runs with fresh data. On the other hand, some groups were observed to ignore measurement anomalies altogether. Many groups who did attempt to be consistent would nevertheless have recorded very compromised measurements. We note this here because Section Four cites literature which recommends that children at earlier stages of investigative skills development are likely to recognise meaningful data patterns more easily when the data are relative or categoric rather than absolute or continuous. This task seems well suited to the implementation of a simpler, more visual, form of data gathering. In Section Six we describe a strategy that teachers can easily use to help young children see the data patterns in the trucks’ capricious travel runs. In the light of the literature discussed in Section Four, we think that developing awareness of patterns of data variability is more important than requiring children to make judgments about the “most accurate” measurement.

Ball Bounce task

Management of variables

Many of the Year 8 students continued to discuss their plans for managing the *Ball Bounce* task at the carrying out stage. Table 12 shows the aspects to which the students attended. Comparing these data with those from the planning stage (Table 6) we see that just 2 additional groups explicitly recognised the need to drop the ball from a consistent height. However, faced with the practicalities of the situation in action, 9 more groups now explicitly identified the orientation of the ruler as an aspect to be managed, and 11 more groups began to discuss strategies for determining the top of the bounce. For these groups, “planning” was facilitated by action. There were more instances of explicit management of variables by these older students but many intuitive actions were still observed, and a number of groups continued to ignore key variables.

Table 12

Comparison of Ball Bounce *variables attended to by Year 8 children at the planning and carrying out stages*

Action	Planning	Carrying out stage		
		Explicit	**Intuitive**	**Ignored**
Keeping the ruler vertical	5	14	20	18
Consistent height for dropping balls	31	33	12	6
Strategy for determining the top point of the bounce trajectory	5	16	4	32
Ball bounced on same place on surface (to avoid cracks between desks)	-	1	41	10
Total instances	**41**	**64**	**77**	**66**

Aspects of measuring

Several contextual features made measurement challenging in this task. Table 13 summarises aspects of measuring that were taken into account as the task was carried out. As was the case in *Truck Track*, aspects of measuring were more likely than management of variables to be explicitly discussed. Again, the challenges presented were strongly linked to the contextual specifics of the investigation.

As already noted, the dependent variable (height of the bounce) posed considerable difficulty because students needed to find a way to "stop" the ball at the very top of its bounce, and the measurement point was fleeting. Most groups recognised that they then had an issue with accuracy and up to 3 students sometimes attempted to read each bounce. However this created new issues when, as often happened, different readings for the same trial resulted. Reading the vertically oriented scale at eye level was seen as one solution to this dilemma by 16 groups, while in other groups children of differing heights took measurements that we could clearly see included errors of parallax. Some groups simply ignored such differences, opting for whichever measurement they thought seemed right. Other groups adopted an "averaging" strategy. Some took the median measurement of their conflicting readings – a crude type of instant averaging process. Some repeated bounces of each ball several times to try and better establish the most commonly occurring measurement – again an attempt at averaging, albeit via a more explicit process. Students were not consistent in these averaging attempts. In many groups really bouncy balls were dropped more times than less bouncy balls.

Seven groups attempted to count the number of times each ball bounced in addition to taking a height measurement. This simultaneous measuring and counting was not easy, particularly with the balls that bounced a lot, or bounced more rapidly towards the end of the bounce, as did the table tennis ball. Since "bounciness" was not defined in the task instructions, this appeared to be a legitimate interpretation of the task that complicated the challenge for these groups. No groups tested the balls in an order that suggested thinking about causes for bounciness. In fact most simply worked their way down the order of balls listed on the Results Sheet

These Year 8 students, like the Year 4 *Truck Track* groups, found the folding ruler a novelty. Although the ruler had different scales on either side, 29 groups kept the ruler consistently oriented, and if at times the ruler did twist, members in the group would bring this to the attention of the person holding the ruler.

Table 13

Aspects of measuring that were attended to by Year 8 children during the carrying out stage of the Ball Bounce *task*

Item – *Ball Bounce*	**Intuitive**	**Explicit**	**Ignored**
Keeping the ruler vertical	20	14	18
Discuss need for accurate measuring	0	43	9
Checking consistency by replicating	12	21	19
Managing data variation – e.g., by use of an averaging strategy	11	13	28
Measures number of times ball bounces	0	7	45
Trials done to decide likely range of the bounce	1	0	51
Total instances	**44**	**98**	**170**

Three groups devised novel scales to measure the ball bounce. Two groups designated readings between 0–100 mm as a “1”, 100–200 mm as a “2”, 300–400 mm as a “3” and so on. The third group devised essentially the same solution, but used centimetres. This creative solution to a tricky set of problems interested us because, in effect, these groups devised a way to collect categoric rather than continuous data, thereby avoiding the vexed accuracy issues with which other groups had to contend. They were still able to compare bounciness, and to quickly run repeat trials to be sure of their emerging data patterns. In Section Six we suggest a modified version of this strategy as an effective means of helping students to recognise data patterns and data variation in “noisy”[6] tasks such as this.

Repetition

Typically, a ball bounce was repeated when children perceived there had been an error in the specific test. For example, a ball might hit the hand of the child holding the ruler or the ruler itself, either on its descent, or after the bounce. Some groups also re-tested when they had been unable to determine the height of a bounce in relation to the ruler, for practical reasons already outlined. As already noted, some groups did attempt a type of averaging strategy for repeat bounces. The end result of this repetition was always to obtain a single “best bounce” measurement. The significance of this type of thinking about data variability is explored in Section Four. For now, we note again that, as in the *Truck Track* task, children could have collected categoric data to make the comparisons required if the task had been structured differently for them.

---

[6] By this we mean the number of distractions presented by the busy context and the problematic data, not the physical noise of the classroom setting.

Emptying Rate task

Management of variables

In this more tightly prescribed and unfamiliar task it was difficult for children to identify variables to control, even once the actual task was underway. This is reflected in Table 14, which reports just 2 variables either explicitly discussed or implicitly managed as the task proceeded. The first number on each case represents children's actions taken while water flows were timed, the second figure represents actions taken while timing detergent flows. Interestingly, one of the two aspects most consistently "managed" is actually marginally relevant once the test run is underway. While accurate measuring did require that children hold the bottle-funnel level, the tilt of the bottle makes no discernible difference to the rate of flow of this small volume of water.

Table 14

Emptying Rate *variables attended to at the carrying out stage*

Water test/detergent test	Intuitive	Explicit	Ignored
Consistent level orientation of bottle	92 / 86	8 / 12	1 / 1
Consistent measurement of liquid to be emptied ("up to the mark")	7 / 4	87 / 88	7 / 7
Total instances	**99** / **90**	**95** / **100**	**8** / **8**

Aspects of measuring

This task required children to carry out a measuring action at both the beginning and the end of the investigation. Most children took explicit care in attempting to measure water/detergent volumes to the pre-specified marks on the improvised bottle funnel. Similarly, most groups explicitly devised a simple protocol to co-ordinate release of the water/detergent flow with starting the stopwatch, and end of the flow with stopping the watch. Perhaps because the stopwatch was another relatively unfamiliar piece of equipment,[7] children's discussions typically explicitly differentiated the flow time as measured by seconds, not minutes. Younger children sometimes struggled to determine which of these units of time measurement to employ, otherwise there were no clear Year 4/Year 8 differences in the children's investigative actions reported here.

Table 15

Aspects of measuring that were attended to during the carrying out stage of the Emptying Rate task

Water test/detergent test	Intuitive	Explicit	Ignored
Protocol for use of stopwatch button	4 / 6	97 / 92	0 / 1
Measuring to the mark/adjusting to the mark?	7 / 4	87 / 88	7 / 7
Measuring units e.g., seconds or minutes	2 / 4	95 / 94	4 / 1
Total instances	**13** / **14**	**279** / **274**	**11** / **9**

---

[7] Some teachers modelled the use of the stopwatch and some did not.

Children tended to repeat a measurement only when they had made an error such as forgetting to reset or start the stopwatch, or when they experienced problems in co-ordinating this with the timing of uncovering the hole in the bottle funnel. No groups at either age repeated measurements for accuracy or “fair testing” purposes. Fifty-eight groups actually chose to ignore errors they had made, seemingly in the interests of task completion: *It doesn’t matter, come on.* Seventy-seven groups (76 percent) sequenced the water/detergent test series as demonstrated on the video and explained on the instruction cards.

The issue of an endpoint

When to *stop* measuring was an issue in this task, especially when children moved from the water to the detergent tests. Because the water flowed quickly and consistently, only 3 groups of students discussed ways to define an endpoint, debating between determining when the water had completely passed through the hole when looking down, or when the water stopped dripping. When detergent was tested, the slower rate of flow and increased adhesion to the container sides made dripping a much more obvious feature of the system. Some groups now talked about the thickness of the detergent as the factor that slowed draining rate and about how it stuck to the sides of the bottle. Consequently, 14 groups now recognised that there was an issue with the endpoint. However, as the following quotes show, the children did not necessarily have strategies to deal with this in a controlled way:

> *It had not stopped.*
>
> *But it was just bubbles.*
>
> *It’s still going.*
>
> *Doesn’t matter. It’s only drips.*

Nor did any of these groups backtrack to discuss the same aspect of the water task. Seemingly this was an isolated feature of only this series of 3 detergent trials. This would appear to suggest these children are not working with an overall “fair test” plan in mind, but rather are moving from one test episode to the next. The significance of such a view is discussed in Section Four.

In 37 groups we observed instances of comments that anticipated the results of the trial that was about to be run:

> *Detergent will take ages.*
>
> *That* [3 cm of water] *will take 3 seconds.* [And it did!]

Some groups drew on the results from the water series to predict draining times for the detergent series:

> *It will take double the time as it* [detergent] *is twice as thick.*
>
> *It’s going to take ages, squeeze hard, it must be two times slower than water*.

Some teams felt that making a correct prediction indicated that they understood the purpose of the task:

> *Detergent will take longer as it is thicker than water. I think I am getting this.*

We saw flip-flopping predictions in one group who were surprised that the detergent initially came out of the bottle more quickly than they had thought it would. Their initial prediction: *"It will go slower"* became: *"No faster, no slower. It will stick to the sides."*

DISCUSSION AND REFLECTION STAGE

At the end of each task, teachers asked questions that challenged children to think about the "patterns" that emerged from the task they had just carried out – that is, to make simple inductive generalisations. In all 3 tasks they then asked the children to make a prediction about a novel aspect of the situation (emptying tomato sauce, testing a "5th cork" height ramp, and bouncing a squash ball respectively). The intention seemed to be to probe for each group's ability to transfer their generalisations to similar contexts. The youngest children were also asked what they could have done to make their *Truck Track* tests "more accurate". Since many of these children did not know what "accurate" meant, teachers often clarified at this point. However this word did not necessarily have the same meaning for the teachers themselves. Some defined accurate as *"exact", "perfect",* or *"more correct"* while some emphasised *"so it is fair"* or *"a fairer test"*. Presumably the teacher who said, *"so you don't cheat"* also intended to emphasise the fair testing aspect.

Truck Track discussion

Table 16 shows that almost all groups could interpret their results to see a pattern. Most described a qualitative relationship between ramp height and the distance travelled by the truck. Only a few groups made reference to the quantitative data they had so laboriously collected as they struggled with the measuring demands that have been described above.

Table 16

Year 4 children's interpretation of the Truck Track *task*

Recognising Patterns and Trends	*Frequency (n)*
Patterns in words only	38
Pattern refers to comparison of numbers/measurements	11

Reflecting on accuracy

Table 17 summarises children's responses to the question: "Could you have done anything to make your results more accurate?"

Table 17

Year 4 children's reflective ideas for making Truck Track *more accurate*

How they would make it more accurate	*Frequency (n)*
Better management of truck release and/or uneven mat surface	12
Measuring: accuracy and/or placement of ruler	10
Contextual comment on ramp arrangement: longer, straighter	10
Confounding factors: suggest changes to make test less "fair"	5
Replication to "double check"	1
Social: sharing, having turns	1
Total no of responses	**39**

Comments relating to better management of variables and/or more accurate data gathering obviously addressed the question of improving accuracy directly. However some of the comments the children made focused on the *context* of the investigation:

> *Use a bigger ruler.*
>
> *Make the ruler more straight.*

While they related to the overall experimental situation, their implementation would not necessarily have resulted in making the test any "fairer" or the data recording any more accurate. Some children, struggling to find something to say about the situation, described contextual changes that could make the test less fair – the opposite of what was intended. For example, some children suggested pushing the corks further under the ramp to make it steeper, but did not say if this positioning should be controlled to be the same for each test. A few children made suggestions that would most definitely confound the fair test, for example, suggesting blowing or pushing the truck down the ramp. One group, reflecting on the challenge of accurate data gathering, responded:

> *Measure really, really, carefully.*

In view of the measuring challenges outlined above, this would have been a hard ideal to live up to. The comment also appears to imply that with sufficient care it is possible to get one "right" reading. It seems likely that this belief is reinforced when teachers focus on the development of young children's measuring skills as the key aspect of "scientific" data collection. However this type of thinking is described in the literature as becoming a hindrance to the ultimate development of scientific thinking about ways to manage data *variability*. We return to this issue in Sections Four and Six.

Making predictions

When the perceived pattern had been described, the children were asked to predict how far the truck would travel both forwards and backwards if there were 5 corks under the ramp. Once they had made their predictions, the children carried out their task to see whether or not they were right. Some groups were helped to do this with teacher support:

Teacher: *Do you understand what I am asking you? Ok, I will show you again — so with one cork have* it [the ramp] *very low to the ground, so not steep, but with four corks the slope is a lot higher. So what did you find?*

Children remain silent.

Teacher: *How does the slope affect how far it* [the truck] *goes?*

She uses the cork to show what she means as she asks the question.

Children: *When it was on the higher slope it went further. When it was backwards it went faster.*

Teacher: *I wonder why? Any ideas?*

Children: *Heavier. More weight in front. We couldn't push it as it would go faster. So we didn't because it is cheating.*

Teacher: *You mean not fair?*

Children: *Yes.*

Without such support, some other groups remained unable to clearly articulate the thoughts they may have had.

Table 18

Year 4 children's predictions for the Truck Track *task*

Prediction Data	*Frequency (n=51)*
Accurate prediction	47
Incorrect prediction	4
Results used to justify prediction	29
Mention of features of context	17

While most groups correctly predicted that the truck should go considerably further when facing backwards down a "5-cork" ramp, some groups had their prediction compromised by obstacles or off-track runs. All trucks were slowed on their forward run because, at this steep ramp angle, the front bumper connected briefly with the ground. Backward runs were not affected because there was no back bumper. One group who initially made an incorrect prediction reasoned that coming off the steep ramp the truck *"would hit the ground and not go anywhere"*. This did in fact happen when this group released the truck to go forwards.

Some teachers drew a range of ideas from the children as they shaped their prediction via a group discussion. As Table 18 illustrates, just over half of the groups used their data pattern to justify their prediction. Seventeen groups discussed their prediction in relation to features of the context such as the slope of the ramp and the distribution of weight in the truck:

> *The truck will go further with 5 corks because the ramp is steeper.*
>
> *The ramp is steeper so when it hits the bottom it hits the front bumper so the truck slows down.*
>
> *The truck has 4 wheels on the back and the front has 2. The back of the truck is lighter than the front.*
>
> *It has more weight on that side* [the back of the truck] (prediction that it would go further when released backwards).

Some groups also introduced contextual knowledge from other sources at this point. They related the truck patterns to other moving objects such as cars, bikes, and trucks:

> *Trucks have to put their brakes on when they go down a hill because they go fast.*

Unsurprisingly, since the investigation began from a "ready-made" question, no groups related their contextual knowledge (slope, "weight") to conceptual ideas of cause and effect (i.e. mass, gravity). At no stage did children plan or carry out this activity to test a causal theory of their own, or a scientist's theory that they had discussed in advance. In this sense, it could be argued that they were not actually given an opportunity to plan "scientifically" at all. This lack of a theoretical component became a significant issue in the *Ball Bounce* task, as outlined next.

Ball Bounce discussion

At the discussion stage 3 groups noticed they had made a mistake with their task, or had not tested all of the balls. These groups started again, and this time paid more careful attention to control of variables and taking measurements. However only one of the groups attempted to use a strategy for determining the point of bounce of the ball — clearly a very challenging aspect of this task. As one group said:

> *They* [the balls] *bounce high and fast so we can't get an exact measurement.*

Two other groups identified that they had also made mistakes in their technique that could explain their results. One group said that:

> *It depends on where the ball was bounced on the desk. When it is on a join it does not go very well.*

Another group thought they had measured incorrectly, but they declined the teacher's invitation to re-do the task.

Despite these difficulties, many groups were able to describe patterns with respect to bounciness. Qualitative descriptions were again favoured by a majority of groups. The data in Table 19 are incomplete because some teachers did not ask about patterns of results as the basis for predictions.

Table 19

Children's interpretation of the Ball Bounce *task*

Recognising Patterns and Trends	*Frequency (n)*
Patterns in words only	25
Pattern refers to comparison of numbers/measurements	15

Qualitative comments often made reference to personal theories of causality. The discussions of these Year 8 children, like those of the Year 4 children, focused on observable features of the context:

> *The least is the tennis ball, so the biggest was the less bouncy.*
>
> *Small balls bounce more than heavier balls.*
>
> *Small ones bounce the highest. The smaller the ball the bigger the bounce, and softness of the ball does not bounce much.*

A difficulty with this contextual approach is immediately apparent. There were several material characteristics that could be implicated in bounce and the children often conflated several of these into one personal theory. Section Four identifies the understanding that there can be interactions between variables as a significant developmental step in learning to investigate scientifically. However these children were not given an opportunity to disentangle their multiple theories of causality, nor to discuss testing strategies to see which really were making the difference. It is highly likely that several of these variables do indeed interact but we do not know for certain which and how. Nor, we suspect, did the teachers have resolved and coherent ideas about this.

The variables that were collectively implicated in bounciness are listed in Table 20. While size and weight were most often mentioned, some groups suggested that solid balls would bounce better than those that had air in them or were hollow because they would not compress or lose their shape when they hit a surface. Rubber was seen as a bouncier material than foam. This posed challenges at the prediction stage because the small, heavy, rubber squash ball presented a combination of variables that generated conflict amongst common personal theories of causality.

Table 20

Children's personal theories of the causes of bounce

Cause for the difference in bounce	*Frequency (n)*
Size	21
Weight	16
Material composition	10
Springiness (elasticity when squeezing the ball)	9
Solid or hollow	6
Other	16

Via extensive studies of children's science investigations, researchers in the UK have identified and described 6 broad types of investigations (Watson, Goldsworthy and Wood-Robinson 1999). "Fair testing" is one of these. Another is "Exploring". It seems to us that this rich *Ball Bounce* context presents great exploratory material. The identification of all the various possible causes of bounciness, and clear clarification of personal theories concerning these, would be an engaging focus for an investigation in its own right. Only once that stage has been carried out would children be ready to shape fair tests to begin to differentiate amongst the many possible combinations of variables. As the literature in Section Four makes clear, this process presents many intellectual challenges, but also rich opportunities to teach in ways that could help children actively learn to think more conceptually and metacognitively about what investigation actually entails. We return to the challenge in Section Seven.

Making predictions

After the discussion of their results (if this actually took place) students were shown a squash ball and asked to predict how high it would bounce. They were able to touch the ball but not drop it during this discussion. Fifteen groups immediately recognised the ball and drew their prediction from their contextual knowledge of the game:

> *Squash balls don't really bounce, you have to hit it hard.*
>
> *The squash ball should be warmed up for it to bounce high.*

One teacher made similar links after the group had tested their predictions and found they were wrong:

> *In squash, you have to hit the ball hard for it to bounce.* (For this task, this was the only teacher reference to contextual knowledge that we observed.)

Students often referred to the previous tests to place the squash ball in order of bounciness with the other balls. Some gave numerical estimates as well as reasons for their predictions:

Student: *Not very high — about 80–90 mm.*
Teacher: *Why?*
Student: *Not very high because it's quite heavy.*

The students were confronted with a dilemma at this stage. Depending on the personal theories they had espoused during their reflection on the results, it was possible to justify both "high" and "low" predictions. Those groups who attended to the weight of the ball tended to correctly predict a lower bounce. Those who focused on the material composition (rubber) or the hardness of the ball typically predicted a higher bounce. Those who selected squashiness as the influencing property sometimes said this made balls more bouncy and sometimes said it made them less bouncy.

> *It would bounce higher because the ball can be squashed and bounce back to its shape.*
>
> *Won't bounce very high as it is soft and squashy.*

After making their predictions, most groups tested these with one drop only and many did not attempt to take a measurement as the ball bounced so little. Two groups identified possible errors in their technique as an explanation for the unexpected result. One of these groups felt their measuring was at fault and that they had used the ruler incorrectly, while the other group said the squash ball landed on a crack and they re-did their test. However we noticed that this stage was usually rushed. Once the prediction was tested the task was over, no matter how astonished the children seemed if they had predicted incorrectly. Thus the conceptual dilemmas we have just described were never addressed.

Emptying Rate discussion

At the completion of the *Emptying Rate* task, the teacher made the scripted response: *"Now I would like you to tell me what you found out."* At this point 84 groups could successfully describe a pattern in words. Some also attempted to identify quantitative patterns, saying for example that the detergent took "*double the time of water*". Some groups simply read their data aloud without attempting to extract a pattern at all.

As part of this reporting back, some groups also gave reasons for the patterns they described:

> *It was the pressure and weight of the detergent.*
>
> *The detergent is heavier than water.*
>
> *Water went faster because it was not as thick as detergent.*

Some groups recognised that the size of the hole was implicated in the draining time. Did they register that this key variable had been controlled for them in the provision of the ready-to-use equipment?

> *If the hole was bigger, it* [the detergent] *would have gone faster.*

This comment seems to suggest that this group only considered this variable in relation to the detergent tests, although in the absence of any clarification via teacher probing we cannot be sure. The next comment also suggests that other groups were not thinking about the whole "experimental space" of the entire test series:

> *The bubbles from the detergent clog the hole. So we need time for the bubbles to pop and turn into liquid before they go through.*

On the other hand some groups seemed surprised at the speed with which the detergent had drained. Here, as in the tasks they were to do next, personal theories of cause and effect were inherent in the discussion. Teachers did not probe or elaborate on these, perhaps because this was the opening task. They had scripts to follow and processes to complete.

Reflecting on accuracy

During the discussion of their results, 5 groups identified an error that they believed had affected their results. For some the issue was determining an endpoint:

> *It depends on when we stop it.*
>
> *Stopped the detergent too early.*

One group then re-did the detergent tests at the 6 and 9 cm marks, this time correcting for an endpoint. Others recognised errors with the use of the stopwatch, starting it too early and/or stopping it a bit too late at times. One group simply said:

> *Did it wrong, we didn't have a plan.*

Now, belatedly, some groups recognised why the order of the tests had been so carefully prescribed:

> *You should tip the water out so it won't mix* [with the detergent].

At this point, one group which was alternating between the water and the detergent realised that they had to wash out their container between tests.

Making predictions

The children were next asked to use their results to make a prediction about whether tomato sauce would be faster or slower at emptying from bottle-funnel. Unlike the tasks that were to follow, students did not test tomato sauce after they made their predictions, nor was tomato sauce available for observation. Fortunately, most groups could draw on a rich contextual knowledge of the relevant material properties of tomato sauce. They combined this knowledge with their test findings, and their personal theories about the cause of those findings, in a number of interesting ways. Table 21 shows that 5 groups confused units of time when making their predictions, but most groups at both year levels were able to predict correctly.

Table 21

Patterns in predictions made for the Emptying Rate *task*

Prediction Data	*Frequency Year 4 (n=48)*	*Frequency Year 8 (n=53)*
Word prediction accurate with appropriate time estimate	39	46
Word prediction accurate but estimate of less seconds/minutes	2	3
Incorrect prediction	7	4

Most students predicted correctly that the tomato sauce would take longer to drain:

> *Tomato sauce would take twice as long as the detergent.*

Material properties mentioned to justify this prediction can be grouped in 3 main categories, although students used a wide variety of adjectives for 2 of these:

- Texture (thick, fat, like a jelly, slimy, sloppy, chunky, has blobs in it)
- Viscosity (sticky, less water, more liquidy)
- Weight (heavier)

Some groups who predicted correctly revealed interesting conceptions about the nature of the material world:

> *It* [detergent] *is mixed with chemicals so it takes a longer time and is thicker.*

There was a sense in a number of the comments made that “the more 'chemicals' a material contains, the heavier it will be”.

One Year 8 girl rehearsed a helpful “thought experiment” to visualise the comparative viscosity of detergent and tomato sauce:

> *If you pour detergent on your plate it would run all over it so it* [tomato sauce] *is thicker.*

Five groups correctly predicted that the tomato sauce would take longer to drain, but got confused over units for the measurement of time and so specified, for example, 10 seconds instead of 10 minutes.

Some incorrect predictions also involved interesting reasoning:

> *It* [tomato sauce] *has no bubbles in it, so no air in it. No bubbles means it can come out faster than detergent.*

Other incorrect predictions seemed to be made to avoid a clear decision:

> *Between the two.*

A SHORT COMMENT ON TEACHER ACTIONS

Although teachers are trained to be NEMP facilitators, we saw differences in approach that seemed to us to impact on the progress of the children’s actions and thoughts. While the focus of their work was on *assessing* children’s abilities with respect to the planned tasks, it became very evident that some teachers’ actions supported children to reveal what they could do and/or did know, and that not all groups got that type of support.

Table 22

Patterns in teacher interactions with groups/tasks

Category	***Emptying Rate*** *(n=101)*	***Truck Track*** *(n=51)*	***Ball Bounce*** *(n=52)*
Interruptions to re-emphasise instructions	26	5	5
Focus of the discussion is on results	28	24	10
Focus of discussion is on the meaning of purpose	3	3	0
Focus is on language (both difficult words, e.g. accurate, and science words, e.g. viscosity)	1	0	1
Teacher makes links to other relevant contextual knowledge	8	0	1
Extra variables introduced by the teacher	No	Yes	Yes

We have already noted that the preliminary planning discussion emphasised roles for children and that this was a scripted feature of each task. Perhaps because it was the first task, and because the method was clearly specified in advance, teachers made many more interruptions to re-emphasise method in the *Emptying Rate* task. Some teachers used hand gestures and/or demonstration runs to give both visual and verbal instructions. This was clearly helpful for some more hesitant groups. Some asked children to demonstrate the use of the stopwatch, even when the children had said they could do this. It was apparent that in fact, many did need help with this.

In all 3 tasks teacher-student discussions were seldom focused on purposes of the tasks, nor on the *meaning* children drew from their results. Table 22 also shows that a focus on the language demands inherent in the tasks (whether less familiar everyday words, or the specialist language of scientific terms) was almost never a feature of the teachers' task facilitation. Teachers who asked probing questions *"Why do you think that?"* or *"I wonder why…"* drew more responses from children.

Some teachers did make links to children's contextual knowledge, especially in the *Emptying Rate* task when predictions about tomato sauce were to be made:

> *We put it* [tomato sauce] *on our chips.*
>
> *What we put onto our hamburgers.*

Some children did indicate that this allowed them to compare the substances mentally:

> *Wattie's tomato sauce. Oh yeah, it would be thicker.*

Some teachers introduced extra variables by the manner in which they facilitated the task. This was particularly an issue for the *Truck Track* task. As already noted, the mat often had semi-permanent bumps or creases because it was folded in the same way after the conclusion of each testing episode. The ramp was not always set up in relation to the mat as directed by the photograph provided. However some teachers actually had the foresight to modify this arrangement to maximise the mat distance for the truck's longer runs. Sometimes the mat was set up over a join in two or more tables so that trucks hitting this bump stopped or slowed down. In some cases a teacher moved the students to the floor when this happened. Other teachers ignored this issue. Ten groups of *Ball Bounce* students also had to contend with cracks between desks where the experimental space was set up over 2 or more tables.

SECTION FOUR: THE DEVELOPMENT OF CHILDREN'S INVESTIGATIVE SKILLS IN SCIENCE

INTRODUCTION

In this research we set ourselves the challenge of clarifying ideas about progression in the development of children's investigative skills. Our initial intention was to compare the actions of Year 4 and Year 8 children, as described in the previous section, with the model of progression represented on the Exemplars Matrix. However, as the research has unfolded we have also searched for reports of investigations from the fields of cognitive psychology and science education in which researchers have aimed to describe and explain various patterns in the development of children's investigative skills.

Our purpose in comparing these different sources of information has been to think critically about what progression actually entails. We also wanted to locate and summarise findings that might give classroom teachers useful ideas for actively developing the investigative skills of the children they work with. These intentions have, however, highlighted a tension that exists in any efforts to exemplify children's work in a staged manner that compares children of the same/different stages and ages. Should such comparisons be based on what the "average" or "normal" child can and does do, or on what is possible for children who are fortunate enough to experience specific types of exemplary teaching?

This tension is not easily resolved. If naturalistic observations are taken as a basis for comparison, then observations of actual children working on science investigations should be the most important source of information. Our NEMP analysis obviously fits this method of investigation, as does the Exemplars project. This "common sense" way of approaching developmental issues is not, however, unproblematic. Contemporary science philosophy emphasises the *theory-laden* nature of observations (Chalmers, 1982). What we attend to when observing children in the process of carrying out investigations very much depends on the theories of progression that drive our thinking. If these theories are taken for granted, we are in danger of assuming an obviousness to development that limits the possibilities for challenging what might be achieved with different types of teaching. To help tease out this issue, some ideas central to theories of progression are briefly discussed in the first sub-section below. Following that, research that does attempt to develop specific and explicit theories of progression in the development of science investigation skills is summarised.

THE NATURE OF "PROGRESSION" IN SCIENCE LEARNING

We expect children to make progress in their learning during their years at school. We expect teachers to "make a difference" to that progress, and to effectively enhance children's natural tendencies to learn. However, we have already noted that there are differing ways in which progression can be viewed. This sub-section highlights some theoretical ideas that underpin various common models of progression, with a particular focus on those views that are likely to be taken for granted as "common sense", and perhaps only recognised in hindsight (Barker, 2000). These are the views that would seem to be in particular need of examination, if ideas about progression are to be rethought by curriculum developers, classroom teachers, and those who monitor and evaluate the results of teachers' work.

Learning theories and progression

Theories of learning frequently inform views of progress. Some views of progression are focused more on the "breadth" and "extent" of children's knowledge — that is, they take quantitative "knowledge acquisition" models of learning. Other views are premised on more qualitative "conceptual change" models of learning (Harrison, Simon, and Watson, 2000). The "acquisition" models align with *behaviourist* (building block) or *developmental* (staircase) theories of learning, while the "change" models align with *constructivist* theories of learning (Biddulph and Carr, 1999). More *ecological* theories of learning (humanistic and enactivist), as also outlined by Biddulph and Carr, raise quite different types of questions about purposes for learning science, and may align with ideas of progression centred around changing *goals* for science education at different stages of schooling (Fensham, 1994). These ecological theories are not discussed further in this section because the focus of this report is very much on the development of traditional types of classroom-based science investigations.

Behaviourist and developmental theories of learning typically derive their progressions from the "internal logic of the academic disciplines of science, as they have developed, particularly conceptually since the mid 20th century" (Fensham, 1994, p. 79). Such progressions are ordered according to what is seen as the logical order of development of subtopics, with curriculum organisation based on criteria such as a move from simple relationships to complex patterns, or concrete experiences to ideas that involve abstract logical thinking. Constructivist theories of learning may lead to progressions based on conceptual change research, so that the focus moves from the logic of science to patterns of children's thinking and subsequent next learning steps (Harrison et al, 2000). Frequently, curricula are structured in ways that present an amalgamation of these two different approaches.

Developmental theories and progression

Progression can be theorised as having a *developmental basis*. In this view children are not "ready" for particular sorts of learning until they have reached some physical and/or psychological maturation stage. This has been described as a "folk theory", albeit one that is still widely held amongst the teaching profession (Watson, 1996). Watson reviews recent research concerning the nature and potential of young children's learning to describe a quite different type of view from the folk idea of "readiness":

> *Readiness, on this [new] view, thus requires that children recognise beliefs qua beliefs — that is, as not synonymous with reality and subject to revision. The revision of belief is a domain-general, metarepresentational ability that lies at the very core of intentional learning. It underlies the child's ability to intentionally enrich or change the core explanatory principles in a naïve theory, and thereby, the ability to learn from formal instruction* (Watson 1996, p. 160).

Here the formal theory-based ways of thinking practised by scientists have been compared in essence with the processes that children follow as they revise and expand their mental models and schemas. This view of "readiness" potentially aligns with constructivist and/or ecological theories of learning, drawing attention to the importance of metacognitive as well as the cognitive thinking. Progression is premised on the concept that children think in “theory-like” ways from a young age – something that has been borne out by recent research with very young children (Watson, 1996).

These new biological insights raise important challenges to widely-held Piagetian notions of *stages* of *mental operations* that have strongly influenced the way in which children’s learning in science has been structured in the past. Piaget’s ideas, in particular the notion of a “formal operations” stage of development, have arguably formed an important, often implicit, part of the theoretical framework within which both curriculum itself and observations of children’s actual learning achievements have been assessed and interpreted since the middle of the twentieth century. However, Metz (1995) mounted a strong critique of the Piagetian position. If students are shielded from experiences that will foster their awareness of warrants for their own beliefs, on the grounds that such ideas about their experiences are as yet too abstract, they cannot begin to gain key skills needed to appreciate the specific ways of thinking involved in “being scientific”. Where teachers do not expect that students could make such progress at a relatively young age, “folk theories” of readiness become self-fulfilling in the limitations placed on children’s learning opportunities.

The point just made returns us squarely to the tension between an “exemplary” model of progression, that demonstrates what children *could* achieve, and a more “naturalistic” one that orders what children typically *do* achieve at various ages as they experience conventional models of teaching and curriculum. The framework presented in Section Five draws partly on research that presents evidence that children can achieve much more than is typically expected of them – but only if they are exposed to the sorts of constructivist models of learning that foster their personal epistemological development. However, we have attempted to side-step the normative/exemplary dilemma by avoiding the use of “level” numbers or age-group alignments, instead giving the 5 identified framework clusters names that we hope fairly reflect their overall place in the sequences of progressions. We have done this partly to avoid confusion with the idea of curriculum levels, but also because we want to emphasise that this is a developmental sequence where any one stage might be applicable to students of a wide range of ages, depending on their previous learning experiences. Equally, any one student is likely to vacillate at the boundaries of two stages during transition times, sometimes showing characteristics of one, sometimes of the other.

Fensham (1994) makes the point that frameworks that seek to unify views of progression across the entire school curriculum are a phenomenon of the late 1980s and beyond. Before that time, he suggests, views of purposes for learning, and thus beliefs about what constitutes progress, were perceived quite differently at different stages of schooling. The Exemplars Matrix mainly seeks to exemplify teaching and learning at primary school levels, and is restricted to Levels 1–5 of *SNZC*. Its developers were certainly aware that the purposes for science teaching perceived by primary teachers can be very different from those perceived by secondary teachers (Coles, MacIntyre, and Radford, 2001). Whether the Exemplars Matrix could be extended right through all the levels of the curriculum is an interesting matter for debate.

The meta-level framework presented in Section Five develops descriptors for learning continua that arguably extend from earliest school learning right up to undergraduate university study. We felt it was important to do this because the interplay between ideas (whether children’s own ideas or scientists' ideas) and the development of children's investigative skills is something that we believe is in need of critical attention. The analysis we have undertaken suggests that thinking and reasoning within coherently developed frameworks of *scientifically* correct theories only become important in the final stage – that is at the upper levels of secondary schooling and beyond. Fensham's alternative model of progression, slanted towards a "science for all" view of purposes for learning, also emphasises a "solid foundation" [of scientifically accurate theories] only at the upper secondary level (Fensham, 1994). This seems to us to free the way for rethinking the emphasis on content at lower curriculum levels, something that has been widely advocated in recent international discussion of reform in science education (see, for example, Millar and Osborne, 1998; American Association for the Advancement of Science, 2001).

THEORETICAL SUPPORT FOR THE FRAMEWORK

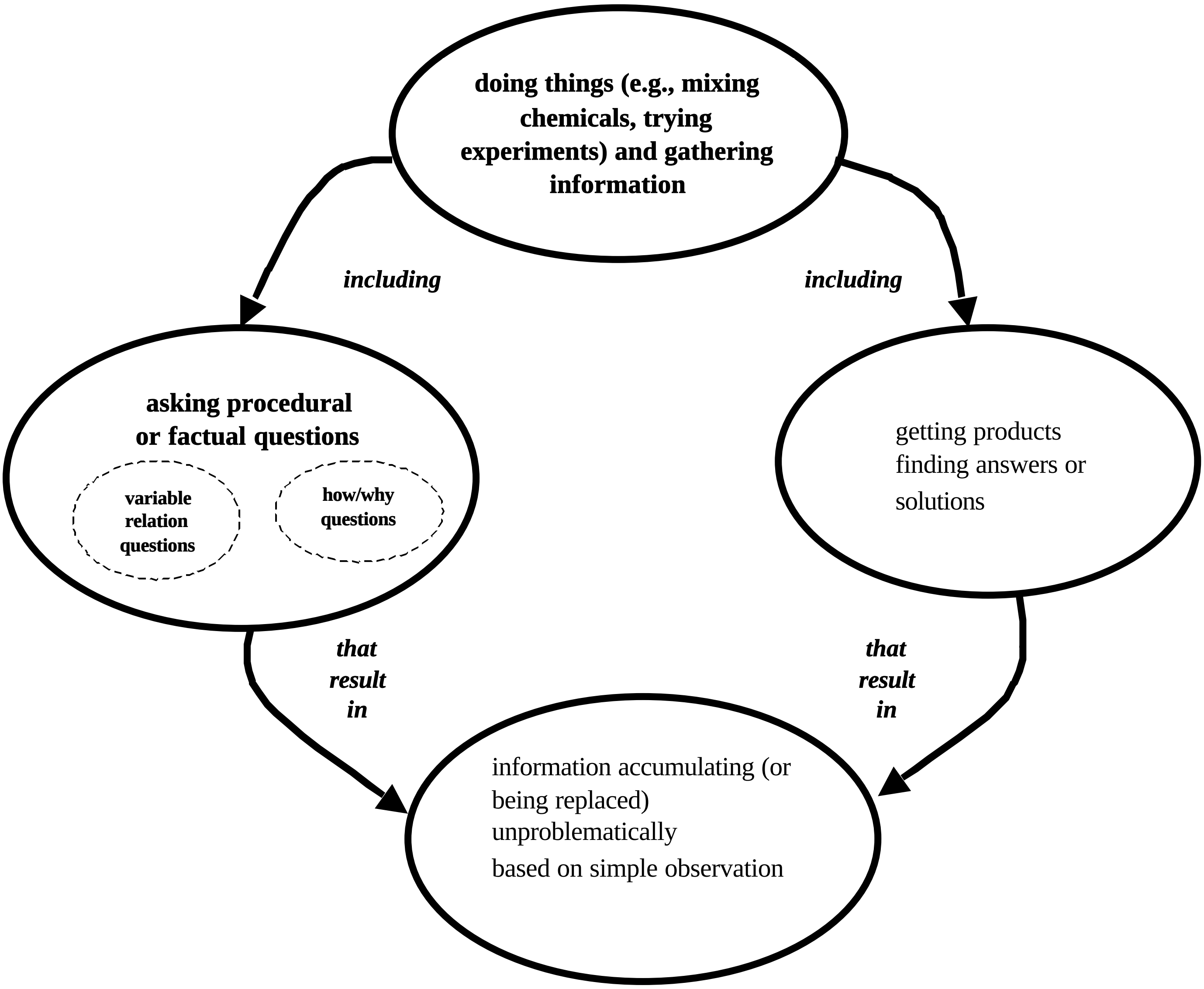
In seeking to design this developmental framework, we have drawn heavily on research. One paper proved particularly helpful in providing a starting point for reflecting on what primary school children can achieve if freed from prevailing expectations of "normal" development. Smith, Maclin, Houghton, and Hennessey (2000) compared 2 grade six classes of American students who were closely matched demographically. One group of children had been exposed to a constructivist science education programme from the outset of their schooling. The same teacher had worked with these students every year, during 3 hours of science learning time per week, with the aim of developing in every student a metacognitive awareness of theory building as personal learning process, as well as the primary goal of scientific inquiry. Teaching that could achieve such aims was described in Chapter Five of the recent literature review on effective teaching to raise achievement for all New Zealand students in science (Hipkins et al., 2002). The other class was taught by a teacher who took a keen but more conventional interest in science learning, focusing on the "Science Fair" as an important culminating goal for student investigations. The research team carried out "nature of science" (NOS) interviews with each child from both classes, using previously validated semi-structured interview questions, and the results were analysed within a careful framework that allowed for inter-rater validation.

From these results, Smith et al (2000) described a developmental sequence of children's understandings of the nature of scientific inquiry. They described 3 levels in the development of children's understanding, with transitions between these characterised as level 1.5, and level 2.5. In their sequence, level 1 represents a naïve "knowledge unproblematic" view of science, and level 3 represents a sophisticated and well-developed epistemology that reflects contemporary "knowledge problematic" thought about the provisional nature of science theorising. The details of this developmental sequence have made a significant contribution to the shaping of the framework presented in Section Five. Specifically, the columns headed "meta-level view of knowledge", "personal knowledge skills", "view of scientists" goals/methods", and "personal meta-values" were initially collated using data drawn from these research findings and the associated discussion. Some aspects of the column "meta-level view of investigations" were also drawn from this research, especially for the first 3 clusters.

Smith et al. (2000) found that students from the constructivist classroom averaged level 2 or above on the developmental sequence they had devised, while those in the traditional classroom averaged between level 1 and level 1.37. They found very little overlap between 2 groups, in any of the categories that they developed. The diagrams they developed to represent a "modal epistemology" for each group vividly illustrate the richness of the nature of science thinking that had been developed by the children in the constructivist classroom, compared with the children in the traditional classroom. We have redrawn these diagrams for presentation below. Note the richer links and deeper NOS ideas of the children in the "science as knowledge building" classroom (Figure 2). The centrally linked idea of knowledge building as a collaborative activity is completely missing in the NOS understandings of the children in the traditional classroom (Figure 1).

Figure 1

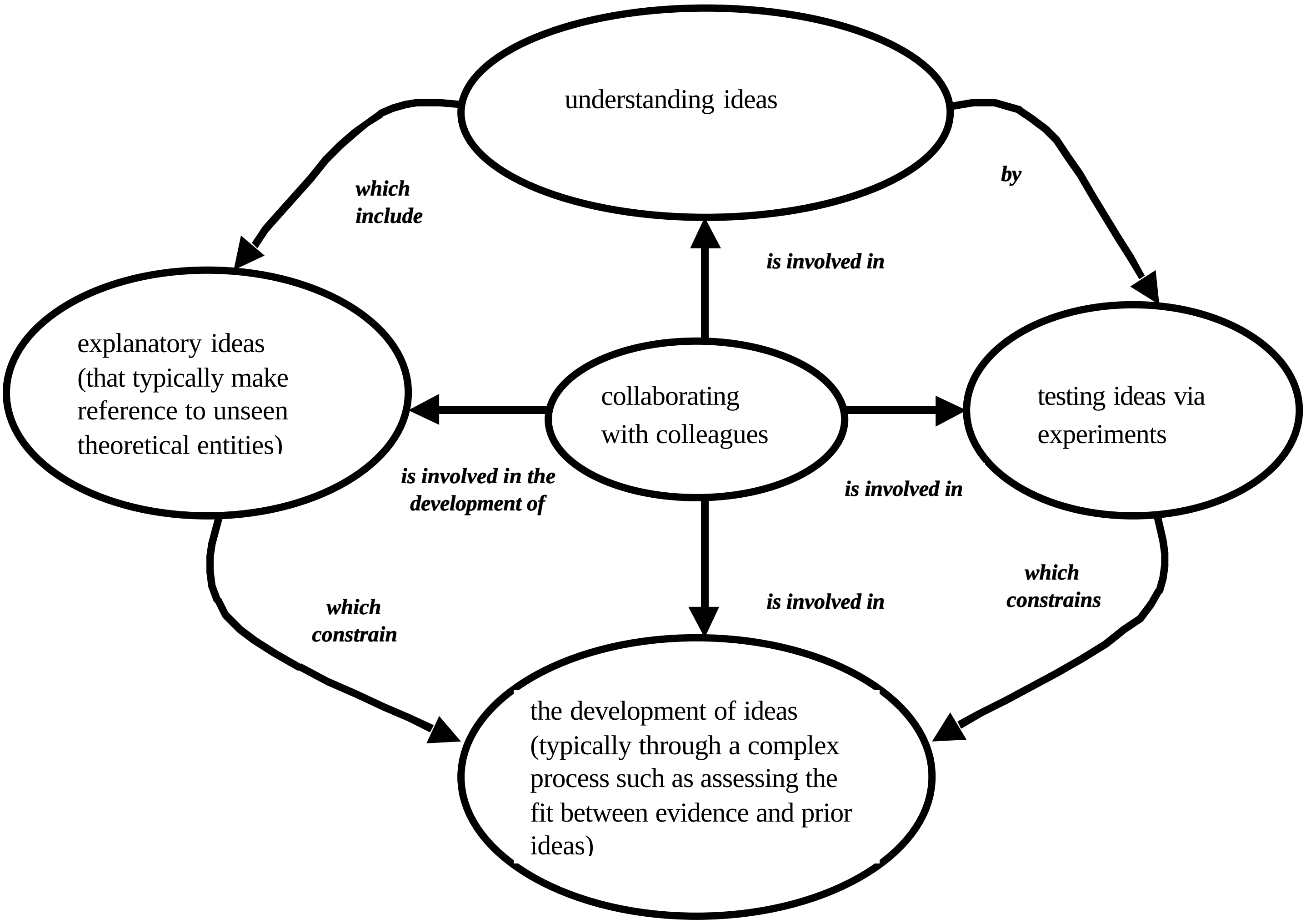
A "modal epistemology" for sixth grade students in a traditional American class



Source: Smith et al., 2000, p. 382

Figure 2

A "modal epistemology" for sixth grade students who had experienced teaching with a specific focus on knowledge building processes



Source: Smith et al., 2000, p. 381

Their findings lead these researchers to suggest that the achievement record of the students in the constructivist classroom "need not be limited to just a few precocious sixth graders, but is well within the grasp of an entire classroom of students" (Smith et al, 2000, p. 398). Level 2 of their sequence is represented as the third cluster of our framework: *A developing sense of knowledge testing*. This was the *lower* level of achievement for one class but above the level of achievement of all the students in the other class Smith et al researched. Tellingly, these researchers suggest that the main factor preventing some of the "constructivist classroom" students from achieving the top level (level 3) of their scheme was that they did not yet hold a coherent and overarching framework of science explanatory knowledge. As they pointed out, students cannot be expected to develop such frameworks at least until their senior years of secondary school. Nevertheless, it is food for thought that these still relatively young students, as a whole class, achieved what traditional views of progression might judge as being beyond many junior to mid-secondary students.

In a critical review of research on children's NOS understandings, Hogan (2000) alerts us to a potential gap in our knowledge of children's development if they are only asked about their ideas of scientists' science. Such knowledge is declarative – a *distal* knowledge that may or may not influence the way in which children carry out their own science investigations (that is, their *proximal* NOS knowledge). The analysis of the NEMP tapes, reported in Section Three, illustrates how children often **do** much more than they **say** when carrying out investigations, so the distal/proximal distinction could make a significant difference to the way ideas about progression in children's learning are interpreted and ordered. With this caution in mind, we note here that, while Smith et al. (2000) do report children's comments on their own investigative work, the interview questions primarily address their distal knowledge – that is "what scientists do", which does not necessarily connect to what children actually do. To address this issue, we next aligned the preliminary framework clusters with research findings from projects that describe progressions in what children actually do when they carry out their own science investigations.

THE ROLE OF MENTAL MODELS IN PROGRESSION

Kuhn, Black, Keselman, and Kaplan (2000) provide a different but complementary perspective to the research of Smith et al. Kuhn and her colleagues were also interested in the impact of meta-levels of cognition on children's progress. They hypothesised that success in learning via active inquiry may be promoted or hindered by the mental model of causality currently held. In this view, constraints on children's ability to progress could be top-down rather than the more usually considered bottom-up constraints posed by the limitations to the actual learning experiences provided. Kuhn et al. (2000) tested their hypothesis via the development of a computer simulation that tracked children's decision making when solving problems that involved discerning which were the causal variables in a multivariable "real world" context. The task involved determining the causes of flooding in a simulated lake by running trials of different variable combinations and levels (e.g. water temperature = hot or cold: soil depth = deep or shallow). The researchers acknowledge that "mental models of any sort remain essentially unobservable theoretical constructs" (p. 516). However, they present the patterns of responses made by the children as evidence that supports the existence of a developmental sequence of mental models of causality.

At the beginning of this developmental sequence children focus exclusively on outcomes – if an event is triggered and a certain outcome occurs, then the two are directly and unproblematically related to each other. With a mental model such as this each investigative episode has its own discrete outcomes and the idea of controlled comparisons between adjacent events has yet to be developed. During the NEMP tape analysis we certainly saw many instances where this type of thinking appeared to be in operation, and we have described some of these in Section Three. While other explanations for this type of "discrete episodes" response can be proposed (for example, limitations of memory space – see below) this mental model does appear to align strongly with the naïve, "knowledge as true facts" view described as the beginning of the developmental sequence of Smith et al (2000). With this juxtaposition we began the alignment of Kuhn et al's research with the emergent clusters of the framework.

Kuhn et al. (2000) contrast the outcomes-focused mental model just outlined with what they call an *analysis* model. As this new mental model gradually develops children come to recognise that there can be a relationship between separate variables, and this in turn is associated with the realisation that it is necessary to think more carefully about the actual cause of a particular outcome. Children will begin to demonstrate their emergent development of an analysis mental model when they correctly **choose** between possible variable combinations to identify unconfounded “fair tests”. Emergent “analysis” from this mental model perspective seems to us to align readily with “explanation” from the epistemological perspective of Smith et al (2000) and so the second framework cluster also falls into alignment. (This emphasis on choosing — as opposed to actually producing/designing/planning — seems itself to be significant as a developmental step and will be further discussed below.)

Kuhn et al (2000) then describe several subsequent developmental stages of the “analysis” mental model of causality. Children progress as they learn to discern the separate effects that two or more variables may exert on an outcome – an *additive* stage. They are beginning to show a “metastrategic understanding” (p. 513) as demonstrated by their ability to give a simple explanation for **why** one potential “fair test” is better than another. We did not see any evidence of this type of thinking during the NEMP task re-analysis. However, we also note that the tasks presented, while having the potential for multivariable analysis, were not developed and presented to the children in a manner that would allow them to display such thinking.

This “additive outcomes” mental model seems to us to align with the notion that scientists need to “test ideas”, as characterised in the next developmental stage described by Smith et al (2000). Later still, the additive mental model can develop into a more *interactive* model as children expand their repertoire of experiences of situations to include more complex episodes where variables affect the outcome in some combined way (as potentially seen in the *Ball Bounce* task). Children who are working with investigations at this level will recognise a need to manage two or more independent variables simultaneously, something that has also been associated with progression by science educators who have taken a “task based” approach to their theoretical reasoning (Gott and Mashiter, 1994). While the link is more tenuous, it could be argued that this stage aligns with the idea that evidence has the potential to **disconfirm** theories – an idea that is identified at the fourth level of Smith et al*’s* epistemological developmental sequence. This potential link will be further explored through the research projects of Chinn and Malhotra (2002), and of Zohar (1995) and her colleagues, that are introduced below.

The final step in Kuhn’s developmental sequence of mental models of causality introduces the development of more complex interactive mental models where multiple possible pathways to outcomes must be taken into account. Kuhn et al (2000) note that this opens up the possibility that the co-ordination of theory with new evidence can proceed in several different ways and hence it is necessary to represent evidence independently of theory as far as is possible – a skill that they identify as “the hallmark of mature or skilled scientific thinking” (p. 519). This seems to us to align with the overarching explanatory frameworks for scientific theories that Smith et al (2000) identify as a key feature of their “level 3” students. In all likelihood, many school students will never reach this level. The nature of appropriate learning opportunities that can help students develop their metacognitive awareness of data evaluation is still a matter of debate in the research community (Chinn and Brewer, 2001) and we have not developed this aspect further in this report.

LEARNING ABOUT "FAIR TESTING"

Researchers from the cognitive psychology field have given quite detailed attention to the manner in which children learn to manipulate investigations using controlled variable strategies (known in this literature as CVS investigations). Such strategies are given considerable emphasis at the primary school levels of *SNZC*, where they are denoted by the use of the phrase "fair testing". This type of investigation has been defined as one in which "students decide to change an independent variable, observe the effect on a dependent variable, and control other key variables" (Watson, Goldsworthy, and Wood-Robinson, 2000, p. 71). All 3 investigations reported in Section Three anticipated "fair tests" of this sort as the appropriate means of carrying out the set task.

The notion of "fair testing" has proved to be a powerful metaphor. This approach to science investigations has been widely adopted amongst teachers. Science education research in the UK has shown that, at least at the primary school level, this has been to the detriment of the development of children's knowledge of other forms of scientific investigations. Furthermore, these researchers report of a very narrow range of contexts in which such investigations are typically carried out, with 30 percent of all the investigations using just 4 contexts: thermal insulation, dissolving, pulse/breathing rate with exercise, and friction (Watson et al, 2000). Watson and his colleagues express concerns about the NOS messages implicit in this narrow focus on just one type of investigation, and point out that such experiences also provide inadequate opportunities for children to learn about the relationships between the development of scientific theories and empirical evidence. These concerns reinforced our belief that fair testing warranted close scrutiny as we shaped our developmental framework. The research reported next has particularly influenced our thinking about how the framework clusters can be used to encourage active teaching that leads to progress in the development of "fair testing" science investigative skills.

Differentiating between choosing and carrying out

When presented with simple **choices**, young children can recognise "fair" (unconfounded) CVS tests even if they are not yet able to justify their choices (Kuhn et al, 2000), and well before they can actually **produce** such tests for themselves. Schauble (1996) cites a range of studies where this has been confirmed. A rare longitudinal study of one cohort of German children, that tracked their development across the primary school years, also confirms this pattern (Bullock and Ziegler, 1999). These researchers suggest that the distinction provides important new insights into children's developing investigative skills:

> *..when children are asked to produce experimental tests, their performance suggests that they do not understand the logic of experimental control. In contrast, when children are asked to choose a controlled test, their performance at least by fourth grade suggests that they do understand the logic of experimental control. Furthermore, and this is important, of those who chose controlled tests, more than 50% of the fourth graders, about 80% of the fifth graders, and almost all of the sixth graders also justified this choice in terms of controlling variables, suggesting that their understanding is also somewhat explicit* (Bullock and Ziegler, 1999, p. 45).

It seems to us that **making choices** has been a neglected factor in descriptions of progression, and we identify it as a focus for the active teaching of investigative skills to children who seem to be located developmentally within any one of the first 3 clusters described within our framework. We cannot report on this ability from our observations of the NEMP tasks because the children were required to actually produce fair tests in all 3 tasks. It seems to us that it would be a relatively simple matter to design NEMP tasks and other learning materials that provide examples where children choose the "fair test" from pairs or sequences provided to them, and then discuss the reasons for their choices. Furthermore, the work of Kuhn et al. (2000) suggests to us that choosing and justifying those choices form a progression of their own, with the increasing ability to explicitly justify choices in terms of CVS strategies linked with the development of more sophisticated mental models of causality.

Thinking about variables during investigations

Clearly, children do also show progression in their ability to **produce** fair tests. Chen and Klahr (1999) investigated changes in children's ability to produce fair tests after explicit instruction. They found that the youngest children in their study (American grade 2) improved marginally after instruction but had difficulty remembering the skills they had been taught when transferring them to other similar tasks. Grade 3 students successfully transferred their new skills to new tasks, while grade 4 children were able to retain the skills for longer, doing better than untrained children of same age on pencil and paper tests 7 months later. What factors might account for this progression? From the perspective of Kuhn et al. (2000) we can hypothesise that the younger children do not yet have the mental models of causality that would prompt them to see the necessity for using CVS strategies consistently. However, the studies reported next suggest a complex mix of contributing factors. It seems to us that this research has immediate potential to inform good teaching practice.

Schauble (1996) and her team compared the investigative skills of 10 children, with an average age of 11, with 10 adults who had had no formal science training. Both groups carried out 2 series of investigations across several weeks, and the researchers tracked the development of their skills as they went along. Thus, this was a *microgenetic* developmental study, designed to track the actual development of each individual across the course of the intervention. One task involved determining the causal variables for patterns of boat movement in a model canal. Some variables had counter-intuitive effects and some interacted with each other, complicating the causal patterns to be unravelled. The second task involved the suspension of objects in water at different depths, using a spring. This task had a similar level of variable complexity to the canal task. In part of a different research programme, Toth, Klahr, and Chen (2000) took a complimentary focus. They investigated children's ability to learn to create their own CVS investigations and to evaluate CVS investigations designed by others. The researchers used carefully structured instruction strategies that had been informed by the earlier research of Chen and Klahr (1999). The practical implications of this work will be reported in Section Six. The areas of congruence between the reported findings from these 2 research programmes are outlined next.

Schauble (1996) and her colleagues found that the children in their study experienced varying degrees of difficulty in making valid inferences about certain types of variables. They identify single causal variables (inferences of **inclusion**) as the easiest to master. Inclusion episodes involve "fair" (unconfounded CVS) tests in contexts that allow children to confirm their own theories with the evidence they generate. Support for this argument is provided by the work of Chin and Malhotra (2002), which we outline in the next sub-section. Toth et al. (2000) similarly found that children can identify and use simple unconfounded CVS tests more easily than other types in the absence of formal instruction.

We saw episodes during the NEMP task analysis where the agreement between the unfolding events and the children's ideas about what they thought should happen encouraged them and gave them confidence that they were working correctly.

Schauble (1996) reports that non-causal variables (inferences of **exclusion**) are somewhat harder to master because they require children to integrate information across several trials, comparing the effects of controlling different variables as they go. Similarly Toth et al found that non-contrastive investigations (where children had to identify an irrelevant variable) were not quite as easy for them to recognise, or to learn to produce with appropriate instruction. It seems to us that the possibility of exclusion is unlikely to even occur to children until they hold at least an emergent *analysis* model on Kuhn's mental models progression, and so we have aligned these ideas in the framework clusters we have developed.

Indeterminate CVS tests – those that require recognition that evidence does **not** exist to support a resolved conclusion – are also more difficult for both children and adults to recognise (Schauble, 1996). Similarly again, Toth et al found that single and multiply confounded CVS tests are the hardest for children to identify – and that, following formal instruction, children did not show as much improvement in evaluating these types of investigations as they did with unconfounded and non-contrastive investigations. By definition, confounded CVS tests will obviously generate indeterminate data, although indeterminacy can also be linked to more complex considerations of theory/evidence links, as explored below. Nevertheless, it seemed to us that these findings align logically with Smith et al's (2000) criteria concerning ability to see inconsistencies in peers/personal thinking, and also with the development of a more *interactive-analysis* mental model from Kuhn et al's (2000) work. Thus, while Schauble does not differentiate degrees of difficulty between exclusion and indeterminacy, we chose to place these in clusters 4 and 5 respectively in our evolving framework.

Progression in planning

Schauble (1996) reports that adults were typically more systematic than children when working through sequences of trials. Her team identified 4 patterns of increasing sophistication in systematic planning and carrying out.

At the lowest level there was no evidence of a plan or system. Children working in this manner would typically draw conclusions from single tests, or sometimes make post-hoc comparisons with previous tests. This clearly aligns with the meta-level understanding of each investigation as a separate episode, already identified as a feature of cluster 1. Schauble next introduces the term *local chaining* (p. 109) to describe emergent attempts at being more systematic. Children who worked in this way focused on pairs of trials, without relating these to a larger plan structure, and this could lead them to repeat trials they had already carried out without recognising that they had done so. This seems to us to represent an emergent recognition of the need to explain investigations in terms of more than one episode, and so we have placed this idea in cluster 2.

Some children did develop a simple plan and initially used a VOTAT – vary one thing at a time – strategy. However, after 5 or more such trials they "lost their place" (Schauble, 1996, p. 109) in the face of the accumulating body of evidence, and so the initial plan was abandoned or forgotten. This finding resonates with reports from the longitudinal study of German primary school children (Bullock and Ziegler, 1999). The German researchers checked for differences between children who could recognise, but not produce, controlled experiments at age 12 and those who could consistently both recognise and produce simple CVS tests. The 2 most important differences they found were having access to explicit verbal knowledge about experimental control and **memory span**. We have used the juxtaposition of these 2 quite different studies to link the ability to plan and carry out sequences of investigations with increasing memory capacity in cluster 3 of the evolving framework.

Schauble (1996) reports that some participants had the ability to produce a “global plan” that “reflected the overall structure of the experimental space” (p. 109). Adults were more likely than children to review such a plan as they proceeded through a series of trials. A plan that can anticipate a whole sequence of trials obviously places demands on the memory. Some researchers have reported success with using “representational scaffolding” in the form of tables, charts, and computer mapping programs, to help children develop a more global view of the whole experimental space (Toth, Suthers, and Lesgold, 2002). Similarly, the researchers in the German longitudinal study found that a simple training intervention that “fostered representation of the entire problem space” (p. 51) made a “dramatic difference” to the ability of 11-year-old children to produce unconfounded CVS tests (Bullock and Ziegler, 1999). In the absence of such scaffolding, the ability to plan globally presupposes a meta-representational awareness of the possibility of relationships between combinations of variables – that is, an *interactive-analysis* mental model (Kuhn et al., 2000). Thus we tentatively placed the emergent use of global planning in the fourth cluster, and its consistent use in cluster 5 of the evolving framework.

Support for this decision came from the findings of research that investigated the causal reasoning of a group of community college students, with an explicit focus on how they actually reasoned about interactions between variables (Zohar, 1995). Noting that the more usual focus on single variable CVS strategies neglects instances of **co-variation** between variables, this research team hypothesised that strategic competence in producing CVS tests is a “necessary but insufficient” (p. 1040), condition for carrying out an effective inquiry. They designed 4 “microworld” investigations with different contexts but similar structures. In each of these microworlds 2 variables were non-causal, 1 had a simple causal effect, 1 had a causal effect only when in interaction with one specific level of other causal variable, and one 3-level variable had a partial, curvi-linear effect. This complexity meant that it was possible to run pairs of CVS tests that, when considered in isolation, appeared to generate conflicting evidence. To illustrate, one task that was adapted from the earlier research of Scauble involved determining the factors that influenced boat speed on a canal. If weight and boat size are considered as the focus variables, the 4 CVS tests that should be carried out are: small boat with weight/small boat without weight; and large boat with weight/large boat without weight. However, in the microworld designed:

> *The two conclusions described here contradict each other; neither of these two sets of experiments considered separately provides a full explanation of the phenomenon under investigation. To resolve this contradiction, a broader picture must be drawn: A second-order investigation needs to be carried out to discover the interactive effect of weight and boat size. To make the interactive inference that the causal effect of weight depends on the size of the boat and that it makes a difference in a small boat but not in*

> *large boat, the investigation requires the full set of four experiments. Only such a double set of comparisons may lead to an interaction inference* (Zohar, 1995, p. 1046).

Thus, controlling variables in simple CVS tests is a necessary but insufficient investigative strategy in this type of situation. Zohar goes on to compare the stumbling confusion of some of the lay participants with the orderly manner in which two "expert" investigators devised meta-strategic plans that allowed them to detect the interactions relatively quickly. These findings clearly align the unprompted use of global planning with a well-developed "interactive analysis" mental model. The use of lay adults in this study is a timely reminder that the fifth cluster of the developmental framework we have devised may never be reached in the absence of opportunities to build a sophisticated metacognitive awareness of investigative strategies. The provision of appropriately structured and supported learning experiences is a huge challenge for teachers and for those who are teacher educators.

ATTENDING TO THEORY/EVIDENCE LINKS

Most of the "fair testing" research reported above has focused on what Klahr (1999) calls the "experiment space" of the investigative process. He points out that there are actually 3 major interdependent processes in scientific discovery, although very few studies of children's investigative skills actually attend to all 3. The other 2 are the "hypothesis space", and "evidence evaluation". Each of the 3 sets of processes can be further subdivided into domain specific and domain general cells, depending on whether the knowledge being drawn on is specific to the focus investigation or more general.

The literature reported here was originally intended to inform our observations of children's skills as we observed these on the NEMP tapes, and to assist us with our analysis of the fit between these observations and the Exemplars Matrix. Each of the 3 NEMP tasks began with a "ready made" investigative situation, in which children were free only to decide on the details of how they would manage a CVS strategy. Although some "planning" was required of the children, this does not constitute thinking in the "hypothesis space", which would have required the children to decide on the parameters of the investigative situation, grounding their planning decisions in their previous thinking, experiences, and questions. Similarly, evaluation implies a discussion of whether or not these initial ideas were supported by the evidence produced, thereby linking evaluation back to purposes. Without the hypothesis space, the evaluation space may also remain meaningless.

As reported in Section Three, many NEMP groups had scant understanding of the purposes of the 3 tasks beyond the most obvious requirement for task completion, and their "discussion" was accordingly very brief. In some cases the teacher seemed not to have a sense of theory/evidence links either, and the discussion was inevitably superficial, or truncated by haste to move the children to the next task. Despite the lack of overt reference to theory in the tasks we observed, in this sub-section we do make some links between the research we report and the patterns we observed because of a subtle but important shift in focus. As this sub-section unfolded it became increasingly apparent that questions of progression in children's knowledge and skills cannot be separated from considerations of progression in the selection and structuring of the **contexts** of the investigative tasks. While the role of contexts in determining progression has not been the direct focus of the research we discuss, some issues that emerge seem to provide very useful guidance for teachers.

Children's access to an expanding knowledge base

In their longitudinal study of the development of children's investigative skills, Bullock and Ziegler, (1999) found that about 42 percent of all linear growth in individual children was explained by just 2 variables – **logical reasoning** and **general knowledge.** Presumably, children with a broad general knowledge can envisage more types of episodes in which to ground their cause/effect inferences. Reflecting on the role of "domain specific" knowledge in making personal sense of investigative experiences, Schauble comments that:

> *There are notable differences in the "libraries" of causal mechanisms cited by children and adults, with adults more likely to propose mechanisms that appropriately applied to the situation and that accounted for the observable data. Even when children and adults had similar concepts, they did not necessarily access them with the same facility in the context of these tasks. Participants' attempts to reason about these knowledge-rich systems thus suggest that the contribution of prior knowledge to scientific reasoning may be deeper and more domain specific than most "experimentation strategy" studies have acknowledged* (Schauble, 1996, pp. 117–118).

Libraries of causal events are prior knowledge that is not "general knowledge" in the usual sense, although we have suggested that the two are likely to be interconnected. Based on the research of Smith et al. (2000) we have already indicated the importance of gaining an expanding repertoire of "scientific" ideas in the "personal knowledge skills" aspect of the evolving framework. Clearly, children need to experience a range of rich exploratory contexts to develop this specific type of "general knowledge". We note here that this was also recommended as one finding of the recent literature review of effective pedagogy to raise achievement in science learning (Hipkins et al., 2002).

The development of observation/theory links

The research of Chinn and Malhotra (2002) provides a more detailed analysis of the links between observations children make during investigations, the ideas they bring to these observations, and the learning that they achieve. In an attempt to better describe the cognitive complexities of conceptual change Chinn and Malhotra investigated children's responses to anomalous data during 4 sets of experimental interventions. They designed deceptively simple investigations where the empirical evidence that addressed the question posed could easily seem ambiguous against a "noisy" background (as, for example, determining whether 2 objects of different mass hit the ground at different times if they are dropped at the same time from the same height).

Chinn and Malhotra found what they called an "asymmetrical bias" in children's observations. Those who initially made correct predictions were highly likely to observe accurately. On the other hand, children whose initial predictions were incorrect did not necessarily make matching incorrect observations. Accordingly, Chinn and Malhotra suggest that children's observations are "schema facilitated and not schema determined" (p. 332). Knowing what to look for can help children observe accurately, regardless of the "noise" in a potentially anomalous situation. They further demonstrated that where children were given opportunities to make predictions explicitly based on correct explanations that were provided to them,[8] their observations were likely to be correct and to prompt conceptual change:

> *Conceptual change is impeded largely at observation, and observations are accurate only when children can apply a perceptual schema that helps them detect the outcome of an experiment against a noisy background. Explanations operated by providing students with schema that they could use to guide observations in an ambiguous-stimulus environment* (Chinn and Malhotra, 2002, p. 338).

These findings again support the assertion that evidence that disconfirms personal theories is more difficult for children to process than that which confirms them. However, Chinn and Malhotra also alert us to the critical importance of context. As part of their decision making to take account of the "readiness" of children, teachers could consider the selection of contexts that can allow the observations that children will be required to make to be "cued" to align with the explanatory theories they are likely to find plausible. Much advice is given about the latter aspect of this combination (i.e. potential progression in children's theory development) in traditional curricula, including *SNZC*. We instead turn our attention to the structuring of inquiry contexts in ways that might assist teachers in matching observation tasks to children's stage of skills development.

Awareness of data patterns

Toth, Khlar, and Chen (2000) identified an interesting type of response of "know-all-along students" to the explicit teaching of fair testing skills. These children were already able to consistently produce CVS tests before instruction, and they continued to do so after instruction. However, they became less certain of the **justifications** they made for the strategies they used after CVS instruction, even as they continued to generate unconfounded CVS tests. The researchers suggest that these children were becoming more attentive to variation in data outcomes, and that this had raised their awareness of potential sources of experimental error, thus rendering them less certain of the justifications they gave for the tests they had run.

Might Bullock and Zeiger's finding that memory capacity limits children's ability to conceptualise the whole experimental space help explain this finding? Perhaps attention turns to more subtle features of an investigation only once memory space is freed from a laborious focus on the CVS design and implementation? Certainly it seemed to us that many of the children we observed on the NEMP tapes saw each separate test as a stand-alone event. Their laborious measuring and recording of each result seems to us to have made it unlikely they carried an awareness that there even could be an overall pattern of data generated. Some attention was given to the broadest of such possible patterns in the limited discussion at the end of each investigation, but we think some specific memory aiding strategies could easily be implemented to help make patterns of data variation much more visible to young children.

Probing children's understanding of error in experimental data gathering (Masnick and Klahr, (2001) point out that error is less important when the goal is to compare a **relative** measure of two situations than when it is to measure **absolute** data. They note that children can more confidently predict and justify overall trends when results are categoric than when data are continuous. In their study, even grade 2 children could describe likely sources of variation in trials that involved

---

[8] The researchers caution against using this finding to justify "teaching as telling". Their explanations were provided in the context of a discussion grounded in children's own thinking.

relative rather than absolute measurements. Masnick and Klahr suggest that children have a "nascent understanding" that main effects should be robust, even when there is variability in individual samples. These insights seem to us to hold considerable promise for helping teachers shape contexts in specific ways that support the development of children's investigative skills, especially at the earlier stages of progression. Accordingly, we have used the criterion of collecting categoric rather than continuous data to differentiate early attempts at producing fair tests (cluster two) from later stages of progression. Once continuous data are introduced, it also seems that careful consideration of the contexts in which measurements are to be made could assist in helping children develop awareness of data variation as a key observational skill, as outlined next.

A juxtaposition of our observations of children working on the NEMP tasks with Schauble's (1996) rich descriptions of her participants' investigations has helped us to specify some contextual features that contribute to the observational challenge of measurements to be made:

- Categoric or continuous data – counting/comparing is easier than measuring to a scale and describing as a number pattern;
- Familiarity with the measuring instrument, and the scale(s) it provides – instruments with multiple scales and/or fine gradations require intense concentration of children and so are likely to separate the investigation into a series of seemingly disconnected events;
- Time available to determine the measurement – a stationary object can be measured at leisure, a moving one must be "stopped" in the appropriate instant, a sequence without a single clear finishing time requires knowledge of research protocols for specifying an endpoint;
- Horizontal or vertical scale to be read — parallax is more likely to be an issue with a vertical scale, as in the *Ball Bounce* task;
- The magnitude of the difference to be observed (effect size) – this impacts on the "obviousness" of the evidence and, as outlined below, can invoke "theory saving" beliefs when differences are small enough to be discounted as errors.

These contextual considerations could be used to better match an investigative task to the level of progression of children's investigative skills. We illustrate how this might have been done with respect to each of the selected NEMP tasks in Section Six of the report. Meeting several of these types of challenges within one investigation should arguably be postponed until students are able to consistently produce unconfounded CVS tests. Accordingly, we have identified these skills as cluster four attributes.

Awareness of experimental error

Both the 1995 and 1999 NEMP science reports commented on children's failure to plan and carry out repetition of individual tests (Crooks and Flockton, 1996; 2000). The research discussed above has led us to question whether this emphasis, implicitly grounded as it is in protocols to manage experimental error, is developmentally appropriate for young children. Might a focus on simple awareness of variations in data patterns be more productive for children whose investigative skills are still in the early stages of development? The research reported next led us to align the management of experimental error with cluster five of our evolving framework.

Rollnick, Lubben, Lotz, and Dlamini (2002) identified a progression in meta-strategic awareness of the necessity to manage measurement errors amongst undergraduate chemistry students. They describe a "point" paradigm in which each separate measurement made is seen as independent of all others that are made and so each individual measurement could potentially represent the "true" measurement of the situation, if sufficient care was taken. We certainly saw evidence of this sort of thinking during the tape reanalysis. ("*Measure really, really carefully.")* Contrasting with this is the more sophisticated "set" paradigm in which individual measurements are seen as approximations of the true value and awareness of variation encompasses ideas such as the spread of data. In this "set" view a number of measurements must be combined and characterised by mathematical operations such as the calculation of the mean and standard deviation for an actual value to be accurately described. Significantly, Rollnick et al (2002) describe an intermediate stage at which students use "set" actions but "point" reasoning. The researchers suggest these students are merely following an algorithmic process that has been inculcated during their previous education. They describe this as unhelpful and "an impediment to moving learners towards a set paradigm in much the same way that alternative conceptions about concepts impede learning of scientific concepts" (p. 3).

In light of their experiences with 2 different groups of students, Rollnick et al. recommend that "in order to become consistent users of a set paradigm, students need to have authentic experiences of data handling where the goal is a finding that does not necessarily correspond to a correct answer in a demonstrator's marking memorandum" (p. 17). While this could be read as a negative comment about certain university assessment contexts, it also seems to imply that such experiences should allow students to consciously explore data patterns that emerge in a specific investigative context. If such a focus is to begin from a younger age, as we have advocated above, teachers should be supported to choose contexts where data variations can be both observed and documented at the appropriate investigative skill levels. We return to this challenge in Section Six.

Different types of scientific investigations draw on differing research protocols as part of their anticipatory management of known types of experimental error. Here, too, context could play an important role in determining progression in awareness of patterns in data variation. Haigh (1999), commenting on Year 12 students' ability to carry out open investigations in biology, notes that sound knowledge of CVS strategies is a necessary but not sufficient basis for managing data variation:

> *Although these students' written work indicated that they knew about the principles of fair testing they did not always demonstrate consistent application of these principles. When planning and gathering data these students had a poor understanding of experimental protocols relating to sample size and replication. They had difficulty identifying and manipulating variables and did not always specify measurements with sufficient precision* (Haigh, 1999, p. 7).

Clearly, the management of experimental error in all its various potential manifestations presents a complex set of challenges whose specifics are unique to each type of authentic investigation. However, we have also suggested that it is possible to simplify data gathering procedures without compromising children's growing awareness of these issues, and that manipulation of the context of the investigation might be one aspect of actively teaching for progression.

Issues of “data distortion” and “theory saving”

Schauble (1996) describes some ways in which her research participants’ personal theories of cause and effect seemed to influence the manner in which data were collected and then either accepted or discounted by both children and adults. She illustrates a process that she calls “data distortion” as follows:

> *…when participants measured the spring length with the ruler, they could raise or lower their line of sight, resulting in slight fluctuations of the observed value. Similarly, participants could produce variability with the stopwatch used to measure travel time in the canal task. Beyond the difficulties naturally associated with timing events to the nearest hundredths of a second, there were opportunities for intentional delays or “jumping the gun”* (Schauble, 1996, p. 114).

Schauble is not suggesting here that her participants deliberately cheated – rather that data distortion during measurement is one way in which they subconsciously attempted to align their prior ideas of cause and effect with the observations they made. In an attempt to avoid this data distortion effect, Zohar’s (1995) team simplified Schauble’s boat task so that only **relative** speed need be observed. They did this by using coloured flags on the side of the canal. Participants stated how many flags a boat passed on each test run rather than measuring the actual distance travelled. With the earlier source of potential error thus removed, Zohar’s team detected a process they called “theory saving” that operated at a later stage of the investigation. When pairs of CVS tests yielded seemingly conflicting results, some participants made limited inferences, accepting only the inference drawn from the paired CVS tests that agreed with their prior causal thinking. For example, if they had a prior belief that weight does make a difference to boat speed, they accepted the evidence generated by the small boat/weight and small boat/no weight pair of tests. However, the seemingly conflicting inference provided by the equivalent test pair with the larger boats was simply ignored when the findings were reported. Obviously, participants who used theory saving strategies were unable to use their investigations to detect and describe the interactive effects of some of the variables.

Some important NOS issues are captured in the theory/evidence interactions described by these researchers. Schauble points out that scientists must also wrestle with the process of drawing correct inferences from their empirical evidence:

> Phenomena like measurement error and data distortion underline the fact that theories and data are inextricably intertwined. Theory guides fundamental decisions such as which independent variables should be considered as potential causes ….With the hindsight provided by a scientifically accepted model, researchers can conceptually separate theory and data in an experimentation task, but for the scientist laboring before the critical conceptualisation is made, or for the experimental participant who does not already understand the problem domain, such distinctions are unavailable (Schauble, 1996, p. 115).

Like Zohar’s “expert” investigators, scientists must proceed from an explicit understanding that evidence may be explained in more than one way, and that testing to eliminate alternative hypotheses is an essential feature of scientific inquiry. We have aligned this aspect of the “view of scientists goals/methods” (see page 62) with the interactive analysis mental model in cluster five of the evolving framework. Both are linked to the availability of coherent overarching explanatory frameworks of science knowledge in the “personal knowledge skills” aspect, and to specific types of personal investigative skills. We reiterate that many school students, and possibly some undergraduate students, will not reach this level, at least with the sorts of investigative experiences to which they are typically exposed at present.

We have already noted the likelihood of strong links between personal theories and the inferences children draw from the evidence generated by their investigations. Absolute measurements inevitably generate measurement errors and Schauble (1996) has noted that children and adults have difficulty in distinguishing variation caused by measurement errors from variation due to true differences between the experimental conditions. In her investigations, when in doubt both children and adults fell back on their prior theories. If they expected a variable to have an effect, they interpreted small variations in measurements as a positive effect. If they did not expect an effect, they were more likely to interpret small variations as error. Here, as in previous sub-sections, the nature of children’s theory/evidence links may well form an important aspect of progression.

SECTION FIVE: A FRAMEWORK FOR PROGRESSION IN CHILDREN'S SCHOOL SCIENCE INVESTIGATIVE SKILLS

INTRODUCTION TO THE META-LEVEL FRAMEWORK

In Section Four we described the manner in which the key findings of the literature review were used to develop a framework for thinking about the types of meta-level epistemological development that might underlie children's observable actions when they undertake science investigations. More specific detail of possible stages in the development of "fair testing" knowledge was also tentatively aligned with each meta-level cluster. Having set out the reasoning process followed, we now present the framework, with some general suggestions for the teaching of investigative skills at each level. We then identify some shortcomings of the evolving framework by outlining components of a science education "for all" (Fensham, 1985) that we think are important but that are not covered within the focus taken here on classroom-based investigative skills.

Section Four described how the framework we have developed draws on various research projects that suggest sequences in the development of various meta-level characteristics of children's thinking. These include their views of the nature of knowledge, their metacognitive awareness, their meta-level views of the purposes for which activities such as "investigations" are carried out, by themselves and by scientists, and their mental models of causality. It seems to us that some interesting alignments of these various characteristics are possible and we present these as 5 "clusters" below. We have resisted the temptation to call them stages because there is, of course, no certainty that they will develop seamlessly in perfect integrated harmony. Indeed, the complexity of children's development patterns suggests this is highly *un*likely. Nevertheless, the clusters as developed may provide a useful basis for taking account of the range of thinking likely to be present within the diversity of a class of children. Dealing with this diversity in a manageable fashion is the daily challenge that confronts every classroom teacher.

Our attempt to develop these meta-level clusters could prove helpful for teachers in that it provides a relatively simple theoretical framework for thinking about **why** children do the things they do, and why they face the specific types of learning challenges that their teachers can detect. Such insights are an important aspect of teachers' pedagogical content knowledge, and help inform in-the-moment teaching decisions. We discuss learning challenges related to the specific teaching of investigative skills, along with potentially helpful teaching foci, for each of the 5 clusters.

CLUSTER ONE: EARLIEST FORMAL SCIENCE LEARNING

Meta-level view of science knowledge
Simple belief in science knowledge as true and **certain**

Meta-level view of science investigations	View of scientists' goals/methods	Personal knowledge skills	Personal meta-values	Mental models of causality
Belief that simple one-off investigations can yield answers in a straightforward manner	Goals of science are simple doing things/gathering information/finding out answers Scientists might change their minds on a whim or as the result of one episode of investigation	Can state *own* ideas Beginning to recognise that **what** they think is different from **why** they think it	Willing and able to share ideas and listen to others Willing to explore, play, and show curiosity	**Outcome** focused (if variable is present and outcome occurs, the variable must have caused the outcome)

The juxtaposition of these characteristics suggests that the very notion of "fair testing" as a formal means of systematically approaching scientific investigations could prove challenging for children whose meta-level views predominantly align with this cluster. Can the relevant investigative skills be actively taught while still keeping open rich opportunities for play, for fostering curiosity, and for extending children's "library" of knowledge of the material world?

A consistent finding of research in this area is that children can **recognise** fair tests long before they can **produce** them via their own planning skills. Accordingly, productive approaches to children's formal introduction to "scientific investigations" could begin by presenting children with concrete forms of direct "fair testing" comparisons, so that they can see, feel, hear, smell the relevant variables and make direct comparisons between matched pairs of tests.

It is important that the **contexts** chosen for investigation will yield results that make intuitive sense to young children – that is, the things that happen support and confirm their existing theories of causality. (An example of this is the longer run distances of cars from steeper ramps in the *Truck Track* task. Children expected this result because of their knowledge of vehicles on hills.)

Teachers could encourage emergent recognition of links between adjacent investigative episodes by using simple visual strategies to record data patterns. Such approaches make no demands on children's memory space because the "evidence" takes shape before their eyes. We describe one very simple way of doing this for the *Truck Track* task (and other similar tasks) in Section Six.

CLUSTER TWO: DEVELOPING SKILLS OF EXPLANATION IN SCIENCE

Meta-level view of science knowledge Growing recognition of knowledge as explanatory				
Meta-level view of science investigations	**View of scientists' goals/methods**	**Personal knowledge skills**	**Personal meta-values**	**Mental models of causality**
Beginning to recognise that finding answers requires thinking/explaining Beginning to explain "Why I am doing this." "What do I want to find out?"	Goals of science are finding explanations from investigations Increasing recognition that this requires thought/effort/ exchanging ideas with other scientists	Beginning to talk about theoretical entities (e.g., atoms/gravity) but these are still treated as factual Adopts new ideas most easily when new experiences align with existing theories of cause and effect	Beginning to try and understand **peers'** points of view before making comment Willing and able to practice simple analytical thinking – e.g., by spotting inconsistencies in thoughts of **others**	Emerging **analysis** model – recognition that more than one variable may influence outcome

This cluster marks a profound shift from the simple certainties of direct links between causes and effects, and unproblematic knowing about the world. Arguably, the very notion of "fair testing" cannot really make sense until children recognise that there **are** alternatives between which choices need to be made. There are correct explanations to seek and incorrect explanations to rule out. In this context, a more systematic approach to investigations can yield rewards of knowing for its own sake, and the personal satisfaction of a sense of ownership of new knowledge built.

This type of awareness is not of course limited to fair testing. The ruling out of alternative explanations can begin with **exploring**, and give shape and direction to **pattern seeking** and **modelling**, for example.

Now children can more consistently recognise simple fair tests, and strategies to help them think about these **in principle** may be appropriately introduced. The ability to retrospectively **explain** "fair test" choices can be encouraged and children may be making significant steps towards **producing** their own simple fair tests by managing **causal** variables.

Again, contexts selected should yield effects that are directly observable and that confirm children's existing causal theories, so that their confidence in their knowledge building skills can be fostered. Results that can be captured as simple visual and/or **categoric** data patterns will encourage recognition of these patterns. In such situations, children may be learning to link pairs or short sequences of tests to propose explanations. Possible strategies are described in Section Six.

CLUSTER THREE: A DEVELOPING SENSE OF SCIENCE AS KNOWLEDGE TESTING

Meta-level view of science knowledge
Science knowledge seen as a collection of **tested** ideas
Beginning to recognise that science ideas can change when new **evidence** is available
Beginning to recognise that prior ideas can constrain new thinking

Meta-level view of science investigations	View of scientists' goals/methods	Personal knowledge skills	Personal meta-values	Mental models of causality
Recognises that the effects of separate variables require explanation Starting to think about reasons for actions: "Why should I do it this way?" Emergent recognition of the need for a **plan** to coordinate linked investigations	Scientists ask questions about why things happen Scientists use investigations to develop or test ideas and work together to develop their understandings Scientists investigate theoretical entities as well as concrete events/objects	Beginning to draw on emergent understanding of **theoretical ideas** of science as well as personal theories Beginning to justify thinking in terms of available evidence, especially when this **supports** existing personal view Emergent ability to use physical **models** and **analogies** as mental models	Willing and able to seek explanations Emergent willingness and ability to think critically about **own** ideas and discuss changes in these over time	**Additive** analysis model – can identify situations where 2 or more variables separately influence an outcome

Again, this cluster marks a profound shift in the types of ideas and thinking activities children can entertain. Explanation *per se* will no longer suffice, because various explanations may seem plausible, depending on the theoretical ideas seen to be relevant. With this increasing awareness of the inherent complexity of knowledge building comes the idea of **testing** personal theories against the available **evidence**.

Children may begin to experience the power of using the evidence generated by their investigations to persuade their peers to accept their explanations and personal theories. At the same time, they are beginning to compare their own theories with those held by scientists, and to recognise that science ideas are tested against the evidence presented by the world around us. Investigations that involve exploring and/or pattern seeking and/or modelling can all extend these types of theory/evidence links.

Once children can use these types of reasoning skills they might be expected to consistently select fair tests, and to appropriately justify their choices. They might also be expected to accurately **produce** simple fair tests and hold a **short sequence** of such tests in the **memory** while working through them systematically.

As their knowledge of science ideas begins to expand, and their experiences of rich inquiry contexts continue apace, so their personal “library” of cause and effect explanations will also grow. This growth could lead to an increasing ability to identify **non-causal** variables, and to exclude these from their investigations as they proceed.

The specific types of data gathering employed could be another attribute of increasing the contextual challenge presented by investigations. Students could be using specific types of measurement to begin to collect and process continuous data, although strategies that encourage them to display data patterns as these accumulate will help them keep track of the whole investigative space.

CLUSTER FOUR: ENHANCING SKILLS OF OPEN-MINDED REASONING

Meta-level view of science knowledge
Science knowledge seen as a collection of tested ideas that are open to **revision** as a result of on-going investigation
(i.e. the process of revision has more clearly specified “rules” than does revision of some other types of knowledge)
Increasing recognition that science knowledge is not easily changed

Meta-level view of science investigations	**View of scientists’ goals/methods**	**Personal knowledge skills**	**Personal meta-values**	**Mental models of causality**
Recognition of need for a **global plan** to coordinate and review linked investigations Recognises importance of considering alternative explanations Can give clear reasons for actions: “Why should I do it this way?”	Scientists use investigations to develop and test ideas about entities (unseen/ observable) Scientists may interpret data differently but they have to justify their ideas to their peers to build consensus. They can influence each others’ interpretations	Draws on growing body of scientific knowledge Beginning to recognise when new evidence **disconfirms** personal theories Explicit use of simple analogies and metaphors as mental models for scientific thinking	Willing and able to clarify, challenge, and identify inconsistencies in own thinking Willing and able to use established criteria to evaluate quality of thinking (e.g., intelligibility, fruitfulness, plausibility)	Emergent **interactive** analysis model – beginning to identify situations where variables impact on each other to collectively influence outcome

There is a clear sense of increasing awareness of the complexity of knowledge building in this cluster of attributes. The sense that the warrants for making scientific knowledge claims should be always open to scrutiny and justification is present throughout the cluster. With this comes the challenge of suspending judgment and considering the various alternatives before deciding on “correct” explanations. The sense of the power of justification of knowledge, through processes that include the consideration and elimination of alternative explanations, can bring aesthetic pleasure.

At this stage students could be beginning to produce more complex sequences of school science investigations that require the management of a number of variables in sequenced tests. Teaching strategies that help scaffold students’ increasing memory capacity will enhance this development.

The choice of investigative contexts in which a number of potentially relevant variables can be easily observed and practically managed will also be essential. The *Ball Bounce* task has already been identified as having this potential, and we explore it further in Section Seven.

The process of eliminating alternative explanations will challenge students to pay increasing attention to data collection processes and patterns, including sampling protocols, endpoint determinations, measures of variance and central tendency, and the identification of sources of experimental error. There may be an associated recognition of situations where the evidence does **not** support the formation of a valid conclusion at all.

CLUSTER FIVE: A "KNOWLEDGE PROBLEMATIC" ORIENTATION

Meta-level view of science knowledge
Science knowledge consists of well tested **theories** about the world
Science knowledge is **tentative** (always open to revision in the light of new evidence) but essentially stable
Clear differentiation made been their personal theories and experts' science theories

Meta-level view of science investigations	View of scientists' goals/methods	Personal knowledge skills	Personal meta-values	Mental models of causality
Recognises the inherent complexity of science as involving multiple levels of chained investigations Recognises that some outcomes are **probabalistic** rather than clearly determined Increasing sophistication in reasons given for actions (e.g., considers complex methodological issues and/or deeper links to theoretical frameworks and/or works confidently with open types of investigations, etc.)	The goal of scientists' investigations is to empirically distinguish between alternative hypotheses via the controlled manipulation of the investigative situation	Personal science knowledge is now being organised into explanatory frameworks with an overarching coherence More advanced level of personal understanding of scientific theories Uses a range of mental models to support own scientific reasoning	Willing and able to challenge and actively modify and/or extend own thinking Seeks opportunities to engage in constructive argumentation processes to build agreed knowledge Adopts a position of healthy scepticism – seeks consistency and generalisability before adopting new science ideas	**Interactive** analysis model – with growing recognition of complexity and number of variables involved Move from simple linear pathways to multiple possible pathways of cause/effect

How may adults systematically explore the warrants for the beliefs they hold, seeing knowledge as inherently problematic? Only once the attributes of this final cluster are in the process of being achieved can we claim that students might be truly beginning to think as scientists think – at least when the latter are undertaking their professional work. The ability to represent evidence independently of theory as far as is possible has been identified as one hallmark of such thinking.

Again, the qualitative differences between this cluster and those that have preceded it are profound. Here, students are coming to an appreciation of science as a vast nexus of interconnected investigation processes and theories about the world. The appropriate and accurate utilisation of scientific theories differentiates this cluster from cluster four. Such reasoning requires students to hold a broad overview of the framework of key science ideas and inquiry methods in the relevant discipline area, and to know how these connect with the experience(s) at the heart of their own investigations. It is likely that many school students will never reach this level of understanding of the nature of science at all.

This cluster is aligned with a sophisticated view of scientific inquiry, even where the contexts of investigations are necessarily constrained by the equipment that is available to the fledgling scientist. Students working within this framework might be expected to systematically compare different possible combinations of variables to detect and describe different types of interactions. They will be paying closer attention to the **quantitative** analysis of data outcomes, including the identification and management of measurement errors. A global plan could be used to justify and coordinate a series of activities within one overall investigation.

LIMITATIONS TO THE FRAMEWORK

The framework we have developed, and the literature we have cited, present a traditional, rationalist view of science that is open to both philosophical and practical critique. In this final part of Section Five we can only briefly sketch areas that these critiques might encompass. We do so to signal that we regard this as a work in progress.

The literature selected explores the types of activities that are most likely to be a part of school science investigations, and also some aspects that may be neglected. One of these neglected areas is the focus on theory/evidence interactions (Driver, Leach, Millar, and Scott, 1996) and so they are not usually seen as contentious. However, these interactions arguably should be addressed if students are to come to understand the nature of scientific knowledge building (Osborne, Erduran, Simon, and Monk, 2001). In the selection and shaping of pedagogical material that is at an appropriate level of conceptual challenge for learners of science, the complexity of messy, genuinely new, scientific questions is unlikely to be in evidence. Students could be misled about the intellectual challenge involved in actual scientific research, so that the role of creative thinking could be inadvertently downplayed. Our developing framework makes no mention of creative thinking at present. Whether it should is an open question.

The literature we have reported on the development of scientific inquiry skills usually draws on concepts from the physical sciences that can be demonstrated in relatively simple classroom contexts. Trucks roll down ramps, and boats float down model canals pulled by gravity, for example. This emphasis on simple mechanistic concepts and contexts has been critiqued as misrepresenting science by neglecting the types of systems thinking that are more common in some of the biological and earth sciences (Mayer and Kumano, 1999). What might be involved in the development of systems thinking? Are there missing components that the matrix does not address? Some recent environmental education literature suggests that the scope of the thinking we have outlined should be extended to address this emergent area of research (Sheehy, Wylie, McGuinness, and Orchard, 2000).

If students are expected to be able to transfer "scientific reasoning" skills to other contexts in their adult lives — as "scientifically literate" participants in a democracy — other critiques of the overrepresentation of the physical sciences apply. Zohar and her team (1995) used 4 micro-world simulations that were isomorphic in terms of the types of interactions between variables that were modelled. Two tasks were set in physical science contexts (one being the boat/canal task described above) and 2 were set in **social** contexts. Interestingly, Zohar reports that participants struggled much more to resolve the social micro-worlds than the physical ones. She speculates that personal theories about social interactions are much more deeply held and so "theory saving" strategies of the type she described are more likely to prevent systematic unravelling of the actual interactions between variables. Do we expect students to develop the skills to make such transfers of contexts? If so, at what stage of development would this be considered appropriate, and what are the implications for curriculum integration? These are questions we could not address.

This report can also be critiqued for the lack of discussion of science as a socially constructed activity/body of knowledge. In the absence of such a focus the only mention of values concerns those that are epistemic and intrinsic to science itself. Values external to science that interact with its knowledge building processes, contribute to considerations of risk and uncertainty, and shape ethical thinking, are not discussed. They are important aspects of the interaction of science with society, so their absence is a significant omission. It seems to us that progression in these areas may also necessitate the development of some cross curriculum links – especially with social studies, and environmental education. Such a project is beyond the scope of this paper. We note that there is some literature in this area (see, for example, Zeidler, Walker, Ackett, and Simmons, 2002) and the inclusion of these aspects within the developing framework would balance the exclusively rationalist focus that has resulted from the project thus far.

One recent research project has suggested that prominent, successful contemporary biologists do not typically derive their hypotheses and research questions from the "pure" theory/evidence thinking of the type that is modelled by the framework. Rather their questions are funding driven and as such, are directly influenced by the funding agency's social agenda (Abrams and Wandersee, 1995). These researchers found that some of their biologist participants were not even able to articulate the manner in which theoretical thinking in their field had influenced the design of the data gathering technologies they used with such enthusiasm. We have not addressed the nature of interactions between contemporary science and technology at all, and so this is another omission that remains to be rectified.

"Success" in the types of investigations we have described could be critiqued as implicitly presenting a teleological message about the inevitable "rightness" of well-conducted scientific investigations. While the framework does address the epistemological understanding that science knowledge and methods change over time and are always open to revision, we note the absence of the messy history of science stories that present a more contextual view of now-settled science questions. Nor have we addressed the development of children's ability to understand that different bodies of knowledge may address the same questions from different perspectives and for different purposes. This issue is important as we struggle to find ways of educating children in increasingly multicultural "global" contexts.

Finally, we note that no framework or matrix, no matter how well developed, can present a completely reliable guide for the assessment of the progress of individual children because:

> Progression can be defined for a curriculum more easily than it can for assessment purposes. A curriculum, unless we are attempting to create an individualised learning package, can prescribe progression by reference to group behaviour, which is likely to be a little more predictable than that of any individual. Assessment, on the other hand, is by definition personal and subject to all the idiosyncrasies of the individual's interests, experience out of school, home background and so on, which we know to be so influential in pupil attainment (Gott and Mashiter, 1994, p. 186).

This thought-provoking note turns the focus to the uses to which matrices and/or developmental frameworks may be put. We hope our evolving framework might be found useful for the provision of guidance for planning and structuring classroom activities that can support and extend the investigative skills the students already have. We see dangers in trying to use it as an assessment rubric by which children are judged against normative assumptions of the "stage" they should be at.

We began this section with a discussion of research which demonstrated that children of the same age may be at very different developmental levels with respect to their understanding of the knowledge building processes that constitute scientific inquiry. However, appropriate skilful teaching clearly does make the difference for most children. In Section Six we turn our attention to supporting primary teachers to develop the necessary skills to take up this challenge.

SECTION SIX: TEACHING INVESTIGATIVE SKILLS

INTRODUCTION

The discussion of the previous sections has identified several areas that would appear to provide a useful focus for the active teaching of investigative skills at the primary school level. However, many of the epistemological ideas raised by the framework presented in Section Five cannot be easily translated into actual teaching strategies in the absence of a well-developed personal understanding of the distinguishing characteristics of scientific inquiry. Some of the literature reviewed in Section Four contributes to a wider body of research which demonstrates that “lay” adults do not necessarily demonstrate such understandings when confronted with the challenge of producing a scientific investigation (Schauble, 1996; Zohar, 1995). The research that prompted the design of this project (Gilmore, 2001) showed that New Zealand primary school teachers may not easily articulate ideas for actively teaching these types of understandings. Nevertheless, it is unhelpful to paint all (or indeed any) primary school teachers into a “deficit” corner. To support those primary teachers who have not yet had opportunities to develop such thinking themselves, the final stage of this research project has involved the development of several strategies that the teachers could pick up and use right away, given the provision of appropriate materials. These strategies have been designed to help children:

- articulate their “nascent understandings” of fair testing, and of experimental error;
- more consistently produce unconfounded fair tests;
- recognise and describe patterns of data variability;
- carry more of the “experimental space” in their memories, so that they can begin to recognise interactions between separate test episodes;
- talk about links between their personal theories and the “evidence” generated by their data patterns.

As Toth et al (2000) point out, strategies that are effective in developing children’s investigative skills in small-scale experimental settings will not necessarily translate easily to the more complex reality of the classroom setting. It is important that teachers see strategies as both meaningful and manageable if they are to be useful in enhancing children’s learning. The strategies introduced in this section of the report have been introduced to some small groups of primary teachers and their ideas about their appropriate use in the classroom are included in this section.

THE TEACHER FOCUS GROUPS

Three volunteer groups of primary teachers took part in 90-minute after-school sessions in which 8 teaching strategies were presented and evaluated. These focus groups were held in different geographic areas and 23 teachers took part in total. They are currently teaching classes across the full primary age range as shown on Table 23. Three teachers from the most rural area taught classes that spanned more than 2 year levels. Twenty-one teachers were female and 2 were male. Most were experienced teachers who had been in the profession for a number of years.

Table 23

Teaching levels of the focus group participants

Year levels	Number of teachers who taught at level(s)
1–2	2
3–4 or 4–5 or 3–5	6
5–6	4
7–8	5
6–8 or 5–8	2
Not specified	4

The teachers were initially asked to talk about their current methods of teaching investigative skills in science. They were then introduced to 8 potential strategies and asked to evaluate these on feedback sheets provided. During this process, they often spontaneously shared ideas about how they might adapt the strategies to their classrooms, and some teachers briefly explored new questions that arose as strategies were demonstrated with actual equipment. One page of a feedback sheet is included as Appendix 2. Teachers' responses in the "tick box" categories of reasons for choosing to use the various resources have not been included in this report. Typically, the teachers were so enthusiastic that they ticked most boxes for most activities and so the exercise failed to sufficiently discriminate between reasons for choosing the various tasks. Instead, comments that exemplify the discussion about each strategy have been included.

TEACHERS' IDEAS ABOUT TEACHING INVESTIGATION SKILLS

All 3 focus groups began with an invitation to the teachers to tell us what they actually do at present to teach fair testing. In all 3 groups most of the ideas advanced centered around discussion of the concept of "fairness". In all the groups teachers described ways of stimulating children's natural sense of "what's not fair" in science contexts. The emphasis was on "having children figure it out for themselves".

One teacher described his method as "working backwards". Children made paper darts in any way they chose and then these were ranged against each other in competition. The ensuing indignation opened up awareness of the relevant variables at the same time as leading to a discussion of "how to make the test fair". Another teacher spoke of linking the concept to basic statistics, using loaded and unloaded spinners to compare the frequency with which a number came up. The typical emphasis of such a discussion was on "changing one thing at a time" — often with considerable step-by-step direction from the teacher. One teacher described this as using a "set of rules". Several teachers spoke of setting up "norms" or "controls", which were basic experimental set-ups from which to begin changing one thing at a time.

All the ideas advanced referred to **setting up** a fair test. This may of course have been an artifact of the way the question we asked was interpreted. Nevertheless, no one gave an unprompted description of any strategy for fair measuring/observation, or for fair interpretation of results, or for keeping track of a sequence of related tests. Yet several of the strategies we subsequently introduced were reported to be familiar by a number of the teachers. It seems the setting-up stages are at the top of teachers' minds when they think about fair testing in their classrooms.

There was very little discussion of problems in relation to teaching fair testing, and those that were raised were of an essentially practical nature. For example, one teacher spoke of the challenges of having different groups running different fair tests in the same context. Sometimes children look across to a seemingly more exciting group and want to change the things that that group has changed, as well as what they are supposed to be changing. On the whole, however, it seems these teachers were confident that they know what fair testing means for their children.

Several of the contexts described seemed to involve pattern seeking or exploration, rather than requiring a fair test *per se*. For example, one teacher spoke of making fair tests when using chromatography to explore the various colours in the sugar coats of small sweets. For her, fairness meant starting with one colour and then systematically comparing others. Another teacher spoke of fairness as "being honest" when results were not as anticipated.

Most of the teachers were unaware of the NEMP science tasks. One asked if they were the same as the Assessment Resource Banks. Another said "Well we might have used them. We don't know what they are."

BUILDING AWARENESS OF FAIR TESTING PROCESSES

The first 4 strategies provided alternative ways of supporting children to initially **choose** fair tests, and to justify their choices. The focus of these strategies was on exploratory and/or planning stages of the investigative process.

Strategy One: Matched pairs of equipment

Drawn directly from the research of Toth et al (2000) this strategy was modelled using the *Truck Track* task. Two side-by-side ramps were set up with same model cars and uniformly sized blocks of wood used to adjust their respective heights. Ways to alter one ramp's settings to make it obvious that just one variable had been changed were briefly demonstrated. The strategy supports children's ability to recognise fair tests by **simultaneously** showing the whole experimental situation.

The strategy was familiar to about half of the teachers (although none had mentioned it in their preliminary discussion) and was seen as being immediately obvious for children of all ages. Three teachers commented that the strategy could be used for making predictions:

> *Clearly visible. Leads to "what if…." questions. Leads to prediction. (Year 1 and 2 teacher who had not seen strategy before)*

Two teachers commented on the opportunity to foster the use of everyday language as children talked spontaneously about the settings. Overall response to this strategy was very positive, as shown in Table 24. However, 2 teachers commented that the strategy was resource intensive and would have implications for their limited science budgets. They saw this being used as a whole class strategy for that reason, although a majority of the teachers thought it could be used either as a whole class or as a small group strategy.

Table 24

Teachers' responses to Strategy One

Response sought	Yes	No	Maybe	No response
Have you seen this strategy before?	11	11	NR	1
Would you consider using this strategy with your class?	23			

Strategy Two: Paired planning sheets

Drawn from work on representational scaffolding of children's fair test planning (Toth et al, 2002), this strategy was also set in the context of the *Truck Track* task. Teachers were given an example of a paired planning sheet based on this task, as shown in Figure 3. The strategy was introduced as one that could allow children to simultaneously view all the variables that need to be managed, reducing demands on their memory space and allowing them to see the overall shape of their plan, and to check that only one thing had been changed. In the example provided, children would circle their selection for each variable, allowing both the children and their teacher to see at a glance whether or not they have used a VOTAT (vary one thing at a time) strategy. Older children could be given the outline table and asked to complete variable levels for themselves.

At first sight, some teachers were more ambivalent about this strategy, especially those who teach younger children. Very few of them had seen this type of sheet before. However the written feedback gave less indication of their reservations than we had expected from the verbal comments, with most of the teachers indicating that this, too, is a strategy they would use. One said they intended to immediately file the example sheet in their science resources folder upon return to school. Responses are shown in Table 25.

Table 25

Teachers' responses to Strategy Two

Response sought	Yes	No	Maybe	No response
Have you seen this strategy before?	4	18	NR	1
Would you consider using this strategy with your class?	18		4	1

Figure 3

Planning sheet for the Truck Track *task — option for younger students*

Variables	***Ramp A*** *(Circle* ***one*** *in each box)*	***Ramp B*** *(Circle* ***one*** *in each box)*
[✓ Steepness of the ramp]	*Number of blocks* *1 [✓ 2] 3 4 5*	*Number of blocks* *1 2 3 [✓ 4] 5*
Car Direction	*[✓ Forwards] Backwards*	*[✓ Forwards] Backwards*
Ramp Position	*[✓ Straight] Crooked*	*[✓ Straight] Crooked*
Table Surface	*Rough [✓ Smooth]*	*Rough [✓ Smooth]*
Ramp Surface	*Rough [✓ Smooth]*	*Rough [✓ Smooth]*

Most teachers made written comments about this strategy, and these covered a range of areas. Four reiterated that the format could be confusing for younger children, but 3 Year 7 and 8 teachers suggested ways to modify it to show a series of sequential tests – a more complex layout! (Something similar had been prepared by us as Strategy Eight, but we had not shown that at this point.) Five teachers stressed the need for clear modelling, although they saw the effort needed to do this as worthwhile:

> *I liked the way it clearly made the fair test variables obvious. After "teacher talk" the children would know the variables and how to test one at a time.* (Year 5 and 6 teacher who had not seen the strategy before)

This strategy was seen to support small group discussion, and to be helpful for reducing reading, language, or writing demands:

> *Children have more time to think and discuss their new learning. Takes away writing time.* (Another Year 5 and 6 teacher who had not seen the strategy before)

A number of comments affirmed the potential of this strategy to make the comparisons between tests obvious, with one commenting that it would also help the teacher to keep track of the teaching point.

Strategy Three: Card pairs

For this strategy children are given a series of cards that show the same situation in various different variable combinations. We modelled this strategy with a set of *Emptying Rate* cards as shown in Figure 4 on the next two pages. From amongst all the cards available to them, children select a pair that would create a fair test for a specific question. For example, these cards could be used to choose a fair test for questions with a range of degrees of conceptual challenge:

- Does the size of the hole make a difference to the emptying rate for water?
- Does twice as much water take twice as much time to empty?
- Which empties faster, water or detergent?

Again, drawn directly from the research literature (Toth et al, 2001) this strategy provides a very structured visual way for children to demonstrate that they can recognise fair test situations. Selection of a pair of cards allows children to check that only one aspect is different, in much the same way as with the paired planning sheets of Strategy Two, but with **visual** rather than purely verbal cues. Individual cards can be manipulated into various trial combinations, making the strategy ideal for small group discussion.

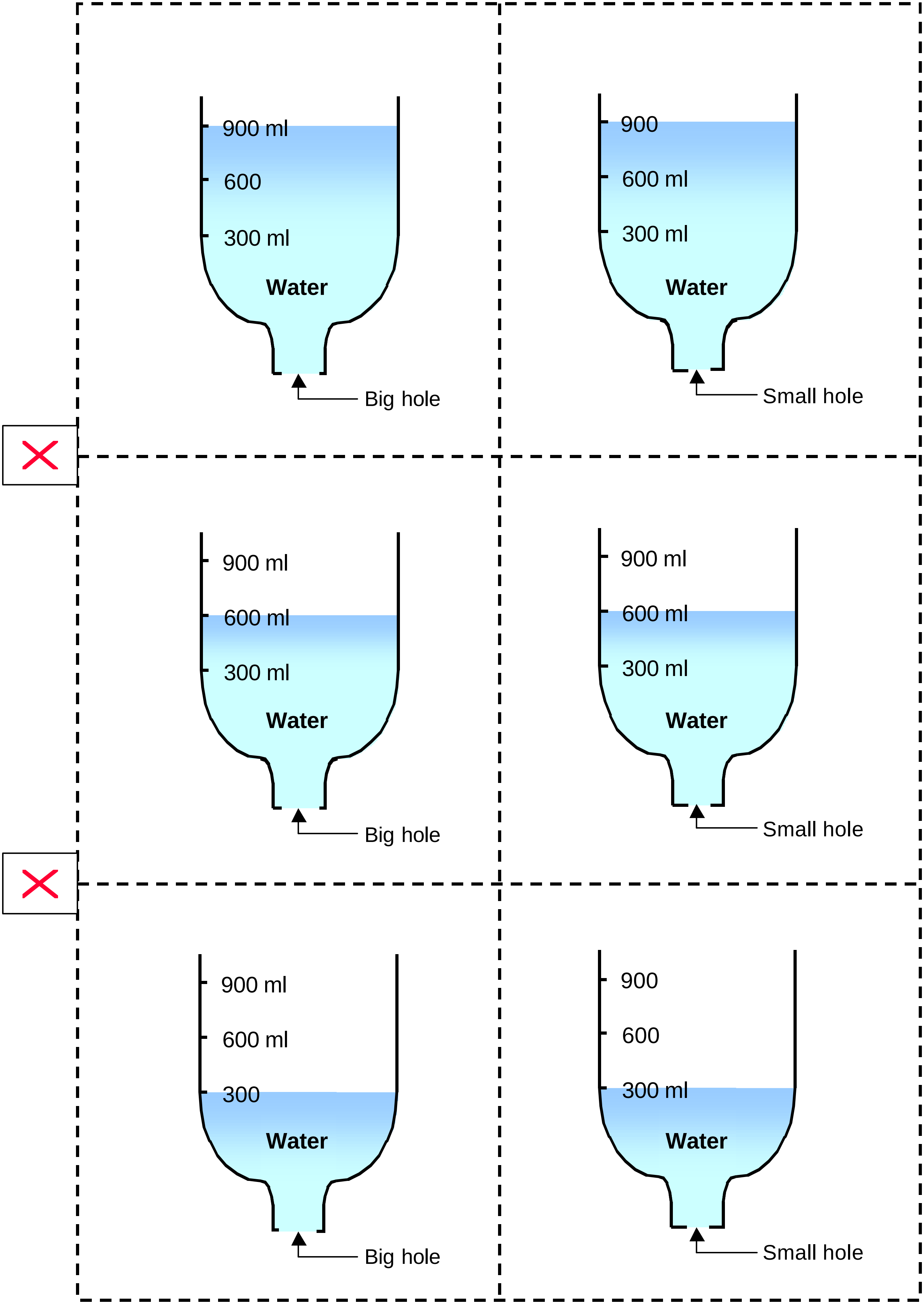
Table 26

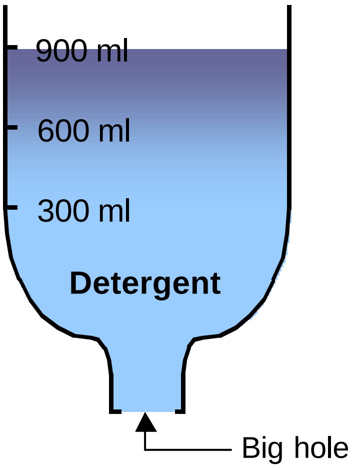
Teachers' responses to Strategy Three

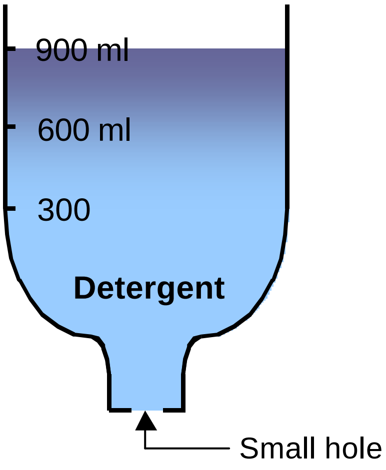
Response sought	Yes	No	Maybe	No response
Have you seen this strategy before?		22	NR	1
Would you consider using this strategy with your class?	17		5	1

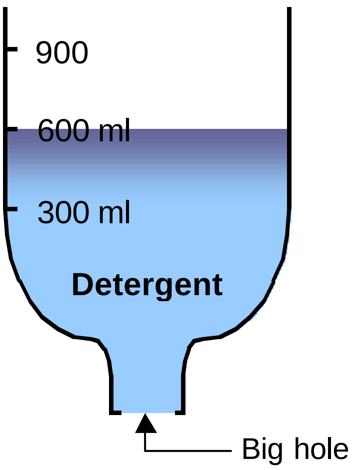
Figure 4

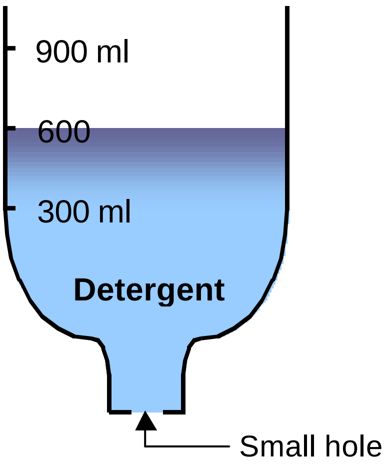
Examples of card pairs

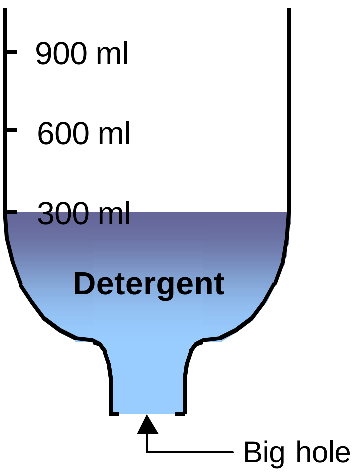



900 ml
600 ml
300 ml
Detergent
Big hole


900 ml
600 ml
300
Detergent
Small hole


900
600 ml
300 ml
Detergent
Big hole


900 ml
600
300 ml
Detergent
Small hole


900 ml
600 ml
300 ml
Detergent
Big hole

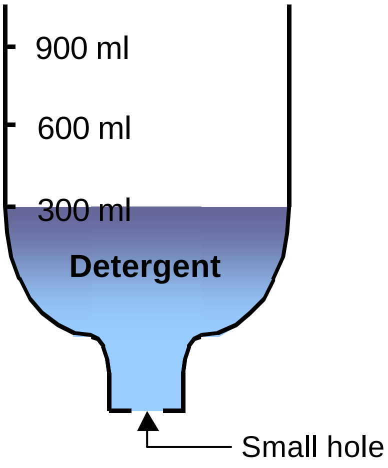

900 ml
600 ml
300 ml
Detergent
Small hole

Table 26 shows that paired card selection was one of the least familiar strategies, and 5 teachers expressed ambivalence about using it. However, the possibilities for using the activity outlined above were immediately obvious to all 3 focus groups:

> *I liked the way it allowed children to select the fair test — this would really help the children to learn planning a fair test. It's a great visual — children can see it and not have to imagine it.* (Year 5 and 6 teacher)

> *I think this might be good to get the children talking and discussing to develop higher level thinking. It could give the teacher an idea of whether children can predict.* [The teacher could also assess] use of language and ability to understand language (no year level given)

Those who indicated ambivalence did so for practical reasons (preparation of the necessary resources) or because they would not wish this to be a substitute for children's experiences of actual investigations:

> *Still would want to give children the actual equipment after they've done the thinking.* (Year 5 and 6 teacher)

> *Allows for predictions. Is a planning tool. Visual learners – very helpful for those with little reading or language skills. It would encourage them to try it in a practical situation.* (Year 7 and 8 teacher)

The latter teacher also suggested making easier versions for children who did not yet have good measurement skills — *if* the development of these skills was not the focus of the activity. This could be done by using terms such as "quarter full, half full" instead of actual measurements. Teachers of younger children suggested a different type of simplification, with less variables depicted on any one card so that choosing test pairs was not as challenging. Modelling the use of the cards alongside a concrete demonstration of the equipment was another suggestion made for familiarising children with the demands of the activity. Two teachers felt children would lose interest if they only had the cards to work with because these would be "too abstract".

Strategy Four: Sequenced coaching

Toth, Khlar, and Chen (2000), recommend a combination of independent exploration and expository instruction. Using the paired sets of equipment suggested in Strategy One, the teacher first allows the children to explore, then demonstrates and discusses "bad" then "good" designs. This discussion with the children is followed by a second "bad" then "good" demonstration round to consolidate the idea. The children are encouraged to contribute their ideas at each stage but the teacher is very much in control of intended learning. Children are then given the opportunity to devise simple fair tests in groups.

The full sequence of coaching steps is presented in Figure 5. Toth et al. (2000) found this an effective sequence for teaching fair testing skills that children would remember and would be able to transfer to other contexts. For time reasons we did not model this whole sequence but rather scanned it on paper, as presented in Figure 5. This may have contributed to the ambivalence teachers expressed, although they were very clear about the reasons for their reservations:

> *Kids would lose interest in the process. Would demonstrate only once at [stage number] 3. Then let them do their own — don't revisit and analyse.* (Year 1 and 2 teacher)
>
> *Used over a period of time. E.g. over a whole unit to build on previous skills/knowledge. I would spend 10–15 minute sessions each time.* (Year 5 and 6 teacher)

Figure 5

Sequence description provided to teachers

A sequence for teaching fair testing

1. **Enhancing awareness of the context**
   With the whole class, show and talk about the equipment being used for the planned investigation, emphasising the relevant variables. Demonstrate how everything works, so that students' familiarity with the context of the investigation is enhanced.

   NEMP examples

   - Show how a ramp can be set up. Ask the students to identify all the things they could potentially change (the variables). Run a car down the ramp and draw students' attention to the relevant aspects of the car travel patterns;
   - Bounce one or more balls against a one-metre ruler that has coloured zones on it. Invite students to present ideas about the factors that could affect the height of the ball bounce.

2. **Exploratory stage**
   In small groups, give students a simple "fair test" challenge and let them play with the equipment involved to explore the best way to make a plan for the investigation. Have students share their ideas, and talk about "fair" and "not fair" plans.

3. **Expository instruction**
   Demonstrate a trial design for the challenge that is "not fair". Ask students to identify what is not fair and to say why the comparison will not answer the question posed.
   Follow this up with a demonstration of a "fair" test for the question posed. Ask students to identify why this comparison would allow them to find an answer to the question posed (i.e. why this arrangement is "fair").
   Check that the children understand by carrying out another "not fair" then "fair" testing round, using a variation on the original question.
   It is important to invite students to contribute their ideas at each of these demonstration /discussion stages, but to keep control on the overall intention of the learning.

4. **Application stage**
   The children again work in small groups practising their fair testing skills on similar tasks to those used by the previous exploratory phase.

Despite these reservations, written feedback indicated that the majority of teachers would consider using the strategy, as shown in Table 27.

Table 27

Teachers' responses to Strategy Four

Response sought	Yes	No	Maybe	No response
Have you seen this strategy before?	19	4	NR	
Would you consider using this strategy with your class?	17	1	3	2

Discussion about the time this sequence would take dominated the verbal feedback. Teachers were more familiar with this than with any of the other strategies but had mixed feelings about the practicality of spending so long revisiting the ideas. Teachers of younger children recognised that their children "just want to do" whilst teachers of older children worried about working with mixed groups where some were able to quickly grasp the ideas while others needed more repetition. One teacher stated that she would do this if she was "convinced" that it would be worth the effort. Another reflected on her own recognition that more attention needs to be paid to children's learning of the skills for shaping of the fair test process:

> *The first stage often gets overlooked due to limited time and resources but I recognise how important it is and will now allow for it.* (Year 5–8 teacher)

STRATEGIES TO MAKE PATTERNS OF EVIDENCE CLEARER

The second sequence of 4 strategies focused on ways of capturing and thinking about the results of fair tests, once these have been carried out. As a consequence of the observations reported in Section Three, we were interested in modelling ways to modify data gathering so that children can see patterns of data variability in **real** time (as opposed to having to carry these in patterns in their memories, or read them retrospectively from numbers). We wanted to model ways to simplify the cognitive demands of the measuring tasks, thus freeing children to think about, and anticipate data variability, and allowing more test runs to be done within the available time. The literature we read had less to say about ways to teach skills at this investigative stage and for these purposes and so we had to devise some strategies of our own.

Strategy Five: Capturing patterns of data variation within repeats of a test

We modelled this strategy in the context of the *Truck Track* task. The considerable variability in individual truck runs has been described in Section Three and this strategy was devised as a means of capturing that variability quickly and visually. We used a large strip of brown parcel paper as the landing surface of the truck travel zone. With whatever variables were chosen as the beginning test, 10 runs were completed in rapid succession and coloured dot stickers were used to mark the resting point of the truck after each run. Each focus group agreed in advance where each sticker would be placed — for example, "in the centre rear", or "beside the right front wheel".

The teachers responded enthusiastically to this strategy, especially as the dot pattern built up. One group spontaneously decided to mark their dots with the run number, creating a sequenced pattern that led them to a discussion about whether the runs began to get longer as the truck wheels were warmed up by friction on the ramp. They recognised this as a question that could be tested using the strategy and other ideas for variations began to flow.

Table 28

Teachers' responses to Strategy Five

Response sought	Yes	No	Maybe	No response
Have you seen this strategy before?	12	10	NR	1
Would you consider using this strategy with your class?	23			

Every teacher said they would use this strategy. Again, although more than half of them said they had seen it before, none had mentioned it during the introductory discussion.

> *This is very kinesthetic and manageable – you could have multiple groups doing the same test for very little cost.* (Year 3–5 teacher who had not seen the strategy before)

As the teachers shared their ideas about unavoidable sources of error and the need to manage these, they also commented on opportunities for children to respond thoughtfully at all stages of the data pattern creation:

> *A fun and user friendly strategy. Predict before testing.* (Year 7–8 teacher)
>
> *A great idea — would get them thinking and they wouldn't really know what was being done.* (Year 4–5 teacher)
>
> *Again, no writng by children. Data can be moved around.* (Year 1–2 teacher)

Older students could measure from the dot centres to the ramp centre to calculate averages. Hence students could begin to learn about mathematical strategies for managing data variability that could arise and strategies for calculating and presenting data.

> *We did a marble run similar. We were trying to make each channel different and were looking to see what variables were needed to create it.* (Year 7–8 teacher)
>
> *Can actually measure if you want to.* (Year 5–6 teacher)
>
> *Creates patterns, allows use of statistics such as averages etc.* (Year levels not given)

Strategy Six: Capturing patterns of "instantaneous" measurements

Observations of the *Ball Bounce* task revealed the extent of the measurement challenge that children face when required to "stop" a split-second mark in a continuing pattern of movement – in this case determining the very top point in the bounce of a ball. Most of the children we watched simply guessed, repeating their measurements only if some gross deviation led them to see a result as erroneous. We have already suggested that collecting categoric data might provide one solution to this dilemma. Strategy Six was our attempt to put this recommendation into action. In this first instance, we wanted to model a way to quickly and easily collect patterns of repeat bounces of the same ball – that is, to make a context-appropriate strategy equivalent to that just modelled for *Truck Track* as Strategy Five.

We began by modifying metre rulers that might typically be used for a task such as *Ball Bounce*. Using different colours of "electrical tape" purchased from a discount chain, we marked 10 cm lengths as various coloured "bounce zones". (One advantage that occurred to us only as we did this was that we could mask at least one measurement scale on the ruler, reducing its potential ambiguity if absolute measurements were to be taken later.) Holding the ruler vertically, it is possible to quickly record at least 10 bounces for each ball, marking the appropriate zone(s) on a tally chart such as that shown in Figure 6. Even so, the top of the **first** bounce may not be clearly spotted and it is necessary to look directly at the likely colour zone to "stop" the highest point.

Figure 6

Model of categoric data table for Ball Bounce

Type of ball	Red zone	Blue zone	Green zone	Yellow zone	Black zone
Tennis ball					
Squash ball					
Foam ball					
Table tennis ball					

The use of such zones allows an overall pattern for each type of ball to emerge, despite the difficulties inherent in making accurate measurements. Most teachers supported the use of this strategy — see Table 29 — and the opportunity to "get away from numbers" and to gather data more visually, was the most commonly cited reason amongst their open responses on the feedback sheets:

> *The colours on the ruler make it that much easier for students to identify the zone! Excellent idea!* (Year 7–8 teacher)
>
> *This would help provide critical visual concepts.* (Year 6–8 teacher)

Table 29

Teachers' responses to Strategy Six

Response sought	Yes	No	Maybe	No response
Have you seen this strategy before?	9	12	NR	2
Would you consider using this strategy with your class?	18		3	2

Some teachers talked about how scientists use technology to "stop" time when measurements have to be determined very quickly. They saw opportunities to combine this more visual preliminary approach with the development of some ICT skills as they video-taped the bounce trajectory of each ball against either zones or a numerical scale. Some teachers thought the "zone" strategy was better suited to younger children.

> *Prediction and "why" for predictions would be interesting. Students would love it. Great problem solving strategies — developing use of ICT equipment for those who want to define the area more.* (Year 7–8 teacher)

While the strategy works well for gathering the bounce patterns of one type of ball, there can be a complication when 2 types of balls bounce in the same zone. This caused some reservations and discussion. Some teachers saw this as an opportunity to revise the sampling strategy to make finer grained distinctions where necessary. Others thought it would be an unsatisfactory result.

> *I see this as creating a pattern and a discussion point – e.g. the "high" white or "low" white – it's easy to integrate with numeracy.* (Year 1–2 teacher)

Strategy Seven: Comparing tests in a series: Patterns on paper

Comparing categoric data for different ball bounce patterns turned the attention of the focus group session to the need to see patterns across related tests in a series, as well as within a specific test. The cognitive challenges for young children have been described at some length in Section Four. It seemed to us that almost all the NEMP groups failed to attend to the overall "experimental space". Few gave any thought to sequencing tests for any reason, let alone in relation to some causal theory. With these findings in mind, we revisited the strategy of placing coloured dots on brown paper, but this time we used 2 sets of dots, to directly compare a pair of VOTAT (vary one thing at a time) tests.

Again, most teachers were very enthusiastic about this strategy. And again, although nearly half said it was familiar, none had described it in the initial discussion.

Table 30

Teachers' responses to Strategy Seven

Response sought	Yes	No	Maybe	No response
Have you seen this strategy before?	10	12	NR	1
Would you consider using this strategy with your class?	20		1	1

Unsurprisingly, given their close relationship, reasons given to endorse the use of the strategy were similar to those given for Strategy Five. Teachers appreciated the visual nature of the resultant data patterns, seeing that these would lead to much discussion, and that they would stimulate predictions:

> *Very visual and able to make comparisons more easily, able to see clusters.* (Year 3–4 teacher)
>
> *To select aspects of predictions and results that can be justified/explained.* (Year 7–8 teacher)

> *Looking at patterns. Why? What changes? Really good to integrate with numeracy.* (Year 1–2 teacher)

Strong links were seen to teachers' numeracy programmes, and to the statistics component of *Mathematics in the New Zealand Curriculum.* As one teacher said "There's your graph really. It's done for you." One teacher suggested beginning with whole class modelling and then letting groups investigate a question of their choice.

Strategy Eight: Comparing tests in a series: Data recording sheets

The research literature revealed that holding a whole series of tests in their memory space can be very challenging for children. For our final focus group strategy we modelled a form of paper-based, relatively conventional representational scaffolding of the type suggested by Toth et al (2002). This took the form of a table on which successive test results could be entered and then directly compared. We modelled this for the *Emptying Rate* task, as shown in Figure 7 on the next page.

As for the Paired Planning Sheets in Strategy Two, some teachers thought this type of recording sheet was "too much information on a page" and that children would need to be carefully taught how to use the approach. Six teachers of younger children rejected the strategy in its present form, although only 2 said unequivocally that they would not consider using it, as shown in Table 31.

> *Never used — emphasis is at higher age level — too much (written) recording — it's boring to read/display/report on if children are working in co-operative groups.* (Year 1–2 teacher)
>
> *Too much for younger children but it could be modified.* (age levels not given)

Table 31

Teachers' responses to Strategy Eight

Response sought	Yes	No	Maybe	No response
Have you seen this strategy before?	11	9	NR	3
Would you consider using this strategy with your class?	16	2	3	2

Some teachers, while remaining cautious, saw possibilities for modification. Two suggested developing a large wall chart along the lines modelled, using this to model the recording of serial results with the whole class.

> *I like the clear way of showing what they have actually done and the results gained. Maybe some would need the sheet partly filled in at Year 5/6.*

Figure 7

Recording sheet for a series of tests

Our question	What we have varied	What we have kept the same	What happened? ✓ = fastest	What patterns did we find?
Which empties fastest – the bottle with the big hole or the bottle with the small hole?	One bottle has a **small** hole One bottle has a **big** hole	Both bottles have **300 mls** of liquid Both bottles have **water** Both bottles are held **straight**	big hole │ small hole ✓✓✓✓ ✓✓✓✓	The bottle with the big hole always empties fastest.
Which empties fastest – water or detergent?	One bottle has **water** One bottle has **detergent**	Both bottles have **300 mls** of liquid Both bottles have a **small** hole Both bottles are held **straight**	water │ detergent ✓✓✓✓ ✓✓✓✓ ✓✓	The water empties the fastest – but they are pretty close.
What will happen if we try the same test with a bigger hole?	One bottle has **water** One bottle has **detergent**	Both bottles have **300 mls** of liquid Both bottles have a **big** hole Both bottles are held **straight**	water │ detergent ✓✓✓✓ │ ✓✓ ✓✓✓✓ ✓	The bottles emptied very fast. Sometimes the detergent seemed to be first but it was hard to be sure. ***We could try using 900 mls of liquid next time.***

SECTION SEVEN: AREAS FOR FURTHER CONSIDERATION

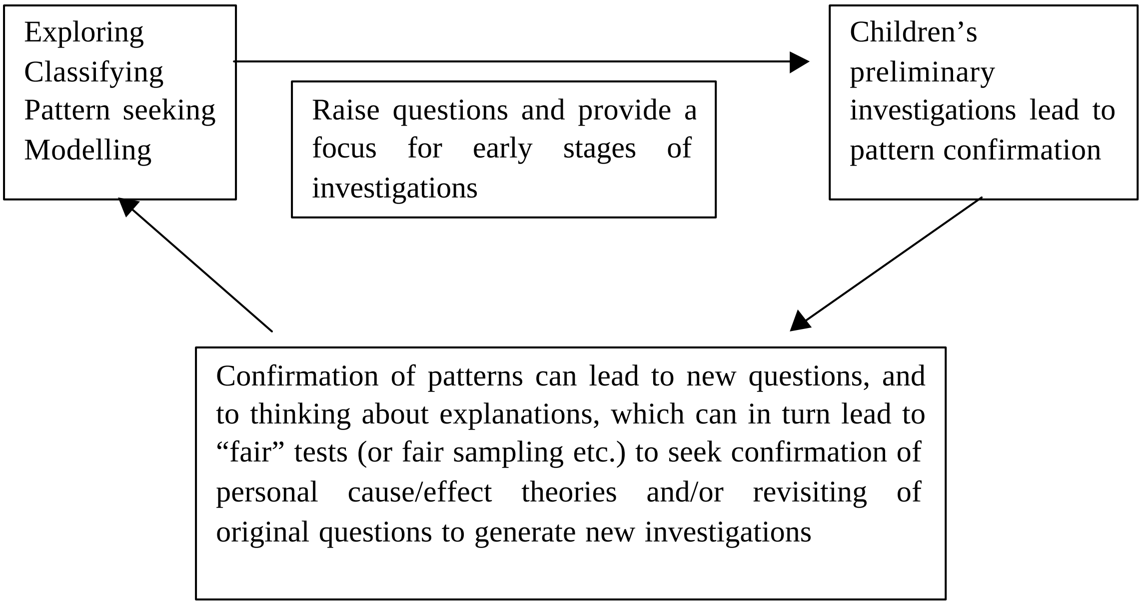
This section is something of an “after word”. It draws several loose threads together to complete the report whilst identifying areas where further work is needed. Areas where our findings might merit further consideration by the NEMP Board are also outlined, along with some suggestions for modified strategies to assess children’s progress in developing their skills of science investigation.

AN AFTER WORD ON FAIR TESTING

Section Four summarised concerns expressed by some science educators about an exclusive focus on fair testing as “the” method of primary science investigations. It seems to us that the strategies we have reported here need not be seen as exclusively supporting such a model, if teachers are encouraged to think more critically about the role of “fair” comparisons in various types of science investigations. If thought of in broader terms, the notion of “fairness” can come to be seen as an approach that helps to eliminate alternative explanations. However, other types of investigations might be better for generating those alternatives in the first place. A mental model for thinking about such relationships could look something like this:

Figure 8

Fair testing in the context of wider investigations



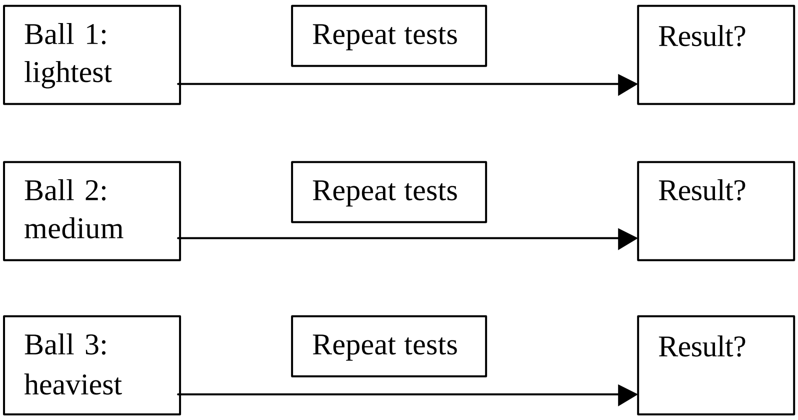
In these wider contexts fair testing becomes a tool that has relevance for a range of types of investigations, rather than an end in itself. The strategies described in Section Six would lend themselves well to this more inclusive approach, where teachers and their students draw on whichever of the various investigation types best suits the particular stage of their research, and where one investigation may well be an iterative, cycling series of mini-investigations. When our focus group teachers expressed a desire to play with the friction/heat impact on the travel of the toy truck just such a sequence could easily have begun!

AN AFTER WORD ON DEVELOPING MENTAL MODELS OF CAUSALITY

The mental models of causality outlined in Sections Four and Five present some very real challenges for teaching investigative skills in science. Typical approaches to fair testing treat each set of comparative tests as an entity complete in itself. Provided variables are managed correctly, appropriate measurements are taken with sufficient care, and the correct conclusions are drawn from the results reported, the student is considered to have demonstrated an appropriate skill set This is as true for secondary students as for primary, as demonstrated by the requirements for Achievement Standard 1.1 – the investigation standard for Year 11 students. Arguably, only the contexts and theoretical ideas that ground the investigation differentiate its skill level from the tests done by primary school pupils. However, as outlined in Section Four, adjacent sets of tests can potentially yield conflicting results when there are interactions between variables. Such results may generate considerable uncertainty until the puzzle that they present is unravelled – and that will demand a careful separation of theory from evidence, and a planned sequence of related investigations.

How important is it that students be given opportunities to confront this challenging aspect of scientific inquiry? Clearly it would be inappropriate for young children to be doing other than gradually building their basic CVS skills in the manner outlined in earlier sections. But should we be expecting students who have mastered the idea of unconfounded CVS tests to be extending their skills in contexts that require them to consider interactions between variables? The literature suggests that such skills are within the grasp of even upper primary students – with appropriate teaching. But what might such teaching look like and how would it differ from the present emphasis on the “stand-alone” fair test? The ideas presented next are set in the context of the *Ball Bounce* task. They are, however, speculative. There was not sufficient time to engage the focus group teachers with the more lengthy sequence outlined. In any case our observations (both of children’s actions and teachers' ideas) suggest that this type of approach might be seen as going well beyond anything that should be expected of most school students.

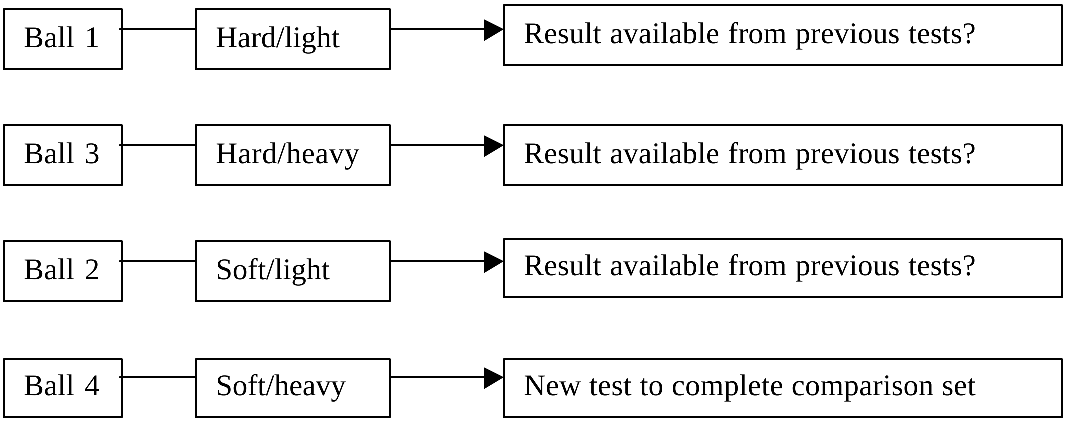
The least sophisticated possible approach to the *Ball Bounce* task is the familiar one taken in the NEMP testing context. Addressing the question of “Which ball is the bounciest?” has no meaning beyond task completion because no theory of causality is invoked or demanded. What would a test series look like if students were encouraged to clarify/explore their personal theories of causality? For example, a group who said “The ping pong ball will bounce the highest because it is the lightest” might be encouraged to arrange the balls to form a continuum from the lightest to the heaviest and check for a corresponding pattern of heights in the bounce zones achieved. If they used just 3 types of ball, this could result in the following 3-test series (assuming each test is repeated to detect data variations as outlined in Section Six):



Taken together these results may well confirm the personal theory expressed. In which case students could be challenged to try more types of balls, learning in the process about the dangers of generalising from a small sample. On the other hand, it is entirely possible that the combination of types of balls chosen will not yield the result expected – some small, dense, foam balls are very bouncy even though they are relatively heavy. If this should happen, students will be challenged to reconsider their theory in the light of the evidence generated. (This in itself would be a really useful skill for students to be encouraged to develop — especially given the critique that suggests teachers do not encourage students to make links between theory and evidence.)

Suppose the puzzled students now decided that hardness or softness could be causal agent. Again, they could be encouraged to form a continuum of ball types for this property – a process in itself that would demand more from them than simply measuring weight, since some test of softness would need to be devised. Again, results might be confirmed – or they might not.

Students would then face the challenge of deciding if both variables were implicated in some form of interaction. This would require them to start comparing 2 variables at once. The question now might become:



As suggested above, this new stage need not require all-new tests, if results have been kept systematically for all previous tests. Rather, evidence generated by previous tests could be reconsidered in a new light. Of course, it is entirely possible that other variables such as the material composition of each ball type are implicated in bounciness, too.

As outlined here, such a sequence would not be inherently more difficult than planning and conducting single tests. It would certainly take much more time and might call for some interesting problem solving along the way, but the basic CVS strategy still sits at the heart of the process.

Challenging teachers to work with children's theories of causality in this way can apply equally well to the single tests devised and carried out by younger children using simpler tasks such as *Truck Track*. It could be modelled in the scripted responses of the teacher-facilitators who deliver the NEMP tasks, especially if they were given opportunities to explore the range of likely responses as part of their training programme.

OTHER RECOMMENDATIONS TO THE NEMP BOARD

It would help teachers to think more widely about types of investigations if NEMP investigation tasks modelled these. Simple, assessable investigations set in the *Living World* and *Planet Earth* strands are no doubt challenging to design, but their absence implies that investigations are most appropriately set in a narrow range of physical science contexts.

Given the clear finding that children can identify and select fair tests well before they can actually design these, it would seem advisable to design some assessment tasks that model this type of approach. The *Emptying Rate* card sets (Figure 4) shared with the focus group teachers might be a good strategy for use in more formal assessment contexts. Design of such sets would be possible for probing understanding of a range of investigations, including those set in *Living World* or *Planet Earth* settings that would require longer time periods if actually carried out. This might be one means of addressing the challenge of curriculum coverage in assessing investigative skills, and it would also subsequently provide interesting new "ready-made" resources for teachers.

Children are not able to demonstrate their planning skills in unfamiliar contexts. Giving children time to play with equipment seems to be essential if they are to demonstrate what they can actually anticipate and manage. Testing planning ideas at the end of a simple investigative sequence, rather than at the start as at present, would seem advisable.

The challenges provided by measurement need careful attention in planning the overall investigative context. The collection of categoric data, or visual data patterns, may assist children to more easily recognise data trends, and hence to more freely talk about their causal theories, and to display the skills of relating these to the evidence generated. However, these types of responses would seem to require scripted prompts because supervising teachers do not typically spontaneously prompt children to think in these "scientific" ways.

An emphasis on repetition *per se* is unhelpful, especially as children seem to repeat because they think they have made a "mistake" and they are actually still seeking to make one "correct" measurement. Again, different strategies for measurement/data collection need to be modelled if children are to be given opportunities to demonstrate their emergent awareness of the need to manage experimental error.

It might be helpful to de-emphasise allocation of role in scripted planning prompts. Such prompts seem to turn children's attention away from the scientific aspects of the task at hand.

REFERENCES

Abrams, E., & Wandersee, J. (1995). How does biological knowledge grow? A study of life scientists' research practices. *Journal of Research in Science Teaching, 32* (6), 649–663.

American Association for the Advancement of Science. (2001). *Designs for science literacy*. Oxford University Press.

Barker, M. (2000). Learning theories in science teaching — hindsight or master plan? *New Zealand Science Teacher, 94,* 32–39.

Biddulph, F., & Carr, K. (1999). Learning theories and curriculum. *Teachers and Curriculum, 3,* 31–35.

Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences In F. Weinert & W. Schneider (Eds.) *Individual development from 3 to 12: Findings from the Munich longitudinal study* (pp. 38–54). Cambridge: Cambridge University Press.

Chalmers, A. (1982). *What is this thing called science?* Queensland: University of Queensland Press.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70* (5), 1098–1120.

Chinn, C., & Brewer, W. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, 19, 323–393.

Chinn, C., & Malhotra, B. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, *94* (2), 327–343.

Coles, B., MacIntyre, B., & Radford, A. (2001). What to exemplify; science education and the cultures of our schools. *Annual Conference of the New Zealand Association for Research in Education*, Christchurch, 6–9 December.

Crooks, T., & Flockton, L. (1996). *Science Assessment Results 1995. National Educational Monitoring Report 1.* University of Otago: The Educational Assessment Research Unit.

Crooks, T., & Flockton, L. (2000). *Science Assessment Results 1999: National Educational Monitoring Report 13.* University of Otago: The Educational Assessment Research Unit.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science.* Buckingham: Open University Press.

Fensham, P. (1985). Science for all. *Educational Leadership, 44,* 18–23.

Fensham, P. (1994). Progression in school science curriculum: A rational prospect or a chimera? *Research in Science Education, 24*, 76–82.

Gilmore, A. (2001). *National Education Monitoring Project: Assessment Resources for Teachers.* Report to the Ministry of Education. Christchurch: University of Canterbury.

Gott, R., & Mashiter, J. (1994). Practical work in science: A task based approach? In R. Levinson (Ed.) *Teaching science*. London and New York: Routledge.

Haigh, M. (1999). Learning to be investigators. *Australasian Science Education Research Association (ASERA) Annual Conference,* Rotorua, July 1999.

Harrison, C., Simon, S., & Watson, R. (2000). Progression and differentiation. In M. Monk & J. Osborne (Eds.) *Good practice in science teaching: What research has to say* (pp. 174–189). Buckingham, Phildelphia: Open University Press.

Hipkins, R., Bolstad, R., Baker, R., Jones, A., Barker, M., Bell, B., Coll, R., Cooper, B., Forret, M., France, B., Haigh, M., Harlow, A., & Taylor, I. (2002). *Curriculum, learning and effective pedagogy: A literature review in science education.* Wellington: Ministry of Education.

Hogan, K. (2000). Exploring a process view of students' knowledge about the nature of science. *Science Education, 84* (1), 51–70.

Klahr, D. (1999). Comment: “Scientific Reasoning”. In F. Weinert & W. Schneider (Eds.) *Individual development from 3 to 12: Findings from the Munich longitudinal study*. Cambridge: Cambridge University Press.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction, 18* (4), 495–523.

Masnick, A., & Klahr, D. (2001) *Elementary school children's understanding of experimental error*. Paper presented at Twenty-Third Annual Conference of the Cognitive Science Society.

Mayer, V., & Kumano, Y. (1999). The role of system science in future school science curricula. *Studies in Science Education, 34*, 71–91.

Metz, K. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research, 65* (2), 93–127.

Millar, R., & Osborne, J. F. (Eds.) (1998). *Beyond 2000: Science education for the future*. London: Kings College.

Ministry of Education. (1993). *Science in the New Zealand Curriculum*. Wellington: Learning Media.

Osborne, J., Erduran, S., Simon, S., & Monk, M. (2001). Enhancing the quality of argument in school science. *School Science Review, 82* (301), 63–70.

Poskitt, J. (2002). National consultation on exemplars: What difference does it make for teachers? ***set**: Research Information for Teachers, 3*, 13–16.

Rollnick, M., Lubben, F., Lotz, S., & Dlamini, B. (2002). What do underprepared students learn about measurement from introductory laboratory work? *Research in Science Education, 32*, 1–18.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32* (1), 102–119.

Sheehy, N., Wylie, J., McGuinness, C., & Orchard, G. (2000). How children solve environmental problems: using computer simulations to investigate systems thinking. *Environmental Education Research, 6* (2), 109–126.

Smith, C., Maclin, D., Houghton, C., & Hennessey, G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction, 18* (3), 349–422.

Toth, E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: a cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction, 18* (4), 423–459.

Toth, E., Suthers, D., & Lesgold, A. (2002). “Mapping to know”: The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education, 86* (2), 264–286.

Watson, R. (1996). Rethinking readiness for learning. In D. Olson & N. Torrance (Eds.), *The handbook of education and human development* (pp. 148–171). Cambridge, Mass: Blackwell.

Watson, R., Goldsworthy, A., & Wood-Robinson, V. (1999). What is not fair with investigations? *School Science Review, 80* (292), 101–106.

Watson, R., Goldsworthy, A., & Wood-Robinson, V. (2000). Beyond the fair test. In J. Sears & P. Sorensen (Eds.), *Issues in Science Teaching* (pp. 70–79). London and New York: Routledge Falmer.

Zeidler, D., Walker, K., Ackett, W., & Simmons, M. (2002). Tangled up views: Beliefs in the nature of science and responses to socioscientific dilemmas. *Science Education, 86* (3), 343–367.

Zohar, A. (1995). Reasoning about interactions between variables. *Journal of Research in Science Teaching, 32* (10), 1039–1063.

APPENDIX I

NEW ZEALAND COUNCIL FOR EDUCATIONAL RESEARCH
TE RŪNANGA O AOTEAROA MŌ TE RANGAHAU I TE MĀTAURANGA

USING NEMP TO INFORM THE TEACHING OF SCIENTIFIC SKILLS

Task Name: Truck Track

Tape Number: ______________________

Team: ☐

Year Level: Year 4 ☐

Gender Mix **Boys** __________ **Girls** __________

Ethnicity: **Non-European** ____________ **European** ______________

A. PLANNING STAGE:

1. Task Roles (within grp) ☐
   - Discussed with each other and decided ☐
   - Used hand, rock, paper, scissors/or rhyme ☐

2. Discussion of what to keep the same ☐

3. Accuracy
   a) ➢ **Position of ramp on mat**
      - i) Intuitive ☐
      - ii) Explicit/addressed ☐
      - iii) Ignored ☐

   b) ➢**Ramp straight**
      - i) Intuitive ☐
      - ii) Explicit/addressed ☐
      - iii) Ignored ☐
      - iv) Ramp on angle ☐

   c) ➢ **Accurate measurement of truck travel**
      - i) Intuitive ☐
      - ii) Explicit/addressed ☐
      - iii) Ignored ☐

   d) ➢ **Consistent set up and release of the truck**
      - i) Intuitive ☐
      - ii) Explicit/addressed ☐
      - iii) Ignored ☐

e) ➢ **Checking consistency by replicating**
  - i) Intuitive ☐
  - ii) Explicit/addressed ☐
  - iii) Ignored ☐

f) ➢ **Measuring units e.g., cm instead of mm**
  - i) Intuitive ☐
  - ii) Explicit/addressed ☐
  - iii) Ignored ☐

g) ➢ **Full member participation**
  - i) All participated ☐
  - ii) 3 participated ☐
  - iii) 1–2 participated ☐

4. Purpose (do they talk about the purpose of the task) ☐

5. Sequencing (do they do the corks in a systematic order e.g., 1,2,3,4) ☐

6. Repetition
   - i) To correct error ☐
   - ii) For accuracy/fair testing standards ☐
   - iii) Ignored ☐

7. Variables
   - a Level of mat-bumps ☐
   - b Changes in surface ☐
   - c Truck pushed instead of let go ☐
   - d Objects in the way e.g., chair ☐
   - e Ruler placed to the side of ramp to keep the truck moving straight/ on a set path ☐

8A. Other ☐

B. CARRYING OUT STAGE:

1. Task Roles (within grp) ☐

2. Discussion of what to keep the same ☐

3. Accuracy

i) **Truck forwards**	a)	i)	ii)	iii)	
	b)	i)	ii)	iii)	iv)
	c)	i)	ii)	iii)	
	d)	i)	ii)	iii)	
	e)	i)	ii)	iii)	
	f)	i)	ii)	iii)	
	g)	i)	ii)	iii)	

i) **Truck backwards**	a)	i)	ii)	iii)	
	b)	i)	ii)	iii)	iv)
	c)	i)	ii)	iii)	
	d)	i)	ii)	iii)	
	e)	i)	ii)	iii)	
	f)	i)	ii)	iii)	
	g)	i)	ii)	iii)	

4. Purpose (do they talk about the purpose of the task) ☐

5. Sequencing (do they do all the corks in a systematic order e.g., 1,2,3,4) ☐

6. Repetition
   i) To correct error ☐
   ii) For accuracy/fair testing standards ☐
   iii) Ignored ☐

7. Variables
   a ➢ Level of mat-bumps ☐
   b ➢ Changes in surface ☐
   c ➢ Truck pushed instead of let go ☐
   d ➢ Objects in the way e.g., chair ☐
   e ➢ Ruler placed to the side of ramp to keep the truck moving straight/ on a set path. ☐

8B. Other ☐

9. Predicting (after the first one) ☐

C. RESULTS STAGE/DISCUSSION

1. Task Roles (within grp) ☐
2. Discussion of what to keep the same ☐

3. Accuracy

i) **Truck forwards**	a)	i)	ii)	iii)	
	b)	i)	ii)	iii)	iv)
	c)	i)	ii)	iii)	
	d)	i)	ii)	iii)	
	e)	i)	ii)	iii)	
	f)	i)	ii)	iii)	
	g)	i)	ii)	iii)	

i) **Truck backwards**	a)	i)	ii)	iii)	
	b)	i)	ii)	iii)	iv)
	c)	i)	ii)	iii)	
	d)	i)	ii)	iii)	
	e)	i)	ii)	iii)	
	f)	i)	ii)	iii)	
	g)	i)	ii)	iii)	

4. Purpose (do they talk about the purpose of the task) ☐

5. Sequencing (do they do all the corks in a systematic order e.g., 1,2,3,4) ☐

6. Repetition
   - i) To correct error
   - ii) For accuracy/fair testing standards
   - iii) Ignored

7. Variables
   - a ➢ Level of mat-bumps ☐
   - b ➢ Changes in surface ☐
   - c ➢ Truck pushed instead of let go ☐
   - d ➢ Objects in the way e.g., chair ☐
   - e ➢ Ruler placed to the side of ramp to keep the truck moving straight/on a. set path ☐

8C. Other ☐

9. Predicting ☐

10. Mistakes in technique used to explain result (e.g., rug not flat) ☐

11. Recognising patterns/trends
   - i) Patterns in words only ☐
   - ii) Pattern refers to comparison of numbers/measurements ☐

12. Group consensus — if yes tick ☐

13. Anything you could have done more accurately now you have done the experiment?
   - i) better consistency in releasing truck ☐
   - ii) more accurate measuring ☐
   - iii) included replication ☐
   - iv) completed all the truck directions ☐
   - v) recorded the results better e.g., use all one unit etc. ☐
   - vi) Other ☐

D **POE**: EXTRA CORK

8D. Other ☐

14. Group consensus — if yes tick ☐

15. Use of their results to make a prediction ☐

16. a) Accurate prediction in right direction ☐
   - b) Incorrect prediction ☐

17. Mention of physical properties (weight distribution/slope of ramp) ☐

18. Gravity ☐

E TEACHER INSTRUCTIONS

19. Planning task (emphasis on **planning – things kept the same**) ☐

20. Planning roles (emphasis on **roles** students do) ☐

21. Interruptions to re-emphasise things the same or sequence of corks ☐

22. Focus of discussion is on results ☐

23. Prompts to check number pattern ☐

24. Focus of discussion is on meaning of purpose ☐

25. Focus is on language/science language e.g., gravity ☐

26. Teacher makes links to other relevant contextual knowledge ☐
Example(s)____

27. Students makes links to other relevant contextual knowledge ☐
Example(s)____

28. Helpful teacher actions ☐
Example(s)____

29. Unhelpful teacher actions ☐
Example(s)____

30. **Extra variables introduced by the teacher**
a) Rug not flat (bumpy) ☐
b) Where the ramp has been set up in relation to mat
i) end of ramp touches the start of mat ☐
ii) corks placed on the end of mat ☐
c) Equipment and mat set up on a table join over two tables ☐

8E. Other
____

APPENDIX II

SAMPLE OF FEEDBACK SHEET FOR FOCUS GROUP SESSIONS

Feedback on strategies for structuring awareness of fair tests

A Matched pairs of equipment

please circle

1. Have you seen or used this strategy before? 1. Yes 2. No

2. (i) Would you consider using this strategy with your class? 1. Yes 2. No 3. Maybe

(ii) If *yes*, what appeals to you about this strategy? *(tick all that apply)*

☐ It makes the skill focus more obvious.
☐ It provides clear teaching point(s).
☐ It is appropriate for the stage my children are at.
☐ Children would find this fun/enjoyable.
☐ It will stretch children's thinking about investigations.
☐ It models small group strategy.
☐ It models whole class teaching.
☐ Other/ comments (explain briefly)

(iii) If *maybe* or *no*, which reasons would influence your decision? *(tick all that apply)*

☐ It would be/is better for older students.
☐ It would be/is better for younger students.
☐ It is too much work to prepare materials/ get the equipment.
☐ I would use it if materials were provided in a ready to use format.
☐ It could take too long to carry out in class.
☐ I prefer children to find planning ideas for themselves.
☐ It is more important to practice actual investigation skills.
☐ It wouldn't interest my children.
☐ Other/ comments (explain briefly)