

Benchmark Lessons and the World Wide Web: Tools for Teaching Statistics

Andrew Schaffner
Department of Statistics
University of Washington
USA
andrew@stat.washington.edu

David Madigan
Department of Statistics
University of Washington
USA
madigan@u.washington.edu

Earl Hunt
Department of Psychology
University of Washington
USA
ehunt@u.washington.edu

Edith A. Graf
Department of Psychology
University of Washington
USA
egraf@u.washington.edu

Jim Minstrell
Mercer Island High School
USA
jminstrl@wnet.org

Martha Nason
Department of Psychology
University of Washington
USA
mnason@u.washington.edu

1. Introduction

With the advances in computing technology over the past 15 years, much of applied statistics has changed. Researchers and students can now easily fit a wide variety of complex models and generate elaborate graphical displays of data with the press of a few buttons on a keyboard. Statistical software with graphical user interfaces does not require a conceptual understanding of the driving ideas behind the applications: design of experiments, data collection, descriptive statistics, and statistical inference. While traditional undergraduate applied statistics courses may provide some background and useful algorithms, they frequently leave a student with a cook-book of algorithmic recipes rather than any true understanding [Hawkins, et al. 1992], [Garfield & Ahlgren 1988]. In addition, while students may be able to memorize some algorithmic details to solve a familiar well-defined problem, they often fail to grasp the underpinnings of the methods and are therefore unable to apply their knowledge to new situations.

Statistics educators stress the need to move towards a conceptual rather than a mechanical understanding of statistics [Madigan, et. al 1995]; [Hawkins, et al. 1992]; [Garfield 1995]; [Moore 1991]; [Hogg 1990]. The ability to collect, organize, display, and interpret data as well as communicate findings are basic inter-disciplinary technical skills in academia and the marketplace. Because high-speed personal computers and statistical software perform analyses more quickly and easily than ever before, we now have an opportunity to focus more attention on the underlying concepts of statistical analysis rather than the mechanics.

Cooperative learning strategies provide an effective mechanism for teaching and learning conceptual information [Reynolds et al. 1995]. The theoretical foundations for cooperative learning derive primarily from four theoretical perspectives: *social learning theory* (teamwork), *Piagetian theory* (conflict resolution), *Vygotskian theory* (community collaboration), and *cognitive science research* on experts and novices [Murray 1990]; [von Glaserfeld 1991]. Building on these cooperative learning theories (with a special focus on cognitive science research), diSessa, Hunt, Minstrel, and van Zee developed Benchmark Instruction in the context of high-school physics. Benchmark Instruction is a genre of teacher instigated full-class discussion aimed at promoting conceptual changes in students' thinking. Benchmark Lessons draw out and engage students' own ideas in a rich context of communal inquiry [diSessa 1993]; [diSessa & Minstrell 1995]; [Hunt & Minstrell 1994]; [van Zee & Minstrell 1994].

In the following sections, we discuss the usage of benchmark lessons and present tools that facilitate virtual benchmark lessons for use in large collegiate classrooms where fully interactive class-wide discussions would otherwise be impossible. In addition, we share our experiences using the virtual benchmark lesson in the context of statistics instruction and report on its effectiveness.

2. Benchmark instruction

2.1 A sample benchmark lesson

Statistical inference often marks the zenith of an introductory statistics course. Here is a benchmark which naturally brings out many important “big ideas” in statistical inference such as paired vs. pooled t-tests, practical significance vs. statistical significance, and costs and decision making.

A baby with excessive red blood cells (polycythemia), may suffer restricted blood flow through tiny blood vessels resulting in a number of serious or life-threatening conditions. A new machine called the HemoCue has been developed to allow nurses to measure hemoglobin in the nursery. (Hemoglobin is the protein that gives red blood cells their characteristic color and that enables them to transport oxygen). Using an optical sensing method, this machine gives a digital readout claimed to be equivalent to the standard hemoglobin (Hgb) test -- and it only requires 2 drops of blood. This machine is relatively expensive to buy and maintain, but it promises efficient Hgb determinations using little blood. In particular, since it can be installed in the nursery it would provide immediate results which could improve infant care.

The data for this study was collected at a Northern California hospital during a six week period in 1992. Forty-three babies were selected from among those born at the hospital

The most immediate question of concern to the investigators in this study is whether the HemoCue machine gives Hgb results similar to the ones performed in the hospital lab. We have both ‘HemoCue’ and ‘LabHgb’ values for the babies in the study. A summary of the data is below:

	mean	s.d.
HemoCue	17.495	2.088
Lab Hgb	17.995	2.292

Because each baby had their blood analyzed by both the HemoCue and LabHgb we also have data on the difference between the two readings for each baby. The difference is defined as ‘LabHgb’ minus ‘HemoCue’. Below is a summary of the differences. Furthermore, for all but one of the babies the LabHgb was higher than the HemoCue.

	mean	s.d.
DifHgb	.5000	.3677

Using the benchmark lesson, we ask the students to participate in three primary phases of discussion (which we discuss in more detail in [Section 3.1]):

1. Presenting initial response to the problem or idea and a justification;
2. Critiquing and discussing the ideas drawn out by the lesson;
3. Reflecting on what was learned and generalizability.

A goal of the benchmark that we emphasize with the students is that we are not simply solving a particular problem, but we are *learning how to solve problems in general*. We encourage the students to be aware of and identify the key concepts of the problem and to *develop their communication skills* while answering the questions and providing understanding to the rest of the group.

Here is a particular student’s initial response which touches on many of these points:

There is a difference. I used the DifHgb data because *each* baby had his/her blood tested by both machines. Thus we can use the matched-pairs t procedure.

I decided to reject the machine based on the following calculations:

[calculations omitted]

The p-value [for $t=8.92$] (with d.f. = 42) is $<< .0005$. Therefore, I rejected H_0 because there is a difference between the two machines. This rejection was too hasty, however. Just because it is statistically significant does not make it practically significant. For example, I originally assumed that the LabHgb method is more accurate, and thus any difference from it indicates a fault. Is this assumption correct? Maybe the difference I detected through the above method is due to the fact that the HemoCue is *more* accurate than the lab

method. I have answered whether there is a difference, but need more info before I can intelligently decide whether to accept or reject HemoCue.

S2 agrees and adds some more comments:

I agree with the test that you used and the way about which you interpreted the result...In order to better assess whether to buy the machines or not we would need to know what an 'excessive' number of red blood cells is before it becomes hazardous to one's health. The HemoCue machine gave almost consistently lower results than the old machine by an average of about 0.5, but if it takes a red blood cell count of much more than the difference between the two machines then it really wouldn't be a significant difference in the results. Furthermore, depending on the budget of the hospital, it might be too expensive and really not worth the price that they need to pay for this new machine...if it doesn't give a practical significant difference in their results other than the benefit of using less blood and perhaps giving faster readings. I would probably want more tests done before I decided to purchase a machine.

S3 realizes that she made a mistake after reading S2's critique:

...I admit that I was a little too hasty in rejecting the new machines. You made a good point about how to interpret the results – are they significant in the practical sense or not?

The above problem and responses raise many issues in statistical inference (i.e., Should you use a pooled or paired t-test? How does the statistical test help you make a decision?) The problem also makes other points to challenge higher ability students (i.e., Are there other tests that you can use? Is the machine better, worse, or simply different from the lab? What other information would you need to make a more informed and economical decision?) In general, this problem allows students of varying ability to each construct a deeper and more general understanding of a particular cluster of statistical concepts.

2.2 Benchmarks in small classes

Benchmark lessons have shown great promise for small classes. High school physics students instructed by Jim Minstrell and by others using benchmark lessons consistently outperform other students in physics tests [Hunt & Minstrell 1994]. We conducted a pilot experiment involving 20 students comparing the effectiveness of benchmark lessons and traditional lecture-style instruction in the undergraduate statistics classroom. Students receiving the benchmark lessons significantly outperformed their counterparts ($p=0.001$) [Madigan, et al. 1995]. Benchmarks are fun to participate in; students get excited; and most importantly, students learn and develop an appreciation for statistics. However, there are obvious logistic problems when we come to a 200 person lecture.

3. Virtual Benchmark Instruction

We developed a virtual discussion technique to accommodate benchmark instruction in large classrooms (i.e., approximately 30 or more students). Virtual benchmark instruction uses the same types of lessons as "live" benchmark instruction, but the discussion takes place outside of the classroom in a virtual environment using the World-Wide Web. At the University of Washington, our typical classroom size for undergraduate introductory statistics is approximately 150 students making full-class discussions impossible. The virtual environment allows us to break the class into smaller workable discussion groups of 6 to 8 students so that everyone has the opportunity to participate and contribute new or supporting ideas to the discussion.

NCSA's HyperNews provides the core functionality for benchmark discussion groups. HyperNews blends together the hypermedia structure of the World-Wide Web with the message posting capabilities of Usenet news. World-Wide Web browsers exist for many systems (UNIX, PC, MAC) so students can use it in classrooms, laboratories, at home or anywhere with an internet connection. HyperNews discussions are easy to follow because HyperNews automatically links messages to each other as they are being posted making navigation and information retrieval

easier and more efficient than using email or Usenet News [1]. [Figure 1] shows how these links graphically appear [2].

3.1 Facilitating a virtual lesson

We impose a fairly rigid structure on the discussion. The virtual benchmark lesson has four primary parts which have evolved over time: (1) initial response and justification, (2) critique, (3) discussion and rebuttal, and (4) reflection and summary. (In fact, the same structure is imposed by a teacher in a classroom discussion [diSessa & Minstrell 1995]). Our most recent virtual benchmarks follow schedule as in [Table 1]. This schedule is demanding as it requires daily participation; however, rewarding discussions must be active and timely.

Initial Response and Justification. The first thing a student must do is commit a response with justification to the problem. We initially blind the students from each other's responses by rendering the posted responses unviewable until everyone has posted and committed an answer and justification. Temporarily blocking the responses forces the students to think independently and to be unbiased by other students' thoughts.

Critique. The second stage, critique, marks the beginning of the benchmark lesson. Once all initial responses have been posted, students read each other's posts and are required to critique at least one other post by either arguing in favor or against it while providing support and examples for their position. We remind the students that the goal of the critique is to be constructive. The student critiquing must communicate where the weaknesses lay in the reviewed post and suggest alternatives and improvements.

Discussion. During the two days of discussion following the previous day's critiques, the students must assist each other in their understanding. We tell our students that they have two primary objectives during the discussion: (1) to help everyone discover how to best solve the problem at hand and (2) to ensure that everyone in the group understands what the problem is, the concepts behind the problem, and the group solution.

Reflection and Summary. We wrap up the discussion with reflection and summary. Again, temporarily blinded to other responses, each student makes a final contribution to their group's collection of posts by contributing a short summary of what was learned in the benchmark lesson. In the reflection phase we ask them to explain why the benchmark lesson was important, and what the general concepts were. This reflection time helps them solidify and digest their understanding as well as serve as a final diagnostic check for the instructor to make sure everyone understood and synthesized the main points.

In addition to the web work, at the end of the week each student generates an entry for a personal journal that more extensively summarizes what was learned during this benchmark lesson. When the course is complete, the journal should serve as a valuable personal repository of statistical concepts and examples.

3.2 The pros and cons of virtual benchmarks

Pros of virtual benchmarks. Each student is given the opportunity to voice their opinions and participate. Often shy students won't speak out in a classroom, or there isn't enough time for everyone to contribute. But, in the virtual environment everyone has a chance and the environment feels less confrontational. We hope that this environment encourages the students to post their ideas without fear of being judged right or wrong. We wish to foster students' value of criticism as well as help them learn to be critical of their work and others.

The virtual benchmark is a forum where students can look to each other for understanding (as did S3 in our earlier example). High-ability students may benefit from explaining ideas to low-ability students, gaining the intellectual

[1] For more information about HyperNews see the HyperNews home page at <http://union.ncsa.uiuc.edu/HyperNews/get/hypernews.html>.

[2] Visit <http://www.stat.washington.edu/andrew/fbl.html> to visit archived and active virtual benchmark lessons.

benefit from teaching while still benefiting from solving new problems and learning new information. Low ability students may benefit from having peers explain concepts in terms closer to their understanding.

Virtual benchmarks are valuable to the instructor. Because discussion proceeds openly without judgment, teachers are able to get a deeper idea of what is and is not understood by the class as a whole. Using benchmarks to open topics up, teachers find out at what level the class is at and where they should begin helping them construct their knowledge on a particular topic.

All of the important aspects of benchmark lessons are incorporated into the virtual benchmarks with the added advantage of increased student involvement and a record of progressing ideas.

Cons of the virtual benchmark. The primary advantage of the virtual environment (that everyone can participate fully) is also the primary disadvantage. With every student participating in many ways and at many levels, it is very difficult for an instructor to give each response the careful and thorough attention it deserves. A course with 200 students and a minimum of 3 posts per student per week adds up to 600 responses! Though this number is large, we have found that with daily monitoring, an instructor and assistants can still actively participate and guide the discussions.

Another problem is that there is no way to ensure that each student reads all of the other postings in the group. They might read only one or two posts before selecting the response that they will critique. However, compared to live discussions, students in virtual discussions will tend to participate and think more deeply since a certain level of participation is required and evaluated for their grades.

4. Conclusion

Using the technology of the World Wide Web, we have extended the use of proven benchmark instruction to large classes. The availability of these lessons to larger audiences can improve students conceptual understanding of statistics, develop critical thinking skills, and exercise communication skills. In conjunction with class lectures, text-book readings, and exercises, the virtual benchmark lesson provides a powerful new environment for statistics instruction.

5. References

- [Bruer 1993] Bruer, J.T. (1993). *Schools for Thought: A science of learning in the classroom*. Cambridge: MIT Press.
- [Cohen 1994] Cohen, E. G. (1994). *Designing Groupwork: Strategies for the Heterogeneous Classroom, 2nd ed.* New York: Teachers College Press
- [diSessa 1993] diSessa, A.A. (1993). Towards an epistemology of physics. *Cognition and instruction*. **10**, 105–225.
- [diSessa & Minstrell 1995] diSessa, A.A. and Minstrell, J. (1995). Cultivating conceptual change with benchmark lessons. In: Greeno, J.G. (Ed.), *Thinking Practices*, to appear.
- [Garfield 1995] Garfield, J. (1995). How Students Learn Statistics. *International Statistics Review*, **63**, 1, 25–34.
- [Garfield & Ahlgren 1988] Garfield, J. and Ahlgren, A. (1988). Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research. *Journal for Research in Mathematics Education*. **19**, 1, 44–63.
- [Hawkins et al. 1992] Hawkins, A., Jolliffe, F., Glickman, L. (1992). *Teaching Statistical Concepts*. New York: Longman Publishing
- [Hogg 1990] Hogg, R.V. (1990). Statisticians gather to discuss statistical education. *Amstat News*. pp. 19–20.

[Hunt & Minstrell 1994] Hunt, E.B. and Minstrell, J. (1994). A cognitive approach to the teaching of physics. In: McGilly, K. (Ed.), *Classroom lessons: integrating cognitive theory and classroom practice*. MIT Press/Bradford Books.

[Madigan et al. 1995] Madigan, D., Clarkson, D.B., Donnell, D., Hunt, E., Keim, M., Minstrell, J., Nason, M., Schaffner, A., Volinsky, C.T. (1995) Facet-Based Learning for Statistics. Under editorial review in *Statistics and Computing*.

[Moore 1991] Moore, D. (1991). Uncertainty. In: L. Steen (Ed.), *On the Shoulders of Giants*. Washington, D.C.: National Academy Press.

[Murray 1990] Murray, F.B. (1990) Co-operative Learning. In: Entwistle, N. (Ed.), *Handbook of educational ideas and practices*. New York: Routledge.

[Reynolds et al. 1995] Reynolds, B.E., Hagelgans, N.L., Schwingendorf, K.E., Vidakovic, D., Dubinsky, E., Shahin, M., Wimbish, G.J. (1995). *A Practical Guide to Cooperative Learning in Collegiate Mathematics*. MAA Notes Number 37.

[von Glaserfeld 1991] von Glaserfeld. (1991). A constructivists view of learning and teaching. In: Duit, R., Goldberg, F., and Niedderer (Eds.). *Research in physics learning: Theoretical issues and empirical studies, Proceedings of an International workshop*. Kiel, Germany: IPN at the Univ. of Kiel, Germany, 129–140.

[van Zee & Minstrell 1994] van Zee, E.H. and Minstrell, J. (1994). Using questioning to guide student thinking. Submitted for publication.

[Yackel, Cobb, and Wood 1991] Yackel, E., Cobb, P., and Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education*. 22, 5, 390-40

6. Tables and Figures

Monday	Tuesday	Wednesday	Thursday	Friday
Initial Response and Justification	Critique	Discussion and Rebuttal	Discussion and Rebuttal	Summary

Table 1: Weekly Schedule of virtual benchmark lesson events.

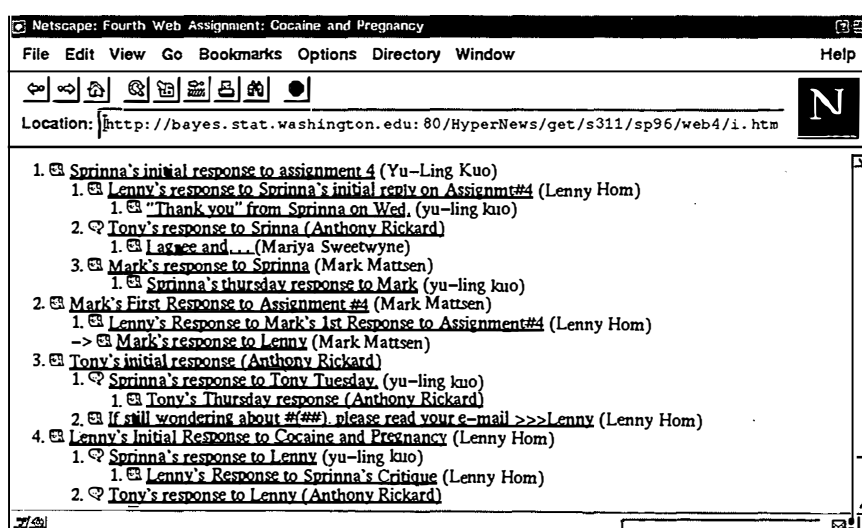


Figure 1: A virtual benchmark lesson in progress.