COMPARING TWO DISTRIBUTIONS: INVESTIGATING SECONDARY TEACHERS' STATISTICAL THINKING ®

Katie Makar and Jere Confrey
University of Texas at Austin
USA

This paper highlights the statistical thinking of teachers in analyzing their own students' high-stakes test data. The research here emphasizes the impact that an immersion model for teachers–doing statistics as statisticians–can provide in raising statistical content knowledge, engendering a mindset of inquiry, developing facility with technology, and enriching understanding of student outcome data. Four constructs – measurable conjectures, tolerance for variability, context, and inference & conclusions – provide the basis of a taxonomy to describe teachers' statistical thinking about comparing two groups in the context of an accountability system.

INTRODUCTION

While many schools are increasing their emphasis on statistics, very few are taking the necessary steps to help teachers master the statistics they are expected to teach. District-based professional development provides mathematics teachers with little opportunity to improve their content knowledge while university statistics courses are rarely aimed at content teachers feel is relevant. The research described here was conceived as a mathematical parallel of the Writers Workshop from the National Writing Project, where teachers learn how to *write* rather than how to *teach* writing. A team at the Systemic Research Collaborative for Education in Mathematics, Science, and Technology in the College of Education at the University of Texas-Austin designed the NSF-funded research. Our long-term conjecture is that when teachers are immersed in content beyond their curriculum in a context they find compelling, this experience transfers into improved classroom practice. In addition, as a result of these experiences, we should see teachers more willing and able to use technology in their classes. The research project had a set of four related objectives to: (1) Strengthen *teacher content knowledge in statistics* by giving them the opportunity to learn statistics well beyond their curriculum; (2) Immerse teachers in *focused investigations* about student data in a high-stakes accountability environment; (3) Build *teacher confidence in using technology*; and (4) Orient teachers with a *healthy mindset about data and inquiry*-the acceptance of uncertainty when searching for solutions, and a critical eye towards the limitations and misuses of statistics and inferential reasoning.

Statistics education is a fairly new field; most studies have focused on student reasoning in statistics. Very little research has been done on *teachers'* statistical thinking. The research project described in this paper is based on an immersion model of teachers doing statistics as *statisticians* and investigates teachers' statistical reasoning about comparing two groups.

METHODOLOGY AND CONTEXT

Texas, where the study took place, has a high-stakes accountability system. Students are tested annually and graduation depends on passing this test. In addition, schools and teachers are held accountable for their students' performances. Urban schools that serve less academically advantaged children are under constant scrutiny to ensure they do not receive unacceptable ratings. As a result, much of these schools' professional development time is spent reviewing test results. While they are given their students' results, teachers seldom have the opportunity to discuss and debate their interpretation in an effort to guide instruction. This context seemed ripe for us to invite teachers to examine their statistical data as investigators.

Our participants consisted of the mathematics department in a mostly Hispanic, urban middle school that feeds into a low-performing high school. The project consists of three phases. Phase I consisted of 2 full day and 3 after school workshops where teachers learned descriptive statistics, the basics of *Fathom* (Finzer, 2001), and became acquainted with their student data. Throughout the Phase II, a two-week summer institute, teachers built a richer conceptual understanding inference through discussion, problem-based investigations with their student data, and simulations in *Fathom*. Sampling distributions were used frequently to provide evidence for differences in groups and to imbue a tolerance for variation. This made it easier to provide a conceptual foundation for discussing and using confidence intervals, p-values, and t-tests without focusing on formulas and rules. In the second week of the summer institute, teachers investigated a problem of their own choosing and presented their findings on the final day of the workshop to their peers. Issues of accountability, reactions to assigned readings, and issues of data were discussed regularly. The third phase, a follow-up to assist participating teachers in campus data projects and classroom teaching of data analysis, is being planned. Classroom observations were made before teachers began the study providing a baseline of their teaching practice. A pre-post test of statistics measured growth in content knowledge. Clinical interviews at the conclusion of Phase II provided important triangulation of additional data collected from the session videos, observations, pre-post tests, and final presentations.

COMPARING TWO GROUPS

As statistics moves to the forefront in education, much interest is developing around the process of comparing two groups (Watson & Moritz, 1999). Several new curriculum projects introduce statistics through the process of comparing two groups (TERC, 1998; Cobb et al., in press). Not only is this motivation to study statistics, but makes it necessary for students to consider not only measures of center, but also the issue of variability: Are the differences in center between the groups *meaningful*? Finally, comparing two groups previews an important concept later developed in introductory college statistics courses: statistical inference. What kind of reasoning is needed to compare two groups? Beyond the computational distinction made through descriptive statistics, comparing two groups requires several important concepts and for our purposes is divided into four constructs: *measurable conjectures, tolerance for variability*, understanding of the *context*, and an ability to draw *conclusions* and/or inferences based on data. Below is a brief summary of each of these constructs.

Creating *measurable conjectures* requires teachers to move from the problematic to a conjecture. This process, often ignored, is essential to using statistics for problem solving. Even teachers who assign projects rarely ask students to consider the messiness of moving from problem to conjecture to data collection. The alignment of these three areas is one of the most difficult skills to master in applied statistics, but provides insight into how data constraints affect problem solving and provides students with a critical eye towards bias in data sources.

Descriptive comparison, much of which requires a *tolerance for variability* (including variability within a group, between groups, and from one sample to the next), requires a mindset very different from the one normally promoted in mathematics. Mathematics tends to focus on universal abstractions and ignores the uncertainty and complexity frequently encountered in life. As a result, traditional mathematics instruction can reinforce a deterministic view of data. The teaching of variation is seen as a grossly overlooked area of instruction in statistics (Meletiou, 2000), even though this neglect has been well documented in research (Shaughnessy, 1992).

A third critical concept needed to compare two groups is an *understanding of the context*. This is a vital area of consideration if the focus on inquiry-based learning is to become a reality in mathematics instruction. Wild and Pfannkuch (1998) argue that consideration of context is key to statistical thinking. Still, there is much focus in statistics instruction on mathematical skills, separating content from context (Gardner & Hudson, 1999). Over-mathematization of statistics could be an effect of the fact that statistics courses are usually taught by mathematics instructors. An example from our research project highlights the importance of context in creating a rich understanding of data (Confrey & Makar, in progress). In an activity early in the project, teachers compared several pairs of graphs without a context. The conversation was at a very superficial level and lasted only about 5 minutes. When the same graphs were examined again in light of a context relevant to the teachers (quiz scores of two different classes), a much more in-depth analysis took place in a discussion lasting 40 minutes. This discussion was the first time variation in a distribution was recognized as relevant. Through context, the teachers were able to begin to gain a more robust understanding of distribution as a tool to make sense of results.

Finally, comparing two groups becomes a powerful tool to draw conclusions and towards a consideration of *inferential reasoning*. Although formal methods of statistical inference are not taught in school-level statistics, an ability to look "beyond the data" (Friel, et al., 2001) and build arguments supported by data is a desired skill at any level. In addition, a basic conceptual foundation of statistical inference can facilitate understanding of one of the most difficult concepts in university-level statistics: sampling distributions (Chance et al., 2000).

These four important concepts-*measurable conjectures*, *tolerance for variability*, *understanding context*, and *a view towards inference*-together provide a portrayal of *mindset* of data. Wild and Pfannkuch (1998) describe statistical thinking as having four basic elements: anticipation of variation, ability to construct and use models, good statistical and contextual knowledge base, and ability to synthesize these elements to produce conjectures and inferences. A mindset of probing, evaluating, and describing, an awareness of contextual constraints involved, and a balance of curiosity and skepticism, they argue, are critical to applying statistics.

Using the above discussion of statistical thinking about comparing two groups, we can begin to think about a taxonomy that might be used to classify this kind of thinking in adults. Other areas of research on graphing data have looked at levels of reasoning, focusing on elementary and middle school students. Friel et al. (2001), describe levels of graph comprehension around students' abilities to "read the data, read between the data, and read beyond the data" (p. 130). A categorization for statistical thinking about comparing two groups (Makar & Confrey, in progress) describes five levels of reasoning that teachers use when comparing two groups. (1) At a *Pre-descriptive* level, no recognition of relationships between datasets is made, except based on individual data points or anecdotal evidence. If conjectures are made at this level, they are unmeasurable. (2) Teachers using a *Descriptive* level focus on summary statistics and make absolute comparisons between datasets with no regard for variability. Conjectures assume data is infinitely available to answer any question. (3) The first holistic view of the data occurs at the *Emerging Distributional* level, where informal qualitative descriptors of the data, along with basic summary statistics, are used to describe two datasets. Teachers begin to understand the difficulty in creating measurable conjectures, but are unable to successfully resolve the conflict and show frustration in attempting to write an appropriate conjecture. Variability, while acknowledged, is not understood beyond a descriptive level. (4) Teachers with a *Transitional View* of the data begin to understand the influence of variability in comparing two groups. More flexibility is shown (e.g. multiple graphical representations, alternative measures of center or spread) in comparing datasets at this level. Conjectures, while questionably measurable, have progressed to show elementary understanding of the difficulty in creating a conjecture that doesn't overly compromise the question at hand, but allows for possible collection of data. The concept of statistical *tendency* becomes part of the discussion and conclusion about data. (5) Finally, at the *Emerging Statistical* level, teachers gain confidence in using standard descriptive statistics to compare data sets, taking into consideration the differences between measures of center in light of the variability in the data and the sample sizes of the datasets. Conjectures demonstrate some ability to frame questions that balance data constraints with the problem at hand. Context and quantified descriptions are well integrated into conclusions and inferences may attempt to draw on statistical models, if relevant.

An important limitation to this model needs to be considered. The ideal situation would be one where the mathematical and contextual areas of statistical thinking develop simultaneously. Most statistical courses taught by mathematics faculty view statistics as an extension of mathematics and focus on formulas, often ignoring technological tools and real data. Data analysis is over-emphasized, and discussion (using mathematical vocabulary) is not embedded in the context of the situation. Likewise, since most teachers regard their practical experience and personal knowledge of their students as more productive than an analysis of data patterns, they may be immersed in context with too much emphasis on anecdotal evidence, ignoring mathematical tools of analysis. The assumption of the taxonomy is that a *balance* between these two extremes is developed in building statistical reasoning in learners of statistics.

ANALYSIS

In clinical interviews conducted at the end of the summer institute, teachers compared the test results of males and females at a local school on the state competency exam. The teachers' greatest competence was in the area of descriptive statistics/graphical representation, where teachers worked fluently with the software to choose appropriate graphs and summary statistics to describe the difference between the two groups. This was also the greatest strength for teachers entering the project, based on the pre-test results. Other areas showed moderate levels of achievement during the interview, consistent with post-test results. At the end of the institute, however, when participants gave a 30-minute presentation of a conjecture they investigated, it was harder than expected to get teachers to provide *statistical evidence* for their conjectures and to portray differences between groups in a non-deterministic way. Usually, teachers chose to support their hypotheses with descriptive statistics rather than more powerful statistical tools available to them. For example, none of the teachers used a sampling distribution to show whether differences they found between groups were significant. Rather, they used their intuition to determine whether the disparity "seemed" big enough for them to conclude that the groups were different. While the pre-post test results show that the teachers who remained in the study throughout gained a significant level of content knowledge, questions remain about the importance of this. First of all, given the opportunity to apply their understanding to their own investigations, two of the teachers regressed considerably on the taxonomy. What does this tell us about their ability to apply statistics in a real situation? Secondly the teachers who remained in the study began with a stronger content knowledge than teachers who dropped out. Of the eleven teachers who began the study, the four remaining were likely already of a mindset that this kind of training hopes to engender. How can we bring in the teachers who need the most change?

CONCLUSION

Statistical literacy has been heralded as a vital skill for citizens while misused, high-stakes accountability systems have the potential of retarding the growth of reform-based instruction. Rather than ignore what is clearly a system that is here to stay, it is imperative to work with the system to better inform it. This serves two purposes: students will be better served by teachers with stronger content knowledge and support for continuing to implement standards-based instruction; secondly, teachers and administrators will be able to make informed choices based on evaluation of data when implementing programs crucial to student learning. Through an authentic statistical experience in a context that is urgent and compelling to teachers, the research study has shown evidence of change in teachers' understanding of data analysis. This study is the beginning of a viable model for working with schools to develop their own literacy while submerged in data of their own students.

REFERENCES

Cobb, G., et al., (in press). *Statistics in action*. Emeryville, CA, Key Curriculum Press.

Chance, B., Garfield J., et al. (2000). *Developing simulation activities to improve students' statistical reasoning*. Auckland, NZ: Technology in Mathematics Education.

Confrey, J., & Makar, K. (in progress). *Secondary teachers' inquiry into data*. Austin: University of Texas.

Finzer, W. (2001). *Fathom dynamic statistics™ software.* Key Curriculum Press.

Friel, S., et al. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education 32,* 124-158.

Gardner, P., & Hudson I. (1999). University students' ability to apply statistical procedures. *Journal of Statistics Education, 7*.

Makar, K., & Confrey, J. (in progress). *Teachers' statistical reasoning about comparing two groups*. Austin: University of Texas.

Meletiou, M. (2000). *Developing students' conceptions of variation: An untapped well in statistical reasoning*. Unpublished doctoral thesis. Austin, University of Texas**:** 336.

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: MacMillan**.**

TERC (1998). *Investigations in data, number, and space*. Columbus, OH, Scott Foresman.

Watson, J., & Moritz, J. (1999). The beginning of statistical inference: comparing two data sets. *Educational Studies in Mathematics 37*, 145-168.

Wild, C., & M. Pfannkuch (1998). *What is statistical thinking?* Proceedings of the 5th International Conference on Teaching of Statistics. Voorburg, The Netherlands: ISI.