Applying Japanese Lesson Study Principles to an Upper-level Undergraduate Statistics Course

Paul Roback  
St. Olaf College

Beth Chance  
California Polytechnic State University, San Luis Obispo

Julie Legler  
St. Olaf College

Tom Moore  
Grinnell College

Abstract

Japanese Lesson Study is a collaborative approach for teachers to plan, present, observe, and critique classroom lessons. Through the lesson study process, teachers systematically and thoughtfully examine both student learning and their own teaching practices. In addition, the process paves the way for a much broader approach to education research by gathering data about student learning directly in the classroom. By piloting an approach using Japanese Lesson Study principles in an upper division statistics course, we discovered some of the challenges it poses, but also some surprisingly promising results for statistics teaching. This case study should provide others considering this approach with information about the philosophy and methodology involved in the lesson study process as well as some practical ideas for its implementation.

1. Introduction

Lesson study, a professional teacher development process incorporating collaborative lesson planning with thoughtful and purposeful observation and reflection on the lessons, has been generating growing
excitement in the U.S. educational community over the past decade. In 1999, Stigler and Hiebert published *The Teaching Gap*, an influential report on the Third International Mathematics and Science Study of eighth-grade mathematics lessons in Japan, the U.S., and Germany. In their book, Stigler and Hiebert make a compelling, evidence-based case for Japanese Lesson Study, hypothesizing “that if our educational system can find a way to use lesson study for building professional knowledge of teaching, teaching and learning will improve” (p. 131). They tout the potential of institutionalizing the lesson study process for bridging the gap in teaching methods between the U.S. and Japan, which in turn could help bridge the well-documented gap between U.S. and Japanese mathematics students in learning and achievement. Over the last five years, an estimated 150 lesson study clusters have been formed in the United States, involving 335 schools in 32 states with over 2300 teachers (Chokshi 2004). This activity has been primarily at the K-12 level; our goal was to implement principles of the lesson study approach at the undergraduate level, specifically in an upper-level undergraduate statistics course.

We began our undertaking intrigued by what we knew of Japanese Lesson Study, but extremely “green” with respect to its implementation. Not only were we newcomers to the ideas shaping lesson study, but we could find nothing in the literature to guide the implementation of lesson study specifically at the college level, especially in an upper-level statistics course. Thus, we embarked on a pilot implementation—a preliminary attempt to assess the feasibility of Japanese Lesson Study principles in upper-level undergraduate statistics courses. We hoped to gain insight into concrete benefits and potential pitfalls. In this manuscript, we will first describe Japanese Lesson Study —its philosophy, its process, its desired outcomes, and early findings from its implementation at the K-12 level in the United States. Second, we will outline the process we followed in implementing lesson study principles in a Mathematical Statistics course at St. Olaf College in the spring semester of 2004. With a prerequisite of Probability Theory, this course was targeted toward juniors and seniors who were mathematics majors or statistics concentrators with no previous course in statistics. Finally, we will report the results of our implementation and offer suggestions and recommendations for others who might consider this approach.

2. Lesson Study: Background

“The expression *lesson study* is a literal translation for the Japanese word *jugyokenkyu*—*jugyo* means *lesson* and *kenkyu* means *study* or *research.*” (Fernandez 2002, p.394) Lesson study, however, is more than merely studying or perfecting a single lesson; instead, it is a process by which teachers systematically examine their practice and their students’ learning to become more effective instructors. At the heart of lesson study are *kenkyujugyo* (study lessons), which are collaboratively planned and systematically evaluated, and which hopefully reveal larger truths about effective teaching and student learning. The professional development process itself is more important than any specific lesson that is developed. Fernandez and Chokshi (2002) end their article with the following statement: “Lesson study is not a vehicle for creating a library of tried-and-tested lessons for teachers to borrow from a shelf and import into their own classrooms. It is a process for creating deep and grounded reflection about the complex activities of teaching that can then be shared and discussed with other members of the profession.”

As outlined in Curcio (2002), the lesson study process involves several important steps:

1. **collaborative planning.** It is recommended (Fernandez 2002) that groups of 4-6 teachers come together for 10-15 hours over 3-4 weeks to carefully plan a single specific lesson (as opposed to a longer unit of material) which will address one or more overarching goals.

2. **teaching and observing.** One member of the group teaches the lesson as designed, while the other group members and outsiders observe the class, taking detailed notes regarding the reactions and
engagement of the students.

3. **analytic reflection.** The teacher, other group members, and observers gather soon after the lesson has been taught to share thoughts and insights, and to evaluate the success of the lesson in meeting its objectives.

4. **ongoing revision.** Based on experience and evidence, the lesson is often revised and taught again, and the process is repeated.

Each lesson study group must identify a broad overarching goal and develop a set of specific objectives. The broad goal contains a vision of the type of student the educational community wishes to produce, identifying gaps between the ideal student and what teachers typically observe. Examples cited include themes such as “be active problem-solvers” or “develop scientific ways of thinking” (Lewis and Tsuchida 1998, p. 14). Once the broad goal is formulated, the group then identifies a specific lesson topic that might address this goal. Selection of the lesson topic customarily involves deep study of the current curriculum, student backgrounds, and other initiatives designed to address the broad goal. The group then forms a second, more specific, set of objectives related to the lesson topic selected. For example, Lewis and Tsuchida (1998) describe a lesson designed with the broad goal of encouraging fifth-grade Japanese students to demonstrate scientific thinking. Specifically, the teacher asks pairs of students to study the effects of three variables generated by the class on the cycle time of a pendulum, with the objective of being able to separate the effects.

A strength of lesson study is the atmosphere of collaboration that it fosters. Teachers bring different perspectives and experiences to a common task. As noted by U.S. educational researcher Richard Elmore, “isolation is the enemy of improvement” (Lewis 2002, p. 11). However, lesson study differs from other collaborative activities because it “makes teacher collaboration concrete and focuses on a specific goal: better understanding of student thinking in order to develop lessons that advance student learning” (Wang-Iverson 2002, paragraph 7). Lesson study guidelines (Fernandez and Chokshi 2002) advise teachers to make the most of their limited meetings, working out fine details of lesson plans and handouts between meetings, while using meeting times for examination of materials, plotting general strategy, and discussion of larger issues.

In lesson study, collaboration does not end with the development of a lesson plan; rather, the collaboration has just begun. As one group member teaches the lesson, the others (and possibly outsiders) observe the class with a careful eye toward how students engage with and process the material, guided by questions posed and objectives stated during the planning process. Thus, the focus of observation is not the teacher, but the students and their learning. Class observers should follow a clear protocol of behavior (Curcio 2002); for instance, they should refrain from interfering in the lesson, e.g., answering student questions, but they should be free to ask clarification questions of students. As discussed in Watanabe (2002), insightful observation does not happen automatically; rather, it is a skill teachers must learn. Part of that skill is the ability to gather meaningful information beyond what can be gleaned from tests, written assignments, or even videotape. Lewis (2002) cites records of student engagement, persistence, degree of interest, emotional reactions, and quality of small-group discussion as examples of meaningful data.

Collaboration then continues as the group reconvenes to reflect thoughtfully on the class period. Again, there are suggested protocols for these feedback sessions (Curcio 2002), which can be summarized by some basic tenets:

- allow the teacher to provide initial reflections uninterrupted, followed by the group members and finally outside observers;
- focus comments on the lesson and the students;
- provide concrete evidence behind points;
- comment on positives as well as areas for improvement;
- listen completely to allow time for processing and to avoid point-by-point rebuttals.

Ideally, the group will modify their lesson plan based on these discussions and a different teacher will teach the lesson to a new group of students. Other group members will observe this class session, and the process will be repeated.

Lesson study, though, is more than just a collaborative activity, and the development of a study lesson is much more than a set of lecture notes. One crucial product created by the group is the lesson study plan. As described in Fernandez and Chokshi (2002), Japanese teachers often use a four-column chart. (Curcio (2002) describes a slightly different four-column chart.) Column One contains the steps of the lesson—the sequence of topics, examples, and questions that the teacher has planned. Column Two contains student activities and expected student responses and reactions for each step in the lesson. Column Three contains points for the teacher to remember, ways in which the teacher might deal with student responses, and ways to tie the lesson together. Finally, Column Four lists methods for evaluating whether each segment of the lesson was successful in achieving its goals.

Through the organization of these columns and the creation of a lesson study plan, some points of emphasis become evident. For instance, most typical lesson plans that we do in our everyday teaching would stop with Column One. The expected student responses in Column Two and the teacher reaction to these responses in Column Three illustrate the focus on student learning—considering a priori how students will be processing information, forming questions, and constructing new knowledge. Through lesson study, teachers develop “the eyes to see students (kodomo wo miru me)” (Lewis 2002, p.12). Furthermore, the evaluation methods in Column Four illustrate the focus on research, as “the classroom becomes the teachers’ laboratory for continuous improvement of teaching and learning,” (Wang-Iverson 2002, paragraph 9) and one can assess objectively the success of the lesson in meeting stated goals.

The reader can find more detailed guides to the implementation of lesson study in the references (see especially Stigler and Hiebert 1999; Curcio 2002; Fernandez 2002; Fernandez and Chokshi 2002; Lewis 2002; Watanabe 2002) and at the following websites of lesson study research groups: www.tc.columbia.edu/lessonstudy and www.lessonresearch.net. Researchers in these groups and elsewhere identify the many benefits noticed in Japan and the U.S. of adopting a culture of lesson study. Frequently cited benefits (Lewis and Tsuchida 1998; Lewis 2002; Lewis, Perry, and Hurd 2004) for teachers and teaching practice include increased knowledge of subject matter, increased knowledge of instruction, increased ability to observe students, increased focus on student learning, stronger collegial networks, stronger support for novice teachers, stronger connection of daily practice to long-term goals, stronger motivation and sense of efficacy, support for taking risks, and improved quality of available lesson plans. Benefits for students include improved achievement, learning more carefully considered content more deeply, enhanced ability to make connections, and a higher level of engagement with the material.

The first lesson study groups in the United States were formed only five years ago, and educational researchers caution that “lesson study is easy to learn but difficult to master” (Chokshi and Fernandez 2004, p.524). Fernandez, Cannon, and Chokshi (2003) describe the development of three new lenses for examining lessons:

1. a curriculum developer lens, to see how to sequence and connect learning experiences;
2. a student lens, to understand student thinking, anticipate student behavior, and learn to build student understanding; and,
3. a researcher lens, to see how to use the classroom as a laboratory for generating data-driven conclusions about pedagogy.
Through these lenses, U.S. practitioners of lesson study have documented challenges which researchers (Stigler and Hiebert 1999; Fernandez 2002; Lewis 2002; Wang-Iverson 2002) maintain must be overcome before obtaining the successful outcomes common in Japan.

Through the curriculum developer lens, one immediately notes that the curriculum in the United States is over-prescribed compared to that in Japan, leaving less time to explore topics in depth. If U.S. teachers choose to lead students to knowledge construction, they run the risk of not finishing the race to complete a lengthy list of topics. This pressure exists in many undergraduate statistics courses, too, as instructors try to pacify client disciplines or other teachers using a particular course as a pre-requisite. Another challenge cited for K-12 educators in the United States (Fernandez 2002) is the lack of common curricular ground, in contrast with the national curriculum strictly followed by Japanese teachers. A successful lesson cannot be planned without carefully considering student backgrounds and the place of a lesson in the larger curriculum. College instructors may face an even bigger challenge in this regard with their considerable freedom to choose material to cover, presentation style, and primary texts. For example, a lesson study group of undergraduate instructors may find themselves debating at length which topics to include and in what order to present them before finally focusing on a specific lesson plan.

Lack of proficiency using the student lens is another barrier to successful implementation of lesson study. U.S. teachers are not often trained to analyze each problem, each question posed, and each choice in idea development from the perspective of the student. For instance, when posing a problem to the class, it is not enough to list potential student solutions; a teacher must consider what each solution says about student understanding and processing. Effective use of the student lens requires teamwork among educators and careful assessment of student learning. One barrier, then, is the independent nature of those attracted to teaching, which may become more pronounced with more experience. U.S. teachers at all levels customarily teach in isolation, not routinely opening their classrooms to outside observers and constructive criticism. Yet observation of students during lessons is essential to the development of a student lens.

From the viewpoint of some researchers (Fernandez, Cannon, and Chokshi 2003), the biggest challenge to the successful implementation of lesson study is the ability of teachers to examine lessons through a researcher’s lens. As Fernandez and her colleagues observed Japanese teachers mentoring U.S. 5th and 6th grade teachers on the lesson study process, they noted that “the Japanese teachers emphasized four critical aspects of good research: the development of meaningful and testable hypotheses, the use of appropriate means for exploring these hypotheses, the reliance on evidence to judge the success of research endeavors, and the interest in generalizing research findings to other applicable contexts” (p. 173). In adopting this researcher lens, a practitioner of lesson study must continually relate lesson steps to overall goals and objectives, carefully consider how to gather evidence to assess whether or not objectives are being met, and reflect on which insights gained might apply to future classroom settings.

### 3. Our Lesson Planning Process

Our initial group of seven undergraduate statistics educators convened for the first time at St. Olaf College on February 17, 2004. Given the typically isolated nature of statisticians at small colleges, we felt fortunate to assemble such a group, helped by a couple of members choosing to spend their sabbatical leaves in the area. Over time, the group contracted to the set of four co-authors; this final cross-institutional group was composed of members with some to considerable experience, both in terms of time teaching at undergraduate institutions and time spent in non-academic settings. Armed with only a novice’s knowledge of the lesson study process—based on readings of a few of the articles listed in the references—we were nevertheless intrigued by its potential and interested in implementing lesson study
principles at the undergraduate level. We decided to meet every other week and to enlist the services of a student recorder to take notes at our meetings.

After spending the first meeting watching a videotape overview of Japanese Lesson Study ([Curcio 2002](#)) and discussing the lesson study philosophy and process, we came to the second meeting ready to brainstorm about big goals and lesson content. The discussion was predictably wide-ranging, and it consumed much of the next few meetings. We discussed the important ideas we’d like students to remember from a statistics class, how to make those important ideas stick, which important ideas students struggle to understand, how to tie several ideas together, and how to manage the level of detail presented. We also mentioned good lessons and activities on which we could build—Cents and the Central Limit Theorem ([Scheaffer, Gnanadesikan, Watkins, and Witmer 1996](#)), the German Tank problem ([Scheaffer, et al. 1996](#)), golf tees inscribed with numbers from different distributions, etc. The concept of sampling distributions became the primary theme we wished to incorporate, in the context of goodness-of-fit tests.

Our target audience was 23 students (primarily juniors and seniors) in Math 312B: Mathematical Statistics at St. Olaf College. This section of Math 312, taught by one of the authors (Roback), consisted of students with no previous course in statistics; the prerequisite was Probability Theory, which the majority of students had taken the previous semester. The required textbook for Math 312 was *An Introduction to Mathematical Statistics and Its Application* by [Larsen and Marx (2001)](#); in addition, S-Plus programming was used on a weekly basis for running simulations, exploring properties of test statistics, and analyzing data (for more detail see the [course syllabus](#)). Weekly homework assignments contained a mixture of mathematical derivation, applied data analysis, and S-Plus simulation. The class met three times a week for 55-minute sessions, which were comprised of lecture and whole-class problem solving, with occasional small group activities. Students were expected to attend every class session so that full participation in classroom activities and take-home assignments could be assumed.

The study lesson on goodness-of-fit tests and sampling distributions was conducted in the next-to-last week of the 16-week semester, immediately after a unit on regression analysis and inference. Another author (Moore) observed the class and took notes of his observations. We also arranged to have both lessons videotaped, but between an absent videographer and marginal video quality, we could not gather as much information from the videotapes as we hoped.

Table 1 contains a short outline of our study lesson plan; more details can be found in the partial four-column study lesson plan in Table 2 and in the complete four-column study lesson plan; handouts from class can also be found at [handouts](#). Specific objectives for our study lesson included:

1. engaging students in an active way with lesson material;
2. having students apply statistical thinking to develop a test statistic;
3. having students suggest the need to examine an empirical sampling distribution (assuming the null is true) to decide if an observed test statistic value is surprising;
4. introducing the theory of the chi-square statistic, distribution, and goodness-of-fit test;
5. extending goodness-of-fit tests from the categorical to the discrete to the continuous case.

The plan in Table 1 is the last of several iterations, and it reflects the efforts from group meetings over 12 weeks, as well as efforts by several individuals between group meetings to fill in details and provide the group rough drafts to discuss.

---

Table 1. Short timeline of the study lesson plan

Page 6 of 20
| Time                      | Steps in Study Lesson                                                                                                                                                                                                 |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Day One                  | Discuss the general problem: How would an M&M manufacturer decide whether the colors of M&Ms are being produced in the correct proportions?                                                                           |
|                          | Discuss potential sample results: How much deviation is too much?                                                                                                                                                      |
|                          | Pass out M&M samples and form groups of two. Each group must devise a test statistic to measure the deviance of their sample from what they would expect if the process is working correctly. |
|                          | Groups of two combine to form groups of four, and each group of four selects and presents one of their test statistics (with rationale) to the class. The rationale should be based on deliberations about what defines a good test statistic. |
|                          | Pose the next problem: Based on their chosen test statistic, would they conclude that their original sample of M&Ms contains convincing evidence that the manufacturing process is malfunctioning? |
|                          | Begin to investigate empirical sampling distributions and p-values with hand calculations from simulated samples generated by S-Plus under the null hypothesis.                                                |
| At-home before Day Two   | Generate an empirical sampling distribution for the group test statistic and also for the chi-square goodness-of-fit statistic.                                                                                      |
|                          | Find empirical p-values for group’s original data and 10 prototype samples designed to illustrate the performance of the test statistic under specific cases.                                                          |
| Day Two                  | Discuss results from in-class and take-home assignments. Think about criteria for good test statistics.                                                                                                                |
|                          | Groups work on the Fumble Problem ([Larsen and Marx 2001, p. 253](#)) – students investigate how to extend the chi-square goodness-of-fit test to discrete probability distributions.                                         |
|                          | Guide groups to think about issues such as estimating model parameters, adjusting degrees of freedom, and avoiding small cells.                                                                                       |
It is important to recognize that lengthy discussion preceded the formulation of this lesson plan. For example, we spent considerable time planning the first few steps of Day One; we wanted to provide motivation with a realistic problem, and we wanted to lead students to develop a test statistic on their own. We ended up using the standard M&M multinomial distribution problem, partly because this Math 312 class was about to tragically complete their first full statistics course without ever eating or counting M&Ms, but mainly because it was a simple problem with a real context.

The introductory portion of Day One represents a unique product of this study lesson that never would have materialized without thoughtful collaboration (and which really never would have materialized under a typical presentation of this material based on Larsen and Marx (2001)). Students, in groups of two, were asked to think about how one could separate sample results into those that favor the null hypothesis and those that favor the alternative hypothesis. While designing their tests, students were asked to consider what properties a good test statistic should possess. Based on these properties, students then had two opportunities to present, defend, and potentially modify their invented test statistics—first as two groups of two came together to compare their respective test statistics, and second as the groups of four presented their chosen test statistics to the class. The four-column lesson plan for this introductory part of Day One is shown in Table 2; the remainder can be found at four-column lesson plan.

| At-home before Day Three | Conduct simulations to see the value in adjusting degrees of freedom in the chi-square distribution when parameters are estimated. |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------------|
|                         | Examine how the simulation extends goodness-of-fit tests to continuous probability distributions. |
| Day Three (not officially part of the study lesson) | Discuss results from Days One and Two, and the take-home assignment for Day Three. |
|                         | Develop the chi-square test of independence for two-way tables. |

It is important to recognize that lengthy discussion preceded the formulation of this lesson plan. For example, we spent considerable time planning the first few steps of Day One; we wanted to provide motivation with a realistic problem, and we wanted to lead students to develop a test statistic on their own. We ended up using the standard M&M multinomial distribution problem, partly because this Math 312 class was about to tragically complete their first full statistics course without ever eating or counting M&Ms, but mainly because it was a simple problem with a real context.

The introductory portion of Day One represents a unique product of this study lesson that never would have materialized without thoughtful collaboration (and which really never would have materialized under a typical presentation of this material based on Larsen and Marx (2001)). Students, in groups of two, were asked to think about how one could separate sample results into those that favor the null hypothesis and those that favor the alternative hypothesis. While designing their tests, students were asked to consider what properties a good test statistic should possess. Based on these properties, students then had two opportunities to present, defend, and potentially modify their invented test statistics—first as two groups of two came together to compare their respective test statistics, and second as the groups of four presented their chosen test statistics to the class. The four-column lesson plan for this introductory part of Day One is shown in Table 2; the remainder can be found at four-column lesson plan.

**Table 2. Four-column lesson plan for first part of Day One**

| Learning Activities and Key Questions | Student Activity and Expected Responses | Teacher’s Response and Things to Remember | Goals and Evaluation |
|--------------------------------------|----------------------------------------|------------------------------------------|----------------------|
| **Day One**                          |                                        |                                          |                      |
| Introduce general problem: How would an M&M manufacturer decide whether the colors of M&Ms are | | A candy manufacturer is told to make 13% brown, 14% yellow, 13% red, 24% blue, 20% orange, and 16% green candies, but he believes the |                      |
| being produced in the correct proportions? | manufacturing process is malfunctioning. | Brainstorm plan | Get students to suggest |
|------------------------------------------|----------------------------------------|----------------|------------------------|
|                                          |                                        | Suggest ways to evaluate the claim | - collecting data  |
|                                          |                                        |                              | - seeing how the data matches the claimed values |
|                                          |                                        |                              | - evaluating the statistical significance of the discrepancy |
|                                          |                                        |                              | May discuss power / sample size… e.g. Is one sample good enough? |
|                                          |                                        |                              | How easily do students consider sampling variability? |
| Discuss potential sample results: How much deviance is too much? | Expect students to be okay with a little deviance from null, but unsure of where to draw the line. | Present possible ways multinomial sample of size 40 could turn out – ask if each one provides significant evidence of malfunctioning. | Do they understand that some variability from what is expected is natural? |
| Introduce data | Each student receives a bag, work in pairs to get the tally for the first 40 M&Ms | Blindly take 20 candies from the big bag of M&Ms. - If asked, tell them to ignore broken ones. - If asked, tell them the bag were randomly purchased from a local store. | |
| Examine sample | Students tally the colors | Look at your sample results. Do they support the manufacturer’s claim? | Do students think beyond the sample? |
| Develop “custom” test statistics: While we expect some discrepancy, how can you decide if your sample is “too different” from expected? How can you measure how “deviant” your sample is? Can you express this as one number? | Students brainstorm ways to measure the deviation. Groups of 2 for 5 minutes. | What are some properties of your measurement technique? Do you expect the results to be large, small? Positive, negative? Pass out Handout #1. Students who want to use z-scores need to combine them in some way to come up with one number. | Which ideas from course do students latch onto? Do their custom statistics separate samples which agree with the null from those which agree with alternative? |
After selecting a final test statistic, students were asked to specify which values of their test statistic would provide strong evidence against the null hypothesis. Until this point in the course, any test statistic the students had considered had magically, after pulling a couple of theorems out of a hat, followed a well-known distributional form under the null hypothesis. Now, with their own creations, students needed to think about empirical sampling distributions and empirical $p$-values. We had defined and discussed sampling distributions at various points, and we had frequently used simulations in S-Plus to investigate issues of robustness, so the groundwork had been laid to explore empirical sampling distributions and $p$-values (or so we believed). By focusing on empirical sampling distributions, we were placing the specific objectives of developing goodness-of-fit tests within the broader goals of promoting statistical thinking and understanding the role of sampling distributions.

Another source of discussion and disagreement was the order of topics for Day One. To transition from the test statistics constructed by students to the chi-square statistic, we attempted to create “prototype samples” to allow students to examine the behavior of their test statistic in specific cases. The prototype samples were designed to illustrate extreme cases and introduce subtle cases that would show the advantages of the chi-square statistic. Table 3 shows some of these prototypes. For example, sample A reflects the most likely multinomial sample under the null hypothesis. Samples B and C were designed to illustrate how the test statistic handles discrepancies of the same absolute size, one in the more abundant categories and one in the less abundant categories. This comparison provided the biggest departure between the chi-square statistic and the most popular student choice (the average squared difference between observed and expected counts). Samples D and E were designed to illustrate the effect of sample size and Sample F was designed to illustrate how extreme results are handled.

Table 3. Examples of prototype samples

| Sample | Blue | Orange | Green | Yellow | Red | Brown |
|--------|------|--------|-------|--------|-----|-------|
| A      | 10   | 8      | 7     | 5      | 5   | 5     |
| B      | 10   | 8      | 7     | 5      | 9   | 1     |
| C      | 14   | 4      | 7     | 5      | 5   | 5     |
Originally, we planned to have students examine the prototype samples immediately after developing their own test statistic, as a way to determine properties, strengths, and weaknesses of their test statistic. However, we decided to follow the development of an invented test statistic with empirical sampling distributions and p-values, allowing the class to spend more time with this fundamental idea. We then introduced the prototype samples, hoping to lead students to see inefficiencies with their developed statistic and motivation for the chi-square statistic. In fact, we decided that the chi-square statistic could be effectively introduced between Day One and Day Two; students could simulate empirical sampling distributions for their statistic and the chi-square statistic, and compare the performance of both statistics on the prototype samples. In this way, we expected some thoughtful, empirically-motivated discussions at the beginning of Day Two with the underlying purpose of addressing the original question about the desired production proportions.

The development of customized test statistics on Day One took longer than expected, so we spent less time than expected (under 15 minutes) examining the empirical sampling distribution. As a result, the instructor gave a hurried summary of empirical sampling distributions and p-values at the end of Day One, and students were responsible for both examining prototype samples and the chi-square test statistic before Day Two, with the guidance of S-Plus template code provided in a handout. The logjam spilled into Day Two as well. The class and instructor spent more time than expected (over 20 minutes) sharing and summarizing what they had learned about empirical p-values and the performance of their test statistic compared to the chi-square statistic, but the discussion was too valuable to cut short. One problem was that, in most of the prototype samples the students investigated, the differences between the two test statistics were too subtle to be meaningful. Students, however, were intrigued with the idea that, by using simulation under the null hypothesis, they were free to employ any test statistic that they deemed sensible. In retrospect, it is not surprising that the subtleties of this new approach (i.e., empirical sampling distributions) took a while to sink in, despite the groundwork in place.

On Day Two, after a reflective discussion of Day One and a little theory about the chi-square goodness-of-fit test, we spent all our remaining time with the Football Fumbles example (see handouts). The agenda for Day Two had been slightly modified when the four authors met after Day One to review successes, surprises, and opportunities for improvement. Instead of the typical “present the formula, then trudge out an example” format, we designed Day Two—goodness-of-fit tests for specific distributions with parameters unknown—as a natural extension of Day One. Students were asked, based on how we attacked the M&M problem, to develop methodology for determining whether the number of fumbles in a game for each college football team could be reasonably modeled with a Poisson distribution. Student groups with a little prodding were able to extend from categories determined by M&M colors to categories determined by number of fumbles in a game. They then hit stumbling points our lesson study group had anticipated, and the instructor was able to direct them to think about issues such as: how do we determine the expected number of teams in each group? how do we handle the unknown parameter from a Poisson distribution? how do we ensure that the expected number of teams in each group exceeds some minimum (after the instructor cautioned about small expected values in light of model assumptions)? how do we determine p-values for our test of hypothesis? Once again, the class spent more time than we had expected on the Football Fumbles example, and we were not able to attack the
Cockpit Noise example, in which students would make a further extension to goodness-of-fit tests for continuous distributions. However, in planning for the at-home activity following Day Two, we illustrated (through S-Plus code) how one might use a goodness-of-fit test to determine if a set of data was sampled from a normal distribution. This illustration was housed in a simulation built to examine the advisability of adjusting the degrees of freedom in the chi-square test statistic when model parameters are estimated from sample data.

Day Three was not part of the study lesson planned by the group, but it used the ideas and activities from Days One and Two to bring the unit on goodness-of-fit tests to a satisfying (and efficient) close. After reflecting as a class on the main ideas from Day Two and the simulation results completed at home following Day Two, the chi-square goodness-of-fit test was extended to two-way tables of categorical variables as a test of independence.

Just as we had following Day One, the four authors met to debrief after Day Three. Beginning with comments from the instructor and the observer, we reviewed the entire lesson, comparing our plans and intentions with how the class actually proceeded and how the students reacted. Many of our observations relating to our original list of goals and objectives are included in the next section. This reflective session was absolutely vital, but it would have been even more valuable if all group members and even some outsiders had been able to observe the lessons being taught. Unfortunately, we were limited to a single observer because of teaching conflicts. Ideally, this reflective meeting would then be followed by the planning and implementation (by a different instructor) of a revised lesson based on insights acquired during the first teaching. In our case, the repeat session fell victim to lack of time at the end of the semester, although repeat sessions in the same semester are inherently challenging since different sections of the same course tend to move at similar paces. We address implications of the lack of additional observers and ongoing revisions in upcoming sections.

4. Reflection on Specific Objectives and Evaluation of the Study Lesson

The success of a lesson study can be assessed, on one hand, by reflecting on the broad goals and specific objectives set by the group. Our five objectives for the study lesson are listed in the previous section.

Objective #1 on student engagement was successfully accomplished. Notes by the classroom observer (Moore) mentioned that the problem set up created interest from the outset and that student groups actively sought solutions to questions posed. Over the previous 12 weeks of Math 312B, student engagement was, with a few exceptions, limited to working out problems in pairs, explaining concepts to partners, and class discussions. So the activities designed for knowledge construction for the study lesson required a higher level of engagement from the students. Although students were actively collaborating at several points during the lesson, we were often surprised at the slow progress of their collaborations. Perhaps our expectations were too high, given that some of us had not previously inserted activities with this level of student content responsibility in an upper level statistics course, but the fact that students were being asked to step out of their Math 312B “comfort zone” in the second-to-last week could have also played a role. For example, the observer noted that groups did not want to write things down, hoping to avoid commitment to written answers, although time limits eventually prodded groups to stick with an answer. Also, the observer noted that the 20-minute wrap-up at the beginning of Day Two was a valuable conversation even though students did more responding than inquiring. A high level of student engagement was still evident; higher quality engagement could be enhanced by implementing lessons such as this throughout the semester, or perhaps by planning a study lesson prior to the start of the semester and implementing it early in the semester.
Objective #2 on statistical thinking about test statistics was the primary focus of Day One, and it was met with fair success according to observer and teacher notes. The observer was asked, in the study lesson plan, to note “How easily do students consider sampling variability?” (Answer: pretty naturally, as they recognized that different sample results could come from the same underlying production process.), and “Do students think beyond the sample?” (Answer: Yes, although no group came up with the idea of examining hypothetical samples which might have occurred.). Groups varied in their abilities to generate ideas, but most ended up with a reasonable test statistic (for example, three of the five larger groups settled on the average squared difference between observed and expected values). We had just completed our regression unit, so many groups leaned heavily on the idea of the sum of squared residuals from that unit. Our predictions for student-generated test statistics (e.g., maximum absolute difference between observed and expected, average z-score for categories) were not realized, in some part by our failure to account for the carryover effect of the previous topic studied. We would expect to find much more variability in proposed test statistics if this lesson was presented earlier in the course. At other times, groups promoted certain over-simplifications, such as the idea that test statistics are only valuable when their distributions are well-known (and preferably normal). Most promising, though, was the observer’s note that, through the lesson, students began to realize that performing hypothesis tests is a process and not just a formula. By being confronted with questions about how to design test statistics and what criteria to use to evaluate them, students began to see beyond simple formulas.

Objective #3—seeing the need for an empirical sampling distribution—was the transition about which we most fretted, and it proved to be a lofty goal. In the Evaluation Column, we asked, “Do students think about sampling distributions? If not, what are their natural inclinations?” According to observer records, the teacher “struggled mightily” to get the students to suggest looking at an empirical sampling distribution to determine if the calculated test statistic provided convincing evidence against the null hypothesis. Even after walking the class through the analogy of one-sample tests of proportions in the context of black and white M&Ms from the Teacher Response Column, the instructor inevitably posed the idea of empirical sampling distributions himself. We should not have been so surprised by this difficult transition. From the teacher notes, the students were looking for “some statistic [which] magically followed a known null distribution.” Upon reflection, we realized that this thinking followed the pattern found in all previous cases during the semester. The idea of a sampling distribution had been defined, discussed, and illustrated through S-Plus simulations at various points during the semester; nevertheless, every important test statistic seemed to, with the introduction of a few magical theorems, follow a known, convenient null distribution. Our planning for and reflection on Objective #3 proved to be one of the most valuable parts of our lesson study-based process; since we believed that students should leave Math 312 with (among other things) a strong notion of sampling distributions, it was apparent that the idea of sampling distributions needs to be stressed and illustrated earlier, more often, and in more effective ways. Indeed, using this lesson earlier in the term would be one effective way to introduce this concept.

The statement of Objective #4 was, in retrospect, poorly written. Although the central statistical content of this unit was indeed the chi-square goodness-of-fit test, our objective, as worded, merely stated that this topic was to be “introduced.” Instead, we sought to provide content and motivation from which the chi-square goodness-of-fit test and its null distribution would naturally proceed. We hoped the students themselves would see the need for and the utility of these ideas. In fact, students first encountered the chi-square test statistic when completing their take-home assignment following Day One. Ideally, in this way, students would be more likely to recall the rationale behind goodness-of-fit tests in general and the chi-square test in particular, which feeds into one of our broad overall goals of making important themes (like sampling distributions) more memorable to students. Teacher notes indicated that students favorably contemplated the idea of empirical sampling distributions and p-values — that we could obtain p-values for their customized test statistic nearly as easily as any classic test statistic.
Finally, Objective #5 on extending the goodness-of-fit tests to various scenarios showed satisfying progress. The first extension, from the categorical to the discrete case, was trickier than envisioned. As the observer noted, “Something about the switch from colors to number of fumbles, defining the categories, caught the groups up. Prompts were needed from the professor.” The next extension, from the discrete to the continuous case, was made by the students themselves for the take-home assignment following Day Two, and a wrap-up discussion at the beginning of Day Three made it apparent students were okay with this extension. The final extension, then, to the chi-square test of independence, seemed almost automatic to the students on Day Three. In considering these extensions, students appeared to be putting the main ideas behind the chi-square goodness-of-fit test together; for example, one student asked insightfully on Day Two, “Now let me get this straight…we’re using a chi-square distribution to test whether or not data follows a Poisson distribution.”

Assessment of our lesson study was also done through the students’ viewpoint. An end-of-semester, online, anonymous evaluation was completed by 13 of the 23 members of the class (a low response, due in part to ill-timed campus-wide computer system breakdowns). One question specifically asked students “Did you like the format we used in class with the Chapter 10 (chi-square) material—i.e., developing ideas in small groups and testing them between classes with S-Plus simulations? What did you like or dislike about these classes compared to others?” Seven of the 13 respondents reported liking the approach in the study lesson, 2 did not like it, 3 were neutral, and 1 did not respond to this question. Those who liked it reported that our study lesson “captured my interest”, “allows for new types of mental connections to be made and to see things in ways different and perhaps richer than before”, and “help[ed] me remember the basic ideas behind chi-squared[sic]”. Others made suggestions and comments such as “lecture main points/ideas at end” (instead of at the beginning of the next class period), “pace seemed a little slow”, “not always clear on what objective was”, “nice balance; difficult to use throughout a course”, and “there was a guy looking over our shoulder and taking notes on us.” (Note: we did discuss the lesson study process and its purpose with the students prior to commencement of the lesson in class.) Since the first 12 weeks of class had featured lectures with examples and some small group activities, but nothing as active and open-ended as our study lesson, it is not surprising that some students were longing for their familiar routine with only one week to go.

In addition, on the final examination, one of the five questions (Question 3 of the Final Exam) was devoted to goodness-of-fit tests, including an S-Plus simulation. Given that no book problems were assigned on this topic, students performed very well, producing a raw median score of 21 out of 25—the second highest of the five questions (see Final Exam Rubric for broad scoring rubric). Unfortunately, these results could not be compared with historical results, since this final examination differed greatly in format from past finals in Mathematical Statistics given by the instructor.

5. Benefits of Implementing Lesson Study Principles

Is lesson study a worthwhile endeavor at the undergraduate level, and, in particular, for an upper-level course such as Mathematical Statistics? Our pilot experience with lesson study principles suggests that the answer is Yes, with proper preparation, faculty commitment, and realistic expectations. Specific benefits we realized included:

1. **Focused and energized collaboration.** The lesson study process brings together interested parties for focused discussions which range from broad pedagogical goals to specific wording of questions for the class. The group must think very intentionally about the objectives of the lesson and how success might be assessed. The lesson study format forces group members to think carefully about the purpose of each idea presented and the likely student responses to each question posed. The range of opinions expressed and solutions offered is truly energizing and
enlightening. Having a structured environment makes it easier to maintain the collaborations initiated, and combining multiple faculty perspectives helps to anticipate student responses in a way that would never transpire with a faculty member working in isolation. In addition, the lesson study process opens up the instructor’s classroom in a non-threatening manner, allowing both the visitors and classroom teacher to benefit from fresh perspectives. We all agreed that the collaboration on a very specific lesson plan produced insights and raised awareness of teaching issues that will influence every class we teach in the future. We all felt the benefits of the lesson study process were worth the time invested, and we carried lessons learned from this experience back to our home institutions.

2. Insight into student learning. All stages of the lesson study process—from lesson planning to classroom observation to reflective analysis—focus intensely and powerfully on the student learning process: How will students attack this open-ended question? Will they connect these ideas? Are they motivated by this example? For instance, in our lesson, we learned that students in Math 312B were comfortable thinking about sampling variability and significant evidence against the null hypothesis, but they were much less comfortable than we imagined with the general notion of sampling distributions. We also learned that, with respect to goodness-of-fit tests, the extension from the multinomial case to discrete probability distributions was not automatic, but subsequent extensions to continuous probability distributions and two-way tables were easier. For further readings on teaching with a student-learning focus, see many of the lesson study references listed at the end of this article, especially Lewis (1995).

3. Development of a strong lesson plan. The lesson plan we developed may or may not make a good “off-the-shelf” plan, since it was carefully tailored to these 23 students, but it contains essential elements that one could model and adapt to future lesson plans. For some of us, this lesson opened our eyes to the power of active, investigative learning even in an upper-level, highly mathematical statistics class. We found that having students form a custom test statistic and ponder the criteria of a good test statistic before seeing the “standard” were effective learning tools. The flow of topics in our lesson was rearranged several times, but eventually led the students successfully and logically to our final specific objectives (the chi-square test of independence) and broad goals of student engagement and statistical thinking. One surprising outcome of our lesson plan is that the instructor spent slightly less time than usual covering goodness-of-fit material from Larsen and Marx (2001), despite using class time to guide students toward the discovery of central ideas. We now believe that by carefully orchestrating student work at home between classes and reducing the number of in-class, teacher-led examples, we were able to achieve a reasonable balance between problem solving and knowledge construction on one hand and coverage of material on the other hand. Finally, the lesson study process also made readily apparent the weaknesses of our lesson plan. For example, lack of student response told us that our prototype samples did not have their desired effect—they did not help students evaluate test statistic performance and highlight differences between students’ test statistics and the chi-square statistic. Of course, realizing that cases which differentiate the chi-square statistic from others are difficult to develop is an interesting revelation itself.

4. Facilitation of pedagogical research. Although for this initial pilot we did not pose experimental research questions and design data collection efforts to use our classroom as a laboratory in the most efficient way possible, the lesson study process did inspire us to take risks (such as introducing new material—even the chi-square test statistic—outside of class), and to think about gathering information to help evaluate our success in meeting objectives. The post-lesson reflection also produced observations of unanticipated developments (such as the struggle to illustrate differences between the chi-square statistic and others) that could be the subject of future, systematic study.
6. Recommendations for Implementation

Of course, along with benefits realized from our efforts to implement principles of Japanese Lesson Study, there were just as many things we wished we had added to the process, done differently, or known before we started. Thus, as a guide to newcomers to Japanese Lesson Study, we attempt to synthesize our experiences and offer the following recommendations to aid implementation:

1. **Read the literature.** Early attempts to introduce lesson study to U.S. teachers included mentoring by Japanese teachers (Fernandez 2002; Fernandez, Cannon, and Chokshi 2003) or the presence of “knowledgeable others” (Watanabe 2003), in order to advise on the philosophy and intricacies of lesson study. We had no such luxury. Although excellent references exist (see the list of references at the end of this manuscript and web sites cited earlier), the list is still small, and most examples are geared toward K-12 teaching. One exception is a recent paper by Garfield, delMas, and Chance (2005). We spent a little time with introductory articles and videotapes, but more time and a more thorough initial understanding would have been extremely beneficial.

2. **Find a committed group to participate.** Finding a critical mass of faculty members with similar teaching interests who are willing to commit to the lesson study project is not a trivial task, especially among isolated statisticians. We were fortunate that a couple of visitors to the Northfield area were interested in joining forces with statistics faculty members already in the area to form our lesson study group. In fact, our original group of seven dwindled to four as time and travel demands caught up to some group members. One might consider options for ensuring ownership and maintaining group membership—for instance, having several different people plan to teach the lesson in their own courses and assigning others specific roles from the start like observing classes or developing lesson materials. Alternatively, one might consider the feasibility of holding virtual planning meetings and broadcasts of the lesson being taught. In an ideal world, participants may even receive a small reduction in their teaching loads.

3. **Be realistic about the time commitment.** For full-time group members, the time commitment is real. We met five times for 60-90 minutes each during the planning stage and two times during the reflection stage. This does not even include teaching and observing two class sessions, outside time spent by individuals working out details of the lesson plan, and time groups may spend revising and reteaching their lesson. Thus, we could have spent even more time preparing, although at some point a deadline must be set. We might have even considered meeting in a more condensed time frame to better harness momentum.

4. **Sketch a timeline for your entire process to unfold.** Schedule regular meetings, and be sure to allow adequate time for the essential “ongoing revision” step—revising and reteaching the lesson. At the undergraduate level, this might have to wait for the next semester, but two teachings in the same semester may be possible with careful early planning.

5. **Develop a broad goal and specific objectives.** In some ways, we went backwards, picking our specific lesson topic, and then identifying the overarching course goals that lesson satisfies. But that’s not so important, and the specific lesson topic is also of secondary importance. Remember that the process is much more important than the individual study lesson plan developed, and do not spend too much time with this step (as we did).

6. **Run efficient meetings.** Conversations at lesson study planning meetings inevitably drift from general to specific issues, and this is just fine. Conversations such as these are valued and encouraged by the lesson study process. Since there are many small details to work through in...
preparing a lesson, we found it valuable to assign tasks at the end of each meeting – e.g., write up a lesson outline based on a certain example and circulate the outline to the group. In that way, conversation was primed for the next meeting, and tiny details such as wording of handouts could be first attempted by one group member and modified by the rest.

7. **Record thorough notes at each meeting.** We asked a student to sit in on our regular meetings to observe (our note taker was keenly interested in both statistics and education), take notes, and occasionally offer a student perspective. Some have recommended videotaping these meetings, but we found written notes to be both helpful and sufficient.

8. **Avoid being overly judgmental in planning meetings.** Our planning meetings were successful not because we were all on the same page at every step, but because we valued opinions from every perspective. No idea was rejected until it had percolated and developed. The success of lesson study requires that planning meetings support a safe atmosphere for collaboration, and we found this collaborative atmosphere helped cultivate and identify unanticipated thoughts and ideas.

9. **Create a detailed four-column lesson plan.** The design of the four columns forces the group to carefully consider the progression of key ideas, likely student responses, thoughtful teacher responses to student concerns, overarching goals, and means of evaluating the success of a lesson. At the same time that it guides the group’s planning, the four-column lesson plan also provides the general framework and detailed script to allow an instructor to deliver the lesson as designed, and to allow observers to comment on the success of the lesson.

10. **Maintain a student focus.** Adopting a student lens, as described by Chokshi and Fernandez (2004), is an essential component of lesson study – understanding student thinking, anticipating behavior, and learning to build understanding. The emphasis on student thinking throughout the lesson study process taught some of us that, although we had designed lessons in the past with student understanding in mind, we had abundant room for improvement.

11. **Embrace the research lens.** In future implementations of lesson study, we would plan for and significantly increase the use of classroom research within the lesson study process. We would consider using more rigorous research principles to test the impact of pedagogical ideas, and we would systematically plan for good data collection to help us more objectively evaluate our goals. Lesson study provides a rich framework for conducting classroom research, obtaining results which will often generalize beyond the specific lesson being taught.

12. **Maintain flexibility with respect to time.** Once we got on a roll with our lesson planning, the ideas started flowing quickly—examples to motivate big ideas, detours to reinforce major points, challenging problems that could be approached in myriad ways, etc. The excitement of our idea-generation overwhelmed us, and we tried to merge everything to form the “perfect lesson.” Naturally, we ran short on class time, and the instructor ended up ad-libbing and cutting material on the spot. We learned that a future lesson study should leave more time for end-of-the-hour summaries and instructions for the take-home assignments, more carefully orchestrate beginning-of-the-hour wrap-up discussions, and generally account for longer times required by active lessons developed under the lesson study approach.

13. **Schedule classroom observers and videographers.** Our lesson study experience would have been greatly enhanced by a larger, diverse set of classroom observers, and by better quality in the videotaping of the classroom sessions. With just a single observer, we were unable to learn from multiple perspectives on the strengths and weaknesses of the classroom session; without observers from outside the planning group, we were unable to gain fresh opinions from someone without
preconceived expectations; and, without good videotapes, we were unable to eavesdrop on small group conversations and observe telling student reactions.

14. **Revise and repeat the lesson.** At least once, suggestions for improvement should be incorporated and the lesson repeated, ideally with many of the same observers, even if the second study lesson occurs in the following semester. Without revising and repeating the lesson, we could only conjecture about possible improvements in our wrap-up meetings, without the ability to document the actual impact of our changes.

15. **Share your experiences.** A specific lesson developed with the Japanese Lesson Study framework will be thoughtful and thorough, and it is worth sharing with the statistical education community, potentially in the form of a web collection of four-column lesson plans. Even more importantly, pedagogical insights acquired both through the lesson study process and through classroom research should also be shared in a similar format.

7. **Conclusions and Future Research**

Our pilot implementation of lesson study principles in an upper-level undergraduate statistics course was not a true, complete lesson study. We had fewer observers than ideal when the lesson was taught, and we did not engage in the revising and reteaching stage so essential to the process. However, we very intentionally followed the lesson study process as much as possible while planning our lesson, providing a uniquely structured environment for setting goals and sharing ideas. Through this process, we believe we developed a lesson collaboratively that was higher in quality and ambition than anything we would have tried ourselves. More importantly, we gained insight and awareness into effective statistical pedagogy and how students develop statistical thinking. Focusing intensely on a single lesson provided an achievable and generalizable means for examining the course as a whole. In this paper, we attempted to share our successes, our failures, and our insights so that others might consider applying Japanese Lesson Study principles to their own professional development and build upon what we’ve done.

Even as we offer our insights, we recognize that many open questions exist about the application of Japanese Lesson Study at the undergraduate level. Some questions for future attention include:

- Will this only work for instructors who feel comfortable using open-ended, active tasks during class?
- Will this also work for large and/or lower level classes?
- Can the same lesson plan work for instructors with different teaching styles?
- Can collaborative lessons be planned virtually?
- Can repeated teachings be done during successive semesters?
- Can lesson study principles be used to effectively train new teachers (see also Garfield, et al. (2005))?
- Can incentives for participating be institutionalized so that efforts in applying lesson study can be sustained over the long term?

True implementation of lesson study is not easy – for this process to be effective, it is essential that instructors feel comfortable devoting sufficient time to the process, sharing their ideas, spending class time on open student investigations, and observing and reflecting on each others’ teaching. Yet, our experience indicates that Japanese Lesson Study principles can be implemented successfully in upper-level undergraduate statistics courses. Despite our inexperience and imperfect implementation, all involved found the application of lesson study principles to be valuable and worthwhile, an experience which has had a lasting impact on our teaching beyond the single lesson on which we collaborated.
Acknowledgements

The authors gratefully acknowledge the Associate Editor and two anonymous referees for their helpful and insightful comments. The authors are grateful for the contributions of Laura Chihara, Carolyn Pillers Dobler, and Martha Wallace during the early stages of the Lesson Study group. Finally, the authors thank Kim Newman for her dedicated note-taking and helpful input, and the Spring 2004 Math 312B class for their hard work and enthusiasm.

References

Chokshi, S. (2004), “Timeline of U.S. Lesson Study,” compiled for NAS/NRC’s Board on International Comparative Studies in Education commissioned report “Impact of Lesson Study”. (www.tc.columbia.edu/lessonstudy/timeline.html)

Chokshi, S., and Fernandez, C. (2004), “Challenges to Importing Japanese Lesson Study: Concerns, Misconceptions, and Nuances,” Phi Delta Kappan, 85(7), 520-525.

Curcio, F. R. (2002), A User’s Guide to Japanese Lesson Study: Ideas for Improving Mathematics Teaching, Reston, VA: National Council of Teachers of Mathematics.

Fernandez, C. (2002), “Learning from Japanese Approaches to Professional Development: The Case of Lesson Study,” Journal of Teacher Education, 53(5), 393-405.

Fernandez, C., Cannon, J., and Chokshi, S. (2003), “A US-Japan lesson study collaboration reveals critical lenses for examining practice,” Teaching and Teacher Education, 19, 171-185.

Fernandez, C., and Chokshi, S. (2002), “A Practical Guide to Translating Lesson Study for a U.S. Setting,” Phi Delta Kappan, 84(2), 128-136.

Garfield, J., delMas, R., and Chance, B. (2005), “The Impact of Japanese Lesson Study on Teachers of Statistics,” paper presented at the Joint Statistical Meetings, Minneapolis, MN.

Larsen, R. J., and Marx, M. L. (2001), An Introduction to Mathematical Statistics and Its Applications, 3rd Ed., Upper Saddle River, NJ: Prentice-Hall.

Lewis, C. (1995), Educating Hearts and Minds: Reflections on Japanese Preschool and Elementary Education, New York, NY: Cambridge University Press.

--- (2002), “Does Lesson Study Have a Future in the United States?” Nagoya Journal of Education and Human Development, 1, 1-23.

Lewis, C., Perry, R., and Hurd, J. (2004), “A Deeper Look at Lesson Study,” Educational Leadership, February 2004, p.18-22.

Lewis, C. and Tsuchida, I. (1998), “A Lesson is Like a Swiftly Flowing River: Research lessons and the
improvement of Japanese education,” *American Educator*, Winter, 14-17 and 50-52.

Scheaffer, R. L., Gnanadesikan, M., Watkins, A., and Witmer, J. A. (1996), *Activity-Based Statistics*, New York, NY: Springer-Verlag.

Stigler, J. W., and Hiebert, J. (1999), *The Teaching Gap: Best Ideas from the World’s Teachers for Improving Education in the Classroom*, New York: The Free Press.

Wang-Iverson, P. (2002), “Why Lesson Study?” in *Papers and Presentations: An Introduction from RBS Lesson Study Conference 2002.* (www.rbs.org/lesson_study/conference/2002/papers/wang.shtml)

Watanabe, T. (2002), “Learning from Japanese Lesson Study,” *Educational Leadership*, 59, 36-39.

--- (2003), “Lesson Study: A New Model of Collaboration,” *Academic Exchange Quarterly*, Winter, 180-184.

Paul Roback
Department of Mathematics, Statistics, and Computer Science
St. Olaf College
Northfield, MN 55057
U.S.A.
roback@stolaf.edu

Beth Chance
Department of Statistics
California Polytechnic State University
San Luis Obispo, CA 93407
U.S.A.
bchance@calpoly.edu

Julie Legler
Department of Mathematics, Statistics, and Computer Science
St. Olaf College
Northfield, MN 55057
U.S.A.
legler@stolaf.edu

Tom Moore
Department of Mathematics and Computer Science
Grinnell College
Grinnell, IA 50112-1690
U.S.A.
mooret@grinnell.edu