Research in Science Education (2005) 35: 63–98
DOI: 10.1007/s11165-004-3434-1

A Longitudinal Study of Children's Developing Knowledge and Reasoning in Science

Russell Tytler and Suzanne Peterson
Deakin University, Melbourne

Abstract

The growth in science understanding and reasoning of 12 children is being traced through their primary school years. The paper reports findings concerning children's growing understandings of evaporation, and their changing responses to exploration activities, that show the complexity and coherence of learning pathways. Children's responses to identical explorations of flight, separated by two years, are used to explore the interactions between conceptual knowledge and scientific reasoning, and the manner in which they change over this time. The paper discusses the particular insights afforded by a longitudinal study design, and some attendant methodological issues.

Key Words: attitudes, science learning

The last few decades have seen an enormous amount of research into students' learning in science. This has involved a theoretical shift away from a Piagetian presumption that children's learning is fundamentally determined by the developmental growth of logical structures, to a greater focus on the conceptual content of children's ideas; their ways of making sense of phenomena. There are nevertheless commonalities between Piagetian and individual constructivist views of learning. As in Piaget's own studies, many researchers into conceptual change use comparisons across age cohorts to describe aspects of children's development (e.g., Driver, Leach, Scott, & Wood-Robinson, 1994). The focus on conceptual dimensions of thinking has been maintained, and possible complexities of individuals' thinking are often lost within a general growth model. The presumption concerning stability in students' thinking has been retained, this time in terms of quasi-theory-like 'conceptual frameworks' which were held to be stable ways of looking at the world, rather than the more general notion of stages in the development of logical structures.

However, for some time now questions have been raised about the characterisation, underlying much of the earlier conceptual change research, that learning and development can be represented by an orderly passage through a sequence of increasingly sophisticated conceptions. The coherence of individual students' conceptual frameworks has been questioned (e.g., Solomon, 1994), and indeed the inconsistency with which individuals use conceptions demonstrated in many studies (e.g., Engel Clough & Driver, 1986; Tytler, 1998a). The presumption that individual learning involves an orderly, sequential growth in conceptual frameworks has not been demonstrated, since research exploring the stability or growth of individuals' ideas across time has been rare, and limited to time scales within a year (Engel Clough, Driver, & Wood-Robinson, 1987; Tytler, 1998a, 1998b). Arzi (1988) and White (2001) have called for some time for longitudinal studies tracing the emergence and development of conceptions. Black and Lucas (1993) argued that a longitudinal research program based on classroom learning "will be necessary before we have a thorough understanding of the way(s) in which pupils progressively build their understandings" (p. 235). Despite these arguments, such studies have been rare in science education (Arzi, 1988) because they are time-consuming and difficult. Studies by Novak and Hellden, described in this volume, which trace details of the nature of individuals' growth in understanding over longer periods, are rare exceptions.

There have been a number of writers calling for the inclusion of broader perspectives than the purely conceptual, on student science conceptions. Still within the cognitive domain, there have been calls for instance to include epistemological dimensions of learning, and students' reasoning strategies and their views on knowledge production and the nature of science, in accounts of conceptual change (Duschl & Gitomer, 1991; Driver et al., 1994; Driver, Leach, Millar, & Scott, 1996). Carey (1985a, 1985b) argued that children's conceptual change and their growth in scientific reasoning are fundamentally driven by growth in domain specific knowledge. Koslowski (1996) also argues the importance of knowledge in constructing a reasoned argument in science.

Strike and Posner (1992) move further beyond the strictly conceptual in arguing for the importance of students' conceptual ecology in framing the change process. The term 'conceptual ecology' encompasses students' broader beliefs and dispositions. Other writers have also emphasised the importance of affective factors in the process. Pintrich, Marx, and Boyle (1993) describe a range of classroom contextual, motivational and cognitive factors that impact on conceptual change. Thus, there is a growing body of opinion that, in order to understand the nature of science learning in classrooms, and provide advice for the effective teaching of science, we must draw on a number of dimensions beyond the strictly conceptual.

Context for the Study

In our own previous research utilising cross-sectional research designs, we have shown the complexity of individual student understandings and learning (Tytler, 1998a, 1998b) and have drawn on such constructs as associative thinking (Tytler, 2000) in attempting to capture the context dependence of conceptions. We have also charted various aspects of student epistemologies and their relationship to conceptual learning, in a number of content areas (Tytler, 1993, 1994, 1998b, 2000). Recognition of the richness and complexity of student thinking, and of the intriguing nature and possible importance of these extra-conceptual elements, led us to the longitudinal study as a potential means for tracing more deeply students' growth of understandings over time. We felt the longitudinal design could more effectively uncover the interaction between different elements of thinking and feeling that impact on understanding. We have become particularly interested in children's epistemological reasoning, its growth, and its relationship with conceptual understanding.

As argued above, there is a need to explore further the links between children's growing scientific reasoning and conceptual knowledge. A longitudinal study that could trace both together, seemed to us likely to provide insights into the relationship. This paper will explore, using results from a longitudinal study of primary school children, the following research issues:

1. The patterns of growth of children's understandings of evaporation, as illustrative of learning pathways in a significant conceptual domain;
2. The nature and growth of children's reasoning in science, and its relationship to growing conceptual knowledge;
3. The particular insights into learning and knowledge opened up by the longitudinal study design;
4. Methodological issues associated with the longitudinal design.

The paper will draw on previously published results, and recent analyses of interview transcripts from the first three years of the study, to explore the nature of children's growth in understanding of evaporation as an example of a conceptual area. Different aspects of the data and analysis will be used to explore the particular insights and issues that are possible with the longitudinal study design. The paper will also describe previously published findings concerning children's scientific reasoning, using these to contrast the types of insights possible, compared with cross-sectional designs. It will then look in detail at new data concerning children's growing scientific reasoning and its relationship to knowledge, to illustrate at a more fine-grained level the nature of the data and the analysis, and the potential insights into learning.

Research Methods

The project this paper is drawing on has been following 12 (initially 14) children through their primary school years. The children are all from a state-run elementary school in a predominantly high socio economic neighbourhood of Melbourne, Australia. While the children in the study come from a variety of socio-economic backgrounds, they are mostly Caucasian, and the majority have parents in professional occupations. We worked with 14 children in one class in the preparatory year (age 5), and they are at the time of writing entering their sixth year of schooling (grade 5, age 10). The children were selected in consultation with their first teacher to represent the full range of abilities in the class. Each year we worked with the children and their teachers in at least two classroom units of work on different topics, gathering data from written responses, tapes of class discussions, interviews of the teacher and interviews of the children. The activities were in each case designed to expose children to a variety of phenomena and to support interpretation, so that they came to subsequent interviews having been exposed to a range of experiences and ideas. During the classroom sequences neither we nor the teacher imposed a definitive science explanation for phenomena, but rather worked through a process of challenging ideas and encouraging deeper level interpretations. The extent to which the explorations and interviews could be viewed as a planned intervention is discussed later in the paper.

During the first year of the project we worked closely with the 14 children who were in one class, from the start of their school lives. We visited the class many times, planned jointly with the teacher, and interviewed each child at least four times on a range of topics as we searched for ways of understanding the children's growing capabilities. We participated in the class for each of four evaporation lessons, and taped class and group discussions. In subsequent years the 14 children were spread over two and then four classes which made this close monitoring practically impossible, and we adopted a procedure of taking a preliminary lesson for a classroom sequence in each class, then handing over to the classroom teacher to complete the sequence using agreed activities and worksheets. We had also realised that the interview transcripts were far more useful for tapping into students' understanding than group discussions or field notes, since in an interview one can probe more thoroughly, so that the data collection and researcher involvement became more streamlined for practical reasons.

In the interviews we have explored, among other things, students' learning from the sequence and their views of previous or related activities. The interviews are conducted in the style of a Piagetian clinical interview, rather like a conversation but with particular agendas followed for each task. The children are encouraged to explore, and to express ideas in their own language, but there were sections of the interview in which they are asked to respond to challenges presented by the interviewer. The interviews are thus partly free, and partly structured. We have developed a comfortable relationship with the individual children that allows us to discuss a variety of their experiences, both in and out of school. Each interview was transcribed and the NUD*IST (Non-numerical Unstructured Data Indexing Searching and Theorising) software analysis program used to organise the data and to generate interpretive dimensions and levels within these to describe children's responses to the tasks. These were validated by a code-recode process (Tytler & Peterson, 2004a, 2004b).

The sequence of classroom units, and some details of the interviews, are shown in Table 1, illustrating the way we have cycled through three core science content areas (changes to materials, force and movement, and animal and plant behaviour and adaptation). We have incorporated investigative design and reasoning into the classroom sequences and interviews, as well as cycling through a range of conceptual content probes based on recent and earlier units.

Children's Understanding of Evaporation

A major part of the first year of the study was the study of children's responses to a sequence of classroom evaporation activities. In the first year particularly, since

Table 1
Sequence of Classroom Units and Interviews.

Year level	Semester	Classroom topic	Interview
Grade prep (1998)	1	Sequence of five evaporation activities.	1. Range of probes in science content areas. 2. Interpretation of sequence, written probes. Mealworm and floating and sinking reasoning tasks.
	2	A range of activities; water boiling, dissolving, construction technology. Air and flight sequence.	1. Classification of minibeasts, review evaporation, air probe. 2. Air and flight review. Whirlybird investigative task.
Grade 1 (1999)	1	Plant unit. Plant experimentation.	Plant questions, air and flight probes; parachute investigation.
	2	Melting and dissolving activity sequence.	Melting & dissolving probes, plants, terrarium probe, animal behaviour and snail investigation.
Grade 2 (2000)	1	Force and motion, floating and sinking and boat making.	Review of force, air and flight, evaporation.
	2	Animal behaviour; activity sequence with crickets.	Animal behaviour review. Air and flight; whirlybird investigation.
Grade 3 (2001)	1	Dissolving sequence, chemical change.	Evaporation, animal behaviour probes. Learning style probe.
	2	Force and motion: rolling sequence.	Review of rolling, and investigative task. Classification & adaptation probes, air pressure.

we spent a lot of time with the children as they became socialised into the processes of schooling, our focus shifted from a concern with their conceptual thinking, to a broader characterisation of the child and their response to the classroom and learning. We could not ignore what was obvious; that there was a consistency and coherence over time in children's responses in interviews, that was not captured by tracing their conceptual understandings. In looking at children's understandings of that evaporation sequence, three major insights were generated (Tytler & Peterson, 2000):

1. Children's understandings were contextually framed and varied from activity to activity in ways not predicted by then accepted conceptual trajectory models;
2. Their response to the sequence had a significant epistemological component in that the children seemed not to appreciate that the purpose of the sequence was to be understood in terms of an underlying conceptual entity (evaporation), but treated each activity as defined by other non-conceptual, often surface purposes; and
3. Children's responses to the sequence could be best understood in terms of narratives related to their construction of identity with respect to the classroom and learning, rather than in simple conceptual terms. Their drawings interpreting an evaporating puddle activity, for instance, indicated they saw the event as essentially social or fantastical in nature (see Figure 1). These social and fantastical elements often persisted into subsequent years more strongly than the conceptual interpretations, more so for some children than others.

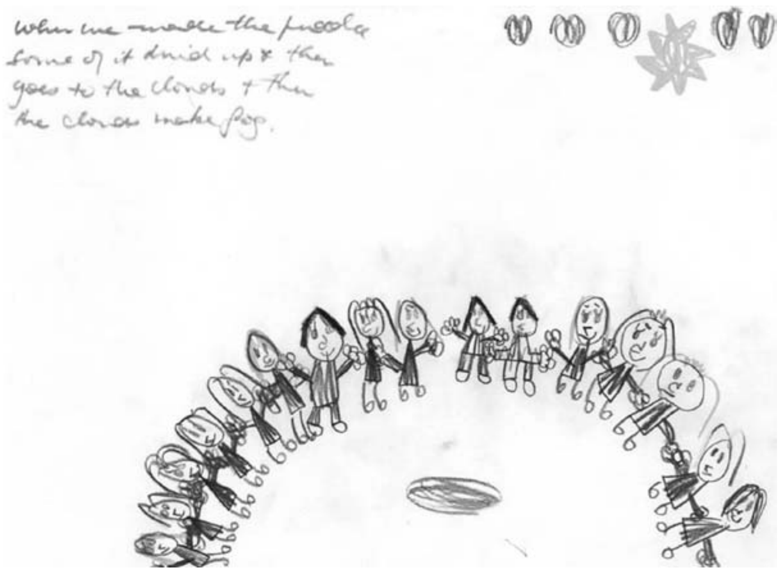

Figure 1: Emily's drawing of the puddle activity, Grade Prep. The explanation was transcribed by an adult: "When we made the puddle some of it dried up and then goes to the clouds and then the clouds make fog."

During the first year a number of emergent features of the longitudinal design were becoming apparent. The nature of the relationship with the children was forcing a different perspective on their learning than we started with, and which characterises mainstream student conceptions literature. We were also beginning to struggle with the possibility of linking these conceptual, personal and epistemological elements together to explore their interactions. We were also being confronted with the increasing difficulty of writing about complexity in a way that is convincing, within the space normally considered decent in the research literature. Data management was also presenting as an issue.

In subsequent years we continued to track these students' ideas about evaporation. During interviews we would present an activity involving evaporation (a wet handprint, boiling water, an open bottle of eucalyptus oil) to the child and ask them to talk about it. In some cases we presented them with their drawings from earlier years to comment on and discuss how their ideas might have changed. An issue we had to consider was the possibility that children would be prompted by memory of previous discussion of phenomena with us, to repeat, or perhaps distort in some way, their earlier responses. We thus recycled similar situations at intervals of at least two years, and introduced variation where possible to guard against mere repetition by memory. In fact, it was clear that some children (and this is very different for each child) could remember their responses from two years previous.

A review of children's responses to evaporation probes over four years has extended our previous findings concerning children's individual conceptual pathways and the nature of their understandings (Tytler & Peterson, 2003b, 2004b) as follows:

1. While children's understandings had progressed, their explanations were very context dependent and there was no simple pathway they followed through a series of increasingly sophisticated conceptualisations (mirroring Hellden's findings on continuity and context, reported in this volume);
2. There was a logic to the way each child approached the construction of interpretations, that could not be described simply in terms of the attainment of a conceptual position. Children have relatively coherent ways of approaching learning and knowledge that are linked to the way they construct their growing subjectivities as learners and school children; and
3. Even where children are ostensibly voicing the same conception (e.g., that water has evaporated up into the air, or clouds) they differ markedly in the way they represent and extend that idea and invest it with meaning. We thus question the view of conceptions as transcendent entities that can be unproblematically identified.

One of the issues associated with a longitudinal design focusing on individual children is the need to find ways of representing complexity. Part of the evaporation analysis involved tracing in detail the learning pathways of two children. While this analysis was supported by data on the changing understandings of the full cohort, it was necessary to narrow the scope in order to illustrate in some detail the complex nature of the learning pathways. We interpreted these through the constructs of "narratives" (and we see this as related to White's (1988) notion of "episodic memory"), and "narratives of the self" which refer to the way these children frame their response to phenomena and to learning as part of their developing subjectivities. The longitudinal design is thus allowing a deeper level of focus on individual learning, but inevitably involves questions of scope in analysis and reporting.

Children's Scientific Reasoning

Over the five years of the study thus far, interview tasks have focused on a variety of aspects of children's approach to investigations, including fair testing, open experiments in class, and interpretation of structured classroom investigations. We used transcript data from the first year of the study to develop a theoretical model for interpreting children's reasoning (Tytler & Peterson, 2004a, 2003a). This is described below. An analysis of one interview task, repeated after 2 years, involving investigations of the flight of whirlybirds, is used in this paper to more sharply investigate the growth in, and interactions between, children's knowledge and reasoning, and to provide a basis for discussing the particular insights into learning and development afforded by the longitudinal study design.

The Interview Tasks

The interview tasks used initially to develop a framework describing children's scientific reasoning involved three different science phenomena – the behaviour of mealworms, what determines whether objects float or sink, and the flight of paper whirlybirds (all included in the grade Prep interviews). The tasks were all concerned with the ability of children to investigate and to deal with interactions between knowledge claims and evidence. They differed substantially, however, in the demands they made in terms of the type of question that was asked, the degree of constraint put on the method of investigation and the associated variables, and the target outcome. Despite this variation, it was found that:

1. There was a coherence in the way each child responded to these very different tasks; and
2. There was considerable variation between children, in their level of engagement with each task, and with ideas.

A number of dimensions were developed to describe children's responses to these tasks, including their approach to exploration, the depth to which they processed ideas, and their handling of variables. The first of these, *Approach to Exploration*, has proved particularly fruitful over time to capture these children's reasoning.

In subsequent years a number of interview tasks were designed to trace growth in these aspects of children's scientific reasoning. A recent analysis (Tytler & Peterson, 2003a) has shown the very different starting points, the coherence, and indications of growth in these dimensions over the first two years of schooling. In the current analysis, we will use data from interview tasks concerning flight, including identical tasks separated by two years, to focus more closely on the nature of changes in children's responses. The tasks are:

1. Whirlybird exploration (November, Grade Prep; October/December Grade 2) – Children were shown six paper "whirlybirds" which spin as they fall. Three had paper clips which make them fall and spin faster, and there were three different wingspans; short, medium and long. They were asked to find which whirlybird fell the slowest, which could be done in an ad hoc or systematic way. They were challenged to give their reasoning and where appropriate to provide evidence to support their assertions.
2. Parachute exploration (June, Grade 1) – Children were shown four parachutes, two with large and two with small plastic bags as canopies, one each of which had a small or a large plastic model of a parachutist. The questioning followed the same pattern as for the whirlybirds, challenging children to separate out the effect of canopy size, and weight. Children's responses to this task will be described briefly only.

Approach to Exploration

On the basis of the first year data we identified four distinct levels that characterised children's different approaches to the exploratory aspects of the tasks; in the way they sought and acknowledged evidence as they generated explanations, and in the nature of their speculations and explanations. These are described in Table 2. These categories apply to the particular sections of the explorations occurring during the interview, and not to the child per se. The essential difference between Levels 2 and 3 lies in the order in which the inference and action occur. For Level 2, the idea arises out of observations. For Levels 3 and 4, the observations are explicitly led by the ideas. This Approach to Exploration dimension is closely related to Chin and Brown's (2000) "Approach to Tasks" category.

Depth of Processing

The Depth of Processing dimension, which was assessed for the first year data, had three levels:

Level 1, Description of phenomena. Children focus on phenomena only. They tend to list or describe without comparison, and with minimal conceptual content. For instance, children focus on the speed of individual whirlybirds and do not seek patterns in whirlybird flight.

Level 2, Pattern identification. Children identify patterns in the data or interpretations that go marginally beyond the data but still lie fairly close to it, for instance,

Table 2
Coding Levels for the Approach to Exploration Dimension.

Level	Whirlybird example
1: Ad hoc exploration	
No systematic observations or comparisons are made, nor use of a guiding explanatory purpose. Exploration at this level is data driven, and restricted to low level interpretation that lies close to observable entities.	Children explore one whirlybird or parachute at a time without explicitly comparing characteristics.
2: Inference searching	
The inference could be about relations between variables, or about theoretical ideas. Children explore on a 'try this and see' basis, without a noticeable sequence, but leading to some hypotheses or inferences. They notice things and comment, with some generation of a conceptual inference. Exploration at this level is data led but with some conceptual interpretation.	Children compare whirlybirds in pairs based on some factor of interest without a plan based on an idea.
3: Hypothesis checking	
The hypothesis could be about relations between variables, or about theoretical ideas. Children carry out focused observations or interventions which involve trying out an idea, or following up a prediction with some conceptual basis. Explorations have a recognisable hypothesis driving them. Exploration at this level is theory led, but does not necessarily separate variables.	Children take the lead in developing a strategy to check ideas (e.g., longer wings slow it down) about what affects whirlybird flight.
4: Hypothesis exploring	
Strategic search for evidence to refine or distinguish between hypotheses or rule out other possibilities. Effectively holds two alternative representations. Setting up checks of ideas generated, and dealing explicitly with the possibility of confounding variables or other limitations on experimental design. Exploration at this level acknowledges data and theory interdependence.	For the whirlybird flight, children spontaneously try out competing ideas, for instance by altering conditions in an ordered way, controlling for variables of weight and wingspan. No child operated at this level until Grade 2.

identifying features of whirlybirds that correlate with the speed of fall ("longer wings fall faster").

Level 3, Explanations. Children generate conceptions with explanatory power. For instance, they explain that whirlybird flight is determined by the interaction of shape or weight with external influences ("longer wings catch the air").

It was found, in the analysis of responses to different tasks the children explored in the first year, that there was an almost complete correspondence between the first three levels in the Approach to Exploration, and the Depth of Processing dimensions (Tytler & Peterson, 2004a). This correspondence is not a logical necessity, since it is conceivable that children could predict a particular pattern of results, without explanatory import, and treat this as a hypothesis (this happened once only). However, it makes sense in terms of the epistemological reasoning underlying these dimensions in that the different levels in each dimension correspond in each case to different ways of looking at making sense of or exploring phenomena. We developed a framework of epistemological reasoning based on the work of Driver et al. (1996), which brings these different dimensions into perspective (see Tytler & Peterson, 2003a) by describing three distinct categories of reasoning represented by these levels. Level 1 represents "phenomenon-based reasoning"; ideas are inherent in phenomena and sense making involves the close observation of individual events. In "relation-based reasoning" (Level 2) ideas are interpretations of patterns in the phenomena. It is only with "concept-based reasoning" (Level 3) that ideas become understood as constructions that are somewhat removed from phenomena directly, that need to be tested against evidence. We will not code for Depth of Processing directly in this paper, but will discuss this issue of the status of knowledge and ideas, as part of the analysis.

Handling Variables

As part of a recent analysis (Tytler & Peterson, 2003a) we developed a dimension focusing on children's ability to recognise and control confounding variables. In the case of the whirlybird and parachute tasks, the control of the variables of canopy/wing size, and weight, was necessary to successfully approach the task. Variable control is considered as the major indicator of science processes in many contemporary accounts, particularly with the ascendancy of the "fair testing" metaphor in the UK and other curricula. This development has not been without its critics (e.g., Tytler & Swatton, 1992; Watson, Goldsworthy, & Wood-Robinson, 1999; Tytler, Duggan, & Gott, 2001). Nevertheless, the ability to recognise and deal with confounding variables is an important aspect of children's ability to reason scientifically, and is a feature of their response to a number of the challenge tasks in this study. Therefore, their responses were examined and a categorisation scheme developed that was capable of describing the range of responses in the various tasks. The coding scheme has been validated through a process of discussion between the two authors. For these three flight investigations, the responses have been coded as:

1. *Level 1.* When faced with confounding variables is unable to separate them, for instance, comparing parachute flight with both canopy and weight varying.
2. *Level 2.* Can recognise, or suggest a means of designing an experiment that avoids confounding two variables, for example, compares parachutes for the effect of canopy size, while keeping weight constant.
3. *Level 3.* Controls confounding variables confidently, and is able to explicitly identify the logic of being able to do so by pointing out the control elements.

In this dimension, as for Approach to Exploration, there were occasions when it was difficult to determine which level was appropriate because the child seemed to move between levels, or the establishment of a higher level was tentative and uncertain, or was achieved only after scaffolding in the interview. In these cases half-levels were assigned. In some cases no level was assigned because of ambiguity in questioning or response, or because the interviewer was judged to be leading the interviewee such that independent action was not possible.

Analysing the Interview Transcripts

In this section, transcripts of five children's interviews are presented for the two whirlybird tasks, as well as the interceding parachute task. Comparing their approach to the same or similar explorations over two years, and their interpretations of the flight based on notions of air, weight and balance, demonstrates the power of the longitudinal study design to pick up key features of individuals' change and continuity. Five children have been chosen to represent a range of responses to the tasks, in terms of the level and type of approach. This seemed to us the minimum number needed in this case to illustrate the diversity of responses. RT and SP are the two interviewers.

Esther

Esther is a very likable and social child. She performs consistently at a low level of reasoning during her first year, but by the end of her second year has grown in confidence and knowledge. At the beginning she tends not to focus on conceptual matters at all, but is readily distracted to other agendas.

Whirlybirds, Grade Prep

Esther seemed to have no strategy, testing randomly unless prompted, and focussing on one whirlybird at a time. She seems unable to focus on wing size or weight as variables with different values in the different whirlybirds. She predicts which will be slowest:

Esther: The short ones.

RT: Okay. And what about the short ones . . . which one would go the slowest?

Esther: That one . . . I like that one the best . . .

RT: That one went the slowest. Do you think any would go slower than that?

Esther: That one and that one.

RT: This has got a paper clip on it and the same wings. Let’s have a look. You were wrong again . . . what do you think now?

Esther: That one.

(Approach to exploration: Level 1; Handling variables: Level 1)

Parachute, Grade 1

Seven months later Esther has a plan to compare and speculate, but it is not a coherent plan. She now is able to focus on the separate effects of weight or canopy size, but she does not appreciate the need to control for the confounding variable. She has moved some distance to being able to interpret her experiment but still focuses on peripheral, non conceptual issues.

SP: Now that one was slow. Why was that?

Esther: I think because it has more canopy . . . and that’s a bit . . . unless . . . I’ll measure the men. They’re both the same size.

SP: So, what do you think makes a parachute come down slowly? . . . Was it the canopy, or something to do with the weight of the men?

Esther: The weight of the men.

SP: And, did you test that? You tried two men together, did you? And they were the same men?

Esther: Yeah? That one was like a builder, and this one was a fixer.

(Approach to exploration: Level 1/2; Handling variables: Level 1)

Whirlybirds, Grade 2

Esther now focuses on the flaps and introduces air as a factor in flight. This seems to provide a basis for suggesting a coherent exploration that did not seem possible 2 years earlier.

RT: So, why do they spin around?

Esther: Ummm . . . coz of the flaps maybe . . . and the air’s making it spin around.

RT: How is it doing that?

Esther: Ummm . . . How it's . . . umm . . . I think maybe how . . . it's making it drop . . . drop . . . and and the . . . it's looks like a helicopter and it's like um its making more . . . sort of . . . when its going down.

Esther is delighted by the comparison with a helicopter, and tries some whirlybirds initially randomly. Then she institutes a controlled comparison.

Esther: I don't know which one would go faster, maybe the smallest . . . we'll have a race against.

RT: OK, well, that's what I want you to find . . . show me which one goes slowest . . . and what's causing that. . .? So you've got three there . . . all without paper clips.

Esther: I already know . . . know how it's um . . . how it's maybe . . . how um . . . how . . . coz it's [all?] moving the blades . . . and these . . . these like wings . . . and it's make . . . whichever one's the shortest goes the fastest and whichever one is the longest goes slowest.

RT: And how are you going to show me that?

Esther: Ummm . . . by dropping them all at . . . the same, nearly the same time . . . .

On prompting about the role of paper clips she predicts, but then constructs a test which confounds the variables.

RT: If I were to ask you . . . you think the paper clip makes them go faster . . . could you show me whether it does or not? (Esther chooses a comparison.)

Esther: Umm, the long . . . the long one went faster . . . with the paper clip

RT: The long one with the paper clip went faster than the short one without the paper clip . . . Oh, OK. If I were to ask you . . . to prove to me that the paper clip makes it go faster . . . could you do an experiment to show me that?

Esther: The . . . umm . . . like I just . . . you know you can . . . I could . . . .

RT: OK, you're choosing both long wings . . . so you're choosing the same wing size? Why is that?

Esther: Ummm . . . Um . . . maybe because . . . mmmm . . . maybe because I could . . . maybe because it could be making it . . . go a bit . . . coz, then it would be a bit fair.

So Esther, with considerable scaffolding, is able to design a valid experiment and articulate her reasons. This is a considerable advance over her previous occasion in terms of her approach to comparing, and also to generating some tentative explanations. There are two gains underpinning her advance; (1) a conception of a controlled comparison (she was doing this in the June Grade 1 parachute exploration) and (2) a more confident grasp of the conceptual factors underpinning the whirlybird flight.

(Approach to exploration: Level 2/3; Handling variables: 1/2)

Calum

Calum is a thoughtful student who sees himself as interested in science, and has a dedication to scientific methods partly through his home experience with a chemistry set and other experiments done in the home. He is very speculative in manner, and thoughtful in his responses. In his first year, when asked about what constituted an experiment, he linked it with the notions of active seeking of answers, and some sense of controlled comparison.

Whirlybirds, Grade Prep

Calum was quite experimental with the whirlybirds but he didn't ever do any sort of variable control. He was looking for confirmation so he tried the three with paper clips in order to see the effect of the paper clip rather than comparing with a non paper clip, and he wasn't controlling for wingspan.

Russell demonstrates a whirlybird. Calum cannot comment on how it works.

RT: (Introducing the variation in factors) Which one will go the slowest out of all of these?

Calum: I don't know, probably the paper clip ones.

RT: Do you want to test. What would you do?

Calum immediately organises a complex comparison test. He doesn't control for wing size or paperclips by pairs, but exhibits a sense of control of conditions by organising the two experimenters to swap whirlybirds. He also organises the paper clip whirlybirds to be dropped by the same person. From the start, Calum has an active and inquiring approach to these tasks.

RT: So you've got a middle sized wing with a paper clip and a long one without a paper clip, I've got the middle sized one without the paper clip and the short wing with a paper clip . . . (They drop them) . . . Which one went the slowest.

Calum: That one.

RT: Why do you think that is?

Calum: We could swap around. . . . You can have the paper clip ones.

They try again, and discuss what makes them go slow. Calum has a complex theory.

Calum: I don't know. Probably because you know how like ... you know how paper is made out of wood, paper is not just one layer and if you know how it weighs um, then paper would a certain amount and some of these are not the same. You know how I said pieces of paper are the same sort of ... they're not the same number of layers ....

He then does some more experimenting, unprompted. He predicts, wrongly, the paperclips will make the whirlybird go slower. In testing this he confounds with wing length.

RT: You think the paper clip would make it go slower. How could we test, what would we do? So you're choosing a long winged non paper clip and a middle winged paper clip? Ready. You drop it ... Ah, the paper clip one went faster ... What do you think now?

Calum: I think that one goes the slowest.

RT: The one with the paper clip.

Calum seems confused. Although he is very exploratory, his lack of an explanatory framework seems to be impeding his ability to design a clear test. He is comparing, but in a confused way.

Calum: I haven't seen these ones ... I hold this and you hold that.

RT: So we're each holding a paper clipped one. Ready? What is this going to show us?

Calum: I want to see which one goes the slowest and which one is the fastest.

(Approach to exploration: Level 2/3; Handling variables: 1)

Parachute, Grade 1

Seven months later Calum has much the same problem with comparing in a controlled way. However, he is able to discuss separately the effect of the canopy and the size of the "man", and has a clearer idea about the effect of air: "it takes more air to fill all of the canopy, so like coming down 'cause it's ... there's more space." He is challenged to predict which parachute will be slowest, and with this support does in fact control for canopy size.

Calum: Um, this one (big canopy, big "duplo" man) might, but it's going to come down sl ... um, faster than the ... (small "lego" man)

(Approach to exploration: Level 2/3; Handling variables: 1/2)

Whirlybirds, Grade 2

Two years after his first encounter, RT introduces the whirlybirds and Calum remembers them. He comes up with an immediate, and scientifically advanced explanation which separates the effects of upthrust on the wings, and the downward weight of the paperclip.

RT: What do you think causes that spin?

Calum: Um, probably ... um the paper clip to pull it down ... and, um ... because of the air pushing up ... instead of them going up it would have to push, the air would have to push it away so it goes round.

This time Calum, when invited to experiment, immediately controls for the factors he has just identified as relevant. It seems that his conceptual casting of the whirlybird flight has provided a ready basis for distinguishing the separate effects of the two variables.

RT: You're going to try those two ... the two long-winged ones, but one without a paper clip and one with. ... (Calum tests) ... Okay – so what do you think?

Calum: Probably it's because the paper clip ... pulls it down. And you can see the one without the paper clip ... spinning ....

RT: And why did you choose the long-winged ones to try?

Calum: Probably they might be ... these ones (he cannot clearly articulate).

RT: Okay, so if you were to tell me that the longer the wings are the slower it went, could you choose two to demonstrate that to me? Okay (commenting as Calum experiments) so we've got two with paper clips but one with long wings and ... let's try this. Okay ... so it's the wing length that makes it go slow ... and the paper clip makes it go faster.

(Approach to exploration: Level 3; Handling variables: 2/3)

Anna

Anna is a verbally competent girl who has many ideas and talks comfortably in interviews. She tends to focus on language as representative of science understandings. This seemed to restrict her ability to establish a flexible understanding of evaporation. In interviews and in her drawings and classroom responses she presents as fundamentally concerned with social issues, particularly her friendship group.

Whirlybirds, Grade Prep

Anna is unable to systematically test the whirlybirds. She is possibly thrown by a commitment to an incorrect idea of the paperclips making no difference. She starts by noting, with some surprise, that the whirlybirds spin. On prompting, she chooses a long winged model as likely to be the slowest.

SP: Why is that?

Anna: Because it's a bit big, and it might not spin around as fast, so, um, it might be a bit slower.

SP: Would you like to try?

Anna begins to drop only one, but SP insists on a comparison. Anna then chooses to compare a large and small whirlybird, one with and one without a clip. She does not control for weight. She predicts the bigger one will be slower, which is born out. SP then challenges her to predict the speed of the smallest winged whirlybird, but she chooses one without the paperclip. She predicts the result but is proved wrong. SP persists in encouraging her to test her ideas. SP then focuses Anna's attention on the paperclip.

SP: So you thought the one with these long one's would be – what about with the paperclip on it? If this one – which one did you think would be slower, the one with the paperclip, or the one without?

Anna: I think it would be still the same.

So Anna's lack of variable control can be related to the fact she cannot see the paperclip as a relevant variable. When they do the trial, and the paperclip whirlybird falls faster, she seems unable to comment on the result. She is confused because her prediction is not born out, and has no explanation.

(Approach to exploration: Level 1; Handling variables: 1)

Parachutes, Grade 1

Anna is more definite in her approach but confounds the two variables of canopy size and weight when constructing comparisons. She seems to have limited conceptual basis for considering canopy size, and weight, as independent factors, so when asked to investigate the effect of canopy size she chooses the same canopy, different weight person.

SP: They've got the same size canopy, those.

Anna: No, but this one is a bit lighter than this one – this one has nothing inside it, and this one does.

Anna organises a comparison, with guidance. She does not focus on isolating canopy and weight. She then organises a comparison of all four parachutes, dropped simultaneously. She doesn't adopt an experimental approach based on variables but seems able to comment meaningfully on the outcomes.

(Approach to exploration: Level 2; Handling variables: 1/2)

Whirlybirds, Grade 2

Anna remembers the physical, personal aspects of the activity, and how they fly. She has clearly done this in her class recently and refers to colouring the whirlybirds and dropping them upside down to make them turn over. SP introduces the task:

SP: (Introducing the different features of the whirlybirds) Now I would like to know which one would drop the slowest. Would take the longest to get to the ground.

Anna chooses the one with big wings. As for the previous occasions, her first focus is on size. However, when "paperclip" is mentioned this seems to trigger a thought and she changes her prediction. She compares two long winged whirlybirds, with and without a paperclip. This experimental "control" seems to come partly from her focus on wings as the dominant effect rather than a design decision. She is now able to articulate the effect of the paper clip.

SP: This one seems to float away a bit, why's that?

Anna: Because it didn't have . . . enough control.

SP: It didn't have enough control with the paper – how does the paper clip control it?

Anna: Because it's heavier and it's pulling it down . . . because it's got more gravity.

SP: Now you've picked the long-winged ones, why did you just immediately pick the long-winged ones?

Anna: Um, because if we did those ones (the short-winged pair) . . . these would (go) faster because they have short wings and they drop just straight through the air.

Anna doesn't seem interested in testing her idea. She seems confident in her explanation. When she is asked how wings affect the flight she is now able to respond:

Anna: Because there's no . . . air pushing them along . . . pushing them here, and then they push . . . and they go slower because there's more air pushing up (on the wings) . . . .

SP: If the air is pushing up, I'm not sure why that would make them ... (turn).

Anna: I think I know, cos they're here and it's going like that because it doesn't have control like an aeroplane and aeroplanes can go straight forward but if they didn't have a control they'd be spinning round in the air or they'd be crashing.

Compared to two years previous, Anna now has a lot more domain specific knowledge to play with, and she has a theory to cover the effect of both the wings, and the paperclips. SP now challenges her to justify her choice. She is quite competent at the long-short wing comparison. She compares a non-paper clip pair, and drops them from the same height. She is able to interpret her observations in terms of ideas of weight, control and air uplift. She still does not lead the investigation using her ideas, responding rather than inquiring actively, but her increased knowledge seems to be the factor that enables her, with support, to control for variables.

SP then asks Anna about previous, seemingly recent experience with whirlybirds. Again she emphasises the social, play and aesthetic aspects.

Anna: . . . we were doing things about play things, we used those . . . we made hot air balloons and boomerangs and, um . . . they made . . . things with wings that were sticking out . . . it looked a bit like . . . umm . . . that, only that was more down. . . . umm . . . Adele and me made a hole in the middle so we could pick them up and we picked them up and they went . . . it looked really pretty because, umm, the wings were . . . the wings were spiralling around and if you made sort of a green colour with a lot of . . . with a circle in the middle, lots of circles . . . um, it would look really good because . . . it would all be spinning around . . . .

(Approach to exploration: Level 2; Handling variables: 2)

Jeremy

Jeremy is a lively child with an active approach to learning and life in general. He is quite capable of independent thinking, outside the mould, and takes charge of the exploration in pursuit of confirmation of insights.

Whirlybirds, Grade Prep

Jeremy's design is constrained by his initial prediction that paperclips make whirlybirds go slower. He first predicts a short winged paper clipped whirlybird will be slower, then changes his mind and drops a longer winged paper clipped whirlybird "because it's got those things". He is initially tentative in exploration but suddenly seems to have an inspiration involving comparison of long and short wings. His following explanation is given with such confidence it seems likely this idea helped him frame the controlled comparison.

Jeremy: I have an idea . . . . Why don't we try it with this one?

SP: What small and big?

Jeremy: Yeah, small and big (he predicts the big one will be slower, and tries).

SP: Oh yes, you were right. Can you come and tell me how this works? How does one of these work?

Jeremy: Um, because it's got longer fins and then it's got longer air and it can catch lots of air, and this one's only got little fins and it can't catch much air, and this can catch lots and lots and lots and lots and lots of air.

(Approach to exploration: Level 3; Handling variables: 1/2)

Parachutes, Grade 1

Jeremy seems to be explicitly leading the investigation based on notions of weight, and air; he is now confident about air as a mechanism affecting flight ("Well the air again gets stuck in here and then it stays in there and it makes the thing go like this so it makes it go down slowly"). He controls variables confidently.

(Approach to exploration: Level 3; Handling variables: 2)

Whirlybirds, Grade 2

Eighteen months later Jeremy has a very specific and detailed view about the action of air on the wings, that echoes but expands on his prep year insight and his parachute explanation. He also seems to have an advanced mechanical sense of the whirlybirds' stability. He explains why it spins round, and his insights drive a very active exploratory sequence.

Jeremy: Um, well, I think, when it's going down, 'cause there's always one piece (?) going up there and one sort of (?) going up there, and one . . . (?) going down, and the air's pushing up them that way and they . . . pushing up at the blades and it's causing it to turn around a bit.

When asked to explore which whirlybird falls slowest and what determines the flight:

Jeremy: It'll be the one without the paper clip . . . . Because the paper clip puts on a bit of weight and also um, the . . . they . . . they have a more longer strip and that . . . it might wobble a bit . . . Because they . . . when it's . . . (?), it doesn't wobble so much.

Jeremy now tries two whirlybirds but this time surprisingly confounds the two variables. Perhaps he has already decided he knows what will happen for the comparison with and without paper clips. He independently decides on a repeat try. SP now challenges him to pronounce on the other whirlybirds and this time he controls for wing size.

SP: So you've got the same size wings? One paper clip. Which one's the slowest?

Jeremy: This one was.

SP: . . . Is it the paper clip that always makes the difference?

Jeremy: I don't think so, not always. Yeah, because I mean that if the wings are long, when the . . . (?) go up a bit and it won't stay . . . (?) so much.

Jeremy is now talking about the independent effect of the two variables. He tries some more comparisons, with encouragement, and talks his way through an interpretation of what happens with long wings and a paperclip. He argues the paperclip's weight could steady the flight which is an echo back to his concern with wobble. It seems that Jeremy may have all along been pursuing a program balancing these different effects! There's a definite sense, with Jeremy, of an active program of trying out ideas.

Jeremy: I'll try it .... Yes, slower ... the paper clip puts weight on it to make it go a bit straight .... And also the wings make it um, well if the wings are really long then it's ... like really long, not middle, and they ... they might work a bit more ... because ... it has more air ... it will have more air in it .... The air ... is going up under the wings and then when it's going down ... it makes it spin more.

Jeremy in fact is pursuing quite a sophisticated idea involving a separation of speed of fall, and spin speed, driven by the weight of the paperclip. SP's questioning, pursuing the notion of the independent effect of the two variables, cuts him back to a simple comparison and seems to interrupt the train of thought. Finally, on challenging, he runs a definitive test.

Jeremy: Yeah, I'll get another one (?) Yeah, I think this one will take a bit more time to get down ... because when the wings go in there, it takes ... it's like a parachute coming down and when ... when a parachute opens it stops from falling. Like if you're falling ... you go a lot faster and where ... and when the parachute lets you go down you get slower ... and slowly and so does that (he tests and his idea is confirmed).

(Approach to exploration: Level 4; Handling variables: 2)

This sequence was one of very few coded at the fourth level of approach to exploration; "hypothesis exploring". In the sequence Jeremy seems to be actively generating his ideas and proposing tests to explore these, sometimes cutting across the interviewer's agenda in the process. This level is characterised by a fluid, dynamic interchange between ideas, and evidence searching, which represents scientific methods at a higher level.

Karen

Karen is capable of quick insights. She works very hard in interviews to generate ideas to please the interviewer. She has, as we shall see, a different model of experimentation in mind than the interviewer. Karen seems to be working in this first sequence with an elimination metaphor derived from sport, rather than a variable control metaphor. In the later whirlybird investigation she is still not structuring the investigation around variables explicitly, and indeed misunderstands a question to this effect. She is quite capable, however, of interpreting variable control.

Whirlybirds, Grade Prep

Karen was really very good at comparing the whirlybirds. She compared them two by two in a program, and she tended to select for particular differences. She predicts the longer winged whirlybirds will go slower "because they've got more", and proposes a test.

Karen: Well we could race two at a time .... And then the slowest ones race. We'll race these two first.

She is highly exploratory but confounds variables as she constructs the comparisons. When challenged to check the effect of the paper clips she suggests:

Karen: We'll race two paper clip ones.

However, when she tests, her interpretation is accurate. Perhaps despite the question she is more interested in pursuing the issue of wing length to tie down the argument.

RT: One with long wings and one with middle wings. Alright then .... So what happened?

Karen: The smallest winged one was ... the ones with longer wings went slower .... Mmmm. I thought little ones go slower but I was wrong ....

She finalises the test cycle and when challenged is able to demonstrate the separate effect of the two variables. Karen's is a rare case of an active exploration without recourse to explanatory ideas.

(Approach to exploration: Level 3; Handling variables: 1.5)

Parachutes, Grade 1

Karen is very fluent in considering the competing variables of canopy size and weight of the model. She is well able to construct and interpret tests that control for these variables, and she focuses on the effect of the air. This time she has a strong explanatory stance and actively demonstrates and interprets after controlling for canopy size. However, she does not seem to quite understand the logic of planning a controlled test which isolates one variable, even though she is very sophisticated in predicting and interpreting findings based on separation of variables.

(Approach to exploration: Level 3; Handling variables: 2)

Whirlybirds, Grade 2

Karen remembers, from two years previous, the different comparisons that were done.

Karen: It was spinning round. I thought we did something with the small ones and the big ones . . . and the different type of . . . (paper clip?)

When she is posed the task of finding which is slowest, and what affects the flight, she immediately drops one whirlybird and starts entering a number in a table on a sheet of paper. This confuses RT until he realises she is counting under her breath to measure the time of fall.

RT: How did you know what to write in it? Oh . . . the time . . . oh, OK, OK.

Karen: Medium with clip . . . small with clip . . . ten seconds . . . .

By this time Karen is unstoppable. She had launched straight into a planned method for measuring and recording. She proceeds to fill in the table showing times for each whirlybird, identified appropriately by wing length and clip/no clip, before reviewing results

Karen: So probably the one without clip takes the longest to go down. And, the small one doesn't with the clip. It takes the small . . . um, it takes the smallest to go down and it um, it wasn't . . . (?) . . . The big one with clip took ten seconds, and so the big ones are probably first, and the mediums could be like second, but the ones with the clips. . . and so it's like . . . 'cause like, the ones without folded up are better than the ones with the clips. 'Cause the clip's heavier and it pulls them down.

RT: OK, so if I asked you from that um, what difference . . . you seem to be saying the long wings fall slower. What evidence have you got for that? How could you justify that?

Karen physically lines up the whirlybirds in order of time of fall.

Karen: Well they're sort of like, they . . . see, this one, it goes slower because it doesn't have a clip to pull it down and it spins more faster and so, it's sort of like a parachute which keeps you up . . . . So I think it's how we did the parachute ones . . . . The big ones are like the slowest . . . and then the medium are second slowest . . . . The medium and the small one without the clip is second slowest and then the small one with the clip is . . . the fastest.

RT: If I asked you to show me what difference the clip makes, how would you use those results to show me that difference?

Karen seems unable to come up with a definitive comparison, but argues on the basis of her line up and common sense insights. RT persists and finally she constructs a test explicitly controlling for wing size to demonstrate the effect of the paper clip.

(Approach to exploration: Level 3; Handling variables: 2)

It was a confused conversation. Karen seemed to be able to balance the two variables in her head but didn't have a natural sense of doing a straight comparison to isolate a variable, working instead within a different experimental design involving trend identification. She didn't seem to understand what RT was asking. Yet, in other tasks, Karen showed an impressive ability to separate competing variables early in her preparatory year.

Children's Style of Approach to Scientific Exploration

These children have very different approaches to exploration that are in many ways coherent across time. We have interviewed these children twice a year for 4 years, and observed them in class. We have found this continuity of approach is true of other aspects of their response to science.

While Anna is quite capable of generating abstract explanations of the whirlybird flight, and indeed shows a facility with generating ideas, she did not seem particularly interested in testing her ideas by a program of exploration. She tended to produce ideas and not move from that position. Anna often takes refuge in decorative descriptions of phenomena, constructing linguistic edifices that serve her well but is perhaps unwilling to become a participant in a program that might probe their inadequacy. She is not confident of her ideas. A major focus continues to be social, and a concern with aesthetic and emotional aspects of play (friends, colouring whirlybirds).

Calum is extremely interrogative in his approach to these and other tasks. He is endlessly speculative, often displays a clarity of vision, and has the ability to go to the essence of an issue. Of all the children, he is comfortable with saying "I don't know". He has been doing science experiments at home since he was very young, and is very self motivated exploring phenomena, paying attention to scientific design, with appropriate controls. On the first occasion while he was quite controlled in his exploration and came up with a high level speculation, he could not put this together with evidence. On the second occasion his whirlybird exploration was extremely sharp, and well coordinated with his understanding.

Esther is not a strong student of science and struggles somewhat with ideas. She is a very likeable girl, and enjoys school. She operates with a restricted epistemology over these three years, but is very interested, particularly in investigations involving small animals. She is very willing to answer questions and explore, but needs to be scaffolded to make appropriate comparisons. She does, however, respond in Grade 2 with the idea of fairness in testing as a reason to control variables, and her domain specific knowledge concerning air is increasing. She tends to focus, even in Grade 2, on emotional or associative aspects of the phenomena; "I like that one the best," fantasy about the lego figures in the parachute exploration; "that one was like a builder, and this one was a fixer," or analogies with helicopters "Oooohhh . . . . Slow! [excited]. And it looks like . . . a real helicopter!"

Jeremy is a highly spatial/visual thinker. He seems to work best when he is playing with the whirlybirds and exploring with his hands and eyes. Perhaps he operates with a visual map inside his head. He is gifted at drawing. He seems to love the sudden insight, and is enthusiastic in his chase after confirmation. On both occasions he enjoyed playing round, but once he had an idea was very self directed. He is very dedicated to chasing deeper causal insights.

Karen on both occasions took charge of the investigation in a very organised and overt way. She is also a very self directed child, very able, and in interviews is very focused on impressing the interviewer and actively probing for what they are requiring. In Grade Prep she planned a program of pair wise comparisons somewhat like a world cup final, paying some attention to interpretation along the way. In Grade 2 she set about constructing a table based on timing, and lined the whirlybirds up in a manner she had experienced three years before in a class activity comparing parachutes, and extracted her arguments from that. She was more confident with her interpretations, but as on the previous occasion the achievement of a rigorous method of comparison was her primary concern.

Patterns of Growth in Children's Reasoning

Each of the interview sequences was coded on the dimensions of *approach to investigation*, and *handling of variables*. The scoring on these dimensions for all the children who were interviewed for at least two of the flight activities is shown in Tables 3 and 4. The missing data occurred when children were not able to be interviewed, or when the transcript did not provide sufficient evidence for assigning a score. The children are arranged in score order to bring out more clearly the consistent growth over time, given the different starting levels. The scores on these tasks are relatively consistent with their scores on other reasoning tasks undertaken at similar times (Tytler & Peterson, 2003a). There are, however, some cases where children seem to regress. This would be expected if there were contextual elements to reasoning. The mean scores on each dimension reflect the overall trend. At the end of their third year at school no child performed on an ad hoc basis, and each child showed some facility with variable control. This was not the case at the end of their first year.

The scores on the two dimensions, based on these three tasks, are strongly correlated. It is not clear from the design whether this relates to variation in underlying general ability, to differences in knowledge, or to different epistemological positions these children tend to adopt. Further work needs to be done on the interrelationship between these reasoning dimensions. In the following section, the role of knowledge in guiding the children's exploration is explored.

The Nature of the Change: The Role of Knowledge in Guiding Exploration

These children were clearly operating more competently, in a more considered way, when they were 7 compared to when they were 5 years old. They each advanced on the *approach to exploration* dimension, either identifying patterns in their findings

Table 3
Children's Scores for the Approach to Exploration Dimension.

	Whirlybird task November – Grade Prep	Parachute task June – Grade 1	Whirlybird task November – Grade 2
Esther	1	1.5	2.5
Anna	1	2	2
Elle	1	2	2
Theo	2	2	3
Sonya	2	2.5	–
Larry	2	3	3
Calum	2.5	2.5	3
Rae	3	2	3
Karen	3	3	3
Walt	3	3	3
Jeremy	3	3	4
Mean score	2.1	2.4	2.9

Table 4
Children's Scores for the Handling Variables Dimension.

	Whirlybird task November – Grade Prep	Parachute task June – Grade 1	Whirlybird task November – Grade 2
Esther	1	1	1.5
Anna	1	1.5	2
Rae	1	1	2
Calum	1	1.5	2.5
Sonya	1	2	–
Elle	1	–	1.5
Theo	1.5	1	2
Jeremy	1.5	2	2
Karen	1.5	3	2.5
Walt	2	2.5	2.5
Larry	–	2	2.5
Emily	–	3	3
Mean score	1.3	1.8	2.2

if they were not before, or exploring ideas in a more focused way. They are identified and controlled variables more competently.

There is a tradition in cognitive psychology of trying to describe levels, and stages in the development, of scientific reasoning as an abstract entity. If one looks at the growth that has taken place here, however, it seems clear that the major gains have been in knowledge of the science relevant to parachutes and whirlybirds. Both at age 5 and age 7 we can find many examples of the way knowledge can constrain, or can shape children's approach to exploration and their ability to identify and control variables. Examples are:

1. Anna's lack of attention to the effect of paperclips at age 5 because she was unable to identify weight as a possible influence on flight.
2. Esther's newly developed notions of weight, and the effect of canopy size, enabled her to propose productive comparisons in the parachute task where there was no movement towards this 7 months earlier in the whirlybird task.
3. Calum's clear exposition of the flight as a tension between uplift on the wings and weight down from the paperclip, seemed to provide a framework for a very efficient program of investigating these two variables independently.
4. Jeremy's understanding of the dynamics of flight, and a possible effect on flight stability of an interaction between weight and the wing length, led him away from the simple two variable investigation into more complex reasoning that involved a fast interaction between evidence and ideas.
5. For all the children, greater access to the concepts of air and weight and their effect on flight, increasingly enabled them to take effective charge of the investigations, based on an expectation of what might happen. In Jeremy's case his understanding undermined his commitment to a simple controlled investigation.
6. In interpreting the whirlybird flight in Grade 2, Karen's lack of focus on conceptual interpretation of whirlybird flight seemed to be limiting her ability to construct a coherent test even though on other evidence she is well capable of this.

One can trace, through the three investigations, children's increase in knowledge concerning the effect of air and weight on flight. Some of this growth may come from their schooling, since in some classes they had experience with flight activities. However, there is not a tradition of physical science topics in the school curriculum, and it is likely their increased knowledge also relates to general life experience. The children supported their increasing knowledge with anecdotes and analogies with parachute and plane flight, and more general falling motion. Carey (1985a, 1985b) argues that growth in children's theories occurs chiefly as the growth in domain specific knowledge. This would seem to be born out by these transcripts. There was no case where, from age 5 to age 7, a child made an advance in scientific reasoning, in terms of the way they framed the exploration, or in their ability to control and interpret variables, without an accompanying increase in sophistication of conceptual knowledge. This makes sense, if one considers that such reasoning cannot occur in a vacuum. It supports Koslowski's (1996) claim that scientific reasoning cannot be reduced to control of covariant relations, but that intuition, and knowledge imported from outside the particular investigation, are properly an essential part of the reasoning of practising scientists.

While conceptual knowledge will affect children's performance on exploration tasks, it is equally true that the ability or tendency to evaluate and rethink explanations in the light of evidence will critically affect children's conceptual growth. From the first year at school, those children who were approaching explorations in a speculative and inquiring manner were also those who showed the earliest growth in conceptual knowledge. This increase in conceptual knowledge then feeds back into an ability to approach and control experimentation. Koslowski (1996) refers to the process of knowledge building as "bootstrapping," involving the interdependence of knowledge and scientific reasoning. She argues that this interdependence implies that science procedures are inevitably "rules of thumb," in contrast to the algorithmic view implied by many covariation studies. That view is supported by this study. We are making the further point that the tendency to approach exploration in a speculative, hypothesis generating way involves the epistemological insight that in science, ideas are generated and used to guide our search for evidence and our interpretation of phenomena. To effectively construct ideas, you need to realise their interpretive role.

Children's Interpretation of the Tasks, and their Reasoning

Despite the children's increasing grasp of conceptual knowledge, it was interesting to note the continuing difficulty some of them had in naturally controlling variables, and in particular in mounting a clear demonstration of the effect of wing length, or of weight. Even if they could do this with prompting, with one exception (Calum) they did not seem to readily understand the question that was asked. Perhaps, while children can intuitively take an idea and seek relevant evidence to confirm it, the more formal process of arguing back from evidence to clearly establish an idea is a more difficult proposition. It may be the case that while an intuitive grasp of variable control is within these children's reach, a formal grasp requires exposure to the discourse of controlling confounding variables as a mode of argument. On the other hand, there is evidence from other tasks that these children undertook, that the idea of controlling extraneous conditions in a single variable investigation (as in "fair testing") is well within their grasp, and has lower order cognitive demand.

It is clear, from the variety of ways these children tackled the whirlybird task at age 7, that the task itself was different for each of them. This was true for their personal response to the task in terms of interest and approach. It was also true at a conceptual level. Each child was engaging with a different problem. Jeremy, for instance, identified quickly the mechanism by which the whirlybird span, and became interested in whether the stability of flight, affected by the paper clip, might be a factor that operates independently of the wing span. Similarly, Karen, rather than setting up a classic variable control design, lined the whirlybirds up in order of flight and proceeded to mount an interpretive argument by looking for patterns. She did not initially understand the question RT posed about demonstrating the effect of the clip because she was working within a different experimental paradigm.

We would argue that, in a real investigation driven by a substantive question, investigative tasks are inherently more complex than a simple variable control model can accommodate. There are other questions, other factors, and outside knowledge that will complicate the issue. The child may in the end have a more sophisticated agenda than the simple variable question anticipates, and it is important to accommodate this possibility. The interplay between ideas and evidence is the key to scientific reasoning.

Reflecting on the Longitudinal Design

We have discussed our findings on a variety of aspects of the study, above. In this section we will particularly consider the methodological aspects; the particular advantages conferred by the longitudinal design, and the issues raised.

The Niche Occupied by Longitudinal Studies

What is it about the nature of the longitudinal research design, that allows special and unique insights into student learning and knowledge? To explore this question we will consider which of the findings are accessible by cross-sectional designs, and which are unique to the longitudinal design.

Firstly, the generation and verification of the reasoning framework would be accessible to cross sectional designs. It was, in fact, generated using data from one year only. Secondly, the general growth trends, evidenced by the growth in mean scores in Tables 3 and 4, are also accessible cross sectionally. These could be duplicated with many more children using different cohorts, to verify the general nature of the growth, although there is a problem with establishing cohort comparability with cross-sectional studies. The proper focus of the longitudinal study is the detailed shape and nature of the change, in this case for individual children. Thus, the change in approach to investigation of Anna and Esther, who started consistently using a phenomenon-based reasoning approach, but are consistently using relation based reasoning by the end of Grade 2, is of interest. The shape of the general rise, with some children appearing to regress, indicates the contextual nature of performance, and this was a major finding of the evaporation study that challenged presumptions of learning as an orderly progression in successively more sophisticated conceptions. Thus, the strength of the longitudinal study lies in its identification of pathways of learning and development with regard to particular understandings or dispositions, with the possibility of providing insight into the complexity of learning pathways. As Hellden (this volume) argues, the particular strength of the longitudinal design is its ability to explore detail within the individual patterns of change, and the nature of the continuity characterising the change.

To make the most of the design, the comparison that must be made is between two points in time, for the one individual, to explore the change pathway and establish the essential logic of the changes. Having 11 other individuals provides some sense of the variation in these pathways and the individual characteristics that give rise to them. Cross-sectional studies look for commonalities and patterns across a cohort, then look to different cohorts to provide perspective. In making sense of the analysis in each case, it is inevitable that different features will be ignored as unwanted "noise," in order to impose some sense on the data. For the longitudinal study, what is retained is a sense of the coherence of an individual's approach to phenomena and conceptual journey and what is secondary is a sense of "average" understandings across the cohort. For cross sectional studies, what is retained is a sense of coherence in the patterns of understandings of a particular age group, compared to other age groups. What is lost is the complexity that derives from individual characteristics not directly related to conceptual (or epistemological or whatever) levels, and, we would argue, the opportunity to construct a nuanced view of learning.

The other particular advantage offered by the longitudinal design, for this study, is the opportunity it offers to look for interactions between different elements of children's thinking and acting, and to focus more broadly on the whole child. If we believe that learning occurs through a rich complex of cognitive, emotional, aesthetic and other factors, then advantage must be offered by the opportunity to study their growth in concert. Thus, with the analysis of the two whirlybird investigations, it was possible to discern both the continuity of children's approach and the effect of increased knowledge on their ability to reason scientifically. The interaction between them becomes more visible. The more complex our view of learning, the more necessary it becomes to study it longitudinally.

Methodologically, also, the closer relationship we have developed with these children has focused our attention much more on the whole child than simply on conceptual issues, and we feel we are able to extract more in interview than if we were strangers to these children. Knowing where a child has come from confers a sense of understanding of their responses, and a way of triangulating to ensure the data is making sense from different, historical and contextual perspectives. It has sometimes been possible, also, to look back over earlier data and make more sense of it, knowing the way the child has developed since. The sense of personal history throws into sharper relief key aspects of their thinking. Thus, Calum's identification of controlled comparisons as an essential nature of science experiments, in the preparatory grade, makes sense looking back from his later fascination and facility with controlled experiments involving measurement. Karen's first performance in exploring the whirlybirds, ignoring any overt conceptual underpinning, is confirmed as a personal approach to ordered comparison by her later exploration of the same whirlybirds. Anna's reliance on vocabulary as an essential aspect of understanding science, and her strong memory of social contexts, has a long history in her approaches to these tasks.

More recently (Tytler & Peterson, 2003c) we have revisited children's early references to "doing science at home," and followed through their references to learning at home, and their later stories of doing science. These resonate strongly with our findings concerning children's approaches to learning and views of themselves as learners, coming out of the interviews. It seems that these children's particular epistemological stances are already broadly shaped when they begin school. We intend to further develop our analyses of this.

Issues with the Design

We have previously raised the dilemma inherent in this study, of where to focus attention to tell the story. An issue that we are sometimes confronted with, methodologically, is whether 12 subjects are "enough" to draw valid conclusions. We believe that for these types of analysis, two subjects are enough to make a valid point. In analysing the data and reporting, we have generally used a two-tiered approach. We have looked for patterns of growth across the 12 children, using content analysis, to explore general growth and the degree of complexity, and used selected detailed case descriptions to illustrate the complexity and unity characterising individuals' learning that we feel to be the real focus of the study. In analysing at this level of detail, 12 are too many rather than not enough. However, having the 12 children representing a range of dispositions and abilities provides the necessary confidence that the study is in some sense capturing the range of children's understanding pathways.

A second issue concerns whether we should regard what we are doing as an intervention. While we have organised or ourselves run sequences of activities on particular topics, they do not sit within the design as interventions we are evaluating. Rather, they provide a focus for discussion and an opportunity to see what children take from them. They ensure that when we question children about a phenomenon, they will have had a chance to develop a position on it. The sequences are similar to what one would expect of best practice in primary school science, but are short and do not push for closure on specific understandings. The interviews and tasks themselves are unavoidably an intervention, and we suspect these 12 children have developed a particular mode of talking to us that focuses more sharply on conceptual matters than it otherwise might. Certainly, our 12 children are the envy of their classmates whenever we appear on the scene to talk with them. It could be argued, therefore, that these children are developing in a somewhat "hothouse environment," changed by regular acts of probing their understanding. While this may be the case, we expect it is not a major effect, and in any case would argue that what we are studying should not be construed as "normal" development and learning in terms of its pace, but rather we focus on the pathways of change, and the interactions and complexities involved in this. Our design lies somewhere between a "naturalistic" study, and an overt intervention. The lines of development, and learning, are blurred. Indeed, we believe they cannot be separated in practice or in principle.

A third issue relates also to the nature of the interview as intervention. There is the possibility that there exists an interference effect from memory, as children's responses to repeated tasks and questions are triggered by the interview situation. For this reason we repeat interview tasks and questions in a 2–3 year cycle. Our general impression has been that children do not seem to be aware of their previous responses, even if we can see a coherence when analysing the transcripts. A further, more general version of this concern relates to the relationship between ourselves and these children. In some senses the interview responses must be read as a co-construction, with children developing over time a particular way of responding to us as representatives of science conceptual pursuits. There is no way to truly disentangle ourselves methodologically from any interview responses, and the problem is exacerbated through the long-term relationships we have built up. However, to minimise and take account of this, the two of us do not consistently interview the same children. We are able to triangulate our analyses by talking through our interpretations of each child, by observing the children in class and by talking to their teachers.

A fourth issue we have been confronted with has been the changing nature of our own understandings as the study progresses. This is partly a result of the literature having moved on over the last five years, but also because we ourselves are understanding issues better, and writing about them. Thus, the way we view and conduct the study inevitably changes, and there are many points at which we wish we had gathered different or better data at the earlier stages. Some of the data we collected has been rendered redundant as our attention shifts elsewhere. This issue also occurs in a more practical sense, in terms of changes in our own professional lives and the time we have available. The trend here is that available time always seems to diminish. It is hard to anticipate in detail, in such a study, where it will go.

A particular problem for us, and we believe for other researchers represented in this volume, is just how we represent our findings theoretically. In studies which throw up new and sometimes surprising data, we are entering new theoretical territory as we attempt to grapple with the layered and individual nature of understanding, and its links with memory and narrative. The design opens up interpretive problems that call for a new theorising. Current models of learning are proving inadequate for our purposes. However, even without such theorising, we believe that this study is providing us with a nuanced view of the learner and learning that is ultimately closer to the way teachers see children and learning than is possible with cross-sectional studies, which may not bring out the richness that is so critical in understanding and breathing life into the learning and development process.

Where to Next?

How long should a longitudinal study be? As with a piece of string, there is no clear answer to this. Already we have data that could be mined for some time to come, concerning children's developing understandings of air, of force and motion, and of animal and plant adaptation. There is the opportunity to examine parallel conceptual development in these areas to look for commonalities in ways of thinking, for instance, that we have not yet had time to explore fully. There is also the feeling that having collected this body of data concerning this group of children, it would be a pity to stop now, since there may be developments around the corner that the earlier data can be linked to. As the children become older, the opportunity will increasingly arise to explore their ideas through writing, about their entry into adolescence, and their view of their own learning and their considered response to school subjects and schooling. Some lines of inquiry that we may pursue include:

1. How does the growth in reasoning continue into the upper primary school, and how is this linked with learning in different science content areas?
2. How do children's developing views of their own learning relate to their earlier positioning as learners in school classrooms?
3. How does a child's changing attitude to science across the primary-secondary transition relate to his or her history of response to science tasks across the primary school?
4. How might children's response and learning outcomes to a carefully structured unit incorporating conceptual challenge and investigations, depend on their previous conceptual histories?

Thus, the longitudinal design could be seen as an unfolding of possibilities, as our conception of the person as a learner and participant in the schooling process broadens to include a wider net of dimensions of thinking and meaning making.

Correspondence: Russell Tytler, Faculty of Education, Deakin University,
221 Burwood Highway, Burwood, Victoria 3125, Australia
Email: tytler@deakin.edu.au

References

Arzi, H. J. (1988). From short- to long-term: Studying science education longitudinally. *Studies in Science Education, 15*, 17–53.

Black, P., & Lucas, A. (Eds.). (1993). *Children's informal ideas in science*. London: Routledge.

Carey, S. (1985a). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1985b). Are children fundamentally different kinds of thinkers and learners than adults? In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2, pp. 485–514). Hillsdale, NJ: Erlbaum.

Chin, C., & Brown, D. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of Research in Science Teaching, 37*(2), 109–138.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7), 5–12.

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

Driver, R., Leach, J., Scott, P., & Wood-Robinson, C. (1994). Young people's understanding of science concepts: Implications of cross-age studies for curriculum planning. *Studies in Science Education, 24*, 75–100.

Duschl, R., & Gitomer, D. (1991). Epistemological perspectives on conceptual change: Implications for classroom practice. *Journal of Research in Science Teaching, 28*(9), 839–858.

Engel Clough, E., & Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education, 70*(4), 473–496.

Engel Clough, E., Driver, R., & Wood-Robinson, C. (1987). How do children's scientific ideas change over time? *School Science Review, 69*(247), 255–267.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.

Pintrich, P., Marx, R., & Boyle, R. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*(2), 167–199.

Solomon, J. (1994). The rise and fall of constructivism. *Studies in Science Education, 23*, 1–19.

Strike, K., & Posner, G. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147–176). Albany, NY: State University of New York.

Tytler, R. (1993). Developmental aspects of primary school children's construction of explanations of air pressure: The nature of conceptual change. *Research in Science Education, 23*, 308–316.

Tytler, R. (1994). Children's construction of explanations in science. In J. D. Novak (Ed.), *Proceedings of the third international seminar on misconceptions and educational strategies in science and mathematics* (published electronically, disk 21). Ithaca, NY: Cornell University.

Tytler, R. (1998a). The nature of students' informal science conceptions. *International Journal of Science Education, 20*(8), 901–927.

Tytler, R. (1998b). Children's conceptions of air pressure: Exploring the nature of conceptual change. *International Journal of Science Education, 20*(8), 929–958.

Tytler, R. (2000). A comparison of year 1 and year 6 students' conceptions of evaporation and condensation: Dimensions of conceptual progression. *International Journal of Science Education, 22*(5), 447–467.

Tytler, R., Duggan, S., & Gott, R. (2001). Dimensions of evidence, the public understanding of science and science education. *International Journal of Science Education, 23*(8), 815–832.

Tytler, R., & Peterson, S. (2000). Deconstructing learning in science: Young children's responses to a classroom sequence on evaporation. *Research in Science Education, 30*(4), 339–355.

Tytler, R., & Peterson, S. (2003a). Tracing young children's scientific reasoning. *Research in Science Education, 33*(4), 433–465 (Theme Issue on Science in the Early Years).

Tytler, R., & Peterson, S. (2003b). The nature of growth in children's science understandings: Insights from a longitudinal study. In D. Psillos, P. Kariotoglou, V. Tselfes, E. Hatzikraniotis, G. Fassoulopoulos, & M. Kallery (Eds.), *Science education research in the knowledge based society* (pp. 191–200). Dordrecht, The Netherlands: Kluwer.

Tytler, R., & Peterson, S. (2003c, July). *Capturing the child: A longitudinal study of children's science learning.* Paper presented at the annual meeting of the Australasian Science Education Research Association, Melbourne, Australia.

Tytler, R., & Peterson, S. (2004a). From "try it and see" to strategic exploration: Characterizing young children's scientific reasoning. *Journal of Research in Science Teaching, 41*(1), 94–118.

Tytler, R., & Peterson, S. (2004b). Young children learning about evaporation: Insights from a longitudinal study. *Canadian Journal of Science, Mathematics & Technology Education, 4*(1), 111–126.

Tytler, R., & Swatton, P. (1992). A critique of Attainment Target 1 based on case studies of students' investigations. *School Science Review, 74*(266), 21–35.

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

White, R.T. (1988). *Learning science*. Oxford, UK: Blackwell.

White, R.T. (2001). The revolution in research on science teaching. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 457–471). Washington, DC: American Educational Research Association.