DYNAMIC TECHNOLOGY SCAFFOLDING: A DESIGN PRINCIPLE WITH POTENTIAL TO SUPPORT STATISTICAL CONCEPTUAL UNDERSTANDING

MADDEN, Sandra
Western Michigan University

ABSTRACT

Fifty-six high school mathematics teachers participated in a four-day technology-intensive professional development experience designed to support their understanding of the big statistical idea of "comparing distributions." Content pretests and teacher interviews informed the hypothetical learning trajectory and design of professional development beyond that from the research literature. Additional data sources included content post-tests and interviews, video-tape of the professional development experience, and teachers' constructed responses to written reflection prompts during the session. Retrospective analyses surfaced a striking phenomenon I have chosen to call dynamic technology scaffolding (DTS) which involves coordination of increasingly-sophisticated technological tools during statistical investigation with the purpose of supporting learners' conceptual understanding of an important statistical big idea.

BACKGROUND

Statistics has emerged as a prominent strand in the secondary school mathematics curriculum (Franklin et al., 2005; NCTM, 2000; The College Board, 2006). As a result, knowledge of statistics for teaching has become essential for high school mathematics teachers if they are to engage students in thoughtful pursuit of statistical ideas. High school teachers typically are ill-prepared in the area of statistics (Ben-Zvi & Garfield, 2004; College Board of Mathematical Sciences, 2001; Shaughnessy, 1992). Makar and Confrey (2004) suggest that *comparing distributions* provides motivation to learn statistics. In a context-rich environment, *comparing distributions* becomes a vehicle to support the investigation of centers and distribution; it provides the basis for informal inference and hypothesis testing; and allows variability to be considered in multiple ways. Researchers have suggested the randomization test may be well-suited as a pedagogical and statistical device when *comparing distributions* to promote the study of sampling distributions and ideas of informal inference (Barbella, Denby, & Landwehr, 1990; Edgington, 1995; Good, 1999; Hesterberg, 2006). This content is relevant to high school curricular demands and potentially relevant to high school mathematics teachers (NCTM, 2000; Franklin et al., 2005).

In this study, *comparing distributions* was broadly conceived to encompass a triadic, multidirectional relationship between the statistical ideas of *distribution*, *variability*, and *sampling distributions,* all potentially supporting informal inference (Rubin, Hammerman, & Konold, 2006). This conceptual framework was used to guide the design of the professional development intervention, assessment instruments, quantitative and qualitative analyses. It was hypothesized that use of the randomization test and dynamic cognitive tools during professional development may support powerful connections between statistical big ideas and representations for teachers. The research question motivating this study was: *What do high school teachers understand about "comparing distributions" and in what ways may statistical professional development impact their understanding of "comparing distributions"?*

THE STUDY

Based upon the dearth of research surrounding statistical professional development of high school teachers and the complexities identified in the literature related to the development of statistical thinking and reasoning, especially in an environment with new and flexible technology, the present study utilized methods of design research (Cobb, Confrey, DiSessa, Lehrer, & Schauble, 2003, p. 9). The study included 56 teachers from 23 school districts involved in a two-year State-funded Mathematics and Science Partnership project. Data collection included teacher pre- and post-content tests and interviews, teacher written reflections, videotape of the 4-day professional development experience, and researcher daily debriefing notes.

The professional development activity system incorporated design principles recommended by Cobb and McClain (2004) and is discussed in detail elsewhere (Madden, 2008). The overarching learning goals for teachers were to: 1) improve understanding of *comparing distributions* by learning to attend to shape, center and spread in distributions, the context of the problem, and multiple representations; 2) begin to understand the power of simulation and resampling approaches to informal statistical inference; 3) improve capacity to use statistical language appropriately; 4) develop facility with graphing calculators, CPMP-Tools (Keller, 2006), and Fathom 2 (Key Curriculum Press, 2005) software to conduct statistical explorations and analyses.

Content test results suggested that a modest four-day, statistics-oriented, technology-rich, professional development session could significantly increase high school teachers' understanding of *comparing distributions* (Madden, 2008). A four-level framework coordinating previous work on students' understanding of *comparing distributions* (Makar & Confrey, 2002), *variability* (Watson, Kelly, Callingham, & Shaughnessy, 2003), and *sampling distributions* (Chance, delMas, & Garfield, 2004; Makar & Confrey, 2002; Watson et al., 2003) was used to measure teachers' levels of understanding of *comparing distributions* as defined in this study. Pre-test scores ranged from 1.02 to 3.50 with mean=1.92, s.d.=0.47. Post-test scores ranged from 1.54 to 3.66 with mean=2.80, s.d.=0.53. Matched-pairs gain scores were significantly greater than 0, ranging from -0.1 to 2.14, mean=0.88, s.d.=0.48, t=13.73, df=55, p<0.0001. The area of greatest improvement in understanding was *sampling distributions*. Given these positive results, one challenge was to determine how professional development experiences with resampling techniques and dynamic statistical tools, as described in the study, shape what teachers know about comparing distributions? While conducting retrospective analyses, the idea of *dynamic technology scaffolding* began to emerge.

RANDOMIZATION TESTING AND RESAMPLING

The randomization test was used as both a statistical tool and pedagogical device with the intent of supporting teachers' understanding of *comparing distributions.* The randomization test protocol for comparing two samples requires first the computation of a statistic of interest (e.g., the difference between sample means). Next the data from the two samples are pooled to simulate "no difference between populations," as the null hypothesis. The pooled data are shuffled and distributed randomly into two groups of the same magnitude as the original samples, n_1 and n_2. Next, the resampled statistic of interest is calculated (e.g., the difference between resampled means). The result is recorded and this process is repeated many times. A histogram of the resampled statistics is constructed. Finally, the original difference is located on the histogram and the proportion of resampled statistics that are at least as extreme as the original difference is determined (this approximates the p-value). Table 1 provides comparisons between the mechanics of hypothesis testing and randomization testing across different technological environments.

Physical simulation

When comparing two samples, physical simulation is straight forward for relatively small data sets (n_1, $n_2 < 10$) using experimental data and index cards representing data values (see Table 1). The process encourages an awareness of sampling variability throughout the resampling procedure. The physical procedure is important to ground students' thinking about what kinds of samples are produced under the assumption that there is no difference between populations, while at the same time allowing for comparison of shape, center, and spread of the resampled data.

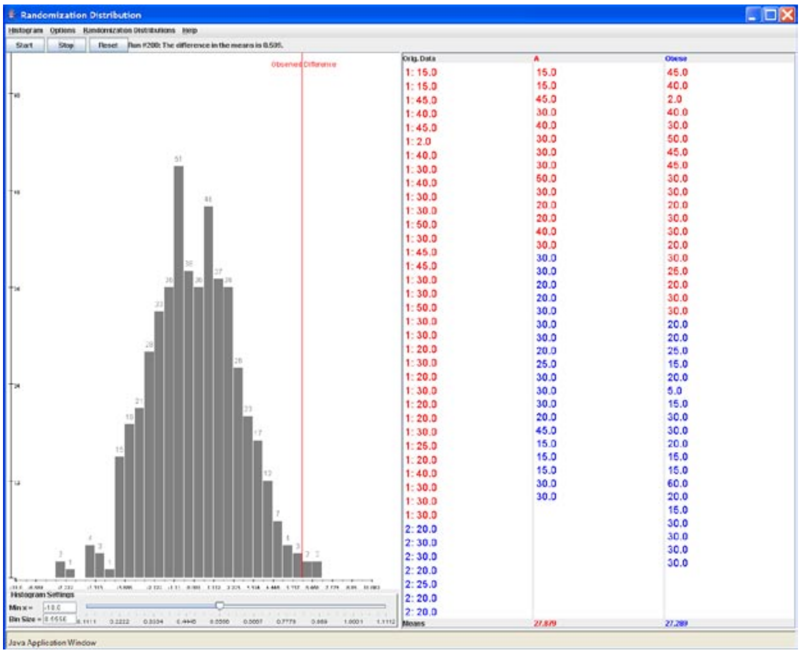
For all its pedagogical merits, physical simulation lacks efficiency. Shuffling, distributing, recalculating, and recording involve a good deal of time and effort for students who may grow weary and become disinterested in this environment quickly. However, after sufficient experience and conversation about the process and its interpretation, moving to an environment in which the process becomes automated but remains transparent is desirable. Ideally, such an environment would minimize time required to construct the distribution yet the essence of the process would remain visible to the learner. Two different technological environments with this potential were utilized in the study: CPMP-Tools and Fathom 2.

Table 1. The Randomization Test across Multiple Technological Environments

	Randomization Testing		
Hypothesis Testing Mechanics	**Physical Simulation**	**CPMP-Tools** (Figure 1)	**Fathom 2** (Figure 2)
Collect experimental data	Write data values on cards	Enter data into CPMP-Tools *(acts like a spreadsheet)*	Enter data in Fathom 2 *(acts like a spreadsheet)*
Determine statistic of interest *(calculate the difference between the measures of interest and record)*	Determine statistic of interest *(calculate the difference between the measures of interest and record)*	Determine statistic of interest *(select mean, median, or standard deviation)*	Determine statistic of interest *(use the formula feature to generate any measure of interest or difference between two measures)*
Identify appropriate sampling distribution under the null hypothesis assumption	Put all cards together, representing the null hypothesis that there is no difference between the two populations of interest	Start animation Watch columns of data change	Stack the data *(option to graph)* New collection is generated **TIER 1**
	Shuffle the cards		Scramble the data *(option to graph)* New collection is generated May manually re-scramble the data **TIER 2**
	Distribute shuffled cards into piles of the same cardinality as the original data sets	Watch columns update *(red and blue font indicates original group)*	
	Compute the difference between the measures of interest	Watch difference between measure of interest display	Create a measure of the difference in statistics between two groups
	Record/accumulate result on a histogram—generates the empirical sampling distribution	Watch randomization distribution update	Collect measures New collection of measures is created Create histogram of measures **TIER 3**
	Repeat multiple times	Watch as animation continues the process; stop after desired number of trials	Use inspector to determine number of trials, select and collect more measures
Compare critical value	Examine where on the resultant distribution the original statistic is located	Locate the line on the randomization distribution representing the difference of interest from the original data	Plot original difference on the resultant randomization distribution graph
Determine p-value	Determine p-value based on the proportion of results equal to or more extreme than the original result	Determine p-value by summing the frequencies of the histogram bars at or beyond the original difference *(may use the drag feature of the histogram to adjust bin-widths for ease of calculation)*	Determine p-value by summing the frequencies of the histogram bars at or beyond the original difference *(may use the drag feature of the histogram to adjust bin-width for ease of calculation)*
Interpret the result	Determine if p-value is small enough to suggest that chance is not the likely explanation for the observed phenomenon (accept or reject the null hypothesis)		

CPMP-Tools environment

Along with other cognitive tools for mathematical and statistical investigation, CPMP-Tools contain a built-in feature specifically designed to conduct randomization tests. Building or growing the sampling distribution is as easy as clicking a button. Replication is a snap. The animation feature allows the learner to focus on the things that are changing and the aspects of the simulation that are remaining approximately invariant. Figure 1 provides a visual, non-dynamic view of the Randomization Distribution custom tool within CPMP-Tools. Data on the right side of the screen are represented in the first column as pooled data (red and blue represent the different original samples); the second column contains a resample for one group (sorted by group); the third column contains a resample for the second group. The randomization distribution on the left represents the accumulation of 1000 resampled statistics (difference between the mean of the two groups in this case) with the vertical line on the graph representing the original statistic of interest.



Run #200: The difference in the means is 0.535.

Orig. Data	A	Obese
1: 15.0	15.0	45.0
1: 15.0	15.0	40.0
1: 45.0	45.0	2.0
1: 40.0	30.0	40.0
1: 45.0	40.0	30.0
1: 2.0	30.0	50.0
1: 40.0	30.0	45.0
1: 30.0	30.0	45.0
1: 40.0	50.0	30.0
1: 30.0	30.0	30.0
1: 30.0	20.0	20.0
1: 50.0	20.0	30.0
1: 30.0	40.0	30.0
1: 45.0	30.0	20.0
1: 45.0	30.0	30.0
1: 30.0	30.0	25.0
1: 30.0	20.0	20.0
1: 50.0	20.0	30.0
1: 30.0	30.0	30.0
1: 30.0	30.0	20.0
1: 20.0	30.0	20.0
1: 30.0	20.0	25.0
1: 20.0	25.0	15.0
1: 30.0	30.0	20.0
1: 20.0	30.0	5.0
1: 20.0	30.0	15.0
1: 30.0	20.0	30.0
1: 25.0	45.0	30.0
1: 20.0	15.0	20.0
1: 40.0	15.0	15.0
1: 30.0	15.0	15.0
1: 30.0	30.0	60.0
1: 30.0	30.0	20.0
2: 20.0		15.0
2: 30.0		30.0
2: 30.0		30.0
2: 20.0		30.0
2: 25.0		30.0
2: 20.0		
2: 20.0		
Means	27.879	27.289

Figure 1. An example of the randomization distribution tool in CPMP-Tools

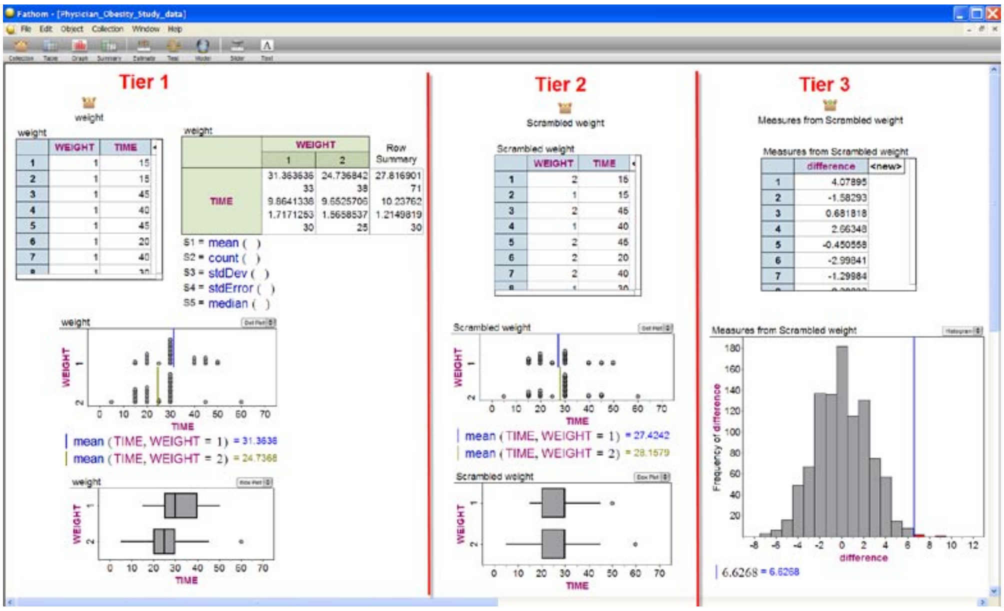
With the animation feature enabled during the simulation, each time resampling is done, the right two columns of resampled data dynamically update, the statistic of interest (mean in this case) for each resample is calculated and shown at the bottom of each column, and the difference between the statistics is plotted on the histogram. In this example, the proportion of resampled statistics greater than or equal to the original test statistic is 4/1000 = 0.004, indicating a very small likelihood that the original difference was due to chance. Hart, Hirsch, and Keller (2007) argue that the use of the randomization test in this technological environment may serve to amplify student learning of probability and statistics and support the conceptual understanding of statistical inference by students.

Following one physical simulation and one brief experience with CPMP-Tools during the professional development session while conducting an investigation, teachers rated the ease of use of CPMP-Tools for conducting the randomization test on a scale from 1 (low) to 10 (high) with a mean=8.3 and median=9 (n=53, min=3, max=10). At the same time, when asked to use the same scale to rate their current understanding of the randomization test, results were mean=6.6, median=7 (n=55, min=1, max=10). Together these data suggest the randomization test process was accessible to teachers and that CPMP-Tools were seen as easy to use.

Fathom 2 environment

In addition to developing statistical conceptual understanding, another goal was to build teachers' capacity to utilize Fathom 2 software in order to support statistical inquiry within their classrooms. It was conjectured that once teachers understood the randomization test, a way to support their learning of Fathom 2 would be through their construction of the mechanisms in Fathom 2 to simulate the randomization test. Prior to constructing this mechanism in Fathom 2, teachers completed two introductory Fathom 2 tours to become familiar with some basic navigational features of the software and observed the instructor demonstrating several dynamic features of the program. One of the things that became quite important in this learning sequence was the ability to utilize the scaffolding afforded from the physical simulation and CPMP-Tools environments to be able to invoke the actions from one or more environment and apply them in a new environment. A comparison of the learner demands when using CPMP-Tools versus Fathom 2 to conduct a randomization test as seen in Table 1 highlights the active construction of mechanisms required to carry out the process in Fathom 2. By contrast, in CPMP-Tools' Randomization Test environment it may be possible for the learner to passively observe and perhaps not fully understand the process. It is conceivable that a learner may be able to use CPMP-Tools to conduct a randomization test and to correctly interpret the results without fully understanding or coordinating the process or the hierarchical objects.

Lane-Getaz's (2006) Simulation Process Model (SPM) is a three-tier graphic pre-organizer to support students' connections between simulation activities and the logic of inference. Its three tiers: 1) population (parameters), 2) samples (statistics), and 3) distribution of sample statistics are consistent with mechanisms which may be seen in CPMP-Tools or developed in Fathom 2. The mechanics of the randomization test process may be summarized as Tier 1—Display Data, Tier 2—Scramble Attributes, Tier 3—Collect Measures, all visible in Figure 2 and reinforced in Table 1.



weight

	WEIGHT	TIME
1	1	15
2	1	15
3	1	45
4	1	40
5	1	45
6	1	20
7	1	40

weight

	WEIGHT		Row Summary
	1	2	
TIME	31.363636	24.736842	27.816901
	33	38	71
	9.8641338	9.6525706	10.23762
	1.7171263	1.5658537	1.2149819
	30	25	30

S1 = mean ( )
S2 = count ( )
S3 = stdDev ( )
S4 = stdError ( )
S5 = median ( )

mean (TIME, WEIGHT = 1) = 31.3636
mean (TIME, WEIGHT = 2) = 24.7368

Scrambled weight

	WEIGHT	TIME
1	2	15
2	1	15
3	2	45
4	1	40
5	2	45
6	2	20
7	2	40

mean (TIME, WEIGHT = 1) = 27.4242
mean (TIME, WEIGHT = 2) = 28.1579

Measures from Scrambled weight

	difference	<new>
1	4.07895	
2	-1.58293	
3	0.681818	
4	2.66348	
5	-0.450558	
6	-2.99841	
7	-1.29984	

6.6268 = 6.6268

Figure 2. An example of the randomization distribution process in Fathom 2

Because of the construction demands in Fathom 2, it is likely that understanding the entire randomization test process is a necessary condition for learners to successfully construct the mechanism. Though the end result of the simulation in Fathom 2 is nearly identical to that in CPMP-Tools, a number of additional dynamically-connected representations can be hot-linked in the construction. With the animation option turned on during simulation, objects in Tiers 2 and 3 dynamically update. Several things are afforded the learner in this environment:

1) flexibility to investigate any statistic of interest (not limited to mean, median, and standard deviation as in CPMP-Tools);
2) constructing a measure to generate a statistic of interest for comparison requires the coordination of fairly sophisticated mathematical symbolization (e.g., *difference = mean(time, weight = 1) – mean(time, weight = 2)*);
3) scrambling attributes (Tier 2) and collecting measures (Tier 3) both generate new collections from which the hierarchical structure of sampling distributions may emerge,
4) making connections between representations of histogram, boxplots, measures of center and spread, and the ways in which changes in one representation are reflected in the others;
5) dynamically-linked representations make visible strong connection to issues of sampling variability under the null hypothesis of no difference between populations.

Evidence from a variety of sources suggests that the scaffolding of dynamic technology provided support for teachers’ evolving understanding of comparing distributions. Teachers’ written reflections strongly communicated the belief they were developing increasing facility with Fathom 2 as well as improving understanding of the randomization test. Teachers’ self-reported understanding of the randomization test increased to a mean=8.30, median=8.75 (min=5, max=10) by Day 4. Their reported comfort using Fathom 2 grew to a mean=7.30, median=7 (min=5, max=10) by Day 4 from an initial mean=6.68, median=6.75 (min=2, max=10) on Day 2. Gain scores were significantly greater than 0 with t=2.973, p=0.002. Teachers reported not feeling overly confident in their ability to use Fathom 2; however, during a task-based portion of post-interviews, all 16 teachers were able to describe the randomization test procedure and apply it to the task. Furthermore, 12 of the 16 teachers elected to build the fairly complicated randomization test mechanism in Fathom 2 to successfully compare two distributions, even when CPMP-Tools were available. Two teachers elected to use CPMP-Tools and the remaining two teachers confidently and correctly compared the distributions without the need to run a randomization test. Teachers’ content tests provided additional evidence that teachers’ understanding of comparing distributions had significantly improved. A representative teacher’s response during the post-interview appears to suggest that the Fathom experience was instrumental to her learning about the randomization test process:

> *I think that this, the Fathom, more replicates what they [students] would do with the cards by hand, and so since they’re building it, it has more impact then having the computer just do it. You know this [CPMP-Tools] is fine once they understand the process and now if I had a randomization with the. . . . I wouldn’t have to necessarily you know, have built it myself, I would know how to read it (Julie, post-interview, August 2006).*

Why multiple technology environments?

As a scaffold for learners wrestling with new statistical terrain as well as new demands from software environments, CPMP-Tools (or other pre-constructed simulation device) provide a vehicle which is elegant in its simplicity. It preserves the physical simulation actions while automating the process. As a cognitive tool, CPMP-Tools relieves the construction burden from the learner and allows focus on the process and the result, potentially highlighting characteristics of the process without overwhelming the learner with technology demands or too many competing representations. According to Zbiek, Heid, Blume, & Dick (2007)

> Cognitive tools play a special role in mathematical activity by externalizing representations (Heid, 1997). Through externalized, though limited, surrogates for a student’s internal mental representations displayed on the surface of the screen, externalized representations become visible phenomena that can be shared and discussed with others (e.g., other learners or the teacher). By bringing such representations literally to the surface, a cognitive tool can allow for unique opportunities or exposing cognitive conflicts (p. 1173).

Many statistics classes offer students opportunities to investigate probabilistic phenomena through the used of physical simulations and computer or graphing calculator simulations. At best, the research into the associated learning due to the use of simulations has been mixed (Mills, 2002). Some have argued that students are able to play a passive role as they watch simulations. Others have seen some benefits through the use of integrated dynamically-linked simulations on student learning (van der Meij & de Jong, 2006). The research reported here suggests possible synergy through the intentional coordination of multiple tools.

DYNAMIC TECHNOLOGY SCAFFOLDING

In addition to developing and applying the randomization test process in support of informal inference, a similar scaffolding of dynamic technology occurred though another sequence of learning activities for teachers. The sampling distribution of the mean was developed first through physical sampling from a finite population, and then demonstrated through a pre-constructed simulation using Fathom 2, and finally with teachers' constructing the sampling distribution mechanism in Fathom 2. The general model in Figure 3 illustrates the flow of activities and their collective relationship to the development of a central idea or concept.

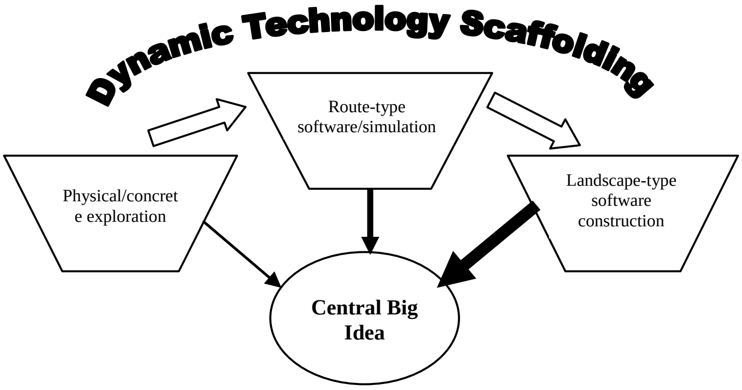


Figure 3. Model of dynamic technology scaffolding

Beginning with a concrete or physical representation and investigation, the basis for a new concept is established. A next level of activity with the potential to inform the development of the concept occurs via a structured simulation or demonstration in which the limitations imposed by the physical environment are removed and a cognitive tool is utilized to automate the process and assist the learner with noticing and connecting essential characteristics (e.g., CPMP-Tools, Java-applets, Fathom simulations). At the third level of activity, the learner is actively engaged in the construction and manipulation of mathematical objects in a flexible technological environment (e.g. Fathom 2 or Tinkerplots) for the express purpose of engineering a mechanism which may possess similar attributes as the physical and/or simulated environment and one which will support additional investigation. Through the development of facility with a flexible, dynamic technological construction tool, the learner gains the additional benefit of the potential to further explore myriad relationships and conjectures. Each activity and associated dynamic technology serves as scaffolding for the development of further conceptual understanding at the next level. From one level to the next, the learner assumes more control over an increasingly sophisticated technological environment with the promise of supporting the conceptual development of important statistical or mathematical ideas and representations.

This progression is similar to the developmental modes of representation described by Bruner (1964): enactive (physical/concrete exploration), iconic (route-type software/simulation), and symbolic (landscape-type software construction). The model reflects the Vygotskian perspective that the use of tools may profoundly impact the ways in which learners come to understand statistical concepts and processes. According to the National Research Council (2001), "*Conceptual understanding* refers to an integrated and functional grasp of mathematical ideas" (p. 118). This paper introduces *dynamic technology scaffolding* as a design principle for developing and sequencing instructional tasks with potential to support teachers' conceptual understanding of challenging statistical ideas.

ACKNOWLEDGEMENTS

The author gratefully acknowledges funding for this research by the Michigan Department of Education with special thanks to the amazing high school teachers who generously participated in this research.

REFERENCES

Barbella, P., Denby, L., & Landwehr, J. M. (1990). Beyond exploratory data analysis: The randomization test. *Mathematics Teacher, 83*(February), 144-149.

Ben-Zvi, D., & Garfield, J. (Eds.). (2004). *The challenge of developing statistical literacy, reasoning and thinking*. Norwell, MA: Kluwer Academic Publishers.

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist*, *19*, 1-15.

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.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13.

College Board of Mathematical Sciences. (2001). *The mathematical education of teachers*. Providence, Rhode Island: American Mathematical Society.

Edgington, E. S. (1995). *Randomization tests* (3rd ed.). New York: Marcel Dekker, Inc.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*.

Good, P. I. (1999). *Resampling methods: A practical guide to data analysis*. Boston: Birkhauser.

Hart, E. W., Hirsch, C. R., & Keller, S. A. (2007). Amplifying student learning in mathematics using curriculum-embedded, Java-based software. In W. G. Martin & M. E. Strutchens (Eds.), *The learning of mathematics: Sixty-ninth yearbook* (pp. 175-202). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Heid, M. K. (1997). The technological revolution and the reform of school mathematics. *American Journal of Education*(106), 5-61.

Hesterberg, T. (2006). Bootstrapping students' understanding of statistical concepts. In G. F. Burrill (Ed.), *Sixty-eighth yearbook: Thinking and reasoning with data and chance* (pp. 391-416). Reston, VA: NCTM.

Keller, S. (2006). CPMP-Tools (Version 1.1, trademark pending). East Lansing, MI: Michigan State University and Core-Plus Mathematics Project.

Key Curriculum Press. (2005). Fathom^{TM} Dynamic Statistics Software (Version 2.0) [Computer software]. Emeryville, CA: Author.

Lane-Getaz, S. J. (2006). What is statistical thinking and how is it developed? In G. F. Burrill & P. C. Elliott (Eds.), *Sixty-eighth yearbook: Thinking and reasoning with data and chance* (Vol. 68, pp. 273-290). Reston, VA: NCTM.

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

Makar, K., & Confrey, J. (2002). *Comparing two distributions: Investigating secondary teachers' statistical thinking.* Paper presented at the Sixth International Conference on Teaching Statistics (ICOTS-6), Cape Town, South Africa.

Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 353-374). Boston: Kluwer Academic Publishers.

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, *10*(1).

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, D.C.: National Academy Press.

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

Rubin, A., Hammerman, J. K. L., & Konold, C. (2006). *Exploring informal inference with interactive visualization software.* Paper presented at the International Conference on Teaching Statistics-7.

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 Publishing Company.

The College Board. (2006). *College Board Standards for College Success: Mathematics and Statistics*.

van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*(3), 199-212.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, *34*(1), 1-29.

Zbiek, R. M., Heid, M. K., Blume, G. W., & Dick, T. P. (2007). Research on technology in mathematics education. In J. Frank K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (Vol. 2, pp. 1169-1208). Charlotte, NC: Information Age Publishing.