Chapter 28
Measuring Levels of Statistical Pedagogical Content Knowledge

Rosemary Callingham and Jane Watson

Abstract The introduction of statistics and probability into the school curriculum has raised awareness of the expectations on teachers who have to teach it. A review of the related field of mathematics education indicates that teachers need more than content knowledge. They must also respond to their students' statistical understandings in ways that move students' current understanding to higher levels. Efforts to measure such statistical pedagogical content knowledge are still in their infancy. Findings from a large-scale Australian study are reported to exemplify these efforts, and the implications for future research are discussed.

1 Introduction

As early as 1988, many of the participants in the International Statistical Institute Round Table in Budapest (Hawkins, 1990) were discussing the lack of adequate preparation of school teachers for teaching statistics and the consequent deficiencies observed. Rubin and Rosebery (1990), for example, described anecdotally the difficulties displayed by classroom teachers that are indicators of the types of understanding that researchers are attempting to measure more formally today. Despite statistics being part of the curriculum in the later years of schooling (Holmes, 1980), the advent of statistics and probability in the broader school curriculum in the early 1990s, including the primary or elementary years (e.g., Australian Education Council, 1991; National Council of Teachers of Mathematics, 1989) raised awareness of issues surrounding the teaching of these topics.

R. Callingham (✉)
Faculty of Education, University of Tasmania, Locked Bag 1307,
Launceston, TAS, Australia 7250
e-mail: Rosemary.Callingham@utas.edu.au

J. Watson
University of Tasmania, Private bag 66, Hobart, TAS, Australia 7001
e-mail: Jane.Watson@utas.edu.au

In this chapter, the understanding of the nature of teachers' knowledge that is needed at the school level and emerging attempts to measure this knowledge are considered (chapters by Burgess; Godino, Ortiz, Roa, & Wilhelmi; Ponte; Pfannkuch & Ben-Zvi; and Makar & Fielding-Wells, in this book present other analyses of teachers' pedagogical and professional knowledge). First, research in the related field of mathematics education is examined as a way of informing discussion about statistics education; second the field of statistics education itself is canvassed, identifying issues in relation to teachers' knowledge. Finally aspects of pedagogical content knowledge for teaching statistics and recent efforts to measure this construct are described to exemplify current approaches, and the implications are discussed.

2 Background

The measurement of teachers' mathematical knowledge has predominantly focused on content knowledge (Hill, Sleep, Lewis, & Ball, 2007), especially arithmetic, with statistical knowledge limited to reading simple graphs. In a seminal paper, Shulman (1987) identified seven knowledge types needed by teachers to be competent in the modern classroom. This work has prompted a variety of attempts to describe, identify, and measure more precisely the nature of teachers' knowledge, in particular in the field of mathematics education. Although the measurement of teachers' mathematical knowledge is well established, often measured by relatively simplistic tests of mathematical knowledge (Hill et al., 2007), more contested has been the level of content knowledge required by teachers if they are to teach mathematics effectively. Ma (1999) suggested that elementary teachers do not need high levels of mathematical knowledge, but do require a "profound understanding of fundamental mathematics" (p. 22), including a deep grasp of the interconnections and relationships among different aspects of mathematical knowledge. It seems likely that such thinking could apply also to statistics.

A different approach to measuring teachers' mathematical knowledge is that taken by the Training and Development Agency for Schools (2009) in the United Kingdom. All teachers are required to take a test of "numeracy" before they are accredited to teach, including many items that could be considered statistical in nature, for example, reading box plots of data from testing programmes. Although addressing some aspects of Shulman's (1987) "knowledge of education contexts", the test remains content-focused.

In recent years, attention has shifted from content knowledge to the description and measurement of "Pedagogical Content Knowledge" (PCK), which Shulman (1987) conceptualised as "that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding" (p. 6). Shulman's ideas have been refined and developed over time. Hill, Rowan, and Ball (2005), for example, described Mathematical Knowledge for Teaching (MKT), which they defined as that "…mathematical knowledge used to

carry out the work of teaching mathematics" (p. 373). Chick (2007) suggested that PCK may be inferred from the nature of the tasks set by teachers, and in particular their use of stimulus material that has a range of possibilities or "affordances" for use in the classroom. Such work suggests that different topics may require diverse aspects of PCK, and the creative and imaginative use of suitable artefacts may play a role in identifying PCK.

Recent work has focused on the formal measurement of aspects of PCK. Hill, Schilling, and Ball (2004) measured teachers' MKT using a multiple choice test based on mathematical content commonly found in elementary school courses, such as number concepts and operations, and pattern and algebra. No statistical knowledge was included. These instruments were then used to link teacher knowledge with students' outcomes (Hill et al., 2005). A recent international study (MT-21) considered both content knowledge and "Mathematics Pedagogy Knowledge" to evaluate teacher education programmes. The content survey included statistics, but the Mathematics Pedagogy Knowledge component was organised around three sub-scales: Curriculum, Teaching, and Students (Schmidt et al., 2007, p. 26). Of particular relevance to the discussion presented here in relation to the types of items employed is the Students scale, which used items that required teachers to respond to student answers and identify the errors made on a variety of topics.

2.1   The Situation in Statistics Education

Measuring teacher knowledge specifically for teaching statistics has a shorter history than for teaching mathematics. At the International Statistical Institute Roundtable in 1992, Begg (1993) proposed a research agenda that began to address such issues at the school level. This agenda focused on professional development and its effectiveness that implied, although did not address explicitly, some measure of teachers' statistical knowledge. Shaughnessy (1992) further indicated the lack of research in this area by placing teachers' conceptions of probability and statistics as one of his seven significant future research questions.

By 2001 however, issues in school statistics were recognised, and issues around teachers' capacity to teach statistical concepts were becoming part of the research agenda. One approach to identifying and measuring teachers' knowledge in statistics was that taken by Watson (2001) using a "profile". This instrument attempted to address multiple domains of teacher knowledge including self-efficacy or confidence in teaching statistical concepts, beliefs about the value and use of statistics, and pedagogical content knowledge using items in which teachers responded to questions based on student survey items. Groth (2007) provided a framework for teaching statistics at the high school level adapting the ideas of "common" and "specialised" knowledge of Hill et al. (2004) and acknowledged that the specialised knowledge area required a growing research base. This is particularly the case in relation to non-mathematical knowledge, which encompasses the pedagogical activities that take place in the classroom.

Watson, Callingham, and Donne (2008) reported on the development of a measure of pedagogical content knowledge in statistics. The instrument used was a pen-and-paper survey similar to Watson's (2001) profiling instrument but extended the items in which teachers were asked to give a response to student work. Using Rasch measurement techniques (Bond & Fox, 2007), the 12-item scale of statistical pedagogical content knowledge had good measurement qualities and provided measures of teachers' performance that could be interpreted in hierarchical levels. One aspect of this study was the relative difficulty teachers had in suggesting suitable "next steps" to move students' understanding forward, even when they could predict students' responses. This finding suggests that instruments identifying PCK may fall short if they only include items asking teachers to identify appropriate responses without any accompanying follow-up action.

In an attempt to refine the measures, Watson, Callingham, and Nathan (2009) used similar items in an interview situation. An initial analysis of the teachers' responses indicated four components of teacher knowledge, including content knowledge, knowledge of students as learners, and two aspects of PCK. One PCK component was characterised by teachers employing content-specific strategies to develop students' understanding; the second involved constructing a shift from a specific to a more general statistical context, such as making explicit connections between different aspects of statistical knowledge through connecting, for example, a data display with notions of probability. This latter category extended the understanding of PCK and embeds specific statistical teaching strategies within the domain of statistical pedagogical knowledge. The interview format, however, has some practical limitations in terms of its use with a large number of teachers, although it does provide insights into the nature of statistical PCK.

PCK for teaching statistics is undoubtedly a complex construct. The study reported in the next section represents an attempt to measure statistical PCK using an instrument sufficiently sensitive to track changes in teachers' understanding over time.

3   Teacher Knowledge for Teaching Statistics

The *StatSmart* project incorporated a professional learning programme over 3 years assisting teachers to appreciate the developmental processes that students go through in reaching statistical understanding (e.g., Watson, 2006). Teachers were also provided with resources including *Tinkerplots* software (Konold & Miller, 2005) and attended a 2-day conference each year at which activities and ideas were presented to help them devise learning activities suitable for their class levels. The project timeline included repeated monitoring of teachers and their students. Design details were discussed in Callingham and Watson (2007). The survey data reported here were collected from teachers after the first professional learning conference towards the end of the first year of the study, and again midway through the project.

The teacher group was composed of 45 very experienced teachers who mostly had limited tertiary-level mathematics study within which none explicitly mentioned studying statistics. They taught middle-year students (Grades 5–10) in a variety of settings or in high school, often including post-compulsory years. Although teaching in the different jurisdictions that comprise education in Australia (i.e., government, Catholic and independent schools) using different curriculum documents, their discussions suggested that they were more homogenous than the background data might imply. Completion of the survey was an expectation of the project and 42 teachers responded to the first survey. The second survey was answered by 26 teachers, of whom 18 had also responded to the first administration. The lower response rate was probably due to the time of year when the survey was sent out, some months after the second professional conference. The eight new teachers had very similar backgrounds to the others.

Only the 12 PCK items from the instrument (Watson, 2001) are discussed here. These items were of three types. The first group (four items) provided teachers with survey questions that had been given to students and asked them to anticipate appropriate and inappropriate responses that their own students might provide. A second group (four items) extended these initial questions by asking teachers to indicate how they might use the item stimulus in the classroom, using Chick's (2007) notion of affordances. An example is the item in the context of a newspaper headline on "odds" shown in Fig. 28.1, in which question a was of the first type (anticipating students' responses) and question b was of the second type (classroom affordance). The stimulus contexts for these items included an incorrect pie chart from a media article and a newspaper report about an association between car accidents and heart disease (Watson, 2006), as well as the "odds" headline in Fig. 28.1.

The third type of item, of which there were four, explicitly invited teachers to respond to particular students' answers. Real student answers from surveys were provided and teachers were asked to say how they would respond to the student. An example of two items employing proportional reasoning in the context of

---

North at 7-2, but we can still win match, says coach

1. What does "7-2" mean in this headline about the North against South football match? Give as much detail as you can.
2. From the numbers, who would be expected to win the game?

The above item was given to a group of students.
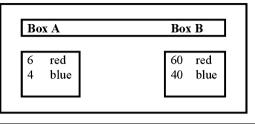
a. What kinds of responses would you expect from your students? Write down some appropriate and inappropriate responses.
b. How would/could you use this item in the classroom? For example, how would you intervene to address the inappropriate responses?

Fig. 28.1 An example of a PCK item asking teachers to anticipate students' responses

Box A and Box B are filled with red and blue marbles as follows. Each box is shaken. You want to get a blue marble, but you are only allowed to pick out one marble without looking. Which box should you choose?

a. Box A.

b. Box B.

Box A		Box B	
6	red	60	red
4	blue	40	blue

c. It doesn't matter.

Please explain your answer.

Consider each of the following answers and explanations to the above problem and explain how you would respond to each answer.

Student 1: (B), because you could get red or blue.

Student 2: (A), because there are only two more reds in A and 20 more in B.

Fig. 28.2 Example of a student response item for addressing teachers' statistical PCK

determining a probability is shown in Fig. 28.2. The other two PCK items were set in the context of a two-way table.

In the second teacher survey in 2008 some modifications were made to the original survey items in order to extend the information obtained into some different statistical content areas. The items on odds, pie chart, and the proportional reasoning item shown in Fig. 28.2 remained the same. The harder proportional reasoning item, using a two-way table, was dropped because many teachers avoided answering it. The association item was modified to make the teacher questions more explicit, reducing the number of individual questions. A new item addressing average replaced the two-way table question. These modifications provided a total of 11 items, of which 4 items addressed anticipated student responses, 3 items addressed classroom affordance, and 4 items asked teachers to respond to students' answers. Of these 11 items, 6 were common across both administrations of the profile (odds, pie chart, and the proportional reasoning task shown in Fig. 28.2) to provide a basis for scale anchoring using Rasch measurement (Bond & Fox, 2007). All items were scored using a rubric based on the increasing complexity of the response. The scores for each item varied from 0–2 to 0–5, where 0 in every instance indicated no response or a totally irrelevant response (e.g., Watson et al., 2008).

3.1 Analysis

In order to explore the construct of statistical PCK a number of analyses were carried out. First, a Principal Components factor analysis with Kaiser normalisation and varimax rotation was undertaken to examine the construct in relation to the four components identified by Watson et al. (2009).

The initial factor analysis using the 2007 data suggested four factors that together accounted for 67.6% of the variance. The first of these included all four items in which teachers were asked to respond to students' work. The second factor addressed anticipated student responses to the items on odds and a pie graph, and classroom use of odds (three items). The third included anticipated student responses to the association item (three items), and the fourth factor included suggested classroom use of two items that included graphing: pie graph and association. An identical analysis was run on the data from the second administration of the instrument in 2008. Only three factors were identified that accounted for 70.1% of the variance, and eight of the eleven items loaded onto more than one factor. The number of teachers ($n=26$) was smaller and 18 of these were completing a second survey. It is possible that the teachers who had been in the project for some time were now interpreting the items in different ways, leading to the different results. The changes did suggest, however, that creating scales based on the initial factor analysis was unlikely to provide consistent results.

Second, the data were analysed using Rasch measurement (Bond & Fox, 2007) in which person ability and item difficulty are placed on the same interval measurement scale. The hierarchical nature of the scoring rubrics provided direction along a single variable of PCK, and the variable hence established had good measurement characteristics. Measures of person "ability" in logits, the unit of Rasch measurement, for all 42 teachers who responded to the 2007 administration were obtained. The second administration, in 2008, was similarly analysed but anchored to the item difficulty of the first so that all measures were against the same "ruler". In this way, direct comparisons of teacher performance could be made across both administrations of the instrument to determine change over time.

3.2   Interpretation of the Construct

The Rasch analysis process spreads the items along the extent of the variable, creating an hierarchical scale where the more difficult items appear at the top of the construct and the easiest items occur at the bottom. By considering the "gaps" between clusters of items and the demands of the items within clusters, four levels of PCK were identified. A description of the nature of teachers' responses at each level of the PCK hierarchy was obtained by a qualitative analysis of the item responses occurring at each level and is described in this section.

Aware level. Teachers are likely only to suggest a single appropriate or inappropriate student response to the items. They display little broader statistical understanding and do not make suitable suggestions for addressing students' understanding. For example, in relation to using the odds question in the classroom a typical Aware level response is, "Great for relating to football betting in real life; Hear this on the news and footy show".

Emerging level. Teachers use some statistical knowledge to suggest several *either* correct *or* incorrect responses for students. Two PCK items, related to odds and

a graph of association, expect the ability to suggest *both* correct and incorrect student responses, for example, in relation to the odds question: "No idea; Not sure; North have a good chance to win; for every $2 placed you receive $7 back or for every 9 matches North would win 7". Generic rather than content-specific strategies are suggested for classroom intervention, which implies good teaching but not necessarily in the context of statistics.

Competent level. The items with the highest codes at this level reflect more traditional and familiar topics in the school statistics curriculum: graphs and probability. Some statistically appropriate interventions are suggested but only in the context of familiar classroom activities. For example, for the proportional reasoning task, "Question: Is it all luck? Would I be more likely to get a red? Why? In both boxes? How could I show my chance of getting a blue in numbers?"

Accomplished level. The highest code for most PCK items appears at this level, demanding the suggestion of *both* correct and incorrect student responses and an integration of appropriate statistical content with student-centred intervention strategies. For example, in using the lung disease two-way table item in the classroom, a response showing an integrated view of statistical PCK is "Research cause and effect relationship for the diseases. Develop understanding of dependent, independent variables, control and how difficult it is to isolate other factors – need to 'read between the statistics.' What other information is needed? If you eat carrots you will die!"

As teachers progress up the scale, they demonstrate more complex and inter-related responses, showing both understanding of appropriate statistical knowledge and of their students.

3.3   Teacher Change

To consider teacher change, a paired sample *t*-test was conducted on the mean teacher ability measures obtained from Rasch measurement from 2007 ($\bar{x} = 0.47$, s.d. $= 0.78$) and 2008 ($\bar{x} = 1.15$, s.d. $= 1.44$) for the 18 teachers who completed both surveys. This test indicated a significant increase in the mean value of teacher ability (df $= 17$, $t = 2.26$, $p = 0.040$) with an effect size of 0.59, indicating that the scale could provide a measure of teacher change.

The four hierarchical levels of statistical PCK identified, however, provide a criterion-referenced approach to considering teacher knowledge that can go beyond considering mean scores. By comparing teacher ability measures obtained through the Rasch measurement approach against the items on the same scale, the number of teachers in each level of the hierarchy can be compared. The proportion of teachers in each level of the PCK construct in 2007 was ($n = 42$): Aware 14% (6); Emerging 28% (12); Competent 50% (21); and Accomplished 7% (3). In 2008, the corresponding figures were ($n = 26$): Aware 15% (4); Emerging 15% (4); Competent 46% (12); and Accomplished 23% (6). Teachers in the Aware level in 2008 were all teachers who had entered the project late. Of particular interest is the increased

number of teachers in the Accomplished level, from three in 2007 to six in 2008. Using this profile approach has the potential to be able to provide more targeted intervention with teachers, by matching professional learning activities to teachers' demonstrated level of PCK.

4   Implications for Teacher Training and Research

The *StatSmart* study represents a new direction in PCK measurement by using items in which teachers have to respond in ways that draw on their classroom practice, which are scored according to the quality of that response. The results suggest that in addition to understanding the statistical content, at high levels of PCK teachers need to be able to anticipate the range of students' likely answers to questions and to respond in ways that will further develop students' understanding. Although this profiling approach may lack the level of detail that might be obtained through interviews and observations, it has the advantage that it can be used for large-scale projects.

The relatively small pool of teachers, and their long experience, limits the inferences that might be drawn in general about teachers' statistical PCK. It is possible that with a larger pool of teachers more levels might be identified because a larger and more representative sample would provide more information about items at the top and particularly the bottom of the scale. Further refining the measure of pedagogical content knowledge would be useful for evaluating professional learning or pre-service teacher education programmes, for example.

Developing measures of teachers' statistical pedagogical content knowledge is an important continuing goal of statistics education. The next stage in research related to measuring teachers' statistical pedagogical knowledge will be to examine the association of teachers' statistical PCK and student learning outcomes. Such research will then prepare the ground for evaluating professional learning programmes for teachers in terms of *change* in teachers' statistical PCK and *change* in their students' learning outcomes.

References

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton, VIC: Author.

Begg, A. (1993). Establishing a research agenda for statistics education. In L. Pereira-Mendoza (Ed.), *Introducing data analysis in the schools: Who should teach it and how?* (pp. 212–218). Voorburg, The Netherlands: International Statistical Institute.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Callingham, R., & Watson J. M. (2007, December). *Overcoming research design issues using Rasch measurement: The StatSmart project*. Paper presented at the Australian Association for Research in Education International Educational Research Conference 2007, Fremantle, Australia. Online: www.aare.edu.au/07pap/cal07042.pdf

Chick, H. L. (2007). Teaching and learning by example. In J. Watson & K. Beswick (Eds.), *Mathematics: Essential research, essential practice. Proceedings of the 30th Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 3–21). Sydney: MERGA.

Groth, R. E. (2007). Toward a conceptualisation of statistical knowledge for teaching. *Journal for Research in Mathematics Education, 38*, 427–437.

Hawkins, A. (Ed.). (1990). *Training teachers to teach statistics: Proceedings of the International Statistical Institute Round Table Conference*. Voorburg, The Netherlands: International Statistical Institute.

Hill, H. C., Rowan, R., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

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

Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). Assessing teachers mathematical knowledge: What knowledge matters and what evidence counts? In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111–156). Reston, VA: National Council of Teachers of Mathematics.

Holmes, P. (1980). *Teaching statistics 11–16*. Slough, UK: Schools Council and Foulsham Educational.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration [computer software]*. Emeryville, CA: Key Curriculum Press.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Rubin, A., & Rosebery, A. S. (1990). Teachers' misunderstandings in statistical reasoning: Evidence from a field test of innovative materials. In A. Hawkins (Ed.), *Training teachers to teach statistics: Proceedings of the International Statistical Institute Round Table Conference* (pp. 72–89). Voorburg, The Netherlands: International Statistical Institute.

Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Gogan, L., et al. (2007). *The preparation gap: Teacher education for middle school mathematics in six countries* (MT21 report). East Lansing, MI: Center for Research in Mathematics and Science Education, Michigan State University.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: National Council of Teachers of Mathematics & MacMillan.

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

Training and Development Agency for Schools. (2009). *Numeracy practice materials*. Online: www.tda.gov.uk/skillstests/numeracy/practicematerials.aspx

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. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.

Watson, J. M., Callingham, R., & Donne, J. (2008). Establishing pedagogical content knowledge for teaching statistics. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and*

Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference. Monterrey, Mexico: International Commission on Mathematical Instruction and International Association for Statistical Education. Online: www.stat.auckland.ac.nz/~iase/publications

Watson, J. M., 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* (Vol. 2, pp. 563–570). Palmerston North, NZ: MERGA.