IT'S NOT WHAT YOU KNOW, IT'S RECOGNISING THE POWER OF WHAT YOU KNOW: ASSESSING UNDERSTANDING OF UTILITY

Janet Ainley[1] and Dave Pratt[2]
[1]School of Education, University of Leicester, United Kingdom
[2]University of London Institute of Education, United Kingdom
janet.ainley@le.ac.uk

Traditional approaches to assessing 'understanding' in mathematics and statistics education tend to focus on the two strands of procedural competence and conceptual knowledge. We take as our starting point the idea that this does not fully capture what it is to understand mathematical and statistical ideas, and suggest a third dimension of understanding which we call utility; that is, knowing why, when and how a particular idea can be used and the power which it offers. We suggest that this is a key feature of statistical literacy, without which knowledge of statistical ideas cannot be effectively applied. In this paper we draw on examples from our current and past research to explore how the assessment of understanding of utility may be approached.

INTRODUCTION

In this paper, we reflect on our previous studies of statistical thinking to argue that commonly proposed models of assessment may not adequately embrace an aspect of statistical activity that we see as fundamentally important. In considering the assessment of statistical thinking, we suggest that it is important to analyse what aspects need to be assessed. To give a flavour of work that has been done in this field we focus on two developments.

Substantial work has been conducted which focuses on trying to build cognitive models for statistical thinking or reasoning. Work in Australia has built frameworks linked to the SOLO model (Biggs & Collis, 1991), a general model of intellectual development that categorises learning into five modes of functioning, each containing cycles of response, reflecting increasing sophistication. The Australian initiatives, for example,

- validated frameworks that described young people's statistical reasoning across a number of constructs (Jones, Langrall, Thornton & Mogill, 1997; Jones et al., 2000), and
- proposed that students' conceptions of probability increased by an average of one SOLO level over 4 years (Watson, Collis, & Moritz, 1997; Watson & Moritz, 1998, 2003).

SOLO-inspired approaches to the task of identifying components of statistical reasoning inevitably produce a Piagetian-like hierarchy of developmental stages. Such frameworks have certain uses in the field of assessment. For example, the average progression of students' conceptions of probability might be of interest to a policy-maker. This observation leads us to an even more fundamental question than asking what is to be assessed; we must also identify who is the assessor and what is their purpose for making the assessment. The SOLO-based approaches to assessment are, in our view, best suited to a macro-perspective of development, where the performance of populations, perhaps over extended periods of time, is in focus.

In a separate initiative, Konold and Garfield have developed the *Statistical Reasoning Assessment* (SRA) tool (Garfield, 1998; Garfield, 2003), a multiple-choice test of 20 statistics and probability problems. Some items focus on calculation (for example, CC3 addresses the correct calculation of probabilities), some on understanding (CC4 assesses understanding of independence) and some on misconception (MC4 probes the common intuition that events of unequal chance tend to be viewed as equally likely). Such an approach is geared towards assessing individual performance, which coincides with our interest on activity within individual classrooms. We are interested in how teachers interact with particular students, attempting to identify the status of their understanding at a particular moment in time with a view to making an intervention in the moment.

In examining the SRA, we notice, in common with Solo-based approaches, a tendency to equate knowledge with representations of statistics. We use the term *representations* very broadly to refer not only to graphs and numerical measures but also the conceptual components of statistics.

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

In this sense, CC3, CC4 and MC4 all assume that knowledge of statistics is knowledge of how to calculate or represent statistical objects or conceptual artefacts.

An exception to this general observation about the SRA is item CC2, which refers to understanding of how to select an appropriate average. Here we begin to get a sense of statistics as an applied discipline, a human endeavour where knowledge of statistics is as much about having reasons for acting on or with data as it is about the logical reasoning basis for various techniques and representations. From a philosophical perspective, we tend to start with an inferential analysis of what it means to know (Brandom, 2002). From this perspective, we believe that CC2 hints at an aspect of understanding neither captured by macro-analyses in which the reasons for activity are regarded as noise amongst a general pattern of human development, nor emphasised in the SRA which is influenced by conventional representational ways of thinking about understanding.

Our own research has not been concerned overtly with assessment. Rather we have focused on the interface between pedagogic design and learning and teaching at classroom level. In reflecting on the challenge of designing tasks that support the learning and teaching of statistical thinking we have attempted to foreground and characterise an overlooked dimension of understanding which is concerned with how ideas are used, drawing on a range of theoretical models. In particular our concern has been to develop a framework which supports an approach to the design of tasks which both creates opportunities for learners to recognize the power of what they know, and also supports teachers in their observation of, and intervention in, children's activity. We are thus inevitably concerned with the ongoing formative assessment of individuals in which teachers engage on a daily basis, and which can so powerfully inform their interventions.

A THIRD DIMENSION OF UNDERSTANDING

It is generally accepted that constructing meaning for a mathematical or statistical idea involves many related elements which must be taken into account in assessing pupils' understanding; a distinction is often made between elements relating to procedural competence, and those concerned with conceptual or relational understanding. Like many researchers and practitioners, we still find Skemp's (1976) seminal ideas on instrumental and relational understanding powerful. However we argue that Skemp's model does not provide a complete framework for thinking about what it means to understand and use mathematical or statistical ideas. We propose a third dimension of understanding, which relates to recognising the power of what you know. We call this dimension the *utility* of an idea, to encapsulate why that idea is useful, how it can be used and what it can be used for. We argue that a rich understanding of a mathematical idea involves procedural, conceptual and *utility* elements (Ainley, Pratt & Hansen, 2006).

In our research on task design over a number of years we have used the notion of *utility* alongside the related idea of *purpose*. In our framework, a *purposeful* task is one which has a meaningful outcome for the learner in terms of an actual or virtual product, the solution of an engaging problem, or an argument or justification for a point of view. This feature of purpose for the learner, *within the classroom environment*, is a key principle informing the design of pedagogic tasks which offer opportunities to engage with the *utility* of ideas.

The purpose of a task, as perceived by the learner, may be quite distinct from any objectives identified by the teacher. The purpose creates the necessity for the learner to use mathematical or statistical thinking in order to make progress towards the satisfactory completion of the task. This progress provides feedback for the learner, rather than this being judged solely by the teacher. Because the mathematical or statistical ideas are being used in a purposeful way, pupils have the opportunity not just to understand concepts and procedures, but also to appreciate *utility*: how and why the mathematical or statistical idea is useful. This parallels closely the way in which ideas are learnt in out-of-school settings, where understanding the usefulness and power of what is being learnt is foregrounded. In contrast, within school ideas are frequently learnt in contexts where they are divorced from aspects of utility, which we believe often leads to significantly impoverished learning, tending to generate a sense of irrelevance and disconnection. A focus on instrumental and relational understanding inevitably starts with mathematics imagined as a collection of connected object-like representations. In contrast a focus on purpose and utility starts with mathematics imagined as human activity involving reasons for engagement.

OPPORTUNITIES FOR ASSESSING UNDERSTANDING OF UTILITY

Because of the constraints of space in this paper, we will focus here on examples from a single task, which was designed as part of a study based on a pedagogic approach which we call *Active Graphing*. The guiding principle here is the design of purposeful tasks in which children are engaged in collecting data about an experimental situation and use graphing as an analytic tool to inform decision-making as they work towards a final product or solution. *Active Graphing* involves an iterative process of collecting data and entering this in a spreadsheet, generating a graph, and using the analysis of this to decide how the next stage of data collection should be focussed. Our aim is to create contexts in which children have the opportunities to experience the *utility* of graphs as tools for the interpretation of data rather than their more common function in classrooms as a communication tools for displaying results.

The episodes described below occurred during a project in which classes of 8-9 year olds worked on a series of such tasks, using spreadsheets to record and explore data. The task we focus on here involves the design of paper spinners, known as *helicopters* by the children. Given a basic design (see figure 1), each group of children chose one feature (e.g., wing length) to vary in order to try to design a helicopter which would stay in the air for the longest time when dropped from a particular height (Ainley, Nardi & Pratt, 1998, 2000).

The children experimented with their helicopters, timing how long each one flew, and recording results on their spreadsheet. After collecting three or four sets of data, they were encouraged to draw a scatter graph of, say, wing length against time of flight, and look carefully at this for patterns in the helicopters' behaviour. On the basis of this discussion, they made decisions about how to proceed with their experiment. For example, if they had already tested helicopters with wing lengths 3cm, 6cm, 7cm, 8 cm, they might decide that they needed to try lengths between 3cm and 6 cm. If they got a sense that helicopters with longer wings were better, they might try even longer versions next.

Normalising

Within a large number of observations in *Helicopters* and other Active Graphing tasks we have identified an activity in which children spontaneously engage; we have called this activity *normalising* (Ainley, Pratt & Nardi, 2001). This involves recognising abnormalities in graphs during ongoing experiments and attempting to adjust and 'correct' the data and the graphs towards some perceived norm. This facility to 'correct' graphs, despite having limited technical knowledge about the structure of the graphs, was a pervasive feature of activity across different tasks and, we believe, provides robust evidence of children's developing understanding of the *utility* of graphing as an interpretive and analytic tool.

We offer one example of *normalising* from the activity of a group of 8-9 year old boys to illustrate this. From their initial explorations, Andreas, Bill and Simon had produced the graph shown in Figure 2. After discussion of what appeared to be a linear pattern in their wing-length data, they used a facility in the software to drop a line over the graph to articulate this. As they added further data points, the researcher asked about the pattern.

Res: What can you say about the pattern?
Bill: The longer wings stay longer in the air.
And: Apart from that one (*pointing to the cross representing a wing-length of 6.5 cm.*)

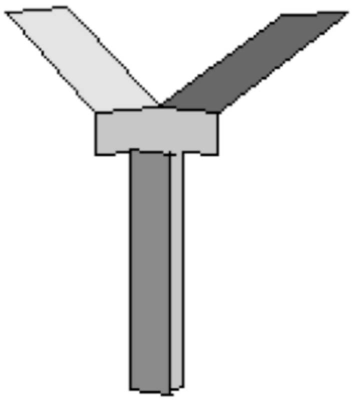

Figure 1. A paper helicopter

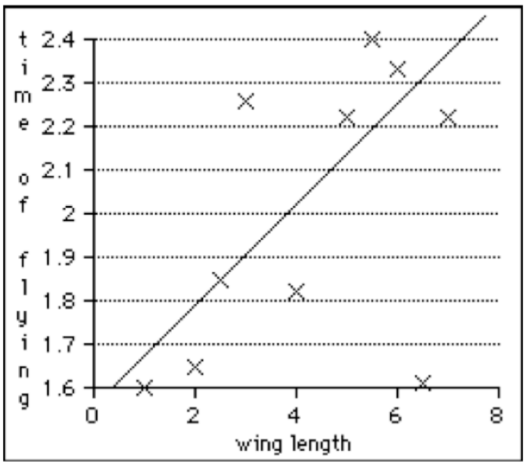

Figure 2. Andreas, Bill and Simon's graph, with a line added to indicate a possible trend

The boys discussed possible reasons for this apparent abnormality, and decided to re-test the helicopter with wing-length 6.5cm, feeling that there may have been some experimental error in their original attempt. Repeating the experiment did produce a result which matched more closely the boys' expectation of the 'proper' appearance of the graph.

We see normalising activity as evidence of children's developing understanding of the utility of graphing data to support their analysis of patterns in experimental data, and of the nature of that data. Importantly, when children engage in normalising, this provides opportunities for the teacher to intervene in a focused way to challenge, extend and further assess their thinking and to introduce new procedural or conceptual ideas. In the instance described here there were opportunities to challenge the children to give explanations for the anomaly in the graph, to re-visit and extend measuring skills, and to check on the children's understanding of decimal notation.

Using average values

The demands of measuring a range of quantities in all the experimental active graphing tasks meant that 'messy data' quickly emerged as a problem for the children. This created opportunities both to discuss ways of improving measuring skills, and to introduce the idea of repeating each experiment and then using an average of the results to create a 'cleaner' graph. Although the children had not yet been taught formally about calculating measures of average, they had some common sense notions of what 'average' meant, and they were shown how to use the *Average* function on the spreadsheet to obtain the mean of several results. Although they did not understand the procedures for finding the average, many children appeared to gain some understanding of its utility for producing better data and clearer graphs, and we saw some groups transfer this idea to work on new tasks, supported by using the spreadsheet's functionality.

We saw further, and perhaps more convincing, evidence of an understanding of this utility amongst one or two groups who apparently had not remembered the detail of how to use the spreadsheet function for the calculation, and devised their own approaches to producing an average measure. One group repeated each experiment in the helicopters task five times, and then used an invented method, which actually produced the median value, as follows. They looked at the five numbers, and crossed out the highest and the lowest. Then they looked at the remaining three, and did the same. (They also introduced a consistent rule for how to proceed if two values were equal). This left them with a single flight time for each wing-length, which they used to plot their graphs. Observing children's explanations of their development and use of this procedure offered opportunities for assessment of their understanding of the utility of a measure of average in a way which would be overlooked in an assessment task based on a hierarchical model, since their procedural and possibly their conceptual understanding in this area would be at a much lower level.

CONCLUSION

We have explained how our research on mathematical and statistical thinking at the micro-level has led to an awareness of the need to integrate utility into our articulation of what it means to understand mathematical and statistical ideas. We have illustrated the notion of understanding utility through two examples drawn from our research on active graphing. The emphasis on utility has been a feature of much of our work in areas other than graphing. For example, our research on young children's understanding of randomness led to the development of a software resource, ChanceMaker, which offered children the opportunity to construct utility for probability distribution (Pratt, 2000). In this study, students were challenged to mend *gadgets*, small virtual simulations of everyday random generators such as coins and dice. To mend the gadgets, the children needed to engage with a feature known as the *workings box*. This was an unconventional representation of distribution that contained the instruction to generate the random outcomes from the gadget. Utility for distribution emerged as the children began to realise that they could predict long-term behaviour of the gadget by inspection of the workings box without needing to generate and analyse results. The children's use of the workings box to generate results had led to a sense of how the workings box *represented* potential outcomes and their likelihoods, which we see as a situated understanding at the root of the utility of distribution.

We note in the ChanceMaker example how purposeful activity led to understanding the utility of the representation. Similarly, intent on finding the 'best' helicopter, the children became aware of the utility of scatter graphs and of averages; again meaningful human endeavour drove the construction of utility understanding. These cases provide a powerful illustration of the connection between an inferential philosophy that knowing is first and foremost associated with reasons for actions and that utility is a key aspect of understanding that can emerge from such activity. We approach task design through deep consideration of what tasks might lead to purposeful activity and look to optimise the approach so that utility of a key mathematical representation might ensue.

In contrast, when teachers and researchers deploy conventional approaches to teaching mathematical or statistical ideas, emphasis is often first placed on introducing representations. It is not difficult to imagine teachers beginning a topic on graphs by describing the many conventions involved in drawing a graph, or on averages by setting out the mechanism by which mode, median and mean are computed, or distribution by offering a definition. Indeed, such approaches are typically consistent with examples in textbooks and curricula. Such an approach is typical of conventional teaching of statistics more generally, where either instrumental or relational understanding is encouraged from the outset by an emphasis on algorithmic calculations and definitions.

The examples we have discussed demonstrate tasks which create opportunities for formative assessment of the understanding of utility. From our inferentialist position, we identify an important challenge to the teaching profession and mathematics/statistics education researchers to develop pedagogic approaches that emphasise understanding of utility by designing purposeful tasks. In fact, statistics educators can lead the way in this respect. Whereas mathematics educators often regard context as an obstacle that can obscure the pure mathematical ideas that lie at the heart of the problem (Cooper & Dunne, 2000; Boaler, 1993), statistics educators recognize that their world is indeed embedded in context. We conjecture therefore that it is easier for statistics than mathematics educators to embrace the inferentialist position and to recognize the power of understanding utility, though that is not to say that the issue is any less important for mathematics education.

As we outlined at the beginning of this paper, much of the focus in statistics education assessment is currently on hierarchical models of understanding, which we argue are best deployed for summative or evaluative assessment. The tasks we have presented have been concerned with the formative assessment of ongoing activity, as takes place every day in classrooms, rather than in examinations or policy-making. It is here, at the micro level that we expect to find evidence of understanding utility and suggest that future efforts to develop formative assessments of utility might be targeted.

REFERENCES

Ainley, J., Nardi, E., & Pratt, D. (1998). Graphing as a Computer Mediated Tool. In A. Olivier & K. Newstead (Eds.), *Proceedings of the Twenty Second Annual Conference of the International Group for the Psychology of Mathematics*, Stellenbosch, South Africa, Vol. 1, 243-258.

Ainley, J., Nardi, E., & Pratt, D. (2000). The construction of meanings for trend in Active Graphing. *International Journal of Computers for Mathematical Learning, 5*(2), 85-114.

Ainley J., Pratt, D., & Hansen, A. (2006). Connecting Engagement and Focus in Pedagogic Task Design, *British Educational Research Journal, 32*(1), 23-38.

Ainley, J., Pratt, D. & Nardi, E. (2001). Normalising: children's activity to construct meanings for trend, *Education Studies in Mathematics Special Issue: Constructing meanings from data, 45,* 131-146.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and intelligent behaviour. In H. Rowe (Ed.), *Intelligence, reconceptualization and measurement* (pp. 57-76). Hillsdale, NJ: Lawrence Erlbaum.

Boaler, J. (1993). The role of contexts in the Mathematics classroom, *For the Learning of Mathematics, 13*(2).

Brandom, R. (2002). Overcoming a Dualism of Concepts and Causes: A Unifying Thread in "Empiricism and the Philosophy of Mind". In R. M. Gale (Ed.), *Blackwell Guide to Metaphysics* (pp. 263-281). Oxford: Blackwell.

Cooper, C., & Dunne, M. (2000). *Assessing children's mathematical knowledge*. Buckingham: Open University Press.

Garfield, J. B. (1998). Challenges in Assessing Statistical Reasoning, *Annual Meeting of the American Education Research Association*, San Diego.

Garfield, J. B. (2003). Assessing Statistical Reasoning. *Statistics Education Research Journal, 2*(1), 22-38. Online: fehps.une.edu.au/F/s/curric/cReading/serj/current_issue/SERJ2(1).pdf.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32*, 101-125.

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A framework for characterizing children's statistical thinking, *Mathematical Thinking and Learning, 2*(4), 269-307.

Pratt, D. (2000). Making Sense of the Total of Two Dice, *Journal for Research in Mathematics Education, 31*(5), 602-625.

Skemp, R. R. (1977). Relational Understanding and Instrumental Understanding, *Mathematics Teaching, 77*, 20-26.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal*, 9, 60-82.

Watson, J. M., & Moritz, J. B. (1998). Longitudinal development of chance measurement. *Mathematics Education Research Journal, 10*, 103-127.

Watson, J. M., & Moritz, J. B. (2003). Fairness of dice: A longitudinal study of students' beliefs and strategies for making judgements. *Journal for Research in Mathematics Education, 34,* 270-304.