DESIGNING TECHNOLOGY-RICH LEARNING ENVIRONMENTS FOR SECONDARY TEACHERS TO EXPLORE AND PREPARE TO TEACH STATISTICS

Sandra R. Madden
University of Massachusetts Amherst, USA
smadden@educ.umass.edu

This paper examines design characteristics refined over seven years through studies with pre- and in-service secondary mathematics teachers in the United States. Teachers simultaneously explore big statistical ideas while learning to use a dynamic statistical software tool that is new to them. Most study participants have taken at least one university statistics course, yet surveys, interviews, and content assessments indicate fragile understanding of statistical content, procedurally dominated learning environments, and little use of technology in courses. Important considerations when supporting these learners include respecting the statistical understanding that is present and connecting to it strategically. Productive design characteristics that have been developed and refined include: 1) focusing on big statistical ideas; 2) providing provocative tasks; 3) attending to dynamic technology scaffolding; and 4) implementing sustained intellectual press.

INTRODUCTION & THEORETICAL PERSPECTIVES

Statistics has achieved a position of status in the Pre-K-12 curriculum (CCSSM, 2010) and the demand for statistically literate citizens continues to increase. Secondary mathematics teachers shoulder the major responsibility for teaching statistics, yet are ill-prepared in the area of statistics (CBMS, 2001; Shaughnessy, 2007). Due to licensing requirements in the U.S, many take at most one undergraduate course devoted to probability and statistics that provides a woefully inadequate basis for teaching statistics at the secondary level. Researchers have been investigating ways to support teachers' statistical knowledge for teaching (Groth, 2007; Lee & Hollebrands, 2011; Madden, 2008; Shaughnessy, 2007), but there is much yet to learn. Additionally, dynamic statistical tools (e.g., *Fathom*, *TinkerPlots*, *CPMP-Tools*) have profoundly transformed the ways in which learners may engage in statistical exploration and reasoning, but many teachers remain unfamiliar with these technological and related curricular innovations.

Highlighting the complexity of the work of supporting teachers, Lee and Hollebrands (2011) introduced technological pedagogical statistical knowledge (TPSK) as framework that extended existing frameworks of technological pedagogical content knowledge (TPCK) (Niess, 2005) and statistical knowledge for teaching (Groth, 2007) by coordinating statistical knowledge, technological statistical knowledge, and TPSK. TPSK includes understanding students' learning and thinking of statistical ideas; conceptions of how technology tools and representations support statistical thinking; instructional strategies for developing statistics lessons with technology; critical stance towards evaluation; and use of curricula materials for teaching statistical ideas with technology. Extended experiences for teachers to explore statistics with technological tools, as learners, is a critical piece of supporting TPSK.

Comparing distributions is a central organizing statistical concept with potential to support statistical reasoning. As conceptualized here, comparing *distributions* may mean *comparing*:

- distributions of samples to one another;
- distributions of samples to population distributions;
- shapes, centers, and spreads of distributions for characteristic signals;
- various measures from distributions;
- experimental and treatment groups;
- variation between and among groups;
- sampling distributions to population distributions;
- sampling distributions with varying sample sizes; and
- bivariate numerical and categorical distributions.

Comparing distributions highlights the intellectual work of making data-based comparisons, while attending to whether detected differences can be considered statistically significant. This framework was used as a curriculum design tool and each task throughout an extended instructional intervention had significant connections to this framework.

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

METHODOLOGY

This ongoing study seeks to explore, understand, and improve secondary mathematics teachers' understanding of the statistical concept of *comparing distributions*. Design research methods have been used across five teaching experiments during which times a hypothetical learning trajectory (HLT) was generated and iteratively refined (Madden, 2008). Data collected included pre and post interviews, content assessments, belief surveys, classroom artifacts, periodic individual reflection statements, and video of classroom activity. Research participants included 56 high school mathematics teachers in the Midwest, USA and 21 pre-service secondary mathematics teachers and mathematics education graduate students in the Northeast, USA. Eleven graduate students participated in the present study, 7 women, 4 men; teaching experience ranged from 0 to 20+ years. Every participant had previously taken at least one undergraduate statistic course. Pseudonyms are used throughout and participants are referred to as students and teachers.

The initial HLT was adapted from a four-day, 20-hour intensive professional development experience to a 13-session, 2.5 hours/session, graduate education course devoted to supporting teachers and researchers experiences with statistics in the secondary curriculum. Retrospective analyses of data sources utilized grounded theory and quantitative methods to understand aspects of learning across multiple environments. Task-sequences-as-enacted have been identified and analyzed for their contribution to supporting TPSK. Analyses focused on the ways in which tasks, tools, and environment ultimately shaped teachers' improved statistical understanding and TPSK, lead to the emergence of constructs of *dynamic technology scaffolding, provocative tasks*, and *sustained intellectual press*, each briefly described next.

Dynamic technology scaffolding (DTS) (Madden, 2008, 2010) is a design principle that involves the intentional sequencing of tasks that first engage learners in a physical environment, then move to a dynamic exploration environment (exploratory but expert designed tool; e.g. java applet, microworld) to a dynamic construction (expressive landscape tool; dynamic statistics) environment. DTS has been utilized repeatedly and successfully to provide learners access to complex ideas and to support technological facility of novice users of dynamic cognitive tools as they navigate new and challenging content and technological terrain.

Statistically, contextually, and technologically *provocative tasks* (Madden, 2011) are instrumental for creating surprise and cognitive conflict for learners, resulting in increased engagement, public discourse, and collective understanding of challenging ideas. Tasks that live in the intersection of statistically, contextually, and technologically provocative appear to be especially valuable for supporting TPSK.

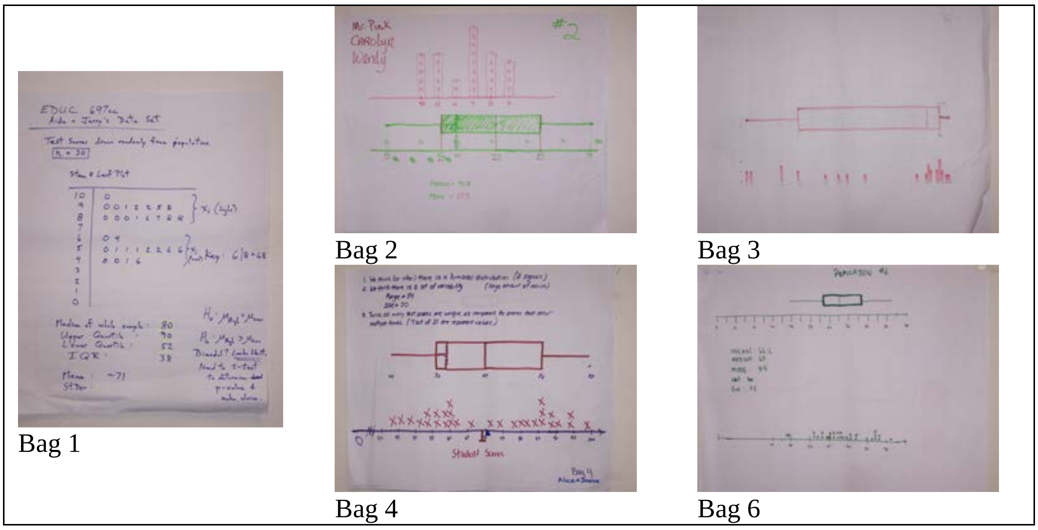
Sustained intellectual press is a pedagogical approach in which a classroom community is co-constructed such that its members come to embrace educative discomfort, cognitive conflict, cognitive overload, undefined endpoints, shared authority for knowledge, individual and group potential, multiple solution paths and ways of knowing, collective intelligence, and sharing not comparing (Madden, 2013). Technologically rich learning environments with these characteristics have been places where learners appear to feel safe to take intellectual risks and ultimately support the learning of all members.

These design tools support task and curriculum development for pre- and in-service mathematics teachers. Two notably different tasks-as-enacted illustrate ways in which teachers' TPSK was perturbed and supported in environments where these design commitments were utilized.

TASK 1: SAMPLING FROM MYSTERY BAGS

Early in the course and prior to use of *Fathom* or *TinkerPlots* (with the exception of one student), pairs of learners were provided a paper bag filled with small slips of paper, each slip representing a student test score from a population of school data and the bag representing a model of the population. The mystery bag task was designed to incorporate the recommendation that learners have experiences of physically drawing samples and then to use technology to investigate sample behavior compared to population behavior (Chance, delMas, & Garfield, 2004).

Pairs drew a random sample of size 30, one at a time and *with replacement* from their bag, constructed a representation of their data on poster paper, and recorded anything about the population they felt confident in concluding. Groups posted their findings (Figure 1) and were asked whether they believed the populations in each bag could have been identical and if not, how many populations may have been represented in the five bags. Initial representations included boxplots, stemplots and lineplots. Between class meetings, all posters were online for students to reason from. Students constructed conjectures in writing as part of an assignment and tended to construct arguments using means, medians, range, shape, IQR, gaps, and clusters. Students' predictions ranged from believing there was one population present—that is, all bags contained the same data values, to two, three, or four populations represented. Sampling one at a time and with replacement created a sense of disequilibrium for students and inconsistencies in students' ways of representing data made comparisons across samples problematic. Each of these issues foreshadowed the usefulness of dynamic technological tools.



Bag 1

Bag 2

Bag 3

Bag 4

Bag 6

Figure 1. Sampling posters made by groups during the sampling bag task.

Multiple students conjectured there were three populations: Pop1—Bags 1, 4, and 6; Pop2—Bag 2; and Pop3—Bag 3, however the level of confidence in the prediction was not high. Nearly half of the students claimed one population could have conceivably generated all samples.

Alex: I think a small sample size with selection by replacement can cause some anomalies. Were we to have selected without replacement with a small sample size, I think the sense of clustering might have been less pronounced and might have more accurately reflected the population as a whole. I see the sampling with replacement as a way of taking repeated measurements and that it should not misrepresent the underlying shape of the data but when combined with small sample size, I think an error factor starts to get magnified. I'm not sure about this, but if these samples all come from the same population, I will be more convinced that small sample size and selection with replacement played a role confusing me.

Another student addressed the difficulty of comparing differently scaled representations.

Kelly: I was able to extract the values and graph them on the same set of axes which makes the comparisons a little easier. … I believe that bag 3 comes from a different population than the others. This is because it has the largest range of values and appears right skewed. Almost 1/3 of its values are below those of the other bags and its median value is above the third quartile compared to the other four distributions. When comparing the distributions of the remaining four bags, I could either argue that there are two different populations being sampled from or that all four of the bags come from the same populations. Two of the bags (1 and 4) appear to have bimodal distributions while the other two bags (2 and 6) appear to have values evenly spread out on the interval. However, given that the sample sizes in each case is only 30, it could be argued that this is due to chance and that the data from these four bags (1, 2, 4 and 6) come from the same population. The range of values for each of these four is approximately the same.

Kelly shared her *Fathom*-generated representation of side-by-side boxplots and histograms, which ignited and fueled stronger statistical reasoning and excitement at the prospect of working with the software. Students refined their comparisons, but not until they were provided the distributions of *possible populations* did their conjectures became more formidable. Students were introduced to navigating in *Fathom* and shown how to sample from a collection in order to use the available populations to recreate samples like they had previously done by hand. Students spontaneously inquired about a way to collect measures of sample means and medians to build empirical sampling distributions (they did not use this language) in order to better understand the behavior of sample measures. Students were soon able to argue convincingly that at least four populations were present and some did conjecture correctly that five distinctly different populations were the basis for the bags. Students demonstrated a profound shift in understanding what a sample could tell them and began to construct a more robust understanding of sampling variability. Individual and collective statistical competence began to blossom. They lived the power of dynamic statistical software to support their learning and this later impacted their lesson designs.

DTS was apparent through students first physically sampling, then being introduced to the process through a technological simulation, and finally through building simulation mechanisms in *Fathom* to sample and construct empirical sampling distributions. The task was provocative as it made fragile statistical conceptions visible, public, and able to be productively perturbed. Sustained intellectual press was evident throughout the extended task as students were challenged and provided with space and support to make statistical sense of the value of samples for supporting inferences. Evidence trumped answers. TPSK was being developed as students simultaneously grew their own statistical understandings, their facility with *Fathom*, and their collective wisdom about the subtle landscape of reasoning while making comparisons in the presence of uncertainty.

TASK 2: ESTIMATING STANDARD DEVIATION

The task in Figure 2 was taken from one state's secondary teaching licensure testing materials. Students had prior conceptual, representational, and computational experiences with various measures of center and variation by this time in the course. The task was a formative assessment of students' understanding of standard deviation and provided a striking example of the difficulty of coordinating computational and graphical aspects of standard deviation. Most students selected distribution D as having the smallest standard deviation, two students selected A, and one student selected C. The computationally correct answer is A.

Alex: If I would have preferred a skewed distribution, I would have picked A before B because there, there's more density around the mean, even though I can't tell the difference between A and D, I can clearly tell that A is better than B (selected D).

Another student contributed:

Karen: I think when I first looked at it, and I thought, OK, I wish I could calculate the mean. I couldn't quickly decide where the mean would necessarily fall for A and B. It was easy for to get a sense of where the mean would fall for C and D. And then I looked at both and decided, all right, there's more concentration, more density around the mean for D than there is for C. But I just quickly dismissed A and B again because it was, it would be kind of difficult to, you know, estimate where the mean would be, so I. Though, you know with A, I think it'd be easier to calculate the mean in A, to pinpoint an estimate of where the mean would be with A than B (selected D).

Students offered a number of estimates for the mean of A and decided on an estimate of the mean between 6 and 7 day. Then another student contributed:

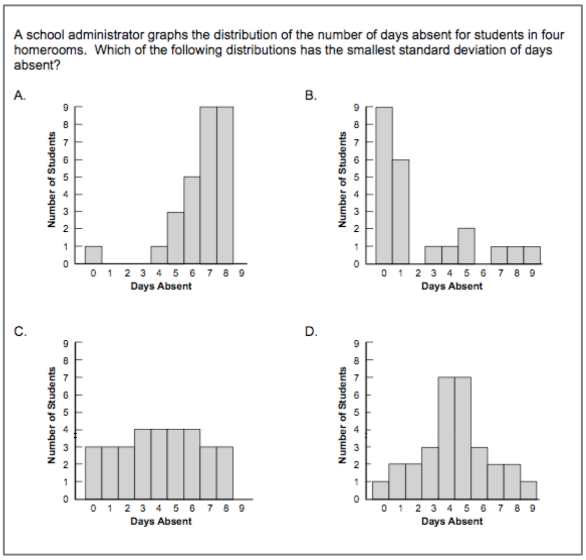
Derek: I have a kinesthetic sense about standard deviation. I think of it as moment of inertia. You know, if you have a barbell, it's very hard to turn a barbell with your hand. But if you had all that weight slid closer into the middle, it'd be a lot easier to rotate (making gestures with hands). So that's a physical sense of, takes into account the … torque really goes with the square of the distance … I ruled out A and B right away because I saw that they did not have a tight

> distribution about the mean, and just looked skewed. Skewed always makes me worried about standard deviation. And then, so I looked at C and D, you can look at C and think of altering C by stealing blocks from the ends and piling them up in the middle. Ultimately if you kept doing that you'd have a great big tall stack smack in the middle and you'd end up with a 0 standard deviation (selected D).

Derek used an engineering concept to reason dynamically, yet incorrectly, in favor of D. Another student validated Derek's approach while respectfully challenging its application.

Connor: But then why wouldn't A be the tightest one, because all you have to do is move one little block. … If you do the thing he's (Derek) talking about, you're talking about something, a weight that's either really long and spread out and trying to twist it or tight in. It's like having a feather way out there. It's still (gesture indicates twisting).

Every student contributed to the discussion and several students revised their decision from D to A, but many remained committed to D. One student then asked his peers about their view of the value of the barbell model and suggested the importance of students having opportunities to physically manipulate models like these to develop intuitions. Peers affirmed the appropriateness of the need to have experiences with physical models like this. Given the student who introduced the model remained committed to distribution D, the instructor elected to confirm distribution A as computationally correct. With some surprise and abruptly, several students suggested the barbell analogy might confound an analysis. A symmetric unimodal distribution suggests balance and balance might suggest ease when rotating, therefore it might inadvertently reinforce the problematic idea that normal distributions have smaller standard deviations than other distributions.



A school administrator graphs the distribution of the number of days absent for students in four homerooms. Which of the following distributions has the smallest standard deviation of days absent?

Figure 2. Standard deviation task

Students were reminded of the *Estimate Center and Spread* tool from *CPMP-Tools* (see www.wmich.edu/cpmp/CPMP-Tools/) that they had seen previously, but used only briefly (see Figure 3). The left figure replicates distribution A. The tool allows the user to estimate the mean of the distribution by moving a small hand. A sense of balance is invoked and the user can then adjust their estimate. Once the mean is located, the user is prompted to drag another hand away from the mean, producing a yellow interval, to the place where approximately 2/3 of the data are covered.

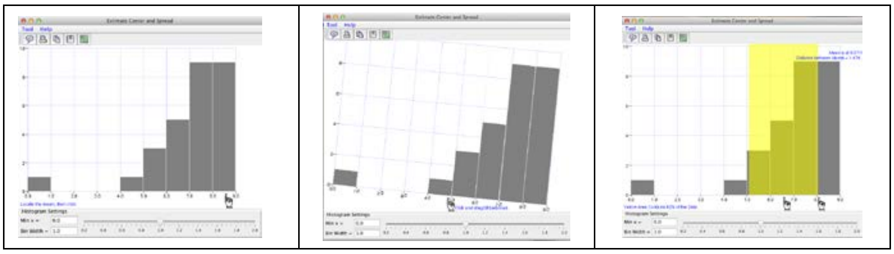


Figure 3. Using *CPMP-Tools* custom *Estimate Center and Spread* tool.

This tool was especially useful in this task setting to disrupt the idea that 68% of the data always lie within one standard deviation from the mean. Explorations of differently shaped distributions with this tool were instrumental for reconciling students' previous ideas about statistical measures.

Comparing distributions by estimating measures of center and spread is present throughout discussion of the task. Students made claims and provided public evidence of their thinking. They offered ideas for others to consider and challenge. Students were engaged and quite animated with the task, making it provocative to them. The practice of beginning with a physical task or model and moving to a dynamic software environment to explore was supported. The instructor practiced pedagogical patience and sustained intellectual press while students publicly wrestled with their ideas and worked hard to clarify, resolve, and develop shared meaning of center and standard deviation using graphical representations. The connection to technology to support learning and teaching was intentional, timely and further indicative of the ways in which these students were challenged to become aware of and facile with a variety of cognitive tools, thus benefitting TPSK.

CONCLUSION

This research alludes to the ways in which tasks-as-enacted in a technology-privileged environment focused on comparing distributions may encourage teachers to challenge and grow their TPSK. Facile tool-users become empowered to pose statistical questions and to interrogate statistical relationships that arise in the context of activity—they discover the provocative nature of tasks and the challenge of reasoning in the presence of uncertainty. This empowerment appears to generate excitement and desire for teachers to provide their own students with similar learning opportunities. With classroom modeling and theory building, the iterative nature of introducing a task, allowing learners to engage with it concretely to generate preliminary ideas, and then letting the technology do the sampling, resampling, representing, and computing, learners with a little knowledge continue to develop. Participants have not hesitated to use dynamic cognitive tools like *Fathom* in lessons with each other or their own students. Statistical meaning making becomes primary and teachers who develop a sense of statistical competence appear eager to engage others.

REFERENCES

Common Core State Standards Initiative [CCSSM] (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association for Best Practices and the Council of Chief State School Officers. Retrieved from www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Boston: Kluwer Academic Publishers.

Conference Board of Mathematical Sciences [CBMS] (2001). *The mathematical education of teachers*. Providence, RI: American Mathematical Society.

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

Lee, H. S., & Hollebrands, K. (2011). Characterizing 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: A joint ICMI/IASE study* (pp. 359-369). New York: Springer.

Madden, S. R. (2008). *High school mathematics teachers' evolving understanding of comparing distributions*. Unpublished dissertation, Western Michigan University, Kalamazoo, MI.

Madden, S. R. (2010). Designing mathematical learning environments for teachers. *Mathematics Teacher, 104*(4), 274-282.

Madden, S. R. (2011). Statistically, technologically, and contextually provocative tasks: Supporting teachers' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1), 109-131.

Madden, S. R. (2013). Supporting teachers' instrumental genesis with dynamic mathematical software. In D. Polly (Ed.), *Common core mathematics standards and implementing digital technologies* (pp. 295-318). Hershey, PA: IGI Global.

Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and Teacher Education, 21*(5), 509-523.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In J. Frank & K. Lester (Eds.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 957-1010). Charlotte, NC: Information Age Publishing.