

COMPUTATIONAL MODELLING AND CHILDREN'S EXPRESSIONS OF SIGNAL AND NOISE

JANET AINLEY
University of Leicester
janet.ainley@le.ac.uk

DAVE PRATT
University College London Institute of Education
d.pratt@ioe.ac.uk

ABSTRACT

Previous research has demonstrated how young children can identify the signal in data. In this exploratory study we considered how they might also express meanings for noise when creating computational models using recent developments in software tools. We conducted extended clinical interviews with four groups of 11-year-olds and analysed the videos of the children's activity through a process of progressive focusing.

In this paper we explain the design of our tasks and report how the children's expressions for noise, supported by the need to communicate with the software, developed from specific values to verbal expressions of uncertainty such as 'around', to offering ranges of values. We consider the opportunities and constraints of such an approach, which we call 'purposeful computational modelling'.

Keywords: *Statistics education research; Active graphing; Exploratory data analysis; Purpose and utility*

1. THEORETICAL FORMULATION

Previous studies (e.g. Makar & Rubin, 2009; Pfannkuch, 2007; Zieffler, Garfield, delMas, & Reading, 2008) have reported how students have used the power of modern software to tabulate and graph observed or given data to unearth possible stories that might account for the data. Broadly speaking, these research studies have made use of pedagogic approaches that exploit the power of digital technology to manipulate, compute and dynamically represent data. These approaches have been based on Exploratory Data Analysis (EDA) (Tukey, 1977). Our aim in this study was more ambitious. We set out to design and evaluate a pedagogic approach in which young students would also account for data which did *not* fit the story. We might regard the 'story' as the signal and the data which does not fit as the noise; we intended to observe students as they attempted to offer accounts of both signal and noise.

Konold and Pollatsek (2002) have argued that the signal and noise metaphor, even though it is rarely exploited in the classroom, is often more accessible than the metaphor of central tendency for analysing data in three different statistical processes, namely exploring situations involving: (i) measuring individuals (for example, when comparing women's and men's heights, the signal might be thought of as the 'normal' height of one gender compared to the 'normal' height of the other, whereas specific heights of individuals contain noise that distances them from the signal); (ii) dichotomous events (such as the rate of contracting or not contracting polio, where the signal is regarded as the propensity for a treatment group to get polio compared to that of a placebo group, while recognising that the frequency of polio victims will be subject to noise and so will not exactly match the

frequency predicted by the propensity probability); (iii) repeated measures (such as when trying to distinguish the ‘true’ value (signal) amongst repeated error-ridden measurements (noise) of the same quantity). Konold and Pollatsek’s use of the terms signal and noise broadens their usage from just contexts in which multiple components of variability are explicitly evident in observable data or in data generated from sampling processes. They claim that there is reason to believe that the metaphor of signal and noise can be transferred with pedagogical benefit even to situations with univariate distributions. Consider measuring the height of individuals. Konold and Pollatsek argue that it is difficult to imagine the average height as a signal (and variation from that average as noise) because the notion of a ‘normal height’ or a central tendency is obscure and does not have a clear purpose. In comparison say to the role of measurement error in a repeated measures process, the sources of noise are not at all obvious in the distribution of heights. Even so, they speculate that when considering the comparative height of men and women, the difference between the averages can be related to the impact of gender and the signal becomes a little more accessible because the comparison provides a purpose and the variation can be partly attributed to gender. We will return to these classifications later but first we discuss studies which have primarily focused on signal.

1.1.ACTIVE GRAPHING

Pedagogic approaches which focus primarily on signal are not uncommon. We have previously reported such studies that, in fact, inspired the current work. These studies had been designed to support the analytical (as opposed to presentational) use of scatter graphs by children aged between 8 and 12 years. We called this pedagogic approach *Active Graphing* (AG) (Ainley, Pratt & Nardi, 2001). Young children conducted experiments that involved the collection of bivariate data. For example, in one task, children explored the mass that could be supported by a paper bridge, where the independent variable took the form of the number of folds in the paper, with the aim of making a bridge that could support a precious china egg. In another task, the focus was on the time of flight of a paper spinner, where the independent variable took the form of the length of its wings, the aim being to make a champion flyer. As they collected data in these kinds of tasks, pupils repeatedly created scatter graphs from a spreadsheet and discussed the emerging patterns within them in order to make decisions about the next stage of data collection. In this way they were able to gain a sense of the relationship between the independent and dependent variables by gradually making sense of the developing scatter graphs.

AG supports students in using graphs to identify the story (or signal) in the bivariate data by recognizing how variation in the independent variable affected the dependent variable. To achieve this, the students needed to be able to control the independent variable. In that way, they were able to make decisions, based on the most recently generated scatter graph, about which values of the independent variable to use as the basis for future trials. AG is therefore highly suited to running sequences of trials of an experiment where each trial involves fixing one quantity and measuring a second quantity. Such experiments are typical in school science. However, in science the emphasis is on managing the noise by standardising all other variables to make a ‘fair test’ and by averaging repeated measures, such as when seeking to minimise measurement errors. These strategies are understandable when the aim is to identify the signal but they reduce the potential of the task in terms of exploring noise.

We have described phenomena such as these, where there is an easily identifiable signal but it is also apparent that the signal is insufficient to account for variation in outcomes, as *partially determined* (Ainley & Pratt, 2014). Although a wide range of phenomena, and

particularly many of those studied in school science, can be regarded as partially determined, educational research on probabilistic thinking has largely focused on phenomena that are generally modelled as totally random and there has been very limited research on students' understanding of partially determined phenomena.

Unfortunately a limitation of AG is that, because it requires an independent variable, it is not an appropriate pedagogic approach when exploring observational data. For example, a typical classroom project is to ask the children to measure themselves and look for relationships within the data, such as a connection between arm length and height. In such an activity, the children select an individual child and measure both variables; there is no independent variable that can be controlled and so the AG approach is not appropriate.

1.2. PEDAGOGIES BASED ON EXPLORATORY DATA ANALYSIS

As we considered how to overcome this limitation, we were attracted to research in which young students were uncovering stories in data through EDA, which was often being used to explore observational data. Probability is generally seen as a difficult topic to teach and learn (e.g. Falk & Konold, 1997; Jones, 2005; Shaughnessy, 2003) but EDA is seen as a technique that could bypass those difficulties and thus facilitate the identification of 'stories' in data (Cobb, 2005). We therefore considered whether EDA might open up the possibility of using observational, as well as experimental, data to explore signal and noise.

We begin by considering an example where young children (age 8) explored data with the aim of finding the typical height of a student in Year 3 (Makar, 2014). In Konold and Pollatsek's classification (2002) detailed above, the students were 'measuring individuals'. The children measured the heights of their classmates and began to articulate notions of what was a reasonable estimate for a Year 3 child's height. As they began to organise their data, the language used for average became 'most common'. The sample was then grown by visiting neighbouring classrooms and statements about middle, medium and modal clump were observed in the children's comments. Towards the end of the process, the children began to formulate notions of the average being representative. It is not clear from the data which of these meanings are close to the notion of central tendency and which relate more to signal. What is clear is that the students articulated a wide range of meanings for average and, although these meanings were set in a situation where there was obvious variation, there appeared to be less attention on that variation than on the central task of identifying the typical height.

In a second example from the literature, the participants were trying to identify the signal when the data were dichotomous, the second category in Konold and Pollatsek's (2002) classification. Hammerman and Rubin (2004) reported how middle and high school teachers and their students drew on the functionality of TinkerPlots to make sense of data on AIDS. They reduced the apparent variability in a data set by grouping the values using numerical bins or cut points and by considering proportions if the bin sizes were unequal. The authors recognised the difficulty in handling the tension inherent in data analysis between reducing variability as a way to deal with the complexity of the data and the risk of making claims that might not be true. A signal in the data may become apparent when the variability is reduced but it is consideration of noise that gives the story-reader some sense of whether the story is a false testimony. We will leave discussion of repeated measures, the third type of data analysis in Konold and Pollatsek's (2002) classification, until the next section on modelling uncertainty.

Some consideration of uncertainty is seen in a study by Ben-Zvi, Aridor, Makar and Bakker (2012), who observed emergent articulations of uncertainty in students aged 10 to 11 years as they used EDA methods to make judgments about patterns and trends in data.

Their expressions of uncertainty first oscillated between certainty-only (deterministic) and uncertainty-only (relativistic) statements but gradually seemed to take on a form that the researchers regarded as the ‘buds’ of probabilistic language.

Although EDA offered a means of analysing observational data, which AG did not, EDA relies heavily upon graphical interpretation, which is itself known to be challenging for students (Curcio, 1987; Monteiro & Ainley, 2004) but which AG addresses. EDA and AG suffer the same limitation in relation to our aims insofar as neither encourages students to reflect on the variation (or noise). In the case of AG, there is a shared purpose focused on a product, such as a strong bridge or a champion flyer. For example, in the task to find the champion flyer described above, students often decided to take several measurements and use the average, which helped to create graphs with clearer stories. This strategy was effective in supporting decisions about the design of the flyer, but missed the opportunity to reflect on the variation within the measurements. In the case of EDA, the avoidance of probability is at the expense of addressing issues around uncertainty in a systematic way. We were concerned that this lack of attention on uncertainty would maintain the focus too tightly on the story, or the signal, without sufficient opportunity to offer accounts of the noise.

1.3. PEDAGOGIES BASED ON MODELLING UNCERTAINTY

It seemed that neither AG nor EDA-based pedagogies gave sufficient emphasis, at least for our purposes, on noise and that we needed an approach that could support students’ accounts of uncertainty. Learning to reason with uncertainty is increasingly recognised as an important element of the statistical literacy needed in adult life and citizenship (e.g. Royal Statistical Society, 2014). Some research on uncertainty has engaged students in modelling data, often through the use of digital technology, and this approach seemed to show some promise for our research.

For example, Lehrer and Schauble (2004) studied a class of twenty-three 10- and 11-year-old pupils. The focus was on the students’ thinking about natural variation. They grew batches of fast-growing plants, observing growth over time, subject to different conditions, including fertilizers. Throughout the two-month growing period, measurements were taken and the students were encouraged to invent and evaluate representational conventions for this observational data. The researchers reported how the students learned to reason about natural variation by generating, evaluating and revising models of data about the plants that incorporated ideas about both signal and noise. The researchers valued the affordances offered by technology to allow flexible partitions of data and ways of viewing cases and aggregates simultaneously. The representations or models created by the students were inscriptions of the actual data or imagined data if the experiment were to be repeated. In this sense, the variation was either observed or imagined to be the result of natural growth.

They also used a specially written program to generate samples and to explore their distributions for different sample sizes and numbers of samples. More recently technological developments in TinkerPlots, an earlier version of which had been used by the students in Lehrer and Schauble’s study, have introduced the possibility for students themselves to create computational models, which incorporate randomness and can generate data. We wondered whether pupils using these tools might account for variation in growth that could not be explained by time or by the differing conditions through randomness, perhaps drawing on probabilistic ways of expressing such noise. This possibility intrigued us. In fact Konold and Lehrer (2008) reported investigations of how 10- and 11-year-olds developed models of errors made when repeatedly measuring head circumference, an example of Konold and Pollatsek’s (2002) third category, ‘repeated

measures'. For example, one pair of students created a model, in which they simulated three typical types of error by spinners that generated random values. The measurement error was taken to be the sum of these three separate errors. A head circumference could be modelled as the 'true' circumference added to the measurement error. When the model was run 100 times, a slightly skewed distribution of head circumferences was generated. This study showed that it might be possible to design tasks based around computational modelling where students needed to address both the signal (the 'true' circumference) and the noise (the measurement error).

Whereas AG and EDA tended to place the emphasis on the signal in the data, we read these studies on modelling data with interest because they seemed to provide opportunities for students to create accounts of both signal and noise. We therefore set out to explore the design of tasks involving a modelling approach as a way to exploit both the strength of AG to support graphical interpretation and that of EDA to handle observational data.

1.4. PURPOSEFUL COMPUTATIONAL MODELLING

We began to explore how modelling in a computer environment might provide a way of placing emphasis on noise while marrying the strength of AG to support purposeful graphing with the strength of EDA to handle observational data¹. We noted the new functionality to create statistical models of phenomena in TinkerPlots 2 (TP2) (<https://www.keycurriculum.com/products/tinkerplots>). Consider again the task for children to measure their arm lengths and heights. Suppose that the focus of the task becomes to model the relationship between arm length and height. Such a model might be understood as a machine to create heights from arm lengths (as in the 'cat factory', Konold, Harradine & Kazak, 2007). A successful machine (i.e. the model) would need to reflect the signal, in the form of a relationship between arm length and height, since arm lengths and heights are not independent, and the noise, as not all children with a specific arm length have the same height. Such a modelling approach creates an artificial independent variable, arm length in this example, much as statisticians model phenomena by conjecturing signal variables to see if they do in fact account for significant amounts of variation in the dependent variable. For example, statisticians build models, such as those embedded in research designs involving analysis of variance, that incorporate signal and noise, in order to ascertain whether the model suggests that the effects identified in the signal are significant compared to the residual error in the noise. The models that the children express might bear some hallmarks of statisticians' methods insofar as the children might begin to articulate, separate and quantify signal and noise. The creation of an artificial independent variable opens up the possibility that AG might be an appropriate pedagogic approach in tasks based on observational data, where it can be combined with EDA methods of investigation through a modelling process.

We have developed a pedagogic approach for young pupils (age 10-11 years) that involves the use of interactive modelling tools, such as those in TP2, within carefully focused data-based tasks. We chose tasks in which one variable is partially determined by another (Ainley & Pratt, 2014), with the signal capturing the variability that can be accounted for by the independent variable and the noise representing the remaining variability. In these tasks, pupils build models that can be run to generate data which are represented in tables and graphs, providing feedback that allows iterative modification of the model. We refer to this as *computational modelling*. Our approach is to harness computational modelling within a setting that gives due regard to stimulating and

¹ This project was undertaken in collaboration with colleagues for the University of Haifa.

purposeful activity. To capture these twin pillars of our approach, we use the term, *Purposeful Computational Modelling* (PCM). In summary, the PCM approach has the following pattern:

- A scenario is presented which relates to the pupils' existing experience, within which they are asked to prepare some kind of report or output.
- A starting point is offered in which pupils either collect data from a simple experiment in the classroom (AG style), or are given a small data set to explore (EDA style).
- As they work on this experiment or exploration, attention is focused on relationships between variables, which can be explored through tables and graphs of the data.
- In order to provide the required output, pupils are shown how they can build a model (usually referred to as a 'machine') in TP2. In creating the model they need to express the relationships they have observed, encapsulating both signal and noise.
- The effectiveness of this model in producing the required output is evaluated by using graphs to compare the generated data with the initial data that the pupils collected or explored, in the style of AG.

Thus the modelling activity does not stand alone, but is intrinsically embedded within tasks that incorporate the following principles drawn from Active Graphing and EDA:

- The tasks are purposeful from the pupils' point of view;
- The tasks involve data that is meaningful because the pupils collect it themselves, or can imagine how it was collected in a familiar context;
- There is a real need for the pupils to use graphs in order to complete the task;
- The nature of the required output provides an obvious purpose for creating the model.

Having set out this novel approach we are now in a position to state more precisely the key research question for this study: How do children express ideas about signal and noise within a PCM pedagogic approach? In elaborating this research question, we expected at the same time to gain insights about how PCM supported or obstructed such expressions.

2. METHODS

In this study we designed and used two extended tasks, which we describe in detail here. The iterative design of these tasks was informed by design-based research (Cobb, Confrey, diSessa, Lehrer & Schauble, 2003). In the first place, each task was designed through discussion amongst the whole research team, comprising the two authors of this paper and colleagues from the University of Haifa². The second stage of development of the extended tasks involved the researchers working through the construction of a TP2 model to inform a hypothetical learning trajectory for the activity of the students as they might work through the extended tasks. Adjustments to details of the tasks were made on the basis of this.

Each extended task then went through two iterations with two different groups of three 11- year-olds, drawn from two state primary schools in England (four groups in all). Between the iterations the researchers reviewed and discussed the video recordings of the pupils' activity and agreed on adjustments to both the practical details of the task and the

² Both tasks were used with students in the UK and in Israel, but we report here only on the UK trials.

use of TP2, and the ways in which the researcher might present aspects of the task. The input of the researcher leading the task was not ‘scripted’ but allowed for flexible responses to the pupils’ activity.

The schools in which the research took place were selected on the basis of personal contacts with the class teachers, who were known to the first named author through professional development activities. One was a small village school, and the other on the outskirts of a large urban area. The pupils involved were chosen by their teachers in order to form groups who could collaborate and communicate well, and were considered to be of an average level of attainment for their age group. They had little or no experience of using TP2, but were familiar with conducting experiments in science in which they had to design ‘fair tests’, take measurements and record data (sometimes using a spreadsheet to do this). These pupils had no previous experience of modelling, and whilst they had some experience of reading graphs, probabilistic reasoning was a relatively small element of the mathematics curriculum.

2.1. THE *ANGRY EMU* TASK

The scenario for this task is a request to help Rovio (developer of the popular *Angry Birds* game) to plan a game with a new bird: an *Angry Emu*. As in the *Angry Birds* game, players use a slingshot to launch the bird, but *Angry Emu* cannot fly in the air and move only horizontally. The children are asked to prepare a data-based recommendation to help develop the *Angry Emu* game in a way that will resemble the real movement of a toy bird (seated on a toy car) launched from a sling made from a length of elastic stretched between chair legs. The activity structure includes four main stages.

1. The researcher introduces the task. The children engage in free play with the equipment. They experience how the bird physically moves, and begin to measure the stretch of the sling, and the distance travelled. They are encouraged to test each stretch a number of times, and to predict outcomes.
2. The researcher shows how to record and graph data in TP2. The children repeatedly generate graphs, make predictions and gather more data as necessary. They experience variation in outcomes, and begin to express ideas about signal and noise. In practice, stages 1 and 2 occupied about one-third of the time spent working on the task in both iterations.
3. The researcher demonstrates how to create a machine (model) using devices offered in TP2. The children create a machine which has dependency links between two devices representing the stretch and the distance travelled. They use this to generate data and produce graphs.
4. The children compare the data generated by their model with their initial experimental data. They predict outcomes for other values of the stretch which they have not explored experimentally, and build this into the model.

The *Angry Emu* task is similar to an Active Graphing task insofar as the children engage with a partially determined phenomenon (distance travelled by the *Angry Emu*) by collecting data. We encouraged children to explore repeated measures by using the same extension of the spring more than once. We wanted to understand how children would account for the *Angry Emu* travelling different distances when the extension was invariant. Would they discuss this in terms of causal factors or in terms of random effects (noise as unexplained variation from a ‘true’ distance)? In comparison with the tasks used in earlier AG research, where the purpose was to make a particular product, signal is not prioritised. The modelling aspect of the tasks requires the children to attend to and express both signal and noise.

2.2.THE 101 DALMATIANS TASK

The scenario is to imagine that a theme park wishes to create a scene representing 101 Dalmatians. The Dalmatians need to be different sizes (as in the original story), and look realistic. The task for the children is to make a ‘Dalmatian factory’ which will provide the data from which 101 model dogs could be made. A sheet containing some data (Table 1) about five Dalmatians, and illustrations to show how the measurements are taken (Figure 1), is provided, and the table of data is also provided in TP2.

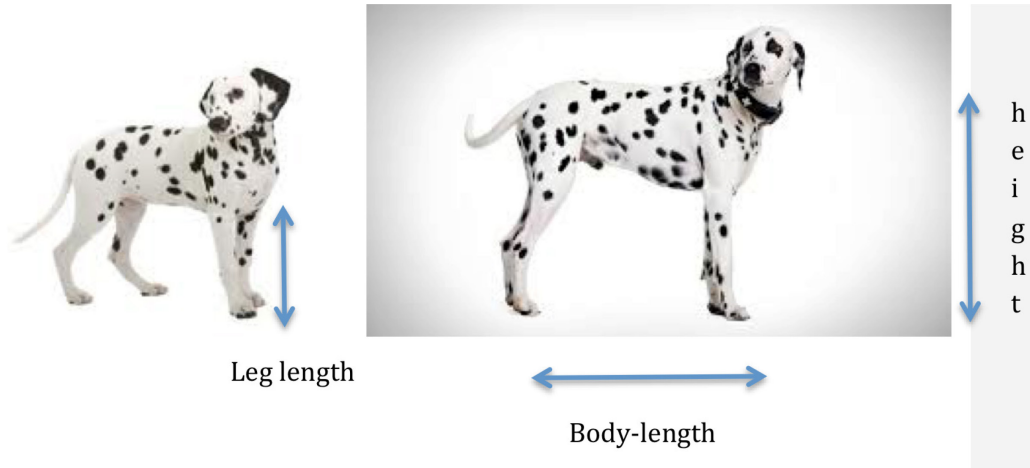


Figure 1. Illustrations to show how the measurements were taken

Table 1. Initial data: Five Dalmatians

spots	eyes	height	tail length	body length	leg length
brown	blue	41	23	40	22
black	blue	37	23	37	18
black	brown	26	13	27	14
black	amber	30	19	30	16
black	amber	30	15	31	17

The data focus mainly on quantitative variables but the colours of the Dalmatians’ spots and eyes were included as simple access points to TP2 devices for the children. In the second iteration of the task, the data for eye colour were removed as this added time to the process of building the model without contributing to the conceptual challenge of the task.

The data set is carefully designed to incorporate the following approximate relationships: (i) body length equals height at shoulder; (ii) leg length is between half and two-thirds of height at shoulder; (iii) tail length is marginally more than half of the body length. Examples are also included where two dogs have the same height or the same tail length but differ in other variables. The planned activity structure, which draws largely on an EDA approach, has four stages.

1. The researcher introduces the task and encourages children to explore the initial data using graphs to compare variables, and to articulate any relationships they observe. In practice this stage took up about one quarter of the time in each iteration.

2. The researcher uses the data on the colours of spots to show how to create a machine and allows children to explore the effects of different devices. The children add devices for height and one other variable (body, leg or tail length) to the machine. When data are generated and plotted the researcher encourages discussion about whether the data are realistic, with the aim that pupils will notice that the machine generates nonsense dogs because no relationship between variables has been built into the machine.
3. The researcher shows how to build dependency links. The children use this method to express their ideas about the relationship between the variables. In doing so they are likely to recognize the need to include some variation, so the researcher demonstrates how to input a range of values. The children use this facility to express their ideas about noise alongside signal.
4. The children use the model to generate larger sets of data (up to 101) and make graphs. They discuss the appropriateness of the data, both in terms of producing realistic dogs and in relation to the initial data. They may then go on to modify their machine accordingly.

We wondered how children might articulate the fact that dogs with the same height or tail length might have differing body and leg lengths. Would the children offer causal explanations or account for the variation in terms of signal and noise? Even Konold and Pollatsek, who adopt the terms signal and noise quite broadly, would regard the sources of variation in this context as obscure, especially as the children would not be comparing different distributions. However, we wondered whether envisaging the creation of dogs through a machine, as in a model, might lead to expressions of signal and noise.

2.3.DATA COLLECTION AND ANALYSIS

Each data collection session was conducted in a school environment, but in an area separated from the main classroom. Both researchers were present, with one acting as teacher/researcher. Camtasia was used to capture video and audio recordings of the children's on-screen activity; this was supplemented by field notes and some video recordings of activity away from the computer, giving rich data sets which allow for detailed analysis. The session with each group lasted around 90 minutes, though the timing was largely determined by the pupils' levels of energy and concentration.

The recordings and supporting data were initially reviewed by both researchers, to identify relevant sections for transcription. Recordings and transcripts were then used together to repeatedly review the pupil's activity, initially focussing separately on the work of each group of children. Themes emerged through discussion amongst the research team, and were then subjected to comparison across iterations and tasks, informed by the reflections which had guided the re-design of tasks. The first phase of analysis, focusing mainly on probabilistic reasoning and expressions of uncertainty (Ainley & Pratt, 2014), involved the whole team of UK and Israeli researchers. Further analysis by the UK researchers has used progressive focussing to explore children's expressions of signal and noise and how these are supported or obstructed by PCM.

3. FINDINGS

We present extracts of data, drawn from all four groups of children, covering two iterations of each task, to sketch the children's developing expressions of signal and noise. These are presented to reflect two stages of the children's activity: their exploration of the

experimental or given data (which for brevity we shall refer to as ‘the original data’), and then the building and testing of the model.

3.1. EXPRESSIONS OF SIGNAL AND NOISE WHILE EXPLORING THE ORIGINAL DATA

In this phase of activity an important element of the task design was that children were encouraged to articulate relationships between variables and to make predictions. Encouragement to repeat experiments with the same stretch in Angry Emus, and the carefully designed initial data set in Dalmatians, were pedagogic devices aimed to focus children’s attention on a level of variation within the patterns that they saw and expected in the original data.

In the early stages of both tasks the children looked at and discussed tables of data and a number of similar graphs that they generated from those data. Despite their lack of experience with using the software, the children generally found reading the graphs unproblematic. They found it easy to associate points with particular dogs, or with physical experiments, and seemed to experience no problems with different scales being produced on the axes. The ease with which they worked with the graphs was evidenced by many examples where children looked at the graph but talked about data or the experiment, a phenomenon, discussed elsewhere as ‘fusion’ (Nemirovsky & Monk, 2000) or ‘transparency’ (Meira, 1998). There is also evidence in the video episodes of children interpreting the data in more sophisticated ways, explicitly recognising and using the overall shape of the data to make interpolations or extrapolations and to identify inconsistencies. We see this as evidence of the children’s expectation and recognition of a signal in the data, based on previous experiences. This is broadly consistent with our earlier experience when studying Active Graphing (Ainley et al, 2001).

Two examples give an indication of how children were able to articulate some relationships between variables, and some apparent inconsistencies, based on their study of the graphs.

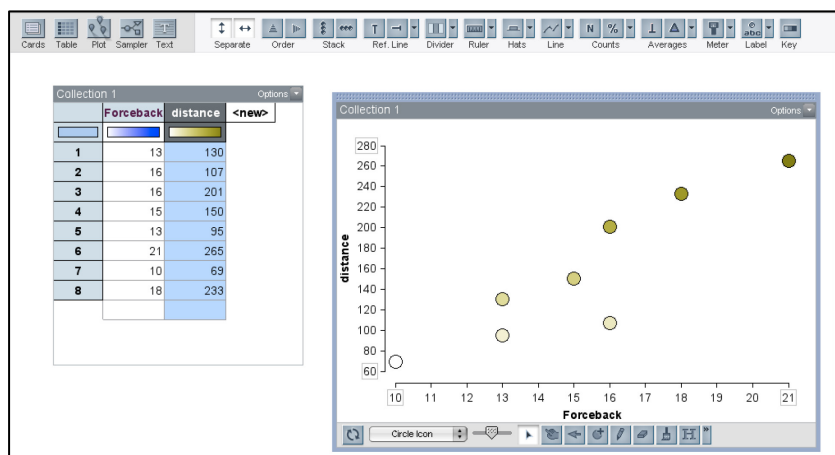


Figure 2. Eight pieces of experimental data for Angry Emus

Figure 2 shows the screen after the children had collected eight pieces of data in Angry Emus. In the discussion below Bella was initially arguing that there was no pattern in the results because of two different outcomes when the stretch (‘forceback’) was 16. However Bob and Dan seemed to be more confident about a pattern, and had a causal explanation for the inconsistency of one point, based on the behaviour of the toy car

carrying the Emu. Although they did not refer explicitly to the overall pattern of the graph, Bob focussed his explanation on the lower value, which he saw as out of line, tacitly assuming that the higher value, which was more in line with others, was more accurate. In the context of Active Graphing, we have identified this behaviour of identifying points which appear not to fit the overall trend, and offering explanations or ways of adjusting them, as *normalising* (Ainley et al, 2001).

- 1 Bella: It was a hundred and seven the first time, the second time we pulled it two hundred and one – and it was at sixteen like in a row
- 2 Bob: But I don't think that hundred and seven was right though because it peered off right over there didn't it?
- 3 Dan: Yeah
- 4 Bella: But it might be to do with
- 5 Dan: There was more turning though
- 6 Bella: It might be to do with the fact that it turned
- 7 Dan: Yeah that's what it is probably

Recognising and articulating relationships in the original data for the Dalmatians task was perhaps more challenging than in Angry Emus, as the children did not have hands-on experience of collecting the data. Figure 3 shows the screen as the children were exploring the Dalmatians data. They had made a graph comparing body length and tail length, and the researcher had asked whether they would expect dogs with longer bodies to have longer tails. While Frankie focussed on the overall pattern in the graph, and expressed the signal in the relationships between the attributes, Emma spotted what she saw as a problem. Like Bob in the previous example, Emma appeared to be implicitly using the overall shape of the graph, to identify a point that appeared to be out of line.

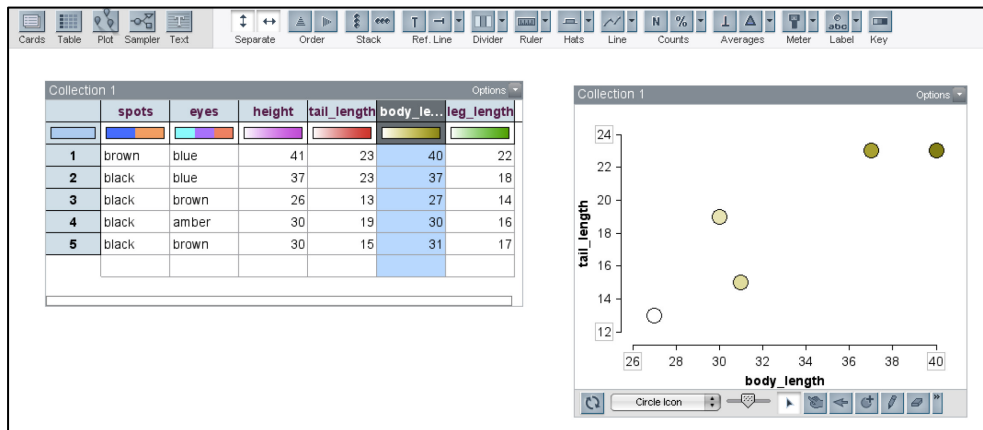


Figure 3. Comparing body length and tail length in the given data

- 8 Frankie: It looks like it does on that graph – because you can see that the dogs that are like the longest ones have the longest tails as well
- 9 Emma: Yeah there's only one though that doesn't do that
- 10 Researcher: Which one do you think doesn't agree?
- 11 Emma: That one [indicating the point at body length 31]
- 12 Researcher: What's wrong with that?
- 13 Emma: So that's 31 – and then 15 but then that one's 30 and that's bigger than 15 – the one above it
- 14 Researcher: Ah right so– what could you say about this dog then here?

- 15 Emma: That it's 15 – it's 31 centimetres long but then it only has a 15 centimetre tail – whereas the one above it has a 19 centimetre tail and it's only 30 centimetres
- 16 Researcher: Ok so that one doesn't quite fit what you expected to happen?
- 17 Emma: Yeah. But the rest do

In the Angry Emus task, children were prompted to make predictions during the collection of experimental data, and later to use the display of graphical and tabulated data to interpolate outcomes for other values of stretch. This focused attention both on identifying a signal in the data, and, as discussed in previous examples, on discrepancies (noise). In the example below the children were using the graph as the basis for interpolation, although on other occasions some showed a preference for working from the tabulated data. It is worth noting that these children had never been explicitly taught to use graphs in this way.

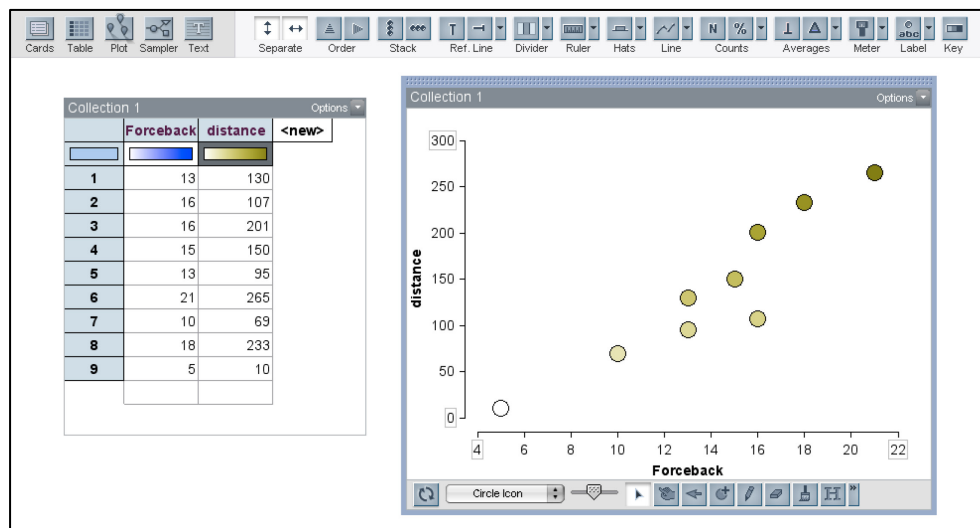


Figure 4. Nine pieces of experimental data for Angry Emus

Figure 4 shows the screen display shortly after the incident described above relating to Figure 2 (transcript lines 1-7).

- 18 Researcher: Ok – If I did ... let's say we did one that you haven't yet done – seven if I pulled it back seven centimetres. What would be your guesses to about how far it would go?
- 19 Bob: That would be ...fifty?
- 20 Bella: About seventy-five I would say
- 21 Dan: Seventy-five's a bit far ...
- 22 Bob: No because that's ten centimetres that's about seventy-five so seven centimetres would be around – sixty
- 23 Dan: I'd estimate round about twenty-five
- 24 Bella: Sixty fifty
- 25 Bob: I'd estimate around I'd estimate it around fifty sixty area

Although the distance which Bob and Bella appeared to agree upon was a little high, given the other data points on the graph, and Dan failed to get his contribution taken seriously, their words and actions, including pointing and referring to the graph, showed a

level of confidence in using the graph as a tool for predicting results of imagined experiments.

In Angry Emus, the children's initial accounts of noise within the signal were generally based on descriptions of the movement of the Emu in the toy car. The following examples come from discussion when the children had tested a stretch ('springy') of 35 cm three times, obtaining distances of 74 cm, 81 cm and 86 cm.

- 26 Researcher: So how would you say it moves when you pull the springy to 35?
 27 Carl: Out of control.
 28 Lily: Yeah he goes out of control sometimes it works and sometimes he does backflips and sometimes he just does spins
 29 Carl: He does spins and goes sideways and like just crashes into things.

When pressed to give a numerical prediction, the children offered single values, although they seemed not to feel any need to agree on a value, and began to qualify their responses with 'about':

- 30 Researcher: If you have to say a number what are you going to say?
 31 Carl: It depends what it does if it flips it'll probably go about 52
 32 Lily: I think it'll go about 85
 33 Karen: I say 82

When asked to explain what they mean by 'about', in the context of prediction for a different value of stretch, the children eventually began to offer a range of values.

- 34 Researcher: [*if the stretch is 10*] What number are you going to say for the distance?
 35 Lily: I think about 20
 36 Researcher: About 20? So not always 20?
 37 Lily: No
 38 Karen: No
 39 Carl: A bit over
 40 Researcher: How are you going to explain to them the 'about'?
 41 Lily: Somewhere between like
 42 Carl: 20 and 30 or ...

This shift from quantifying predictions as single values, to verbal expressions of variation, and finally to ranges of values can be traced on many occasions in the children's discussions as they became more confident in expressing uncertainty and articulating some sense of the partially determined nature of the variation. In the example below the researcher had again asked for a prediction but this time about a value of stretch that the children had not yet tried.

- 43 Bob: I'd estimate around I'd estimate it around fifty sixty area
 44 Researcher: Ok ok and you're saying estimate so do you think it will definitely go that distance?
 45 Dan: No – because it could go we could be looking at it completely wrong and it could
 46 Bella: It could barely go anywhere
 ...
 47 Researcher:: Ok so you feel you can't be sure about how far it's going to go?
 48 Dan: No but
 49 Bob: But you could ...

- 50 Researcher: But is it completely unpredictable?
 51 Bob: No you can have a strong you can have a really strong guess at it
 52 Bob: Because round about twenty [*for the stretch*] though you would expect
 them [*the distance travelled*] to be round about two hundred

The notion of being able to ‘have a really strong guess’ at a prediction seems to us to be a good way of expressing the quality of partial determination.

3.2.EXPRESSIONS OF SIGNAL AND NOISE WHILE BUILDING AND TESTING THE MODEL

Once the children had become familiar with the original data in either task the researcher introduced them to the modelling tools within the software, with the objective of building a ‘machine’ which could generate quantities of data to either simulate the movements of Angry Emus, or provide dimensions for making 101 Dalmatians.

The activity of building the machine was rather different in the two tasks. For Angry Emus, the model was relatively simple as it involved only two attributes. Further, the dependency between the variables was very clear in the structure of the task: the children had had practical experience of how different values of the stretch affected the distance that the Emu traveled. Transferring this to the structure of the model, each value of the stretch that was chosen needed to be linked to a separate device for generating the distance travelled.

The structure of the ‘Dalmatian Machine’ was more complex, since there were more attributes involved. The first attributes, spot and eye colour, were qualitative, and did not have dependency links to the other attributes. Although children had explored relationships between the quantitative attributes in the first phase of the task, it was less obvious than with Emus that the machine needed to contain dependency links between the various measurements. Indeed, it was part of the design of this task that the researcher would not suggest the need for such links unless the children were to mention it, as interpreting the data generated by a machine without such links would be a valuable experience of the utility of graphs for evaluating the model. In practice, time did not allow the children to make complete ‘Dalmatian Machines’ including all the attributes.

When the children began to build models, interacting with the software had the effect of establishing the idea of a range of values to express partially determined outcomes more securely. In building their machine for Angry Emus Lily’s group was guided by the researcher to use a counter to generate the values of stretch, as this ensured that values would be repeated several times, mirroring their physical experiment. They then chose mixers to generate the values for distance, and realised that the ‘range’ button allowed them to create the mixer contents more quickly than adding balls individually (see Figure 5). In effect, this provided a template for the construction of this part of the model. While creating the device to generate values of distance when the stretch was 20, Lily commented: ‘in here that’s where we put the something to something’. The group then engaged in an extended discussion of what the range of values should be, referring back to the values they had recorded during their experiment. Lily explained their rationale for eventually choosing 40 – 75; ‘43 is like over 40 so if we put it at a 10s number and round it that’ll make it easier and 75 because 80 would be a bit too much so we can round it to 75’.

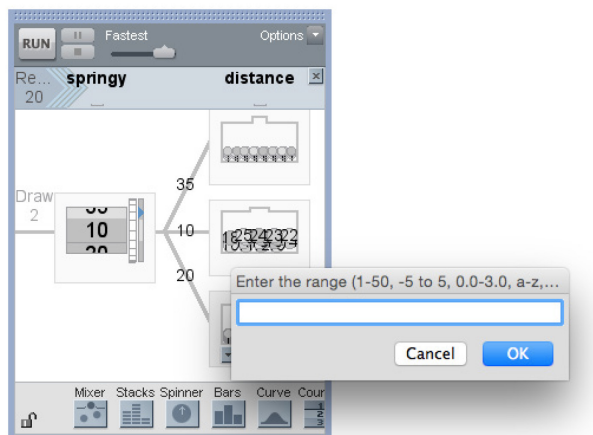


Figure 5. The screen showing the range tool

The experiences that children gained of using graphs in their explorations of the original data, particularly identifying points which appeared not to fit the signal in the data, supported their confidence in looking at the graphs of data generated by the models.

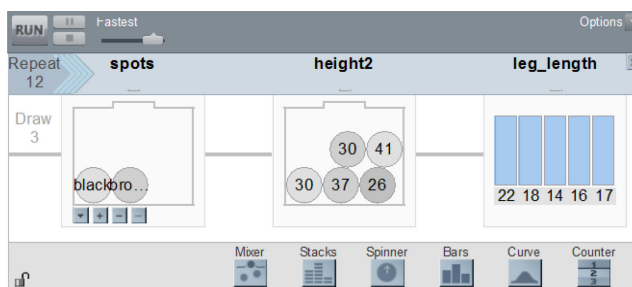


Figure 6. The first 'Dalmatian machine'

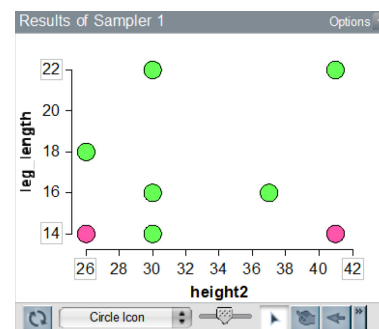


Figure 7. Graph of data generated by the machine

Figures 6 and 7 show a first attempt at building a model to generate data about the Dalmatians. The extract of dialogue below begins before the machine had been run and the graph produced. When the researcher asked for their predictions, all three children struggled to articulate the relationship they expected between height and leg length. Although they used terms like 'equal' and 'the same', it emerged in Emma's comment that these were not quantitative judgments, but rather qualitative descriptions about how they expected the attributes to relate. They clearly anticipated a signal in the data to be reflected in the appearance of the graph, and this prompted Emma's immediate response when the graph appears.

- 53 Researcher: So what would you expect to happen with height and leg length?
 54 Emma: That the –
 55 Billy: They'd kind of be equal
 56 Frankie: The higher they are –
 57 Emma: I think they'll be around the same because you wouldn't have a big dog with small legs but then you wouldn't have a small dog with big legs
 58 Researcher: Is that what you meant by equal?
 59 Billy: Yeah

- 60 Researcher: Yeah – ok so shall we see if we’ve got that? [*They run the machine and produce the graph*]
 61 Emma: Oh no that one’s a bit different. It’s got a big dog with small legs that’s a bit weird
 62 Researcher: Where’s the big dog with small legs? Can you point with the mouse?
 63 Emma: There isn’t it [*the lowest dot for height 41*] A very big dog with very small legs
 64 Researcher: Ah what about that green one up there? [*the highest dot for height 30*]
 65 Frankie: That’s quite a small dog with very big legs

3.3.SUMMARY OF THE CHILDREN’S EXPRESSIONS OF SIGNAL AND NOISE

In this section we summarise key points arising from the data in relation to the children’s expressions of signal and noise, in the context of the principles underlying the design of purposeful computational modelling tasks.

The meaningful contexts for both tasks, and the use of graphs for exploring the data, supported the children’s recognition of signal within the data. This was expressed in two related activities. Children implicitly used the signal within the overall shape of the graphs to *identify anomalies* in the original data (lines 9 – 15). The central role of the model within each task focused attention on the accuracy and completeness of the models they created. The expectation of seeing a signal triggered further identification of anomalies in the model data, for example in lines 61 – 65.

Children’s developing understandings of signal were also expressed through their use of interpolation and extrapolation to *predict values* for imagined data (lines 34 – 42).

Pedagogic devices to focus attention on noise were built into the task designs, in the form of encouragement to conduct repeated trials in Angry Emus, and the structure of the data set given in Dalmatians. In response to requests for predictions, children moved from specific values to verbal expressions of uncertainty such as ‘around’, to offering ranges of values. The need to communicate with the software environment in order to build a model secured the use of a range of values in expressing noise.

3.4.MODELLING AS A COMPLEX ACTIVITY

The summary above has outlined how purposeful computational modelling can support young children’s expressions of noise as well as signal. However, we observed some contra-indications where the modelling setting seemed to place obstacles in front of the children.

Throughout this phase of the tasks the children were engaged not only in interpreting graphs of the data generated by the models they had built, but also in comparing them with the original data and corresponding graphs, creating at times a confusingly complex visual display.

The process proved to be challenging for several reasons. The first is that the models could generate larger samples of data than the children had worked with in the first phase of the tasks, and the resulting graphs therefore looked very different from those of the original data. This was particularly noticeable in Dalmatians, as in the example in Figure 8. The children’s difficulties in making sense of the graphs of the model data were increased because of the large (by default) overlapping dots (the size of which could be adjusted). Furthermore, the children’s use of the range tool to express noise in effect modelled noise around each height as a uniform distribution, creating unrealistically large spreads for the error around each height. Thinking about the nature of distributions would be an obvious next step for these children.

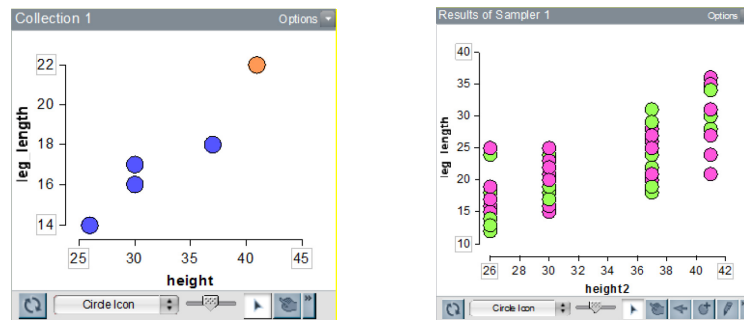


Figure 8. Graphs of the original data (left) and model data (right)

A further issue which emerged in the children's attempts to compare graphs of the original and model data relates to their understandings about the nature of the model and the data it generated.

Although the children in all four groups were able to describe what the model was doing to produce the data, there were instances when it emerged that they did not have a secure understanding of the nature of the model data. We see this as relating to the potential for different kinds of criteria that might be used to judge the success of a model. There seem to be at least three possibilities:

1. Has the model worked in terms of doing what the children intended? This might be evaluated by comparing the output to the structure of the model. Because the children were unfamiliar with the software they needed to reassure themselves about this.
2. Has the model worked in terms of producing data which is like the original data? This might involve comparing the output from the model with the original table/graph of data. This was easier when the children took smaller samples from the model because of the difficulties described above.
3. Has the model worked in terms of producing realistic data? This applies more in Dalmatians, where the initial models produced 'impossible dogs'.

There were small hints within the children's discussion of the model data that they were thinking about this as though it was another source of 'real' data. An example of this confusion about the nature of the model data is given in the extract below. After a first attempt at building a Dalmatian machine, which produced impossible dogs, the children added dependency links as shown in Figure 9. In the process of adding these links, the children have explicitly used a relationship that Adam had identified in the original data, that is that the leg length is around half of the dog's height.

- 66 Researcher: If it [*the device for height*] chooses 42, so Lynne what numbers
do you think we should put in there? [*the mixer for leg length*]
67 Lynne: I think from like kind of like from 20 - no
68 Molly: What's half of it?
69 Adam: 21
70 Lynne: So from 20 up to about 25

They then ran the model, producing the graph shown in Figure 8, and Lynne commented on what happened.

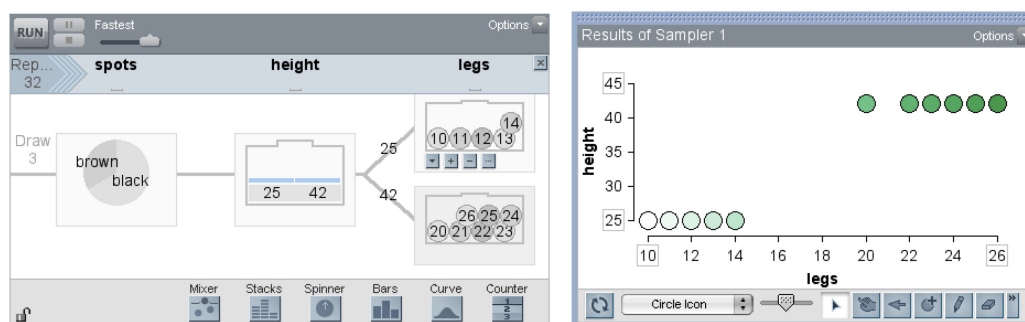


Figure 9. Revised Dalmatian machine and graph of data generated by the machine

- 71 Researcher: Ok let's run that and see what happens
 72 Lynne: It's going - terribly wrong - oh 25 and 10 - it's not that bad actually
 73 Researcher: So why is it not that bad? Say what's worked about it Lynne
 74 Lynne: It's because like up - they're up there they are about 42 or 43 and then their legs are about half of that almost so half of their height - their height will be 45 and half of that would be the leg length so like what Adam was saying

Lynne's first comment that it was going 'terribly wrong' seemed to be about the unexpected appearance of the graph. In the comments which follow she appeared to talk about the 'rightness' of the data as though they were independent of the instructions she had given when building the model, even reading the value of '42 or 43' from the graph, when 42 was clear in the model. She seemed to need to check that that the values in the graph did follow Adam's rule which she had used to choose the range of values in the mixer.

In the original design of both tasks it had been our intention that the 'machine' should generate data that resembled the original data as closely as possible, and that part of the children's activity would be to modify their machines to achieve this in response to comparisons between the generated data and the original data. However, this did not happen as we had planned for (at least) three reasons. A major factor was time constraints, as we only had one session with each group of children, and understandably they became tired after a period of intense concentration. The appearance of the graphs of larger quantities of data also proved to be a considerable barrier to children gaining feedback about the effectiveness of their machines, which we had not fully anticipated. Finally, in order to engage children's attention on the need to modify the machine there was probably a need for a clearer requirement within the overall scenario of the task.

4. DISCUSSION

In this study we set out to design a pedagogic approach which would support learners to express meanings for signal and noise while combining strengths of both Active Graphing and Exploratory Data Analysis. In the previous section we have offered an extended description of the development of children's expressions of signal and noise in response to our main research question. The approach that we have developed, *Purposeful Computational Modelling (PCM)*, has been shown to be effective in that it has allowed us to look closely at the role that modelling plays in supporting probabilistic reasoning. In this section we discuss some broader issues raised by the study.

4.1. EXPLOITING THE STRENGTHS OF AG AND EDA

The PCM approach encompasses both observational and experimental data. Within the findings we have presented there is evidence of the children working confidently to identify and examine relationships and ‘stories’ in the data, for example discussion of the ‘impossible dogs’ (lines 61 – 65). This activity was supported by their fluent use of graphs as tools during the data collection and analysis, which allowed them both to interpolate and extrapolate to extend the range of data they could consider, and to identify anomalies. The richness of the children’s activity across observational and experimental data, and the speed with which they were able to tackle new ideas, indicates a level of success in combining the two original approaches.

4.2. THE ROLE OF MODELLING

Building the model, which was a central feature of both tasks, created the need for children to express their ideas of noise more precisely, and specifically to quantify the range of variation that they thought appropriate. In contrast to conventional situations where noise is a nuisance in terms of the purpose of the task, creating a model provides a situation in which it is necessary to consider both the signal and the noise. We have demonstrated the use of PCM in two contrasting tasks. On the one hand, Angry Emus is an example of a partially determined situation with a strong signal, which can be modelled by the effect of an independent variable that is controllable by the children. Discrepancies from that signal were accounted for both causally and through various types of expressions of noise. On the other hand, modelling in the Dalmatians task allowed the introduction of an artificial independent variable to a situation that really involved observational data. Relationships within the observational data transformed into a main effect of the artificial independent variable on another observed variable and discrepancies from that relationship were accounted for through expressions of noise.

We have commented in earlier studies how it is challenging to design tasks that are both seen as purposeful by children and offer utility for the target mathematical or statistical ideas (Ainley, Pratt & Hansen, 2006). This study was no exception. The Angry Emus and Dalmatian tasks drew on the idea that a model might be seen as a prescription or template to capture the description of a design. The template of course needed to be communicated to someone and so we invented the roles of Rovio and a new theme park. We believe the extent of the children’s engagement with the tasks provides evidence that the tasks were indeed seen as purposeful but it is not clear from the data whether the children gained a holistic sense of the utility of modelling as a template. In fact, we tended to refer to the model as a machine and so we did not reinforce the idea of a template. The complexity of modelling discussed above may also have obscured that message.

4.3. CATEGORIES OF ‘NOISY PROCESSES’

We referred earlier to a categorisation by Konold and Pollatsek (2002) of three kinds of statistical processes in which the metaphor of signal and noise may be appropriate. The Dalmatians task clearly falls within their first category of measuring individuals. Through our focus on relationships between attributes of individual dogs, we created a situation which we see as analogous to Konold and Pollatsek’s discussion of the comparison of heights of men and women. They see gender as a partial determinant of height. As the children build a machine to generate dogs, they focus on a particular value of one attribute (e.g. body length) to model how it partially determines another attribute (e.g. tail length).

As a result, PCM seems to facilitate the signal and noise metaphor even in the case of measuring individuals.

Even so, we note a difference from the situation they describe in that our focus is on relationships between the attributes of an individual. An effect of this difference is that, rather than searching for signal within a noisy process, the children were expecting a signal (tall dogs will have long tails), and were surprised by encountering the noise built into the data set. Konold and Pollatsek suggested that the comparison of groups might facilitate a signal/noise metaphor when measuring individuals. We agree but we would add that, when more than one measurement is available, a signal/noise metaphor might be facilitated by looking for relationships in bivariate data.

The Angry Emus task, however, does not seem to fit exactly within any of Konold and Pollatsek's categories. It does involve making repeated measurements, but these are of different trials under the same conditions, rather than of the same object. In the case of repeated measures of an object the signal is the 'true' measure and the noise is the variation around the signal caused by other sources. In Angry Emus, the signal is the 'true' distance travelled for a given extension, and the noise is the variation caused by other sources.

Konold and Pollatsek discuss the advantages of students conducting experiments. Whilst the situation of children conducting such 'fair test' experiments within statistics classes may be relatively unusual, it is much more common within science classes. Indeed the idea for Angry Emus was developed from an experiment which is common in primary school science. We would therefore extend Konold and Pollatsek's category of repeated measures to include exploring situations where it is possible to repeat the value of an independent variable in order to measure a dependent variable.

4.4.FINAL COMMENTS

We have set out a pedagogic approach based on AG and EDA in which young children engage with modelling using the computational features of TinkerPlots. We have intentionally embedded that approach in tasks which proved to be engaging for the children, thus setting out a pedagogic approach which we have called PCM. The study reported here was exploratory in nature, using a relatively small number of clinical trials, which allowed attention in detail to how those children expressed ideas about signal and noise within a PCM pedagogic approach. We have commented above on the difficulties that modelling presented to the children as well as how it facilitated articulations of signal and noise. We are struck by how our tasks seemed to lead to the children having a sense of the utility of modelling to both generate and communicate ideas. The latter utility embraces not only specific ideas about signal and noise (such as relationships within the data and discrepancies from those relationships) but also a wider narrative pertaining to the stories about how emus move and dalmations are 'built'. A model as a machine and as a means of communication seems to us to hold some promise.

Building on this starting point we see a number of different directions in which further research might be developed. As mentioned during the discussion above, in these tasks the children expressed noise in their models by using a range of possible values generating uniform distributions. A challenge would therefore be to design new tasks (which could be extensions of the two we describe here) which could support the children's thinking about the possible shape of the distribution as well as the range of likely values. This involves thinking of the possible outcomes to be used in the model as values around the signal rather than as the range of values in the original data. We did see beginnings of this kind of thinking in some of the children's comments about the Dalmatian machine (see lines 66-70) and as they predicted the outcomes for values of stretch that they had not tried

experimentally in Emus (see lines 18-25). However we did not explore how these beginnings might be developed into thinking about the shape of distributions.

In order to develop PCM as a viable pedagogic approach for wider use at least two further issues require research. One is to apply the task design principles to develop sequences of tasks aligned to statistics curricula. The second is to look at how an approach that has been effective in a small group situation can be scaled up to whole class teaching without distortion of the design principles, and in ways that would be accessible to the whole attainment range.

ACKNOWLEDGEMENTS

The data presented here are part of the project *Children's interpretation of graphs through active statistical modelling* funded by the British Academy (SG112288). The tasks were developed with our collaborators in this project: Dr. Dani Ben-Zvi, Hana Manor and Keren Aridor from the University of Haifa.

REFERENCES

- Ainley, J. & Pratt, D. (2014). Expressions of uncertainty when variation is partially determined. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute. Downloaded on July 7th, 2015 from http://icots.info/9/proceedings/pdfs/ICOTS9_9A1_AINLEY.pdf
- Ainley, J., Pratt, D. and Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, 32(1), 23-38. doi: 10.1080/01411920500401971
- Ainley, J., Pratt, D., & Nardi, E. (2001). Normalising: Children's activity to construct meanings for trend. *Educational Studies in Mathematics*, 45, 131-146. doi: 10.1023/A:1013822512833
- Ben-Zvi, D., Aridor, K., Makar, K., Bakker, A. (2012) Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM: The International Journal on Mathematics Education*, 44(7), 913–925. doi: 10.1007/s11858-012-0420-3
- Cobb, G. (2005). Foreword. In J. Garfield (Ed.), *Innovations in teaching statistics* (MAA Notes number 65), p. vii. Washington, D.C.: Mathematical Association of America.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R. & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9-13. doi: 10.3102/0013189X032001009
- Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5), 382-393.
- Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, 104(2), 301-318.
- Hammerman, J. K. & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal*, 3(2), 17-41. [http://iase-web.org/documents/SERJ/SERJ3\(2\)_Hammerman_Rubin.pdf](http://iase-web.org/documents/SERJ/SERJ3(2)_Hammerman_Rubin.pdf)
- Jones, G. A. (Ed.) (2005). *Exploring probability in school: Challenges for teaching and learning*. Dordrecht: Kluwer Publishers.

- Konold, C., Harradine, A. & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning*, 12(3), 217-230. doi: 10.1007/s10758-007-9123-1
- Konold, C. & Lehrer, R. (2008). Technology and mathematics education: An essay in honor of Jim Kaput. In L. D. English (Ed.), *Handbook of International Research in Mathematics Education* (2nd ed., pp.49-72. New York: Routledge.
- Konold, C. & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259-289.
- Lehrer, R. & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, 41(3), 630-679. doi: 10.3102/00028312041003635
- Makar, K. (2014). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics*, 86(1), 61-78. doi: 10.1007/s10649-013-9526-y
- Makar, K. & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82-105. [http://iase-web.org/documents/SERJ/SERJ8\(1\)_Makar_Rubin.pdf](http://iase-web.org/documents/SERJ/SERJ8(1)_Makar_Rubin.pdf)
- Meira, L. (1998). 'Making sense of instructional devices: The emergence of transparency in mathematical activity. *Journal for Research in Mathematics Education*, 29(2), 121-142.
- Monteiro, C., & Ainley, J. (2004). Exploring the complexity of the interpretation of media graphs. *Research in Mathematics Education*, 6, 115-128. doi: 10.1080/14794800008520133
- Nemirovsky, R. & Monk, S. (2000). 'If you look at it the other way ...' An exploration into the nature of symbolizing. In P. Cobb, E. Yackel & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools and instructional design*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education*, 2(3), 149-167.
- Royal Statistical Society (2014). *The data manifesto*. Downloaded on May 4th, 2016 from <http://www.rss.org.uk/Images/PDF/influencing-change/rss-data-manifesto-2014.pdf>
- Shaughnessy, J. M. (2003). Research on students' understandings of probability. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A Research Companion to the Principles and Standards for School Mathematics*, (pp. 216-226). Reston, VA: NCTM.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58. [http://iase-web.org/documents/SERJ/SERJ7\(2\)_Zieffler.pdf](http://iase-web.org/documents/SERJ/SERJ7(2)_Zieffler.pdf)

JANET AINLEY
 School of Education, University of Leicester
 21 University Road, Leicester LE1 7RF
 UK