HOW A CURRICULUM MAY DEVELOP TECHNOLOGICAL STATISTICAL KNOWLEDGE: A CASE OF TEACHERS EXAMINING RELATIONSHIPS AMONG VARIABLES USING FATHOM

Hollylynne S. Lee and Jennifer Nickell
NC State University, Raleigh, NC, USA
hollylynne@ncsu.edu

We report results from 74 preservice secondary mathematics teachers' [PSMTs] examination of a multivariate dataset using dynamic statistical software (Fathom^{TM}). The PSMTs were enrolled in similar courses at four US-based institutions, using the same curriculum materials. Adapting Lee & Hollebrands' (2011) framework, we present an analysis of aspects of Statistical Knowledge [SK], Technological Statistical Knowledge [TSK], and Pedagogical Statistical Knowledge [PSK] the materials were designed to teach, and aspects of SK, TSK, and PSK evident in teachers' work.

INTRODUCTION

In school curricula in many countries, mathematics teachers are responsible for understanding and teaching significant statistical content to secondary students (ages 12-18). Across international settings, mathematics and statistics educators have worked to create experiences that can improve teachers' comfort and expertise with the statistical content they teach (e.g., see several chapters in Batanero, Burrill, & Reading, 2011). Since technology tools can enhance the ways in which one engages with data analysis, statistics, and probability simulations (e.g., Chance et al., 2007), many teacher education efforts have had a particular emphasis on preparing teachers to better understand, and be able to teach, statistics with dynamic software (e.g., Madden, 2011; Makar & Confrey, 2005). When given opportunities to learn statistics with such dynamic software tools, we wonder: *How do teachers use dynamic software tools to engage in statistical ideas? How may their work be preparing them for pedagogical decisions when teaching students statistics in their future classrooms?*

This study closely examines the learning opportunities in curricula materials (Lee, Hollebrands, & Wilson, 2010) for prospective secondary mathematics teachers [PSMTs] to learn to use dynamic statistical software tools to engage in statistical investigations, for themselves and their future students. We then use key aspects from the analysis of curricular learning opportunities to analyze teachers' work on a statistical investigation with a multivariate data set. To situate our work in curriculum design and research, we provide a framework that highlights important aspects of statistics teaching and learning.

FRAMEWORK FOR CURRICULA DESIGN AND RESEARCH

Building from prior work on pedagogical content knowledge and technological pedagogical content knowledge (e.g., Mishra & Koehler, 2006), Lee and Hollebrands (2011) proposed a framework characterizing three aspects of teachers' knowledge related to teaching statistics with technology. The guiding framework for development of the Lee et al. (2010) textbook, and for analysis in this study, is an extension of Lee and Hollebrands' (2011) framework. The guiding framework characterizes four types of teachers' knowledge: (1) statistical knowledge, (2) technological statistical knowledge, (3) pedagogical statistical knowledge, and (4) technological pedagogical statistical knowledge. Figure 1 illustrates the belief that teachers' statistical knowledge is foundational to developing the other three.

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. iase-web.org

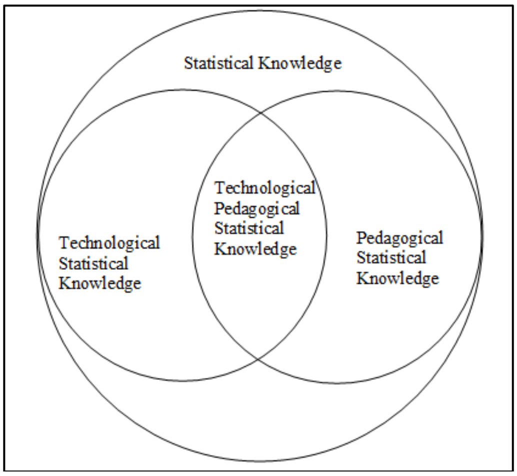

Figure 1. Relationships among knowledge for teaching statistics with technology.

Statistical Knowledge (SK)

The importance of statistical knowledge for teachers has been underscored by Groth (2007) and others. The materials in Lee et al. (2010) are not intended to teach all aspects of statistics that teachers need to know. Instead, the materials build from research on students' learning of a few major topics and ideas that are particularly enhanced with technology tools. To assist in the development of teacher' statistical knowledge, teachers are: (1) engaging in exploratory data analysis [EDA], (2) focusing on distributions and describing data as an aggregate, (3) coordinating measures of center and variability in distributions; and (4) considering key differences between statistical and mathematical thinking.

For example, the materials engage teachers as doers of statistics through exploratory data analysis in rich contexts (e.g., vehicle fuel economy). This promotes asking questions from data within context, and the technology facilitates using a variety of representations to explore data in novel ways. Teachers are given opportunities to consider trends in data (e.g., linear models), along with residuals and correlation, and be able to make interpretations of their models to support predictions and claims.

Technological Statistical Knowledge (TSK)

Engaging teachers in tasks with dynamic statistical software can simultaneously develop their understanding of statistical ideas and technology skills. The materials aim to develop teachers' TSK by focusing on elements proposed by Chance et al. (2007) for technology to be used for: (1) automating computations and graphs, (2) exploring data with a variety of representations, (3) visualizing abstract concepts, (4) simulating phenomena, and (5) accessing large data sets.

Automation of statistical computations and the generation of graphs facilitate a focus on exploring data and visualizing abstract concepts, such as using dynamic dragging capabilities to illustrate effects of an outlier on correlation and least squares regression line. Teachers learn how overlaying statistical measures (e.g., means, regression lines) on a graph can change the way one conceptualizes these measures in relation to a distribution (Biehler et al., 2013).

Pedagogical Statistical Knowledge (PSK)

There are particular pedagogical decisions that arise when teaching statistics. In the Lee et al. (2010) textbook, opportunities are designed for teachers to develop PSK in areas such as: (1) planning for group projects and discussions about data, (2) supporting students in making statistical arguments based on appropriate evidence, and (3) considering the contexts used for teaching statistical ideas. For example, teachers learn to engage students in statistical investigations in a variety of contexts in which they need to make decisions and arguments, and are pushed to consider how respond to different conclusions among groups during discussions (Shaughnessy, 2007). This requires teachers to be comfortable with various approaches and representations that students may use to investigate a statistical task and to expect variation in results

Technological Pedagogical Statistical Knowledge (TPSK)

The ultimate goal in the preparation of teachers of statistics is to develop the specialized subset of knowledge for teachers representing TPSK, which includes: (1) understanding students' learning about statistical ideas with technology, (2) conceiving of how technology tools and representations support statistical thinking, (3) developing instructional strategies to use in statistics lessons with technology, and (4) critically evaluating and using curricula materials for teaching statistical ideas with technology. In designing learning experiences, teachers draw upon TSK, PSK, and SK to appropriately use technology to improve students' learning of statistics.

METHODS

Context and Participants

This paper uses data collected as part of a larger study of PSMTs' use of representations in statistical problem solving while learning to use dynamic statistical software tools in the Lee et al. (2010) materials at eight institutions (see Lee et al., 2012). For this study, we examined data from secondary mathematics teacher education courses at four large teacher preparation programs at US universities. Faculty at these universities received professional development and all were using the Lee et al. (2010) materials in their courses. What is reported here is based on analysis of PSMTs' work on a task from Chapter 4 using *Fathom* (Finzer, 2002, v. 2.1) (see Figure 2).

Task: Explore several of the attributes in the 2006 Vehicle data set.

a) Generate a question that involves examining relationships among attributes that you would like your future students to investigate.
b) Use *Fathom* to investigate your question. Provide a detailed description of your work and your response to the question posed in part *a*. Include copies of plots and calculations.

Figure 2. Chapter 4 Task

At each of the four institutions, the PSMTs worked individually to complete the task and submit a document describing their work, including screenshots, their thoughts along the way, and their final claim. A total of 81 documents were collected and blinded. Seven documents were removed from the data corpus because it appeared the PSMTs interpreted the task to mean they should create an activity aimed for students. Thus, 74 documents from four universities were analyzed (n=16, 9, 23, & 26, respectively).

Data Analysis

While the Lee et al. (2010) book provides opportunities to develop the four aspects of teachers' knowledge in Figure 1, the sample of their work analyzed for this paper could only provide insight into three of the four. Thus, to analyze opportunities for PSMTs to develop understandings related to SK, TSK, and PSK, we examined and coded Chapter 4 of the book (pp. 69-101) and the facilitators' guide. We made note of the big ideas under each category and looked for themes until the big ideas could be collapsed into as few as possible. Once these were identified, we chose a few critical big ideas as a focus for coding the PSMTs' actual work. Table 1 includes examples from our text analysis.

Table 1. Examples from textbook analysis

Focused Big Ideas for Analysis	Examples from Text Materials (p. 80-81)
Statistical Knowledge ● Engaging in exploratory data analysis; ● Focusing on distributions and describing data as an aggregate; ● Coordinating measures of center and spread in distributions	Statistical Knowledge Q16. What does the value of the correlation coefficient imply about the relationship between City and Hwy mpg? Q19. Based on where you placed the moveable line, use your equation with the specific values of slope and y-intercept to predict Hwy mpg for a vehicle with a City mpg of 31.
Technological Statistical Knowledge ● Automating computations and graphs ● Exploring data	Technological Statistical Knowledge Q17. Insert a movable line and adjust it so that you feel it best models the data. Describe the method you used for determining where to place the line.
Pedagogical Statistical Knowledge ● Planning for projects and discussions that use multivariate datasets that allow for different approaches by students	Pedagogical Statistical Knowledge Q22. Think of two strategies you could employ to help students understand that differences in solutions are acceptable and expected in the context of trying to estimate a linear model. How could you capitalize on this difficulty should it arise?

In coding each document, teacher's work was initially analyzed in order to identify cycles of EDA in four phases: Choose Focus, Represent Data, Analyze/Interpret, and Make Decision. By initially identifying each phase, and indicating how many times a teacher may cycle through phases, we were identifying important aspects of SK. Within each phase, several codes were then used to characterize teachers' work (e.g., questions posed, type of representations and statistical measures, type of graphical augmentations, what was noticed, interpretations they offered, and claims they made). Within TSK, our focus is on how teachers used technology to engage in EDA to: 1) create graphical representations of data, 2) examine and visualize measures (e.g., mean, regression equation), 3) use graphical augmentations (e.g., change case icons, show squares on least squares line), 4) examine subsets of data (e.g., remove outliers or filter cases with certain characteristics), and 5) link multiple representations.

RESULTS

Teachers' TSK

We chose to focus on two categories of TSK - *automating computations and graphs*, and *exploring data*. Automating includes creating graphs, and computing and displaying statistical measures. Exploring data includes using multiple representations and graphical augmentations.

Automating computations and graphs. To describe the ways in which PSMTs' automated the creation of graphs and computation of measures, we coded the number and type of graphical representations, the number and type of statistical measures, and whether measures were added to a graph or summary table. The most common graphs created were scatterplots (76%), box plots (32%), and dot plots (36%). Most PSMTs added statistical measures to graphs, with the most common (60%) being a least squares line, not surprising given a task analyzing relationships. PSMTs (42%) also added measures to summary tables, the most common being correlation.

Exploring data. To describe PSMTs' ability to use multiple representations, we coded the total number of representations, the ways in which they augmented graphs, their use of dynamic linking, and whether they examined subsets of data. Most PSMTs (76%) created one or two unique types of graphical representations and between two to four total graphical representations during their investigation (64%). Only a few PSMTs used augmentations to illustrate variation from a linear model through showing squares (12%) and residual plots (7%), while 19% added moveable lines to estimate a linear model. The use of changing the color or symbol of case icons to examine a third variable was done by 23% of PSMTs. We found it interesting that a small group of PSMTs (12%) used advanced techniques such as filtering to examine a subset of data, and only 24% of PSMTs dynamically linked representations.

Teachers' SK

To understand the PSMTs' SK we examined how they tended to use various aspects of TSK discussed above in their statistical investigation, including engaging in exploratory data analysis, focusing on the aggregate, and coordinating measures of center and spread. A large proportion of teachers (72%) asked a broad question that was open and allowed for different approaches, or a focused question that was still open to various approaches but specified exact variables to explore. About 82% of teachers engaged in two or more cycles of investigation, regardless of the type of question they posed, illustrating a sense of how to explore data with technology. For the majority of PSMTs, the graphs and measures they generated were aligned with questions they posed; however, since some did not mention these explicitly, it was unclear how their graphs and measures were used for analysis.

Of the PSMTs that used dynamic linking, most used this action to draw a conclusion about a relationship among variables; typically using techniques explicitly taught in the book to link subgroups of data in univariate and bivariate graphs, as well as linking tables and graphs. However, some PSMTs showed evidence of exploring relationships through linking, but did not comment on a conclusion. The use of changing case icons in a graph seemed to increase PSMTs' sophistication in their analysis to consider the role of additional variables. However, we also noticed that PSMTs' use of moveable lines and showing squares from a linear model seemed too routinized as a demonstration of a technique they learned in the book, and not a tool to offer better analytic perspectives on the data as an aggregate.

An additional focus for our analysis was how well PSMTs attended to the context of data in making claims. About 13% never considered context in their work, and 46% of PSMTs reported the relationships they observed using variable names, but did not make strong connections to the context. Stronger connections to context were observed in the work of 41% of teachers. This may indicate that many PSMTs did not keep the purposes of their questions, which seek to understand relationships in a context, at the fore when conducting an investigation and making claims.

Teachers' PSK

With a specific focus on PSMTs' readiness to plan for and implement a statistical investigation with students, we can gain insight from their work at the beginning of an investigation (question posing), middle (analysis techniques), and end of an investigation (making contextual-based claims). The majority of PSMTs (72%) are able to pose questions for students that are open ended to allow their students to engage in EDA. Most of the PSMTs also illustrated the ability to engage in multiple cycles of an investigation, using more than one representation (with some linking among them), computing appropriate measures, and utilizing graphical augmentations that could support their analysis. However, as noted above, their analysis was often shallow and did not make strong connections to context. Thus, it is not clear that these PSMTs have illustrated, in this *one* task, the development of strong PSK.

CONCLUSION

PSMTs using Lee et al. (2010) had opportunities to develop aspects of SK, TSK, PSK, and TPSK. Their work on the task seemed to indicate a strong developing TSK through appropriate uses of representations and measures in EDA, and a good sense of how to pose open-ended questions. For PSMTs first learning to use dynamic statistics software, it is not clear they can use their new technology skills to fully support their own investigation of relationships among variables. This may indicate they also would not be ready to engage students in EDA using dynamic and advanced features of *Fathom*. The development of TPSK requires a rich interconnected understanding of how to draw from one's SK, TSK, and PSK to engage in learners in learning and doing statistics with technology. Further research on how PSMTs develop and implement statistical investigations with dynamic software tools in classrooms could add significant insight into TPSK development.

REFERENCES

Batanero, C., Burrill, G., & Reading, C. (Eds.) (2011). *Teaching statistics in school mathematics-Challenges for teaching and teacher education*. Dordrecht, The Netherlands: Springer.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick & F. K. S. Leung (Eds.), *Third International handbook of mathematics education* (pp. 643-689). New York: Springer.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1). http://escholarship.org/uc/item/8sd2t4rr

Finzer, W. (2002). *Fathom Dynamic Data Software* (Version 2.1). [Computer Software]. Emeryville, CA: Key Curriculum Press.

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education, 38*(5), 427 - 437.

Lee, H. S., Kersaint, G., Harper, S., Driskell, S. O., & Leatham, K. R. (2012). Teachers' statistical problem solving with dynamic technology: Research results across multiple institutions. *Contemporary Issues in Technology and Teacher Education, 12*(3), 286-307.

Lee, H. S., & Hollebrands, K. F. (2011). Characterising and developing teachers' knowledge for teaching statistics with technology. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education* (pp. 359-369). Dordrecht, The Netherlands: Springer.

Lee, H. S., Hollebrands, K. F., & Wilson, P. H. (2010). *Preparing to teach mathematics with technology: An integrated approach to data analysis and probability*. Dubuque, IA: Kendall Hunt.

Madden, S. (2011). Statistically, technologically, and contextually provocatively tasks: Supporting teachers' informal inferential reasoning. *Mathematical Thinking and Learning, 13*, 109-131.

Makar, K., & Confrey, J. (2005). Variation talk: Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27-54.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017-1054.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1009). Charlotte, NC: IAP.