SIMPLE TEACHING STRATEGIES TO USE DURING PRIMARY SCIENCE INVESTIGATIONS

Rosemary Hipkins and Natasha Kenneally
New Zealand Council for Educational Research

Paper presented at ASERA Conference, Melbourne, July 8-12, 2003

ABSTRACT

Videotaped sessions of students working on ‘fair testing’ investigation scenarios were analysed (N = 204). The analysis suggested avenues for the development of strategies that could help develop teachers’ pedagogical content knowledge for teaching the skills used *during* a simple ‘fair testing’ investigation. Most conversation during each activity focused on specific aspects of measuring or task completion so that the overall investigation appeared to become a series of disjointed episodes. Simple strategies intended to help students hold more of the overall investigation in their memory space whilst laying a foundation for later learning about experimental error and data management were devised and tried out during three focus groups sessions held with experienced primary school teachers. The data patterns generated were seen by the teachers to link science and mathematics learning, and to be accessible to those students who struggled with language and/or literacy issues.

INTRODUCTION

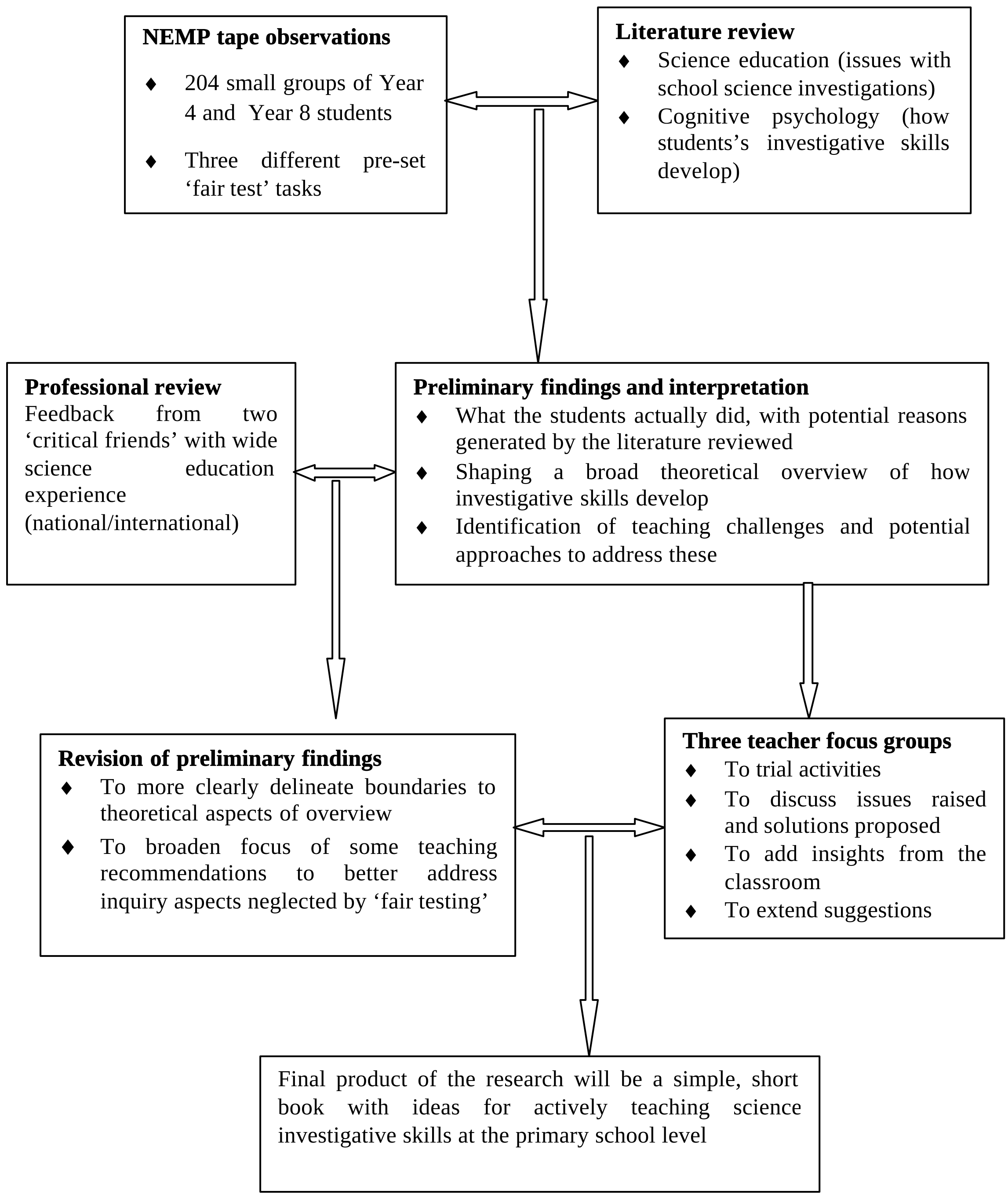
This paper reports on one thread of a project that was designed to inform discussion in New Zealand about effective approaches to teaching science investigation skills to primary school students. Previous research found that many teachers do not appear to actively teach students the key objectives of the ‘Developing Scientific Skills and Attitudes’ strand of New Zealand’s science curriculum (Gilmore 2001). We set out on this project with the intention of seeking ways to help teachers build bridges between simple investigation scenarios and their actual classroom pedagogy. We used a multifaceted approach to the research, as summarised in Figure 1.

During the first stage of the research we observed 204 videotaped episodes of small groups of Year 4 or Year 8 students carrying out simple science investigations. We anticipated that patterns we might see in the students’ investigative actions could reflect areas in which to develop teachers’ pedagogical content knowledge (PCK) (Shulman 1986). Barnett and Hodson (2001) list four different aspects of PCK for science teaching. In the context of investigations these include: curricular saliency; alternative ways of representing the subject matter of investigations; a knowledge of teaching and learning strategies that will be effective in investigative contexts; and a knowledge of learners’ existing understandings of investigations. We hoped to contribute useful advice for teachers in all of these areas except curricular saliency. To this end, the analysis of our observations was informed by a review of existing research on the development of students’ investigative skills.

During the second stage of the research we drew on the first-stage analysis to design some simple teaching strategies that we thought might assist teachers to actively develop students’ investigative skills. We planned from the outset to include a dialogue with some primary teachers as part of our overall plan for the research because we knew that any strategies we suggested would need to be seen to 'work' if they were to have any chance of being adopted (Appleton & Doig 1999). Three focus groups of teachers were convened to test and refine these strategies.

The actual 'doing' stage of a science investigation has emerged from the research as a potentially fruitful focus for further attention for the development of teachers' PCK. Focus group teachers did not spontaneously describe teaching strategies to use during this stage and the students' focus was typically on task completion or on the act of measuring and recording data. In the light of these findings we report on two simple teaching strategies we devised to streamline the process of capturing data patterns for the whole investigation, and we make suggestions about their wider potential.

Figure 1: The overall inquiry process



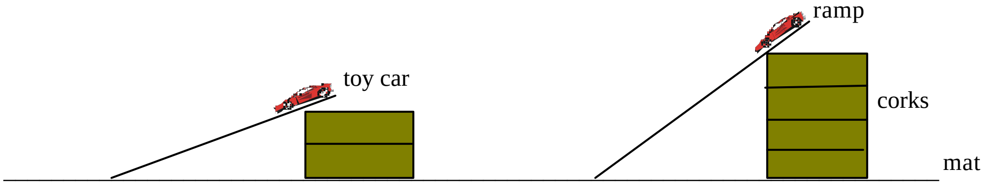
Observing students' investigative actions

The videotapes of small groups of students carrying out simple 'fair tests' were originally recorded as part of New Zealand's *National Education Monitoring Project* (NEMP). Once tasks devised for NEMP assessment purposes are no longer needed they are published and teachers are encouraged to use them as resource material (Gilmore 2001). From the two sets of published science assessment results (Crooks & Flockton 1996; 2000) three broadly similar 'fair testing' tasks were selected for detailed observation. Approximately 50 small groups of students were observed at each year level carrying out each task (actual numbers are shown with the task summaries and vary slightly because of small variations in the content of the tapes provided from the NEMP archives). Usually, each group carried out two successive tasks, beginning with a task called *Emptying Rate* that was used with both Year 4 and Year 8 students. After that Year 4 students completed a task called *Truck Track* and Year 8 students a task called *Ball Bounce*. Brief descriptions of each of these tasks follow.

Figure 2: The investigative tasks observed

Truck Track

A small toy truck was raced down a flat ramp. Year 4 students were shown how the slope of the ramp could be altered by the use of one, two, three or four supporting corks and they were asked to investigate and compare the distance the car travelled at each slope setting. The students were told they could also 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. (N = 51 Year 4 groups)

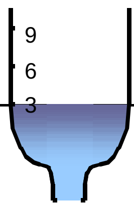


Ball Bounce

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. (N = 52 Year 8 groups)

Emptying Rate

Students 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 on the sides of the funnel away from the lid. Using a stop watch students were asked to determine the time it took water to drain when the funnel was filled to each of these three marks, and then to do the same for detergent. (N = 48 Year 4 students and 53 Year 8 groups). Students were shown a short introductory video of the equipment to be used in this task.



Once the tasks had been selected, the analysis began with the creation of an observation schedule for each of the three tasks. Two researchers together watched three to five tapes for each task and 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. After ten observations for each task had been completed, the observation schedule was re-checked and adjusted if necessary.

Although the students were asked to do some planning and interpreting, our focus in the next section is on the actions that took place in the 'experimental space' in between these other two types of activities (Klahr 1999). Because students were presented with a predetermined investigation to carry out, 'planning' was somewhat circumscribed. Similarly 'interpretation' was of limited scope because students had not had the opportunity to design tasks in the light of previous discussions about inherent causal mechanisms. Interpretation was typically rather rushed as teachers sought to maintain busy schedules of assessment events. A more positive reason for focusing on students' actions *during* the investigation is that we saw rich opportunities here to help teachers extend their PCK.

WHAT ACTUALLY HAPPENS WHEN STUDENTS INVESTIGATE

During all three investigations students were expected to demonstrate their ability to plan for and manage obvious variables. They were assessed on their ability to do this at both planning and carrying out stages of the investigation (Crooks & Flockton 1996; 2000). Accordingly, we took note of any explicit discussion of variables, at both planning and doing stages. Our initial observations alerted us to instances where students would carry out a simple action to manage a variable (for example straightening up the truck ramp if it got bumped) without making any comment as they did so, and without having stated during 'planning' that they intended to do so. It seemed to us that these students could 'do' more than they could 'say' and we shaped our observation schedule to capture data about this.

Table One reports on patterns of Year 4 students' actions and conversations during the 'doing' stage of the *Truck Track* investigations with a forward facing truck. Table Two reports on patterns of Year 8 students' actions and conversations during the 'doing' stage of the *Ball Bounce* investigation. These tables show that both Year 4 and Year 8 students were more likely to discuss the process of making the required measurements than to discuss the management of variables as they worked, although the difference is more pronounced at Year 4.

When data from all 51 Year 4 groups was collated (Table 1), there were just 14 instances where students talked about adjustments as they made these to keep variables constant, although they actually made these adjustments in 93 other instances as they worked. By contrast 54 explicit comments were made about aspects of measuring. The high number of these comments related to units of measurement is indicative of the challenges the students faced with the mechanics of the task. The students found the act of measuring and recording data laborious and the unfamiliar measuring instrument created uncertainty and distractions. (It appeared that many students had not handled a three-fold builder's rule before.)

Table 1

Patterns of Year 4 students' actions and conversations during the Truck Track task

Focus of students' actions	**number of groups observed N = 51**		
	Action made without comment	**Action taken and discussed**	**This aspect ignored**
Management of variables			
Consistent positioning of ramp and mat on table	31	4	16
Ramp straight on to mat	33	-	17
Consistent set up and release of the truck	29	10	10
Total instances	**93**	**14**	**43**
Measuring			
Accurate measurement of truck travel	5	31	11
Awareness of measuring units e.g., cm/ mm	3	22	24
Total instances	**8**	**54**	**35**

As they carried out their investigations the Year 8 students actively managed a greater number of variables that the Year 4 students, and they were more likely to explicitly discuss their actions. However Table 2 shows that many of these older students still ignored variables that were key to a 'fair' investigation. Year 8 groups were faced with a more difficult measuring challenge because each ball needed to be 'stopped' at the split second of the highest point of the bounce. While some groups ignored this challenge, most discussed what to do and a range of strategies emerged. For example 'averaging' actions typically involved taking the middle reading when students of different heights read the same measurement differently because of parallax. This decision was almost as likely to be taken without any explicit discussion as to be talked about and planned for.

The *Emptying Rate* task required two sets of measurements to be taken – water or detergent volume at the outset, and time elapsed during the actual trial. Because there were very few differences between the actions of the Year 4 and Year 8 groups that carried out this task, Table 3 reports on the overall sample. Actions taken when measuring the emptying rate of water are recorded first since students were asked to complete all water trials before beginning on the detergent trials. (In fact some groups ignored this directive, thereby introducing the variable of detergent contamination of the funnel.)

Table 2

Patterns of Year 8 students' actions and conversations during the Ball Bounce task

	Number of groups observed N = 52		
Focus of students' actions	**Action made without comment**	**Action taken and discussed**	**This aspect ignored**
Management of variables			
Keeping the ruler vertical	14	20	18
Consistent height for dropping balls	33	12	6
Strategy for determining the top point of the bounce trajectory	16	4	32
Ball bounced on same place on surface (to avoid cracks between desks)	1	41	10
Total instances	**64**	**77**	**66**
Measuring			
Keeping the ruler vertical	20	14	18
Discussing need for accurate measuring	-	43	9
Replicating when unsure of reading	12	21	19
Managing data variation – e.g. by use of an ad hoc averaging strategy	11	13	28
Measuring number of times ball bounces	-	7	45
Trials done to decide likely range of the bounce	1	-	51
Total instances	**44**	**98**	**170**

The pattern observed in the *Emptying Rate* task changed very little from the first test series (water tests) to the second (detergent tests). Most students took explicit care in attempting to measure water/detergent volumes to the pre-specified marks on the improvised bottle funnel, as modelled during a short introductory video that they had watched. 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 a relatively unfamiliar piece of equipment, students' discussions typically explicitly differentiated the flow time as measured by seconds not minutes. Younger students sometimes struggled to determine which of these units of time measurement to employ, otherwise there were no clear Year 4/8 differences in the students' investigative actions. Students 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 year level repeated measurements for sampling 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.*

Table 3

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

Focus of students' actions	Number of groups observed N = 101		
	Action made without comment	**Action taken and discussed**	**This aspect ignored**
Water test			
Protocol for use of stop watch button	4	97	0
Measuring to the mark/adjusting to the mark	7	87	7
Measuring units e.g., seconds or minutes	2	95	4
Total instances	**13**	**279**	**13**
Detergent test			
Protocol for use of stop watch button	6	92	1
Measuring to the mark/adjusting to the mark	4	88	7
Measuring units e.g., seconds or minutes	4	94	1
Total instances	**14**	**274**	**9**

When to *stop* measuring was an issue in the *Emptying Rate* task, especially when students moved from the water to the detergent tests. Because the water flowed quickly and consistently, only three 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 end point. However, as the following quotes show, the students 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 three detergent trials. This would appear to suggest these students were not working with an over-all 'fair test' plan in mind, but rather were moving from one test episode to the next. The significance of such a view is discussed next.

WHAT SENSE DID STUDENTS APPEAR TO MAKE OF THESE INVESTIGATIVE TASKS?

Having reported on the patterns of students' actions, we turn now to aspects of the analysis that we linked to findings from previous research on students' investigative skills. We hoped to generate potential explanations for the students' actions that we could then use to target fruitful strategies for teaching investigative skills. Aspects of the meaning that students appeared to take from their investigative experiences are discussed. Another aspect that did not arise in our observations, but which it seems important to take into account when thinking about the impact of teaching these aspects of investigative skills is also introduced.

A focus on 'doing'

We have noted that many students could 'do' more than they could 'say' and that this pattern was most pronounced at Year 4. Other researchers have found that younger students can select a fair test from a range of possible scenarios without necessarily being able to justify their choice. Learning to say *why* a test is fair appears to be an important developmental step that in turn precedes being able to design and produce fair tests independently (Bullock & Ziegler 1999; Toth, Klahr & Chen 2000). In these situations 'doing' similarly precedes 'saying'.

A sense of purpose

In all three tasks, the act of measuring, followed by written recording, seemed to partition sequential tests into distinct episodes that did not appear to be seen as parts of a whole, coherent test design. The literature suggests that such patterns are predictable. Even when they have planned a series of tests, students may 'lose their way' and deviate from their intended plan part way through an investigation (Schauble 1996). Students' memory capacity has been linked to their investigative ability (Bullock & Ziegler 1999) and so perhaps exacerbates this effect when memory space is still developing.

This does not necessarily mean that the whole investigation is meaningless. The opportunity to demonstrate an understanding of the overall investigation came when students reported that they had completed the task, and the supervising teacher returned to ask them about 'what they had found out'. Unsurprisingly, since the contrasts were rather obvious, most groups could describe in words the main effects that they had observed. However fewer made reference to their data to justify these patterns, even when prompted. This does seem to rather defeat the purpose of carrying out measurements at all.

Making sense of repetition

Both NEMP science cycles have reported that few New Zealand students repeat trials for fair testing purposes (Crooks & Flockton 1996; 2000). Indeed our observations would suggest that the teachers assessing this aspect of the NEMP tasks might have been rather generous in what they counted as constituting a repetition episode. For example, in the *Ball Bounce* task, where the greatest number of instances of 'repetition' were observed, trials were only repeated when students perceived there had been an error in the specific test. If a ball hit the hand of the child holding the ruler or the ruler itself, either on its descent, or after the bounce, the test might be rerun. Some groups also re-tested when they had been unable to determine the height of a bounce in relation to the ruler, because it was hard to 'stop' the moment at which the descent began. The end result of any repetition episodes was always to obtain a single 'best bounce' measurement. No groups repeated to obtain a data series that could be statistically manipulated although we have already noted that some ad hoc averaging strategies were used by some groups during the *Ball Bounce* task

Asked what they might do better another time, one group of Year 4 students thinking about the *Truck Track* task said: *Measure really, really, carefully.* This comment also appears to imply that with sufficient care it is possible to get one 'right' reading. This type of thinking has been described as becoming a hindrance to the development of scientific thinking about ways to manage data variability because it does not focus attention to establishing a pattern of measurements as the best means of managing inevitable errors of measurement (Rollnick, Lubben, Lotz & Dlamini 2002). Rollnick et. al. suggest that reinforcing the importance of single careful measurements actually creates a misconception about data management that is at least as hard to change as the more widely reported misconceptions of scientific phenomena.

If this situation is to change, students need opportunities to see the data patterns that emerge from sequentially repeated tests. However we have already noted that the laborious act of measuring and recording distracts from the immediate visibility of such patterns, even between different tests. Asking most students to explore such patterns within single tests, that then become part of bigger patterns of sequences of tests, seems to be a big ask. This was a key issue that we addressed in the activities we designed for the focus group sessions.

Distinguishing main effects from experimental error

The magnitude of the difference to be observed in comparative trials (effect size) impacts on the 'obviousness' of the evidence. Research has shown that both students and adults may invoke 'theory saving' beliefs when differences are small enough to be discounted as errors (Zohar 1995; Schauble 1996). Other research has shown that students who are already confident creators of fair tests may become less certain of the justifications they made for the strategies that they used after explicit teaching about controlling variables, even as they continue to generate appropriate fair tests (Toth *et al.* 2000). Toth et. al. suggest that students who respond in this way are becoming more attentive to variation in data outcomes, and that this raises their awareness of potential sources of experimental error, thus rendering them less certain of the justifications they give.

In all three of the tasks that we observed the main effects were quite obvious and so these issues did not arise. Nevertheless we note them here because it seems to us that the strategies we introduce next could provide an avenue for bringing the multifaceted challenges of dealing with measurement error into the open during discussions with students.

ADDING TO TEACHERS' REPERTOIRE OF STRATEGIES FOR SCIENCE INVESTIGATIONS

Toth et. al. (2000) point out that strategies that are effective in developing students' 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 students' learning – they want 'activities that work' (Appleton & Doig 1999). We introduced the strategies that we identified as a result of our analysis to three small focus groups of primary teachers who between them were teaching across the age range of the primary school from Year 1 to Year 8. Their ideas about the appropriate use of these strategies in the classroom are included in this section.

The notion of ‘fair testing’ has proved to be a powerful metaphor and 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 students’ knowledge of other forms of scientific investigations (Watson, Goldsworthy & Wood-Robinson 2000). Watson and his colleagues expressed concerns about the ‘nature of science’ (NOS) messages implicit in this narrow focus on just one type of investigation, and pointed out that such experiences also provide inadequate opportunities for students to learn about the relationships between the development of scientific theories and empirical evidence.

Our focus group findings suggest another point of ‘narrowing’ in the aspects of investigations that that are actively taught in school science. We began each focus group by asking the teachers to describe how they currently went about teaching fair testing. The ideas that were offered focused on the setting-up stages of a fair test. No teacher described strategies that addressed ‘fair measuring’ or data recording/data patterns, or the sequencing of a test series – all of which were identified by us as being problematic aspects of the students’ investigations. While we did not neglect the planning and setting up stages during the focus groups, this paper explores the dynamics of the actual ‘doing’ stage. The literature that we read had less to say about ways to teach skills at this stage of the investigative process and so we had to devise some strategies of our own.

Focusing on specific measuring challenges

A juxtaposition of our observations of students 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 challenge generated by the measurement activity required of students:

- 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 by students and so they 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 end-point;
- 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.

These ideas could be used as a check list to help teachers plan for tasks of suitable levels of complexity, or they could be a beginning point for the explicit teaching of measuring skills in science investigations.

Looking for data patterns

We introduced a simple visual strategy for capturing and thinking about the results of fair tests in situations where the measurement point is static rather than moving. We wanted to model a manageable way to modify data gathering so that students could 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). Our criteria for a good strategy included that it would: simplify the cognitive demands of the measuring tasks; free students to think about, and anticipate data variability; allow more test runs to be done within the available time, stimulate discussion and questioning; and be easily manageable in the classroom setting.

Capturing data variability in repeat trials

We devised a strategy that we modelled in the context of the *Truck Track* task. During the NEMP tape observations we had noticed the considerable variability in individual truck runs and our strategy was devised as a means of capturing that variability quickly and visually. We used a large sheet of brown parcel paper as the landing surface of the truck travel zone. With whatever variables were chosen as the beginning test, ten 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.

Every teacher said they would use this strategy and as they shared their ideas about unavoidable sources of error and the need to manage these, they also commented on opportunities for students 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. (Year 4-5 teacher).
>
> Again, no writing by students. Data can be moved around. Year 1-2 teacher).

Some teachers suggested that 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. They also discussed possibilities for moving the paper sheet to a wall display position where it could be annotated with students' comments and the pattern generated could be reflected on.

The strategy demonstrably 'worked' and this prompted the flow of suggestions that ensued.

Capturing data patterns for comparative trials

We have noted above that the students in the NEMP groups failed to attend to the overall 'experimental space'. With these findings in mind, we extended the strategy of placing coloured dots on brown paper, but this time we used two sets of dots, to directly compare a pair of tests. Again most teachers were very enthusiastic about this strategy.

> 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).

These teachers saw strong links to their 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.

There is evidence that New Zealand classrooms have a relatively high proportion of students born overseas, or who do not regularly speak English at home, compared with other nations (Sturrock & May 2002). The focus group teachers were very aware of the potential for providing such students with visual learning experiences. One commented that the strategy would help new arrivals to simultaneously learn science and language skills.

Creating data patterns when the object keeps moving

The strategy outlined above would not work for the *Ball Bounce* task. Observations of this task revealed the extent of the measurement challenge that students 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 students we watched simply guessed, repeating their measurements only if some gross deviation led them to see a result as erroneous. We wanted to model a way to quickly and easy 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* and we thought that collecting categoric data might provide a solution.

We began by modifying meter rulers that might typically be used for a task such as *Ball Bounce*. Using different colours of 'electrical tape' 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 ten bounces for each ball, marking the appropriate zone(s) on a tally chart. 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.

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 who supported the use of this strategy again commented on the opportunity to 'get away from numbers' and to gather data more visually:

> 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).

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 that the 'zone' strategy was better suited to younger students.

> 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 two 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 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).

Once again the teachers’ enthusiasm was linked to their identification of features that would allow this strategy to work in their classrooms, with their students, inside the parameters of the curriculum they knew they needed to deliver. This was another ‘activity that worked’.

CONCLUSIONS

Our research identified the collection and recording of data during a fair testing investigation as an area where teachers might benefit from suggestions for helping students actively develop their investigation skills. From the literature we identified the challenge of holding an overall plan for the investigation in the working memory. From our observations we identified the laborious way students actually measured and recorded data, in effect dissociating what should have been a series of tests into separate episodes. From the literature we recorded the clear challenge to help students develop knowledge of patterns of data generated by scientific tests, so that they could learn to manage measurement error when they are older. From our observations we saw that students typically ignore measurement error, and when they do act on it they appear to draw on the ‘single perfect measurement’ misconception. These findings gave direction and purpose to the deceptively simple activities that we shaped for primary teachers to try out.

As the teachers tracked developing data patterns when a test was repeated, they began to see the influence of more subtle variables such as friction, and to be much more explicit about how they were managing these (or not). We have noted that students ‘do’ more than they can ‘say’ when carrying out trials. By raising the teachers’ awareness of the dynamics of variable management during tests, it is possible that at least some will return to their classrooms and attend more carefully to the progress of the tests themselves. Such teachers will be better placed to encourage students to also talk about the actions they are taking to control variables during the test(s).

In their earlier paper Appleton and Doig (1999) suggested that ‘activities that work’ appear to be a *substitute* for teacher PCK in science. We suggest that such activities can be effective in *building* PCK when they are introduced in the context of clearly specified teaching and learning challenges.

ACKNOWLEDGEMENTS

This work was financially supported by the Ministry of Education, via the Unit for Studies in Educational Evaluation at The University of Canterbury (USEE). The videotapes of the children’s investigations were provided by the National Education Monitoring Programme (NEMP) unit at the University of Otago, New Zealand.

REFERENCES

Appleton, K. & Doig, E. (1999) Primary school teachers' perceptions of science activities that work. *Australasian Science Education Research Association*. Rotorua, July 8-11

Barnett, J. & Hodson, D. (2001). Pedagogical context knowledge: Toward a fuller understanding of what good science teachers know. *Science Education, 85*, 426-453.

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

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.

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

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

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.

Shulman, L. (1986). Those who understand: Knowldge and growth in teaching. *Educational Researcher, 15*, 4-14.

Sturrock, F. & May, S. (2002). *PISA 2000: The New Zealand Context.* Results from the Programme for International Student Assessment. Wellington: Comparative Education Research Unit, Ministry of Education.

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.

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

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