Biased Sampling and PCK: The Case of the Marijuana Problem

Jane M. Watson
University of Tasmania
<Jane.Watson@utas.edu.au>

Erica L. Nathan
University of Tasmania
<elnathan@utas.edu.au>

As part of an interview protocol investigating teachers' pedagogical content knowledge (PCK) in statistics, 40 teachers were presented with a newspaper article reporting a phone-in survey about the legalisation of marijuana. The article and a question about the reliability of the sample had earlier been used in student surveys, and three student answers to the question were also presented to the teachers. Teachers' PCK was assessed based on responses to questions about the big ideas in the task, potential student appropriate and inappropriate answers, and how teachers would intervene in relation to the three student answers. The wide range of responses provided evidence for the potential of the task in a cross-curricular secondary classroom.

As statistics has assumed a higher profile within the mathematics curriculum, interest has increased in teachers' knowledge for teaching the subject matter to their students. Growing out of the general work on knowledge for teaching by Shulman (1987), Watson (2001) developed a protocol to measure the various aspects defined by Shulman, including content knowledge, pedagogical content knowledge (PCK) and knowledge of students as learners. Ball and colleagues (e.g., Ball, Thames, & Phelps, 2008; Hill, Schilling, & Ball, 2004) focussed on a collection of traits they labelled "mathematical knowledge for teaching" and Groth (2007) adapted this model for teaching statistics. Based on a survey protocol derived from Watson (2001), Watson, Callingham, and Donne (2008) used Rasch (1980) analysis, extended to a partial credit format (Masters, 1982), to obtain a measure of teacher ability in relation to PCK. For that study the focus of PCK was on teachers' content knowledge, its reflection in knowledge of their students' content knowledge, and their PCK in using student answers to devise teaching intervention.

Following the research of Watson et al. (2008) based on teacher surveys, it was decided to interview the same cohort of teachers with the aim of extending the detail and richness of teachers' PCK. This was done with an interview protocol that included general questions on teaching statistics at the beginning and end, with three sections in the middle based on items from student surveys. After the statement of the problems as presented to students, teachers were asked to say what they thought were the 'big ideas' behind the problem and to provide appropriate and inappropriate answers that they might expect from their students. They were then shown two or three responses from students and asked how they would intervene to assist the students to improve their understanding. Detailed analysis of the first task (Watson, Callingham, & Nathan, 2009), based on a pictograph item, refined the construct of PCK to suggest four components. The first two, "*Recognises Big Ideas*" and "*Anticipates Student Answers*" reflected directly the questions asked of the teachers in order to display their own content knowledge and knowledge of students as learners. The third, "*Employs Content-specific Strategies*" encompassed appreciating the nature of the student's answer, beginning at that point, and suggesting appropriate strategies with respect to the answer that demonstrate opportunity to move the student forward. Examples of teacher behaviour included encouraging questions to clarify and explain student answers, constructing sequences of questions based on personal understanding, offering alternative data sets or scenarios, or formalising a discussion. The fourth component, *Constructs Shift to General*, referred to an appreciation of the many

L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia.* Fremantle: MERGA.

statistical ideas related to the initial task and the ability to explore and expand these with the class based on opportunities provided by student answers. It included linking the student answer to these other related statistical ideas and introducing an awareness of language. Specifically related to the pictograph context, examples focussed on revealing the difference between the pictograph as a statistical model and as a vehicle representing real data, exploring the concept of "majority," exposing the limitations of the data collection, and experimenting with alternative representations.

The four components were then used as a basis for analysing the third task in the interview protocol (Watson & Nathan, in press), a two-way table problem with data on smoking and lung disease (Batanero, Estepa, Godino, & Green, 1996). For that analysis a detailed hierarchical rubric was devised for each component of the framework for PCK. The outcomes suggested that distinguishing between "*Employs Content-specific Strategies*" and *Constructs Shift to General* is a useful way of differentiating teacher performance, in that teachers displayed much more ability to handle the content-specific strategies than the generalities. The components and rubrics were then applied to the second task in the interview protocol based on a newspaper article about a survey on legalising marijuana, which is the focus of this paper.

The research questions for this study hence are the following. (1) Given the different context based on sampling rather than a pictograph representation or more a mathematical two-way table, does the framework of four components of PCK provide a comprehensive way of describing teachers' ability to use the task in their classrooms? (2) If so, what is learned about teachers' understanding of the teaching of sampling?

Methodology

Sample. Forty teachers from three Australian states were interviewed: 14 from each of two states, and 12 from the third. They were involved in a professional learning project in statistics for the middle school (Callingham & Watson, 2008). Teachers taught in Grades 5 to 12, had teaching experience ranging from 2 to 25 years, and had a wide-range of previous tertiary study in mathematics and statistics.

Task. The portion of the interview protocol analysed for this paper is shown in Figure 1. The original article appeared in *The Mercury* newspaper in Hobart, Tasmania ("Decriminalise," 1992) and the item for students was used across several research studies (Watson & Moritz, 2000a, 2000b; Watson & Callingham, 2003), including the one in which the interviewed teachers participated. The student answers presented to the teachers were from students involved in the same project as the teachers.

Analysis. The rubric in Table 1 for the first PCK component, *Recognises Big Ideas*, applied mainly to question Q2M, and the rubric for the second PCK component, *Anticipates Student Answers*, applied mainly to Q3M. Occasionally information from Q4M was assessed if it addressed these two features. The rubric for the third PCK component, *Employs Content-specific Strategies*, was based on Questions Q5M, Q6M, and Q7M combined, while the rubric for the fourth component, *Constructs Shift to General*, was assessed across all questions in Figure 1. The responses were coded independently by the authors using the rubric in Table 1 and discrepancies negotiated until agreement was reached.

Decriminalise drug use: poll

SOME 96 percent of callers to youth radio station Triple J have said marijuana use should be decriminalised in Australia. The phone-in listener poll, which closed yesterday, showed 9924 – out of the 10,000-plus callers – favoured decriminalisation, the station said. Only 389 believed possession of the drug should remain a criminal offence. Many callers stressed they did not smoke marijuana but still believed in decriminalising its use, a Triple J statement said.

Is the sample reported here a reliable way of finding out public support for the decriminalisation of marijuana? Why or why not?

Q2M. What are the big statistical ideas in this problem? (Probe: What answer would you give?)

Q3M. Please can you give an example of an appropriate response and an inappropriate response that your students might give? (Probe: Can you explain why it is appropriate/inappropriate?)

Q4M. What opportunities would this problem provide for you teaching? (Probe: Where would you place it in your lesson sequence? Or in your school's curriculum sequence?)

Q5M. ***Show student response: Yes, because 10,000 people is enough to get an accurate average of the view of the public.*** A student gave this answer. How would you move this student's understanding forward? (Probe: What would be the next step in learning?)

Q6M. ***Show student response: No, because it is not everyone in Australia voting.*** (Same as Q5M)

Q7M. ***Show student response: No, because some people could be lying.*** (Same as Q5M)

Figure 1. Interview protocol extracts for Marijuana problem.

Table 1

Rubric for Responses to Marijuana Interview

Code	Description
Component 1: *Recognises Big Ideas*	
0	Responses confused and/or incorrect
1	Response implied and/or understanding revealed beyond initial question
2	Immediate grasp of idea, language specific
Component 2: *Anticipates Student Answers*	
0	Response irrelevant
1	Appropriate or inappropriate but not both, or unclear
2	Distinguishes both appropriate and inappropriate
3	Demonstrates understanding of students' reasoning
Component 3: *Employs Content-specific Strategies*	
0	Response absent or indicates misleading content
1	Content knowledge of sampling requisite to initiate a discussion
2	Demonstrates questions or knowledge that might *structure* a discussion about sampling
3	Extends discussion by illustrating/ referencing beyond the marijuana survey
Component 4: *Constructs Shift to General*	
0	No shift to general evident
1	Considers elements of sampling design in general terms, e.g., size changeable with purpose; profiling of sample population; census vs sampling; accounting for invalid responses; social sensitivity
2	Extrapolates from survey to consider one or more statistical concepts, e.g., random, representation, average
3	Relates survey construction to wider context of argument and/or introduces an awareness of language, e.g., lying, "public"

Results

Research Question 1

Table 2 contains a summary of the codes for the PCK components described in Table 1 for the 40 teachers interviewed.

Table 2
Number of Teachers Receiving each Code for each PCK Component (n=40)

	Code			
PCK Component	0	1	2	3
1 *Recognises Big Ideas*	1	11	28	—
2 *Anticipates Student Answers*	0	8	28	4
3 *Employs Content-specific Strategies*	1	17	16	6
4 *Constructs Shift to General*	12	18	7	3

Examples of teachers' responses at the various levels for the four components of PCK exemplify the hierarchical nature of the rubric. For the first component, Recognises "Big Ideas," teachers generally (70%) displayed an immediate grasp of why the researchers had set the task in relation to the type of sampling involved, e.g., "The biggest thing is about looking at the randomness of, how do we get the random sample and it's not a good example of that at all." One teacher, however, appeared to miss the point responding, "…we would probably take the numbers… So taking the numbers that stand out which is 10,000, 9,924 and 389 and working out, from different perspectives, what those numbers actually mean." A few others struggled initially but eventually suggested reasonable ideas behind the task, e.g., "Well it is sampling, conducted a poll and they have collected data."

For the *Anticipates Student Answers* component, no teachers gave totally irrelevant responses, but a notable few (20%) gave responses that were difficult to distinguish between being appropriate or inappropriate and the teachers were not explicit: "I think a lot of them would probably believe it and that's concerning me… Or they would just look at it and say that most people favoured decriminalisation of marijuana."

Most teachers (70%) could distinguish between appropriate and inappropriate response, e.g.

> …well it is a very limited sample that they have taken the question from. So therefore its reliability is not reliable because it is not a full population sample – not a full age population sample – only listeners, Triple J audience which may be just a particular sort of lifestyle audience…. Or then they would just generalise and sort of say well therefore everybody says the marijuana should be decriminalised because 96% said such. So they are making a big generalisation from that particular response which would be a totally inappropriate thing.

Only 10%, however, displayed a clear understanding of student reasoning, e.g.

> I reckon a lot of the students would probably say that using JJJ as your only source for the survey would show something doesn't represent the population. Whether they go any further and get into more detail, they probably wouldn't. I think 96% would trigger a lot of them to just react from that. So a lot of them do go straight for a percent cause they like the idea of percents… They skim it, so they won't read it, stop, and actually try and understand what's been written. I think most of them would just look for big things that they can base some comment from… A lot would see that and make a personal comment about it rather than a statistical. So, they would, wouldn't actually say

anything about statistics or about the survey, they'd say, you know, that people support it or not. And then they put their own personal opinion in.

When perusing Table 2, it is clear that teachers found it easier to suggest content-specific strategies in response to student answers presented than to construct a shift to more general concepts. However, where only one teacher failed to suggest a strategy for PCK Component 3, 30% failed to show a capacity to shift to general for PCK Component 4. Considering first the responses for Component 3, *Employs Content-specific Strategies*, responses at Code 1 displayed the content knowledge to initiate a discussion based on the article, but did not go further. The range of suggestions for example included the following for the three student responses.

Q5M (10,000 enough): But is that a reasonable snapshot of the entire population or is it a specific demographic? Q6M (not everyone): I'd say fair enough but under what circumstances are you going to be able to have the funding to get everyone to vote? Q7M (lying): Well you're going to take that chance no matter what survey and technique you do.

At Code 2, the increased ability to structure is shown from several perspectives and illustrated in the following detail.

Q5M (10,000 enough): …they have some sense of size being a contributing factor to sampling but also you would have to then obviously go through the whole point of the biased nature of who those 10,000 people are and where they come from … that they don't represent the public as a whole. Q6M (not enough): …You'd have to talk about the fact that, is it realistic to sample everyone when you want to find out something and so therefore sampling is valid and important but in the right format. Q7M (lying): …why would they ring up and lie? So that might be an interesting insight into … what they think is happening there … you'd have to get a sense of where the kid's coming from I think before you could respond to that.

An ability to extend the discussion beyond the marijuana survey (Code 3) was demonstrated by several teachers, including the following.

Q5M (10,000 enough): …How can you talk about the whole sampling process? So for instance, ok, this is a phone-in listener poll. What about people with really valid thoughts that couldn't phone in for whatever reason? Ok, now so it's all about randomness here. And once again we're talking about a youth radio station, well, how well does that represent the population of people? Like I said, great that you've mentioned it's a large sample size, but you're missing the fact that it's random, or not random in a sense. Q6M (not enough): Hearing that straight up, I'd say, listen that's a great point, you're right, everyone in Australia hasn't voted, so do we really know? But could everyone in Australia vote? I mean, how viable is it that we could actually get every single person in Australia to register some sort of a vote. We have to do some sort of sampling don't we? And the key is how we go about that sampling process. That's how I would address that answer there. So they've got a point but they just need a bit of elaboration. Q7M (lying): Once again, that's a great answer. How do we know? How do we know whether some people are lying? We don't. … You like to take their word for it. Some people could be lying. But then again, how would I address that? I don't know whether it's relevant but you'd probably go into, in this case, why is that they would want to lie? … I mean, it's a sensitive issue. Yeah, I'd find it hard to probably justify lying … but I would probably look at the actual question. I mean, it doesn't seem personal … It's not like you're doing a face to face interview, so you're not getting a change in body structure or anything like that where you can tell that this is, oh does this person smoke marijuana themselves, or something. It's a phone in, so the scope for lying here would probably be minimal. You'd think that if people had a genuine idea they'd probably would register a vote and would probably be the truth simply because it's a phone in poll.

With respect to the fourth component *Constructs Shift to General*, the Code 1 responses suggested a range of more general extensions related to the survey described in the article, including the following.

> Oh – so it is really – opening the issue of question, census and sampling, sort of makes … some of them realise just how much bigger this country really is. And I would say it is not supposed to be a census because that's the word I would use for everybody, this is supposed to be a sample, which is a small part of [it] and that's – it's usually reported as a sample rather than saying – suggesting the whole population.

Code 2 responses included consideration of some of the statistical concepts involved.

> What is an average and what are you basically saying by an accurate average? … we might go through how they got averages and work out can we get enough to get an accurate average view, of the view of the public? Can we get an average view for this particular question? Yeah, so probably draw attention to that first. 10,000 is enough … and some people would say that's a high sample, so, but again 10,000 people of what particular age group or particular demographic? If you took 10,000 children, would they say that computer use should be restricted, I'd try and draw attention to those particular social issues of kids. And work statistically on working out what is an average. What does the average of the view mean? … Again it relates back to what you regard as a random sample.

Finally three teachers related the survey and its construction and application to wider public issues, such as the following from one of the three.

> And then I would probably pose some questions in other contexts, so you know, if we were trying to figure out what the public response was to say, you guys wanting a new skate park to be built on our island, which is a big issue at the moment. How would we go about that? What problems might there be, how could that, you know, be similar here. What would happen if they come up with things like, well we just ask all the kids in the school. I said, is that the public? Why should we, why should we ask anybody else? Why should we get a general public opinion? Well, you know, they're the ratepayers, they pay the Council rates and the Council spends the money. So there's all those social sorts of ideas that can be discussed in there as well as bringing the idea of well how do you get a representative sample? What is a representative sample? How many would you have to ask and all of that stuff … Ok, so then we have to discuss the difference between a sample in the census and that has formed a reasonably large part of my teaching. And we've done that, particularly with, to move this thinking on, that idea that you don't get quality information if you don't hold a census; and we've used information from the census online, where we've taken samples and usually what I do is I take a large sample and then I divide it up into smaller samples from that large sample and give them to different groups in the class to analyse and we might go as small as say a sample of 10. … And in terms of what do we do if someone's deliberately lying, there's not a lot we can do. They have to accept that … We have to take that consideration into account when we are relying on the results from our analysis of the data

Research Question 2

Having given examples of teachers' levels of ability on the four components of PCK, what is learned more generally about teachers' understanding of teaching about sampling? Based on the responses, the opportunities suggested by teachers are categorised into several groupings. More than half of the teachers suggested a wide range of other examples of surveys for their students to consider, which either displayed similar bias or gave students the opportunity to devise non-biased sampling. Examples included Australian Idol, Morgan polls, surveys in supermarkets, school-based surveys, and Census@School.

The specific language and concepts suggested for discussion included census and survey, population and sample, the "public," being representative, bias, and random. One teacher specifically mentioned the literacy of mathematics in relation to this task and another teacher noted the importance of asking students multiple questions to "get behind students' meaning" in reference to their use of language. Several teachers, however, mentioned the literacy requirement for some students in reading and interpreting the text

and the difficulties it would pose. One teacher felt the problem was more appropriate for literacy than mathematics and another suggested it be linked to health.

The student response about lying elicited a broad range of suggestions for classroom discussion. Some teachers just explicitly stated the equivalent of "it doesn't matter" as it would be a small percentage of a large sample and no data are completely fool proof. Reasons not to worry included the anonymity of calls, there being no repercussions for answers, and the apparent yes/no type of question. Reasons the people might lie or bias the survey included peer pressure, people wanting a particular voice or lobby group to be heard, and realisation of the possibility to phone in more than once. A couple of teachers suggested asking who might be lying; the callers or the radio station. At another level, one teacher suggested asking how one could check on lying. Others suggested devising different survey questions or conducting interviews where body language could be observed. One teacher suggested asking a student for evidence of lying and how it might be obtained. Finally two teachers referred to their students' experiences with Census@School data sets and the erroneous data sometimes obtained as laying a foundation for students appreciating the fallibility of data for various reasons.

Issues of motives behind conducting the survey were raised by some teachers and others suggested students write some other questions for the survey, either biased or unbiased. Another teacher suggested students consider what kinds of questions people would respond to. On the pragmatic side one teacher noted that one had to be listening to the station at the appropriate time, whereas another noted that only those with credit on their mobile phones could phone. In relation to the student response about everyone not answering, most teachers just noted the impossibility of collecting census data, with some noting sampling practicalities and others cost. A few teachers of younger students (e.g., Grades 5 or 6) felt the problem was too complex, either in terms of the context or the statistical understanding for their students.

Discussion

The framework of four components of PCK did provide the researchers with a comprehensive way of describing teachers' ability to explore the problem of sampling in their classrooms. Coding of four non-hierarchical components encouraged a dissection of the teaching process with the objective of understanding the working constituents of statistical PCK. The fact that 28 out of 40 teachers scored at Code 2 for the first (*Recognises Big Ideas*) and second (*Anticipates Student Answers*) components of the rubric confirmed teacher confidence in the content knowledge of statistical sampling. Almost half the teachers were able to identify the potential of student responses and structure a discussion concerning aspects of the sampling problem which suggests that this knowledge was "active" teaching knowledge, and that there was a capacity to meet, lead and shape student understanding. However, teachers seemed less able to situate the problem of sampling within the wider world of statistics, to link it to associated concepts, and to appreciate its limitations, as well as strengths, as an analytical tool. This is evident from the coding results of the fourth (Shift to General) component. Almost half were able to establish the complexity of sampling design. However, the idea of sampling itself as part of a wider statistical construct was pursued by very few. The indication that most teachers have difficulty in this task of conceptual extension has repercussions for professional development.

Although the teacher responses did not score strongly across all four components, there was a considerable range of responses within the confines of the specific sampling

problem. This is perhaps best demonstrated in the discussion points prompted by Q7M (lying). What might have seemed like a straightforward student answer with a narrow opportunity to build upon actually sparked a surprising array of teacher responses. It seems likely that this abundance of responses was in large part a by product of the teachers' sound performance for the first and second component of the PCK Statistics framework. Q7M has been selected for particular attention, but the range of responses to all three student answers highlighted the appropriateness and importance of statistics being taught across the curriculum.

Acknowledgements.

The research was funded by Australian Research Council grant no. LP0669106. Thanks to Julie Donne and Elena Zema, who conducted some teacher interviews.

References

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*, 389-407.

Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*, 151-169.

Callingham, R., & Watson J. M. (2008). Overcoming research design issues using Rasch measurement: The StatSmart project. In P. Jeffery (Ed.), *Proceedings of the AARE annual conference, Fremantle, December, 2007*. Available at http://www.aare.edu.au/07pap/cal07042.pdf

Groth, R. E. (2007). Reflections on a research-inspired lesson about the fairness of dice. *Mathematics Teaching in the Middle School 13*(4), 237-243.

Hill, H. C., Schilling, S. G. & Ball, D. L. (2004). Developing measures of teachers' mathematics for teaching. *Elementary School Journal*, *105*, 11-30.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press (original work published 1960).

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.

Watson, J. M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education*, *4*, 305-337.

Watson, J. M., & Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3-46.

Watson, J., Callingham, R., & Donne, J. (2008). Establishing PCK for teaching statistics. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICME/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* (Proceedings of the ICMI Study 18 and the 2008 IASE Round Table Conference, Monterrey, Mexico, July, 2008). Available at http://www.ugr.es/~icmi/iase_study/

Watson, J., Callingham, R., & Nathan, E. (2009). Probing Teachers' Pedagogical Content Knowledge in Statistics: "How will Tom get to school tomorrow?" In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides* (Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia, Wellington, NZ, Vol. 2, pp. 563-570), Palmerston North, NZ: MERGA.

Watson, J. M., & Moritz, J. B. (2000a). Developing concepts of sampling. *Journal for Research in Mathematics Education*, *31*, 44-70.

Watson, J. M., & Moritz, J. B. (2000b). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, *19*, 109-136.

Watson, J., & Nathan, E. (in press). *Assessing the interpretation of two-way tables as part of statistical literacy*. Refereed paper to be presented at the 8th International Conference on the Teaching of Statistics, Ljubljana, Slovenia, July, 2010.