The Use of Intervals to Help Coordinate Understandings of Center and Spread: A Preliminary Report

Hollylynne S. Lee
North Carolina State University
hollylynne@ncsu.edu

J. Todd Lee
Elon University
tlee@elon.edu

Background and Statement of Research Issue

As data analysis, statistics, and probability are becoming more important components in middle and high school mathematics curricula, university teacher educators are faced with a challenge of how to best prepare preservice mathematics teachers to teach these concepts. The challenge is exacerbated by the fact that many of these preservice teachers do not have meaningful opportunities to develop an understanding of central statistical ideas. The GAISE project (Franklin et al, 2005) and the 2008 joint ICMI/IASE study are examples of focused efforts to help preservice teachers develop a deeper understanding of data analysis and probability, and the ability to use simulation and data analysis tools (e.g., Batanero, Godino, & Roa, 2004; Konold & Higgins, 2003; Stohl, 2005). Although simulation and data analysis tools (e.g., graphing calculators, spreadsheets, *Fathom*, *TinkerPlots*, *Probability Explorer*) may be available in K-12 classrooms, there is a need for high quality teacher education materials. Such materials can help teacher educators become comfortable with, and incorporate tools for, teaching probability and data analysis. However, the materials need to primarily aim for preservice teachers to develop a specific type of knowledge related to statistics that includes a deeper understanding of: (1) data analysis and probability concepts, (2) technology tools that can be used to study those concepts, and (3) pedagogical issues that arise when teaching students these concepts using technology.

The authors of this paper have been part of the project team engaged in an NSF-funded teacher curriculum development project to create modular course material to immerse technology and pedagogy instruction in various mathematical contexts[1]. The project intends to create three modules that could be distributed separately and used in college mathematics education methods courses, mathematics or statistics courses, or in professional development workshops to prepare teachers to teach mathematics or statistics with technology. The modules are not designed to be used directly by teachers with their age 11-18 students. Rather, it is anticipated that when teachers complete the modules they will have the knowledge needed to create their own activities to meet the needs of their students. The three modules focus on the teaching and learning of: 1) Data Analysis and Probability, 2) Geometry, and 3) Algebra.

The first completed module is focused on data analysis and probability. This module is aimed to support a broad audience of preservice teachers of age 11-18 students, and thus we made a purposeful decision to lay foundational ideas that would support understanding inferential statistics but to not include formal inferential statistics in the materials. For many preservice teachers, engaging in statistical thinking is a different process than that which they have been engaged in teaching and learning mathematics (e.g., delMas, 2004). Thus, it is important to engage preservice teachers as active learners and doers of statistical practices. The first module incorporates four big ideas that can be supportive of learning to teach data analysis and probability: (1) engaging in exploratory data analysis, (2) attending to distributions, (3) conceptually coordinating center and spread in data and probability contexts, and (4) developing an understanding of, and disposition towards, statistical thinking as different from mathematical thinking. For this paper, we focus on the third big idea and pose the following guiding question:

In the context of designing teacher education material, what elements may support preservice teachers developing a coordinated view of measures of center and spread?

Background Literature and Theoretical Perspectives

Developing Specialized Knowledge for Teachers

Teacher education and research on teachers has been greatly influenced by Shulman's (1986) idea of teachers' pedagogical content knowledge (PCK). More recently, Koehler and Mishra (2005) and Niess (2005) have described technological pedagogical content knowledge (TPCK) as the integration of teachers' knowledge of content, pedagogy and technology that is needed to effectively use technology to teach specific subject matter. Following from a model of the components of TPCK, H.S. Lee, Hollebrands, and Wilson (2007) propose a model that integrates mathematical and statistical content, technology, and pedagogy, with a focus on student thinking. The notions of TPCK are often displayed through a Venn diagram with a focus on the intersections of knowledge about content, pedagogy, and technology. Lee and Hollebrands (in press, 2008) proposed a different diagram to represent the development of teachers' technological pedagogical statistical knowledge [TPSK] as layered circles with a foundation focused on teachers' statistical thinking (Figure 1).

As depicted in Figure 1, the inner most layer consisting of elements of TPSK is founded on and developed with teachers' knowledge in the outer two circles. Developing knowledge in the outer two layers of statistical thinking and technological statistical knowledge is essential to, but not sufficient for, teachers having the specialized TPSK. The elements in each layer in Figure 1 are descriptors of the major foci of the knowledge, thinking, skills and dispositions we aim to develop as teachers' TPSK in the materials.

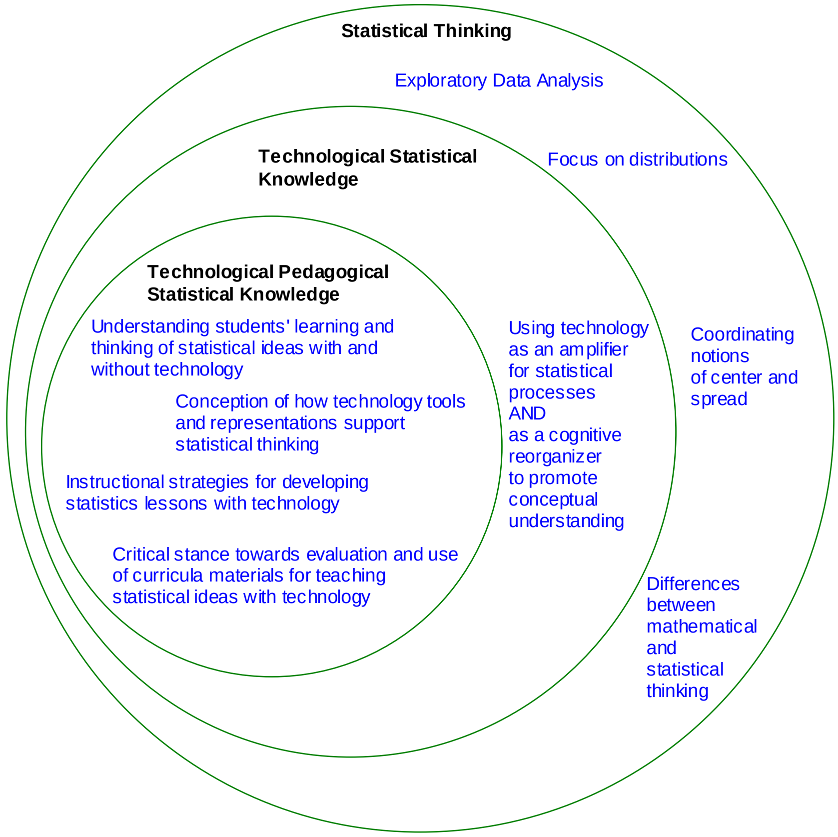


Figure 1. Framework for developing teachers' technological pedagogical statistical knowledge. (Lee & Hollebrands, 2008, p. 2)

Focus on Center and Spread via Intervals

For this paper, we focused on preservice teachers conceptually coordinating center and spread of a distribution in data and probability contexts, using technology tools that can help in this coordination, and developing the appropriate TPSK that would help their future students conceptually coordinate center and spread. The use of a single point value as an indicator of the center of a distribution (e.g., mean or median) or of an expected value in probability contexts has been over-privileged in both mathematics curricula and statistical research methods. Such single point estimates may not be accurate "signals" of what is likely an underlying "noisy" process (Konold & Pollatsek, 2002). Instead, Konold and Pollatsek propose a shift in attention to the importance of conceiving a sample as being drawn from a global process containing many factors and "noise", rather than drawn from a static population that remains unchanged. The “global process is a probabilistic one, unpredictable at the micro level” (p. 266). Thus, a sample can be conceived as a snapshot of the current stability of the evolving process—a perspective that also promotes thinking about the fundamental variability within the global process and its probabilistic nature.

Many argue that a useful foundational purpose of probability in statistics is to attend to the variation in results from a repeated probability experiment (e.g., Franklin et al, 2005; Reading & Shaughnessy, 2004; Saldanha & Thompson, 2002). Understanding variability within a sample and across samples can be fostered through attending to intervals as a way of describing the typical spread around some signal as a measure of a center. There have been many voices of concerns throughout research communities on the limitation of null hypothesis significance testing (which rely on point estimates). The medical industry has taken major moves towards examining and reporting data through alternative tools, confidence intervals being foremost (Gardner & Altman, 1986; ICMJE, 1997). Other areas, such as psychology and ecology (Fidler, 2006), as well as mathematics education (Capraro 2004) are also moving in this direction.

Just as measures of spread are useful in describing a distribution of data, in a probability context, it is useful to consider intervals that represent a reasonable range of values in describing distributions of data. In particular, intervals can be useful in describing variation from expected values within a sample, and variation of results across samples (Shaughnessy, 2006). The inverse relationship between sample size and variation from expected values is often counterintuitive for teachers and students. This relationship is important for teachers and students to develop and to use when considering the likelihood of a given result.

Design Elements

The following are several of the elements in the curriculum materials that can facilitate an understanding of measures of center and spread in a coordinated way:

1) Emphasize the theme of center and spread throughout each chapter in the material, with the coordination between the two explicitly discussed and emphasized with focused questions covering both content and pedagogical issues.
2) Use dynamic technology tools to explore this coordination.
3) Place the preference for intervals above that of single point values. Of course, the construction of these intervals is often reliant upon measures of center and spread in a confidence/credible interval fashion.

The authors of the materials (H. S. Lee, Hollebrands, & Wilson), with consultation from the advisory board and the content expert (J. T. Lee), attempted to attend to these elements, along with other design elements aimed at developing preservice teachers' TPSK.

Methods

In the research associated with these materials for preservice teachers, the author team has followed curricular design and research method cycles as proposed by Clements (2007). The materials have been through four iterations of classroom field-testing with preservice teachers, analysis of field-testing data, and subsequent revisions to materials. The module has been implemented in the 400-level course "Teaching Mathematics with Technology." This course serves middle and secondary preservice teachers and a few beginning graduate students with little experience using technology, with typical class sizes of 13-19. During the five-week unit on Data Analysis and Probability, the instructor in Fall 2005 (not one of the project PIs) used the pre-existing established curriculum for the course to serve as a Control group. The students took a pre-test and post-test designed to assess content, pedagogical, and technology knowledge related to data analysis and probability. Questions on the content section were selected from Garfield (2003) and others from the ARTIST database (app.gen.umn.edu/artist/index).

In each of the subsequent semesters (Spring 2006, Fall 2006, Spring 2007, Fall 2007), the five-week unit on Data Analysis and Probability was taught by the same instructor as Fall 2005. In addition, in Spring 2007, the Module was implemented in a section of the course taught by a different instructor. In the first three implementation semesters, written work was collected from students and pre- and post-tests were given. During the first two implementation semesters in 2006, the class sessions were videotaped and several students were interviewed.

For this paper, we are using several sources of data for our analysis of how preservice teachers may be developing a conceptual coordination between center and spread in data and probability contexts, particularly interval reasoning. Our data sources include: (1) examples of text material from the module, (2) a video episode from the first implementation semester (Spring 2006) in which preservice teachers are discussing tasks in Chapter 5 concerning conducting probability simulations, (3) preservice teachers' work on a pedagogical task that followed Chapter 5 material, and (4) results across all implementation semesters from selected questions of the pre/post-test.

Analysis and Results

Emphasis in Materials: Opportunities to Learn

To begin our analysis, we closely examined the current version of the materials for opportunities for students' to develop a coordinated conceptualization between center and spread. In data distribution contexts, the materials are designed to help preservice teachers understand measures of center like mean, median, and midrange with respect to individual deviations from those measures and for attending to variation (Chapters 1 and 2). In Chapters 3 and 4, the materials build from this notion in a univariate context to help students consider measures of variation in a bivariate context when modeling with a least squares line (e.g., residuals, sum of squares). Just as measures of spread are useful in describing a distribution of data, in probability contexts in Chapters 5 and 6, the materials emphasize intervals that represent a modal clump of values in describing distributions of data collected in a context with random variables, particularly variation from expected values within a sample, and variation of results across samples. The last two chapters help preservice teachers conceive that samples with smaller sample sizes are more likely to have results that vary considerably from expectations, while larger sample sizes tend to decrease this observed variation.

We closely examined the current version of the text materials to identify instances where we considered there was an *explicit* emphasis placed on coordinating center and spread in written text and technology screenshots, content questions, and pedagogy questions. We also specifically marked those places that emphasized a particular focus on interval reasoning (see Table 1). As an example of instances coded as focused on interval reasoning, consider those in Figure 2.

Table 1. Instances in Module of Coordinating Center and Spread

	Occurrences			% with focus on interval reasoning
	Text	Content & Tech. Task	Pedagogical Task	
Ch 1: Center, Spread, & Comparing Data Sets	3	5	2	50%
Ch 2: Analyzing Students' Comparison of Two Distributions using *TinkerPlots*	0	2	2	0%
Ch 3: Analyzing Data with *Fathom*	2	5	3	0%
Ch 4: Analyzing Bivariate Data with *Fathom*	5	3	3	0%
Ch 5: Designing and Using Probability Simulations	4	13	4	76%
Ch 6: Using Data Analysis and Probability Simulations to Investigate Male Birth Ratios	1	15	1	59%

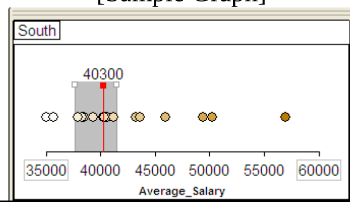
Written Text and Screenshots	Content and Technology Tasks	Pedagogical Tasks
Students may attend to clumps and gaps in the distribution or may notice elements of symmetry and peaks. Students often intuitively think of a "typical" or "average" observation as one that falls within a **modal clump**…Use the divider tool to mark off an **interval** on the graph where the data appear to be **clumped**. (Chapter 1, Section 3, p. 11)	Q17: Use the Divider tool and the Reference tool to **highlight a clump of data** that is "typical" and a particular value that seems to represent a "typical" salary. Justify why your **clump** and value are typical. (Chapter 1, Section 3, p. 13) [Sample Graph] 	Q19: How can the use of the **dividers to partition** the data set into separate regions be useful for students in analyzing the spread, center and shape of a distribution? (Chapter 1, Section 3, p. 14)
In our context, we are interested in how much the proportion of freshmen returning to Chowan College will vary from the expected 50%. To examine variation from an expected proportion, it is useful to consider **an interval around 50% that contains most of the sample proportions.** (Chapter 5, Section 3, p.102)	Q11: Given a 50% estimate for the probability of retention, out of 500 freshmen, what is a reasonable interval for the proportion of freshmen you would expect to return the following year? Defend your expectation. (Chapter 5, Section 3, p. 100) Q16. If we *reduced* the number of trials to 200 freshmen, what do you anticipate would happen to the interval of proportions from the empirical data around the theoretical probability of 50%? Why? Conduct a few samples with 200 trials and compare your results with what you anticipated. (Chapter 5, Section 3, p. 103)	Q19. Discuss why it might be beneficial to have students simulate the freshman retention problem for several samples of sample size 500, as well as sample sizes of 200 and 999. (Chapter 5, Section 3, p. 103)

Figure 2. Examples of instances in materials coded as promoting interval reasoning.

As seen in Table 1, every Chapter contained both content and technology tasks as well pedagogical tasks that emphasized center and spread. This coordination was discussed in the text along with any diagrams and technology screenshots in all but Chapter 2 (which is a video case with minimal text), with slightly heavier emphases in Chapters 4 and 5. Chapters 5 and 6 have the most content and technology tasks focused on center and spread. Of particular importance is that an explicit focus on interval reasoning only appears in Chapter 1, 5, and 6, with Chapter 5 containing a particularly strong emphasis. Thus, although there is evidence to suggest that the design of the materials does attempt to build understanding of center and spread throughout, the attention to this is uneven, particularly with emphasizing interval reasoning.

Classroom Episode from Chapter 5

Since Chapter 5 contains the highest focus on coordinating center and spread via interval reasoning, we analyzed a 2.5 hour class session from the first implementation cycle (Spring 2006). It is not possible to present a detailed analysis of the session; however we present here classroom discussions around several of the interval reasoning tasks shown in Figure 2.

In answering Q11 (see Figure 2), several students propose three ranges they consider to be reasonable for how many freshmen they expect to return the following year [225-275, 175-325, 230-270]. The teacher asks a student to explain his reasoning for the third interval.

T: Can you tell me why you widened the range?
MSt2: I didn't, I narrowed it
T: Tell me why you narrowed it
MSt2: 500 is a big number. So I thought it might be close to 50%.
T: So you thought because 500 is a big number it would be closer to
MSt2: half
T: To half, Closer to 50%. So, MSt1, why did you widen the range? This [225-275] was the first one thrown out, why did you make it bigger?
MSt1: Well it's all according to how long you're going to do the simulation.
T: Out of 500 student how many, what range of students will return? Do you think it will be exactly 50% return
MSt1: Probably not
T: So for any given year, what range of students might return, if you have 500 for every year?
MSt1: 175-325
T: Ok. So can you tell my why?
MSt1: Without knowing anything I wouldn't go to a tight range.
T: Because you don't have enough information.
FSt3: Its like the coin flips, you have some high and some low, so it might not fall into the 225-275.
FSt4: I'd say it will most likely fall into that first range, but it's not a bad idea to be safe and say it can go either way.

First, all intervals were given in frequencies, rather than proportions. This is likely an artifact of the wording of Question 11 during that implementation cycle. At that time it did not specifically use the word "proportion". All suggested intervals are symmetric around an expected 50% of 500 (250). Two of the intervals are within 5% of 250, while the second proposed interval (175-

325) suggests a variation of ±15%. While one student reasons that 500 is a large enough sample to expect values close to 50%, another is much more tentative and casts a wider net due to an uncertainty about the number of times the simulation would be run. This student, and the two females that respond afterwards, may be trying to capture all possible values, rather than consider a reasonable interval that would capture most (i.e., a modal clump). Only one student justified an interval by explicitly reasoning from the expected value, and there were no justifications, or questioning from the teacher, as to why the intervals were symmetric.

After about 30 minutes of exploration in using a calculator to run simulations, the teacher asks each student to run two simulations of the "50% retention rate of 500 freshmen" and compute the proportion of freshmen returning. The teacher collects and displays this data as a dot plot in *Fathom.* This is the second time during this lesson the teacher used *Fathom* to collect data from individual's samples and display them as a distribution. This teacher's move was not suggested in the curriculum materials; however its value in indicating a public record and display of pooled class data is duly noted.

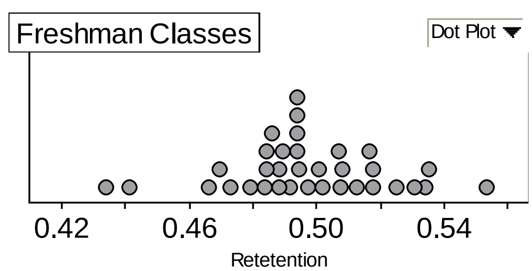


Figure 3. Distribution of 34 sample proportions pooled from class.

The plot in Figure 3 appears quite typical for what might occur with 34 samples, with a modal clump between 0.48-0.51. The teacher recalls the predicted intervals and asks:

T: Here are your predictions from earlier on the number of students you might see in a range. Our proportion range is about from 0.44-0.53. Think any of these ranges for the students are too wide or too narrow…?

What we see as key in her question, is her use of the word *range*, rather interval, and that she appears to be drawing their attention to the entire range of proportion values, rather than on a modal clump. It appears that both the teacher and students are interpreting our request for a "reasonable interval" as the range of all sample proportions that are likely to occur, or that do occur, rather than our intended focus on an interval representing a modal clump.

The discussion continues as the teachers has them use an algorithm to convert the proportion range (which she re-estimates as 0.43-0.55) to frequencies [215-275] so they can compare to the predicted intervals. They briefly note the similarity of the sample range to two of the intervals, and that the range is not symmetric. The teacher then asks a question based on Question 16 (see Figure 2) and two questions that follow in the text. We use this to consider how students reason about the relationship between sample size and variation from expected center.

T: So, lets say instead of doing 500 freshman, we would decrease this set to 200. How do you think the range might differ, or if we increased to 999 how might the range of proportions be different?
MSt2: It would be narrower
T: Narrower for which way, if we reduced to 200 or increased to 999?
MSt2: 999
T: Why do you think it would be narrower?
MSt2: The more trials there are, the closer it will be to the true mean.
T: [asks students if they agree, about half the class raise their hand]
… …. [other students make similar comments]
T: If we decrease to 200 trials in each sample from 500 do you expect the range to be similar or do you expect it to be wider or narrower or similar??
FSt: Wider. With a smaller sample you will have more variability
T: So you are going with the idea that a smaller sample will have more variability. Does everyone agree or disagree? [many students say agree]

This episode suggests that at least the vocal students have an understanding of the relationship between the number of students in the freshman class size and the variation in the distribution of sample proportions from repeated samples.

Pedagogical Task following Chapter 5

The ultimate goal of the materials is to develop preservice teachers' TPSK. Thus, there are many opportunities within the curriculum materials to engage in pedagogical tasks. One such task followed preservice teachers' work in Chapter 5. As a follow-up to our examination of the classroom interactions for Chapter 5 in Spring 2006, we examined how these same preservice teachers may have applied their developing understandings in a pedagogical situation[2]. The task gave a particular probabilistic context and then asked:

> *Explain how you would help students use either the graphing calculator, Excel, or Probability Explorer to simulate this context. Explicitly describe what the commands represent and how the students should interpret the results. Justify your choice of technology.*

Of particular interest to us was whether preservice teachers would plan to engage their students in repeated sampling, use proportions rather than frequencies to report data, and whether they would promote or favor interval reasoning or point value estimates. The majority chose to use a graphing calculator (10 of 17), only 5 of the 17 preservice teachers planned experiences for their students that incorporated repeated samples, and only 7 used proportions. In addition, 10 focused on a point estimate, 1 used both a point and interval estimate for interpreting a probability, while 6 of the plans were not explicit enough to tell what they intended. Thus, the majority of the preservice students planned for students to simulate one sample (of varying sizes across plans, but many were less than 50) and to make a point estimate of the probability from that sample.

The preservice teachers did not provide much evidence that during the week immediately following their discussion of the material in Chapter 5, they were able to transfer their developing understandings of interval reasoning in a probability context to a pedagogical situation. It seems that, for most, any progress made during the class discussions did not have a transference effect into their TPSK.

Select Pre-Post Assessments

It is desirable to have some quantitative measures as indicators of possible conceptual changes occurring with preservice teachers being taught via this material. From the developed pre/post test, we examined four of the questions that address various aspects of our focus on the coordination of center and spread and the alternative use of intervals (see Figure 4).

Table 2. Correct responses and response rates on four test questions.

		Control Fall 2005 n=15		**Spring 2006 n=18**		**Fall 2006 n=15**		**Spring 2007 n=32 (2 sections)**	
Q#		**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
3	**d**	47% 7	47% 7	44% 8	50% 9	53% 8	53% 8	53% 17	53% 17
10	**b**	40% 6	80% 12	44% 8	89% 16	33% 5	80% 12	38% 12	66% 21
11	**d**	53% 8	20% 3	11% 2	22% 4	33% 5	20% 3	25% 8	25% 8
15	**b & d only**	47% 7	40% 6	56% 10	67% 12	40% 6	33% 4	41% 13	56% 18

Across all the implementation semesters and the control group, preservice teachers made little to no improvement in their ability to interpret the accuracy of a 70% probability in data as an interval around 70% (Question 3, answer d), with only about half of them correctly choosing the interval. Across all semesters, there was also little change in preservice teachers' ability to recognize the two reasonable distributions for a distribution of outcomes from repeated samples of 50 coin tosses (Question 15, answers b and d). As shown in response to Question 10 (answer b), preservice teachers appeared to improve in their ability to recognize sampling variability with respect to sample size, in so much as they typically became more likely to recognize that Hospital B with the smaller sample size had a higher probability of having a percent of female births much higher (80%) than an expected 50%.

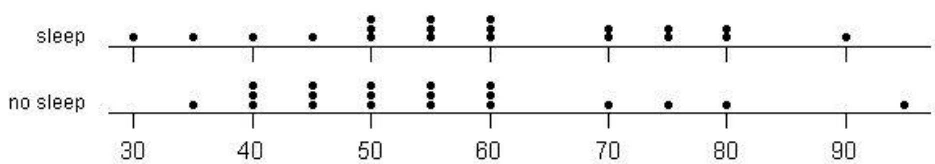
3. The Springfield Meteorological Center wanted to determine the accuracy of their weather forecasts. They searched the records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days. The forecast of 70% chance of rain can be considered *very* accurate if it rained on:

a. 95% - 100% of those days.
b. 85% - 94% of those days.
c. 75% - 84% of those days.
d. 65% - 74% of those days.
e. 55% – 64% of those days.

10. Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

a. Hospital A (with 50 births a day)
b. Hospital B (with 10 births a day)
c. The two hospitals are equally likely to record such an event.

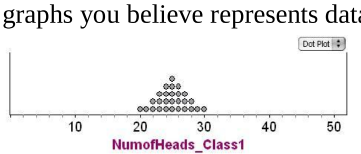
11.Forty college students participated in a study of the effect of sleep on test scores. Twenty of the students volunteered to stay up all night studying the night before the test (no-sleep group). The other 20 students (the control group) went to bed by 11:00 pm on the evening before the test. The test scores for each group are shown on the graph below. Each dot on the graph represents a particular student’s score. For example, the two dots above 80 in the bottom graph indicate that two students in the sleep group scored 80 on the test.

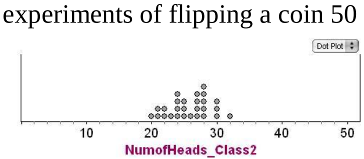


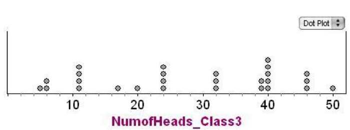
Examine the two graphs carefully. From the 6 possible conclusions listed below, **choose the one with which you most agree.**

a. The no-sleep group did better because none of these students scored below 35 and a student in this group achieved the highest score.
b. The no-sleep group did better because its average appears to be a little higher than the average of the sleep group.
c. There is no difference between the two groups because the range in both groups is the same.
d. There is little difference between the two groups because the difference between their averages is small compared to the amount of variation in the scores.
e. The sleep group did better because more students in this group scored 80 or above.
f. The sleep group did better because its average appears to be a little higher than the average of the no-sleep group.

15. Each student in a class tossed a penny 50 times and counted the number of heads. Suppose four different classes produce graphs for the results of their experiment. There is a rumor that in some classes, the students just made up the results of tossing a coin 50 times without actually doing the experiment. **Please select each** of the following graphs you believe represents data from actual experiments of flipping a coin 50 times.

a.


b.


c.


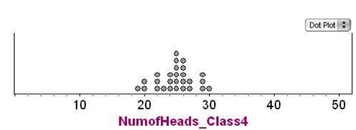
d.


Figure 4. Sample Pre-/post-test questions on center, spread, intervals, and variability.

Since the control group made similar gains on Question10, than those who had engaged in using the new materials, it appears that engaging in learning about data analysis and probability may be helpful in one's ability to correctly respond to that question, regardless of curriculum material. For Question 11, there was very little change in the percent of preservice teachers who correctly chose d to indicate that there was little difference between the groups with respect to center and the large spread, and in fact most chose f, a comparison done only on a measure of center. It is disappointing that more preservice teachers did not demonstrate a coordination of center and spread with this task on the post-test. It is interesting that in the control group, about half initially reasoned correctly but that after instruction, the majority chose f. Perhaps the control curriculum placed a greater emphasis on measures of center.

The main lesson we take from examining these pre-post questions in particular, is that the materials do not appear to substantially help preservice teachers correctly reason about center, spread, and intervals, as measured by these particular questions.

Implications

How do these preliminary results help answer our question about curricular design elements that support preservice teachers developing a coordinated view of measures of center and spread? One design element was the deliberate and consistent focus on the theme of the coordination of center and spread. The module covers a broad range of material, written by three authors through many iterations and reviews from external advisors. Though the theme of coordination was maintained throughout the material, the emphasis is found to be quite inconsistent across chapters. Even more sporadic was the illustrated preference of intervals over point values; with half the chapters ended up leaving this theme behind in the final revisions.

Even though the focus on intervals and modal clumping is consistent in the probability/simulation chapters, a few of the relevant test questions do not indicate any gains beyond those from general exposure to data and probability. To ascertain if these themes can strengthen the intuitions of “clumping” over the seemingly ingrained point-value intuitions, the message must be reemphasized throughout the material. Also, these goals need to be emphasized to the course instructor through different avenues, such as an instructor guide. The classroom discussion revealed that the vocabulary, such as “range” versus “interval”, needs to be clarified in the text. To be able to reliably use the pre-/post-test, questions need to be created and used that more clearly assess understanding of interval reasoning and use of clumping. Much of this may also be an artifact of the curriculum design, implementation, and assessment process. Although the design and implementation evolved over four semesters, the assessment was created prior to completion of the first draft and remained static across implementations for comparisons.

The design element of using dynamic tools is not only heavily apparent in the material; the pedagogical use of such tools is a major focus of the course. However, the technological decisions the preservice teachers made in creating lesson plans to explore a probability concept demonstrate a weak transference of the ways technology can be used to explore coordination of center and spread. For this preliminary report, we did not examine other sources of possible evidence of their development of TPSK related to center and spread. Such data may include preservice teachers responses to a variety of pedagogical questions posed throughout the 6 chapters and 2-day lesson plans required as a course project.

Finally, there is an unaddressed issue of retention; what do these preservice teachers actually take to their class rooms when they lead data analysis explorations? Longitudinal studies are clearly needed to see effects of any teacher education materials on teachers’ practices with their students.

Notes:

[1]The *Preparing to Teach Mathematics with Technology* project is supported by the National Science Foundation under Grant No. DUE 04-42319 awarded to North Carolina State University. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation. More information about the project and materials can be found at http://www.ncsu.edu/project/ptmt.

[2]The data from this pedagogical task were initially analyzed and reported by H.S. Lee, K. F. Hollebrands, P. H. Wilson, and G. F. Mojica at the Research Presession of the National Council of Teachers of Mathematics, April 2007.

References

Batanero, C., Godino, J. D., & Roa, R. (2004) Training teachers to teach probability. *Journal of Statistics Education 12*(1), [www.amstat.org/publications/jse/v12n1/batanero.html].

Capraro, R. M. (2004) Statistical significance, effect size reporting, and confidence Intervals: Best reporting strategies. *Journal for Research in Mathematics Education*, v35 n1 p57-62 Jan 2004

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula". *Journal for Research in Mathematics Education 38*(1), 35-70.

delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79-95). Netherlands: Kluwer Academic Publishers.

Fidler, F. (2006) Should psychology abandon p-values and teach CI's instead? Evidence-based reforms in statistics education. *Proceedings of International Conference On Teaching Statistics 7.* Retrieved October 15, 2006 from http://www.stat.auckland.ac.nz/~iase

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2005) *A curriculum framework for K-12 statistics education.* GAISE report. American Statistical Association. Retrieved September 3, 2006 from www.amstat.org/education/gaise

Gardner M. J, & Altman D. G. (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)* 292, 746-50.

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal 2(1),* 22 – 38. Retrieved July 27, 2005. http://www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf.

International Committee of Medical Journal Editors. (1997) Uniform requirements for manuscripts submitted to biomedical journals. *Annals of Internal Medicine 126* (1), 36-47. Retrieved March 1, 2007. http://www.annals.org/cgi/content/full/126/1/36.

Lee, H.S., & Hollebrands, K. F. (in press). Preparing to teach data analysis and probability with technology. *Proceedings of the Joint Study of the ICMI/IASE*. Monterrey, MX, June, 2008.

Lee, H.S., Hollebrands, K.F, & Wilson, P.H. (2007). The use of research-based methods and materials for preparing to teach mathematics with technology. In D. K. Pugalee, A. Rogerson, & A. Schinck (Eds.), *Proceeding of the Ninth International Conference, "Mathematics Education in a Global Community* (pp. 383-388), Charlotte, NC.

Koehler, M.J., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research, 32(2),* 131-152.

Konold, C., & Higgins, T. (2003) Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 196-215). Reston: VA, National Council of Teachers of Mathematics.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education 33*(4), 259-289.

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

Niess, M.L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and Teacher Education, 21,* 509-523.

Reading, C., & Shaughnessy, M. (2004) Reasoning about variation. In J. Garfield and D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer Academic Publishers..

Saldanha, L., & Thompson, P. W. (2002). Conceptions of sample and their relationships to statistical inference. *Educational Studies in Mathematics, 51,* 257–270.

Shaughnessy, J. M. (2006) Research on students' understanding of some big concepts in statistics. In G. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth Yearbook* (pp. 77-98). Reston, VA: National Council of Teachers of Mathematics.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Stohl, H. (2005) Probability in teacher education and development. In G. Jones (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 345-366). Kluwer Academic Publishers.