Chellgren Proposal
April 2008

Submitted by
William Rayens
Department of Statistics

Project and Course Proposal

Project Summary

It is somehow fitting on this late April weekend that marks the 100th birthday of the legendary journalist Edward R. Murrow to recall something he said in a now-famous diatribe, disguised as a speech, in Chicago at the 1959 annual convention of the Radio-Television News Directors Association. Murrow, in the midst of his own uncomfortable adjustment to delivering the news by way of television instead of radio, noted:

> "It is not necessary to remind you that the fact that your voice is amplified to the degree where it reaches from one end of the country to the other does not confer upon you greater wisdom or understanding than you possessed when your voice reached only from one end of the bar to the other."

In the area of statistics education, particularly when the audience is a general education audience that is typically less quantitatively literate as a whole than, say, engineers or bench scientists, it has become fashionable to use electronic manipulatives to facilitate encounters with mathematically heavy concepts that are nonetheless critical to an understanding of the role statistical science plays in everyday life. I was part of the first wave of statistics educators nationwide to create and/or archive such manipulatives, and to develop individual and group activities based on them. As discussed in detail below, I am no longer convinced they actually help and the purpose of this proposed project is to design an experiment to judge whether these manipulatives are any better than their cruder, simpler, concrete predecessors. Murrow's words remind us that fancy is not the goal. To have the students understand the concepts is the goal.

In particular, I intend to identify 3-5 critical, abstract concepts that need to be surfaced in a general education statistics classroom. A set of electronic manipulatives will be identified as relevant to facilitating this surfacing, and a similar set of concrete manipulatives will either be constructed or purchased, depending on availability and appropriateness. An assessment focused on the identified conceptual constructs will be developed, pilot tested, and revised if necessary. An assessment focused on the level of student engagement will be similarly developed. A controlled experiment will then be deployed to test the null hypothesis that there are no differences in these assessment measures with respect to manipulative type. Other demographic data will also be collected and used as part of a larger exploratory study.

Introduction

At its March 17, 2008 meeting the University Senate unanimously approved the document "Design Principles of General Education." The fifth principle reads as follows: "a revised general education curriculum will pay explicit attention to developing students' communication skills and quantitative reasoning abilities." Importantly, the approved subtext in that document reads:

> In the context of general education, quantitative reasoning is an interpretive activity, employing mathematical and statistical methods to analyze, evaluate and draw conclusions from empirical evidence. Seen in this light, quantitative reasoning prepares students for active and informed lives as citizens in a technological society. Courses intended primarily as mathematical pre-requisites for disciplinary majors might not always be optimal for general education purposes.

I would contend that the term "quantitative literacy" is at least as ill defined as is "liberal arts education" and that it tends to take on the character of those discussing it. In a sense, the (above) subtext is an attempt to articulate a working definition, or at least to surface an interpretation of the phrase that we should use to guide our new curriculum design here at the University of Kentucky. Unfortunately, after spending many hours with the subtext and, frankly, almost two decades thinking about how to best do statistics education, it is evident that distinct ideas are being confounded with this wording. With apologies to the reader, we will have to sort through some of this pedantry in order to appreciate the focus of this proposal and to fairly evaluate its importance.

For instance, the view that quantitative reasoning is an "interpretive activity" could refer to any of the following:

- The *reflection* or interpretation that takes place once a mathematical answer has been arrived at. Questions like "what does this mean in the context" typify this idea of interpretation.
- The *ambiguity* present in data and images. The same numbers and figures can be used to simultaneously support and refute evidence of global warming, or the impact of a proposed tax break. Questions such as "make the case from both sides using this evidence" would be candidate questions here.
- The *rationale* behind the science of formal statistical inference and the sense in which this science addresses uncertainty. Questions like "what is the nature of the risk involved in my inference" and "how has statistical science addressed that risk" are sample relevant questions for this higher level interpretation.

I would claim that the "reflection" is a low-level, pre-university activity, which remains essential, so we still have to still require it, but more in the sense that we require the students to put their names on their papers. It is a reflection of their expected maturity at this level and not something that we design a course around. The issue of "ambiguity" is important. Recent debates in Washington centered around a single set of data that was used to argue both for the middle and the upper class as being the chief beneficiaries of a tax cut. Students have to be sufficiently numerate to see these supposed paradoxes for what they really are (in this case political posturing), and that is a skill they won't always have from high school. But it is tricky to teach. The resolution is sometimes a resignation to the presence of ambiguity, and even when it is not, the skill required to straighten it all out often may only involve the ability the simple manipulation of percentages and an understanding of the nuance of language.

Far harder, and, I contend, far more appropriate to have emphasized in a college classroom, is the "rationale" behind formal inferential reasoning. In what sense, does a confidence interval instill confidence? If a study tells us that a particularly common surgery is not statistically more effective with respect to outcome than a placebo surgery, then what is our take home message? What is the role of sampling, sampling distributions, and the infamous bell curves in the creation of useful, everyday inferences? In my view, this is what our general education students are going to get asked to do on a daily basis - consume the inferences that others have made. Most will not be given a set of data and asked to carry out the deductive steps required to test a hypothesis. The hypothesis will already long-since be tested and the conclusions communicated in words appropriate for a less numerate audience.

Likewise, "to analyze, evaluate, and draw conclusions from empirical evidence" could refer to any of the following:

- The *skills and methods* required to reach those conclusions. If the wording is really intended to be "draw conclusions" and not "make inferences" then the work will be largely deductive (mathematical) and/or visual. Questions such as "what can I learn from the proper distillation of these numbers or from these charts" are going to be typical questions to ask.
- The *pedagogy* behind fact or theorem discovery by way of well-guided, empirical-based guessing. Viewed this way, the instructor might be asked to provide empirical evidence, arising from a concrete or electronic manipulative, that would bring the student into a closer encounter with a difficult or obtuse, but important concept or mathematical theorem.

Focusing on "skills and methods" suggests an emphasis on working out problems and acquiring a set of practical skills for analyzing data. I just don't think this is the proper focus for a general education course in statistics. Most students will have seen a whole laundry list of basic tools (largely descriptive and graphical) long before we see those students. It is common to first see this material in middle school and is, in fact, part of a strand required by KERA requirements, though this strand is usually relegated "time available" in an algebra course. But for a general education audience, this is a low-level activity.

The type of "pedagogy" mentioned above might be better called "inquiry based learning" or "hands on learning," and it is an act of informal inference that has been shown to increase student interest and perhaps even assist in content retention. It can be practiced in many disciplines, certainly in mathematics, statistical science, and in the bench sciences. Indeed, I have the privilege of being a co-PI on "Newton's Universe", an NSF supported endeavor that is attempting to take the important inquiry based workshops conceived and taught by Dr. Joseph Straley, from our own Department of Physics, to middle school teachers in the state, with the hope of improving student learning outcomes.

Modern discovery methods are already being employed in avant-garde middle school mathematics classrooms (e.g. at SCAPA in Lexington) where students use a variety of manipulatives to "discover" important mathematics truths, such as Pythagorean's theorem or an expression for the general term of a mathematical sequence or series, one that may emerge later to have a standard "name". Likewise middle school teachers employing an empirical-based pedagogy will use simple data sets and simple graphical tools such as bar charts and histograms (called, unfortunately, "line plots" in middle school), or simple summaries (such as sample means and sample proportions) to surface level-appropriate encounters with outliers or bell shapes.

Indeed, there are two types of inferential reasoning afoot here. One is informal and surely synonymous with the important activity of discovery and inquiry based learning. In an informal sense, this is reasoning from data. The other sense of inference, though, is formal and tantamount to an understanding of what statistical science is all about. It doesn't have to be understood from a mathematical perspective, but the logical process of this formal inference has to be understood before one can successfully read the morning newspaper. The danger is that if we focus solely on the informal sense of inference, then we aren't really engaging statistical science at all and we risk leaving university students in a place no more evolved intellectually than they were in as they emerged from middle school.

This proposal focuses on the effectiveness of inquiry-based instruction using different classes of manipulatives, with the goal of successfully encountering and understanding some of the important stops along the road that a general education course should take toward communicating the tenants of formal statistical inference.

<u>Manipulatives in the Classroom</u>

I have been teaching STA 200 for over 20 years and I've seen a lot of changes in that time both in what I do in the classroom and what the students respond to. Indeed, as I matured in my profession and began to have deeper insights into my classroom work, I developed similarly mature ideas about what was and was not important for a well-educated, non-technical person's understanding of statistics. A couple of years ago I mapped some of these pedagogical changes I had experienced over time as part of a very useful self reflection in my teaching portfolio and I have retained that mapping for the current version.

It is opportune to point out that I started using manipulatives in my classroom 15 years ago, once I was able to dust off the chalk dust that came with my formal mathematics graduate training. In my summer classes I would take advantage of the small class sizes and make use of concrete manipulatives such as spinners or urns of M&M's, etc. As class sizes grew and three-lecture formats gave way to two lectures and recitation breakouts, I developed a detailed series of electronic (applet) manipulatives, often personally adapting some of the Java source code that was then freely available on the web. Eventually repositories developed around the world, notably at U.C. Berkeley, but elsewhere as well. I developed a set of links to the better sites and the majority of my recitation sessions were then, as they are now, devoted to some sort of discovery that was intended to be made easier by using these manipulatives.

In one sense, this has seemed the perfect approach, the most appropriate evolution. The electronic manipulatives are always available, and one only needs a web browser and a computer that hosts some basic Java to access them. TAs can be easily trained and more material can be covered in one recitation. Constructivist theory supports students taking an active role in their own learning and this sort of applet-based instruction is consistent with that. And even though Piaget's "concrete operational stage" calls for the active manipulation of physical materials, it is not a huge leap to view the electronic manipulatives as just modern versions of what Piaget could not imagine in 1960.

The problem is, these electronic manipulatives don't really seem to be accomplishing what I want them to accomplish. This has been very disheartening as I have continually revised and refocused recitation exercises that depend on them, trying somewhat in vain to better connect the students with the learning objectives for those recitations.

I have my own theory of what is wrong and that is the basis for this project proposal. I have to admit up front that my theory is not based in some deep understanding of development theory or learning styles. Mostly it is based on my significant number of years of experience in the statistics classroom and my never-ending revision, refinement, and reformulation of my classroom pedagogy. But here's what I think.

First, the concepts I am trying to motivate, the encounters I'm trying to engineer are subtle and significant. For instance, all of the simple inference that a general education student is likely to see can be traced back along a path that includes probability sampling, the Central Limit Theorem, bell-shaped distributions, and a painful confrontation with the difference between deriving an answer (mathematics) and seeking evidence against a hypothesis (statistical inference). Although the few engineers who end up misadvised into STA 200 would disagree, it is a mistake to try and flesh out the role of the Central Limit Theorem using mathematics. Rather it is important to simply understand that samples, when taken a certain way, admit statistics that present particular patterns as the act of sampling is repeated.

There are now many good applets that allow the student to do the required repeated sampling in a matter of minutes and we've worked out nice details like having those results drop directly into an Excel spreadsheet to allow the subsequent plotting to follow with equal speed. It should work, but it doesn't. Students still tell me at the end of the semester that the upshot of this sort of exercise is that whenever you can you should sample over and over again. Worse, many still tell me that the bell-curve that they learn to manipulate is an assumption that has been imposed from without, and not something that we are guaranteed by the Central Limit Theorem provided we sample the right way and look at a particular class of statistics. I have similar experiences with applets for confidence regions, or p-values, or even for surfacing less abstract ideas such as how one might use boxplots to defend or refute the well-known claim that the 1970 Vietnam draft was statistically faulty.

These deep connections, that are not intended to be made mathematically in a general education course, but which need to be confronted for a university-level understanding of what statistical science really brings to conversations as common to the well-educated as what surgery is "worth having" and what it means to say that Hillary Clinton is still in the hunt for the Democratic nomination for President, based on a poll that said 53% of likely voters would vote for Obama and 47% for Clinton. This is where the science is. This is what they have to be getting. Otherwise they are getting a repeat of what they saw in their middle school classrooms and that is a disservice to the students as well as to the science.

Second, I have some suspicions that the electronic manipulatives have gotten too passive, too slick, if you will. While I can't possibly "do" as many activities when I use concrete manipulatives, I have seriously started to believe that it might be more prudent to reduce the number of learning outcomes I have for the course, and become more intentional, focusing on fewer, but more truly active, hands-on manipulatives, This is, in a way, a position of near treason for me. I was the first faculty member in my Department to use electronic manipulatives and part of the very first wave nationwide. I'm not saying I don't like them or that I don't believe in them. I am saying that I don't believe they are fully getting the job done and I'd like to see if we can do better with concrete manipulatives.

There has already been some work done in this area, though most of what I have been able to find has been focused on K-8 classrooms. Phyllis (2001) compared computer and concrete manipulatives for teaching probability concepts to elementary school students and found mixed results. Teachers involved in that study were not convinced that either should be used as the full expense of the other. Klahr et al (2006) did a similar comparison in middle school classrooms on an engineering design project and found no differences in learning assessments based on type of manipulative used. The endorsement we all have, still, for electronic manipulatives is based in how easy they are to access, how one is always available for every student, and how they can give you instant feedback. We have a national expert on concrete manipulatives on our own campus. Dr. Joseph Straley, our colleague from the Physics Department has developed his incredibly innovative "physics petting zoo." Dr. Straley feels very strongly that "people like puzzles" and he has used his self-constructed physics puzzles to facilitate encounters with a wide variety of physical laws and principles. Indeed, these manipulatives are the basis for the NSF-sponsored Newton's Universe grant that, in part, is training middle school teachers how to better incorporate these sorts of activities into a practice of inquiry science in their classrooms. The reader is referred to http://www.pa.uky.edu/sciworks/physicspettingzoo/ for a tour of this amazing set of manipulatives. We will not need such intricate construction, or almost surely will not, for this project, but Dr. Straley has agreed to provide assistance as needed in helping me think through my use of concrete manipulatives.

Proposed Timeline

This timeline assumes, I hope not irrationally, that any new curriculum reform will be not be online at the University until the fall of 2010. If changes are made more quickly than anticipated then some changes will have to be made to this timeline.

Fall 2008

- Utilize colleagues at U.K. and at other (to be specified) universities to identify 3-5 concepts that are agreed upon as essential in a general education environment for understanding the role of statistical science in our everyday lives.
- Utilize colleagues at U.K. and at other (to be specified) universities to begin to design (or purchase) concrete manipulatives to address the concepts mentioned above.
- Identify, archive, and prepare to deploy electronic manipulatives to address the concepts mentioned above.
- Begin the IRB approval process for the study.

Spring and Summer 2009

- Complete construction or purchase of the concrete manipulatives.
- Recruit and train graduate student for study (from within our existing enrollment) in the use of both types of manipulatives.
- Pilot test the concrete and electronic manipulatives on Statistics Department TAs.
- Carefully design the appropriate parts of the STA 200 curriculum so that that either set of manipulatives can be employed, depending on treatment section.
- Finalize the IRB approval for the study.
- Begin to design an assessment (emphasis on simplicity of design) to address the content and the classroom environment.

Fall 2009

- Begin study. I will teach a 72 person section of STA 200 set up as one lecture and three breakouts. All sections will use computer-based manipulatives. I will teach all the recitation breakouts personally, but will be assisted by the trained TA.
- Finish assessments and pilot test the content assessment using Statistics Department TAs and colleagues.
- Implement assessments as part of the syllabus in course being taught.
- Collect data from this part of the study.
- Finish construction of concrete manipulatives.

Spring and Summer 2010

- Continue study. I will again teach a 72 person section of STA 200 set up as one lecture and three breakouts. All sections will use concrete manipulatives. I will teach all the recitation breakouts personally, but will be assisted by the trained TA.
- Implement assessments as part of the syllabus in course being taught.
- Collect data from this part of the study.
- Analyze data and compare the content assessment scores, and environmental assessment scores in the two treatments.

Fall and Spring 2010

It is not possible to know exactly what we can do here. If some clarity emerges from the first two years, then the idea would be to work to infuse this process into the new statistics course that is anticipated to be part of the new general education curriculum. This would involve a substantial amount of curriculum work, as well as a substantial effort to create more concrete manipulatives (duplicates), and to train TAs to use the materials.

References

1. Phyllis, H. (2001). The effects of using computer manipulatives in teaching probability concepts to elementary school students. EdD Columbia University Teachers College. Supervisor: Brice R. Vogelli.
2. Klahr, D., Triona, L., and Williams, C. (2006). “Hands on What? The Relative Effectiveness of Physical Versus Virtual Materials in an Engineering Design Project by Middle School Children.” Journal of Research in Science Teaching; Published online by Wiley.