IDENTIFYING A RESEARCH AGENDA: THE INTERACTION OF TECHNOLOGY WITH THE TEACHING AND LEARNING OF DATA ANALYSIS AND STATISTICS

Susan N. Friel
UNC-Chapel Hill

This paper provides an overview regarding the need to identify a research agenda by addressing the following questions: (1) What do we know about the content of data analysis and statistics to be developed at different levels, K–12? (2) In what ways can technology tools enhance current and new directions in teaching and learning data analysis and statistics? (3) What is the role of empirical research in clarifying the interactions between software development and use and the teaching and learning trajectories K–12 in data analysis and statistics? (4) What are the needs and directions that can help frame a research agenda?

Data analysis and statistics have emerged as major topics in primary and secondary (K–12) school mathematics curricula during the 1990's (NCTM, 1989; NCTM, 2000). Statistics has lacked definition at the K–12 levels. The lack of clarity about what content to address has resulted in initial work focusing on how we might take more traditional statistics and translate the content for use with younger students. Increased attention has been given by researchers and curriculum developers to setting better directions for what we want K-12 students to know and be able to do with respect to data analysis and statistics and to defining the nature of instruction needed to support these directions.

More recently, the interaction of technology with efforts to redefine both the content and instructional practices regarding data analysis and statistics in K–12 has provided new directions. Educational technology affords us a greater variety of strategies for *teaching* statistics and, at the same time, offers us new ways of *doing* statistics (Garfield & Burrill, 1997). The role of research must be addressed now, and the opportunity for defining and teaching a new content area with this kind of technological support must be grounded in research as this content is incorporated into school curricula.

What do we know about the content of data analysis and statistics to be developed at different levels, K-12?

Statistics is a vital, albeit relatively new, part of the K–12 curricula. Since the *Standards* (NCTM, 1989) have been in place, statistics and probability have become recognized topics in the K–12 curricula. Before 1989, most statistics and probability coursework and research occurred at the post-secondary levels, with an emphasis was on research about the understanding of probability concepts (Shaughnessy, 1992; Shaughnessy, Garfield, & Greer, 1996). The *PSSM* (NCTM, 2000) contains recommendations about specific expectations for each of four grade ranges (pre-K–2, 3–5, 6–8, 9–12). Schaeffer (2000) summarizes recommendations for content across K–12 curricula from NAEP and adds, as well, his own suggestions. In the elementary grades, there is an emphasis on a process of data analysis, making graphs, and using measures of center. At the middle grades level, the emphasis continues on these topics, with more sophisticated uses of graphs and introduction of association and sampling. At the high school level, content includes distinctions between univariate and bivariate data; regression coefficients, regression equations, and correlation coefficients; sample statistics versus population statistics; simulation; and integration of data analysis with content such as algebra.

Recent attention and emphasis has been given to characterizing and defining the big ideas that need to be considered (Figure 1). Data analysis may be characterized as an iterative, four-stage process that includes asking a question, collecting the data, analyzing the data, and forming and communicating conclusions. Within the context of data analysis, there are several big ideas related to statistics that must be considered, several of which are detailed in Figure 1.

In addition, technology is hinted at in recommendations, but assumptions about its appropriate use and availability to support teaching and learning have been limited by vision, versatility of software, and accessibility to hardware. Most school-appropriate technology tools used for data analysis and statistics fall into the category of spreadsheet software (e.g., Appleworks, Excel) or graphing tools (e.g., graphing calculator, Cricket Graph) that offer similar functions, including limited plotting, graphing, and analysis capabilities. They restrict the user to conventional displays, and emphasize numerical over categorical data.

In what ways can technology tools enhance current and new teaching and learning data analysis and statistics?

While technology has long been available to analyze statistics, the role of technology in teaching and learning statistics at the K–12 levels is still in its infancy. Serious integration of technology data tools in teaching and learning statistics provides a catalyst for an array of other changes, including changes in curriculum, classroom discourse, and students' ways of learning.

With the increased capabilities and availability of technology tools, it is important to consider their most appropriate use in facilitating students' learning of statistics in different situations. Ben-Zvi (2000) describes how technological tools are now being designed to support statistics learning:

1. Students actively construct knowledge, by 'doing' and 'seeing' statistics.
2. Students have opportunities to reflect on observed phenomena.
3. Students develop their knowledge about their own thought processes, self-regulation, and control.

Bakker (2002) has articulated the need to distinguish between software used to do data analysis and software used to learn data analysis. He points out that professional statistical software packages are not suitable for use by students when they are learning data analysis. How can a user choose among histograms, box plots, or circle graphs if they do not yet understand what these representations are and when each would be useful? For this reason special software needs to be and is being designed (e.g., Tinkerplots, Fathom, Tabletop, Minitools) that enhances learning.

Ben-Zvi (2000) sees computers as cognitive tools, that is, as tools that help transcend the limitations of the human mind. Cognition tends to be situated in context; cognitive development involves both the individual mind and the development of knowledge through socially structured activities. He makes the point that this concept leads to specific ways of using computers in education (pp. 139-143):

Technology is an amplifier of statistical power: In learning environments that are not based on the use of technological tools, graphs or tables are either presented to students or constructed by students according to prescriptive instructions. With the use of multi-representational technological tools, many of the standard data manipulations are automatic operations. Students produce a variety of different representations, ones that often reflect their emerging understandings of the data and the context in which the data are situated.

Technology as a reorganizer of physical and mental work: The appropriate use of technology has potential to bring about structural changes in the system of students' cognitive and sociocultural activities, rather than just to amplify human capabilities. Such powerful tools bring about reorganization of physical or mental work in a variety of ways.

What is the role of empirical research in clarifying the interactions between software development and use and the teaching and learning trajectories K-12 in data analysis and statistics?

Integrating Research: Software Development and Students' Statistical Thinking

Research on statistical thinking with students in grades K–12 has been sparse if nonexistent for quite sometime. However, in the more than 10 years since the release of the Standards (NCTM, 1989), research in statistical thinking has emerged as an exciting option and has begun to yield models of students' conceptions that are detailed enough to have practical, pedagogical implications. Further, powerful new software tools designed explicitly for statistics education could make statistical thinking accessible to students in K–12 in ways not before considered.

The design of technological tools and contexts for their use to support statistical reasoning and learning is easier said than done. To design computational environments well requires the "intertwining of many different threads of thought" (Resnick, 1995, p. 31). Resnick identifies three major threads (p. 31).

- *Understanding the domain knowledge*: What are the knowledge and skills that make up the domain of knowledge? How might we approach this domain knowledge differently? In what ways do technological tools recast areas of knowledge, thus providing new ways of thinking about the domain's concepts and allowing learners to explore previously inaccessible concepts?
- *Understanding the learner*: What is the learner's existing framework? How will the learner integrate new experiences into this framework? In what ways might learners construct new concepts and new meanings and how might the technological tool provide direction and scaffolding to support this process?
- *Understanding computational ideas and paradigms*: Technology is not only the medium the computational designer uses to craft artifacts. The computational environment, itself, involves a set of powerful new ideas for students. So, the command structures and the relationships among actions in such software tools as Tabletop, Fathom, or Tinkerplots highlight some of the more generalized understandings of how one might function as a data analyst.

Rubin (2001) suggests two research methodological perspectives and their integration and interaction can be used to frame this work.

> Software development as a research-based endeavor: As argued by Clements and Battista (2000), the state of the art in both models of thinking and software design make it possible for "research and software design to be a more intimately connected, mutually supporting process" (p. 762). The research/design cycle can create a synergy that enriches and accelerates progress in both fields.
>
> The role of conjecture-driven research: As described by Confrey and Lachance (2000), conjecture-driven research begins with a "means to reconceptualize the ways in which to approach both the content and pedagogy of a set of mathematical topics." Such research is most often carried out in the context of a teaching experiment, during which the conjecture is continually revisited and modified in response to students' questions, discoveries, and insights. Conjectures are generative, not restrictive; they lead to more sophisticated conjectures, not necessarily to the proof of a hypothesis.

Selected Findings Related to Software Development and Use

The possibility of students' forming intelligent partnerships with technology in studying statistics gives them the potential to work at a level that may be impossible without technology (Jones, 1997). Clearly, the argument for integrating software development and research of students' knowledge and understanding is persuasive. To a great extent, recent work on the use of technology has addressed this need with varying levels of specificity. Fifty articles (ICOTS-5, ICOTS-6 proceedings; Conference on Research on the Role of Technology in Teaching and Learning Statistics; full reference list to be provided at PME) focused on the use of technology in teaching and learning statistics and data analysis have been reviewed. In addition, for selected articles that appeared to have a focus on student learning, the main findings have been summarized. The major content areas addressed in the research contributed to framing the "big ideas" summarized in Figure 1 (those that occurred most often). A variety of software tools were reported to be the focus of research.

When these articles were reviewed, reported evidence indicated that computer and calculator tools allow students to:

- Rapidly graph and display data for easier analysis.
- Easily access displays, multiple linked representations, simulations, and animated and/or interactive demonstrations of statistical concepts.
- Easily access large amounts of organized data from official sources.
- Problem-solve and receive immediate feedback.
- Use larger and more complex data sets than feasible when work done by hand.
- Creatively develop ideas and learn about structuring data by experimenting with tools before being presented with conventional methods.
- Focus on concepts instead of doing complex calculations by hand.
- Solve problems without having to know complex calculations.
- Rapidly simulate data for modeling.
- Rapidly transmit and share data.
- Explore real-life applications of statistics.

This list provides a good summary of the ways technology contributes—on a surface level—to students' access to and learning of data analysis and statistics. It is also more of behaviors and actions than the development of statistical understandings. For example, statements focused more substantively around big ideas were not the norm, such as a *possible* statement that might emerge on students' actions and thinking about lines of best fit supported by explorations using technology:

> The use of a movable line on a scatter plot helped students explore what finding a line of best-fit might mean. Student discussions following this experience were focused on the clustering or lack of clustering of points around an apparent line and what this might mean for the relationship between the two variables being investigated.[1]

However, there are bits of information that would link research findings related to software use back to the development of big ideas, for example:

- *Software tools with ready-made methods influence the way a subject matter problem is conceived of and is transformed into a "statistical problem" and into a "problem for the software" (Biehler, 1997).*
- Students working in pairs on problems dealing with graphing are more able than their classmates working alone to make the critical inferences crucial to learning from problem-solving experiences (Jackson, Edwards, & Berger, 1993b).
- On a project, students used more mathematics, and took less time, largely because technology took care of graphing and laborious calculations (Erickson, 2002).
- The graphing facilities of *Microsoft Works* are often overwhelming and confusing to students, as when the basic "new chart" menu command is selected, the program immediately confronts the user with a large number of simultaneous choices, thus requiring much advanced planning (Jackson et al., 1993).
- With *Cricket Graph*, the first choice made by the user, the type of graph, is irrevocable. Many students find this frustrating, as they are still learning about how to choose an appropriate graph, and they don't want to start from the very beginning again (Jackson et al., 1993).
- Many students will choose the default representation in the software without thinking about their choices (Jackson et al., 1993).
- Certain sequences of actions (e.g., sort the data, re-graph, move or resize, choose a different graph type) taken by students had correlations to whether the resulting graph turned out basically good or poor (Jackson, Berger, & Edwards, 1992).

Issues about student thinking and use of technology in teaching data analysis/statistics also surface:

- There is a need to make the interface visually simple and clear so that the students pay less attention to the tools and more to the task and target concepts.
- When presented with many choices, such as in an open construction tool, students may experience cognitive conflicts
- Ideas learned may be too deeply connected to specific tools (i.e., learning the tool rather than learning the concepts to be learned through the tool)
- Students often use the display methods offered by the software rather than thinking through the best representations for the purpose of the investigation
- The graphical representation used strongly affects a student's reasoning about a set of data
- The improved performance of a student who is in a computer partnership is necessary but not sufficient to demonstrate learning.

What are the needs and directions that can help frame a research agenda?

---

[1] This statement is hypothetical; there is no evidence for its validity in any research reviewed although there are software options that make this kind of investigation of student understanding a realistic option.

Identifying the questions

We have just begun to understand the questions we want to ask about what it means to know and be able to do data analysis and statistics at the K-12 grade levels. With a shift away from the automatic transfer of traditional content from post-secondary to K-12 and a move toward reconceptualizing this content and how it is learned, we find ourselves in a state of flux. When we add to this the need to consider the interactions between technology, content and learning, given the new kinds of software being developed, there are rich opportunities for framing questions. Suggestions will be provided during the session with opportunities for the audience to add to the list.

Knowing what we know

Part of the dilemma with respect to identifying questions includes the need to know what we know. In a recent funded proposal to NSF, Rubin (2000) provided a short but rich summary of the literature to date focused on the concept of variation in the context of technology use. We need to have these kinds of summaries available to the broader research and software development communities. This need can be met, in part, through access to reviews of the literature related to the big ideas (e.g., Friel, Curcio, & Bright, 2001; Meletiou, 2002), but such reviews are not quickly completed and updating to reflect new research must be a continual process.

Locating the reported research (that includes or excludes the integration of technology) is problematic as well. Much of the current research is reported in conference proceedings (e.g., *ICOTS-5, ICOTS-6, Garfield & Burrill, 1997*) that are sometimes not easily accessible. While more of these publications are being made available online, anyone looking for this information must have a knowledge of the organizations and ways of sharing information that currently exist within the community.

The ideal would be to have available an annotated bibliographic electronic database organized around the big ideas of statistics—a repository of abstracts of research (e.g., Huntley, Zucker, & Estey, 2000) or a more comprehensive database upon which to build a resource bank of related research references and resources. Such a resource would serve as an evidence-based repository that could be used to inform research directions and could be updated to reflect results of new work as it is added to the field.

New issues and ideas emerge when technology is used

Beyond addressing the basic needs of sorting out research questions, providing summaries of the research related to the big ideas in data analysis and statistics, and making the current literature readily accessible that would help in framing a research agenda, the interaction among technology, content and student thinking emerges as an arena that is rich in possibilities for research. The use of the technology itself surfaces perplexities about content and student thinking only because students now can work in such rich investigative environments (e.g. Fathom, Tabletop, Tinkerplots). The very capabilities that these software tools provide raise a multitude of questions about content and how to think about the big ideas in the domain of data analysis and statistics. Examples will be provided with demonstrations of software as part of this session.

We are in a state of flux about exactly what is the content to be addressed at what levels, K-12. We are only now gaining access to new software tools that will push for understanding of substantially richer conceptions of the big ideas of data analysis and statistics. There is a need to include research that clarifies *the impact on students' knowledge*. The process of theory building contributing to practice and vice versa is very much a real phenomenon in the arena of data analysis and statistics education.

Data Distribution Variability Trend Covariation Sampling Model Representations
Measures of central tendency or location Measures of spread and dispersion
Actions on data

Figure 1: Selected Big Ideas: Data Analysis and Statistics
(Cobb, 1999; Cobb, McClain & Gravemeijer, In press; Cziko, 2002; Garfield, 2001; Hancock, Kaput, & Goldsmith, 1992; Konold & Higgins, 2002, in press; and Konold & Pollatsek, 2002)

References

Bakker, A. (2002) Route-type and landscape-type software for learning statistical data analysis. ICOTS 6

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1&2), 127-155.

Biehler, R. (1997). Students' difficulties in practicing computer-supported data analysis: some hypothetical generalizations from results of two exploratory studies. In J. Garfield and G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (176-197). Voorburg.

Clements, D. H. & Battista, M. T. (2000) Designing effective software. In A. E. Kelly & R. A. Lesh (Eds.) *Handbook of research design in mathematics and science education* (pp. 761-776). Mahway, NJ: Lawrence Erlbaum.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning, 1*, 5-43.

Cobb, P., McClain, K., & Gravemeijer, K. (in press). Learning about statistical covariation. *Cognition and Instruction*.

Confrey, J. & Lachance (2000) Transformative teaching experiments through conjecture-driven research design. In A. E. Kelly & R. A. Lesh (Eds.) Handbook of research design in mathematics and science education (pp. 231-266). Mahway, NJ: Lawrence Erlbaum.

Cziko, G. (Fall, 2002). Educational Psychology 390: Elements of Statistics. Web-CT course from University of Illinois at Urbana-Champaign, web site address is http://www.ed.uiuc.edu/courses/EdPsy390A/notes/l04.htm.

Erickson, T. (2002). Technology, statistics, and subtleties of measurement: bridging the gap between science and mathematics. In B. Phillips (Chief Ed.) *Developing a statistically literate society: Proceedings of the sixth international conference on teaching statistics.* Voorburg.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education 32*, 124-58.

Garfield, J. B. (2001) *The big ideas of data analysis*. Notes from presentation to *Tinkerplots* Advisory Board Meeting, July, Madison, WI.

Garfield, J. B. & Burrill (Eds.) (1997) *Research on the role of technology in teaching and learning statistics* (Proceedings of the 1996 IASE Round Table Conference, Granada, Spain*)*. Voorburg, The Netherlands: International Statistical Institute.

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337-364.

Huntley, M. A., Zucker, A. A., & Estey, E. T. (January 2000) *A review of research on computer-based tools (Spreadsheets, graphing, data analysis, and probability tools), with annotated* *bibliography*. Arlington, VA: SRI International, Project PO3377. (http://www.sri.com/policy/ctl/assets/images/Tools_review.pdf)

Jackson, D., Berger, C., & Edwards, B. (1992). Computer-assisted thinking tools: problem solving in graphical data analysis. *Journal of Educational Computing Research,* 8(1), 43-67.

Jackson, D., Edwards, B., & Berger, C. (1993). The design of software tools for meaningful learning by experience: flexibility and feedback. *Journal of Educational Computing Research*, 9(3), 413-443.

Jones, P. (1997). Examining the educational potential of computer-based technology in statistics. In J. Garfield and G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (238-246). Voorburg.

Konold, C. & Higgins, T. (2002). Highlights of related research. In S. J. Russell, D. Schifter, & V. Bastable, with C. Konold & T. L Higgins, *Developing mathematical ideas: Working with data*. Parsippany, NJ: Dale Seymour Publications.

Konold, C. & Pollatsek, A. (2002) Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33,* 259-289.

Meletiou, M. (2002). Concepts of variation: A literature review. *Statistics Education Research Journal, 1*: 46-52, (http:/fehps.une.edu.au/serj).

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

NCTM (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Resnick, M. (1995). New paradigms for computing, new paradigms for thinking. In A. A. diSessa, C. Hoyles, & R. Noss with L. D. Edwards (Eds.). *Computers and exploratory learning* (pp. 31-43). New York: Springer-Verlag.

Rubin, A. (2001) *Visualizing statistical relationships: An exploration of the role of interactive visualizations in the understanding of statistical relationships.* ROLE Proposal funded by National Science Foundation.

Scheaffer, R. L. (2000). Statistics for a new century. In M. J. Burke & F. R. Curcio (Eds.) *Learning mathematics for a new century (NCTM 2000 Yearbook)* (pp. 158-173). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. Grouws (Ed.) *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillian.

Shaughnessy, J. M., Garfield, J. and Greer, B. (1996). Data handling. In Bishop, A. J., Clements, K. Keitel, C, Kilpatrick, J. and Laborde, C. (Eds.) *International handbook of mathematics education* (Part 1, pp. 205-237). Dordrecht, The Netherlands: Kluwer.