Faculty course evaluations:

What is really going on in your class?

Presentation by:
Larry H. Ludlow, Ph.D.

Boston College
Lynch School of Education
Department of Educational Research, Measurement and Evaluation
Chestnut Hill, MA 02467
Ludlow@bc.edu

Prepared for:

St. Mary's College (Marino)
Dublin, Republic of Ireland

May 2005

I. The Imagine Scene

Imagine you have just presented a lecture on a tough topic for students—for me this would be on maximum likelihood estimation and the operation of the Newton-Raphson iteration technique to my doctoral students. You have presented this topic many times in your career. This time, however, you feel you have **actually "taught"** the students how the technique works and **why it is important**. You even believe you have made the topic interesting, if not exactly exciting. Bottom-line, you have answered the ultimate questions, "**So what**?" and "**Who cares**?"

As you look at the students, awaiting their applause, a thought occurs, "**What are *they* thinking about what you just did**?" You think abut this question as they leave. There's no applause, wave, or cheers, not even a "Nice job, Doc!" **You wonder** how they describe this class to their friends. What **visual images** do they construct for their audiences?

Over the next couple of days you ask a few students what they thought about the lecture?

- Did they **understand it** reasonably well?
- Was it **clear**?
- Did it make **sense**?
- Where were the **tough parts**?
- Where did they begin to **lose it**?

Their responses are **non-descript--**it was fine, it made sense at the time, it was **challenging**. Their responses, while somewhat supportive of your efforts, **don't leave you satisfied**. So you decide to try something unusual in the next class. At the start of the next class session you ask them to complete the following task:

> **Think of a "typical" classroom teaching experience this semester with me. Now draw as best as you can, that classroom experience.**

You hear **giggles**. Students look at one another. **Puzzled** expressions are exchanged. Whispers and **groans** are heard. Some of them look at you as if you have gone really weird on them this time. Eventually they begin to draw.

Imagine your reaction to getting this drawing: "Yoda" (NOW PASS OUT HANDOUT): Figure 1

This figure reflects the kind of **positive** classroom experience I wish I had on students every night I stand before them. The statements make sense. I come across as a person, even a personality-- "Yoda"—the **MASTER** teaching There is a **discernible** message on the chalk board. Most dramatic of all, however, is the student's role. "Luke" –the **APPRENTICE** is experiencing the "ah-ha" rush of **insight** that I have never seen represented as effectively as it is here. **THE BEAM OF KNOWLEDGE—YOU FEEL GREAT, WORTHWHILE, SATISFIED, ACCOMPLISHED!**

Now, imagine your reaction to getting this drawing: "Scream"

Figure 2

What are you thinking and feeling now? This figure is **powerful and disturbing, even embarrassing**. No words need to be written on this figure. **No words** written on a standard course evaluation form could adequately express the emotions of this student. This image is **not** one that we would like students to take away from our classes. **YOU DON'T FEEL SO GOOD AND SATISFIED NOW!**

Now that I've caught your attention, I'd like to explain how I got to this point in using drawings from students in classes as a form of course evaluation information.

My general research interests have been in the **development and creative application** of statistical models. In particular, I've always been interested in **tracking change**, largely through using **graphical displays** of complex data. I find the use of plots and graphs to usually be more interesting and revealing than tables of numbers and statistics

My particular interest in teaching evaluations came about because of my dissatisfaction with a **dean's comments about a low rating** I received in a specialty course 10 years ago. It angered me as somewhat unfair because of the circumstances that semester—cancelled classes because of snow, master's students were allowed to enroll in my highest level stat course. So I wanted to know, in general for all my courses, how my ratings looked over time—did they look like they were improving, dropping, or staying steady—**regardless of what anyone else's looked like**.

Now at BC, and most other US colleges and universities I am familiar with, student ratings are used for salary, promotion, and tenure decisions. One of the **higher-level aims**, of course, is to understand one's teaching capabilities and a systematic statistical analysis of the ratings can be useful when used to look at long-term career patterns.

Two examples of typical course evaluation questions are provided in **Figure 3 and 4**.

SHOW FORMS A AND C

Form A is generally described as a generic personality assessment—I don't care for this particular form. Form C focuses on specific classroom teaching practice—I like this form, it asks about things I'm interested in knowing about my classes. Faculty have the option of choosing among a variety of different such forms. And the students use the standardized scan sheet presented as Figure 5—note that there is opportunity for students to submit written comments.

We then receive the results of the evaluations presented in **Figure 6**. These are **course evaluation summary sheets** with the percent of students agreeing or disagreeing with a variety of classroom related questions. These are sent to us on paper with no personal electronic records. There is no University maintained and useable data file.

SHOW THE COMPUTER PRINTOUT PAGE

Unfortunately, many **institutions provide little, if any, guidance** to faculty in how to make sense of this information in any systematic way. And, in general, little systematic or in-depth analysis is performed by instructors upon the reports they receive. Many **faculty, perhaps most, look first at their overall rating**—what percent of their students marked them “excellent” or “poor”? After that they tend to ignore the rest of the report and proceed directly to the student narratives, if any are provided.

Part of the **lack of respect** for ratings has been attributed to a belief that students rate highly only those faculty who are **easy graders and are personable**. Another reason is that a typical institutional report **compares an individual’s ratings to an aggregated result**—such as the combined undergraduate or graduate school results. This type of comparison, however, is often rejected by faculty as too confusing, confounding, irrelevant, and inappropriate.

Finally, another common problem is that annual reviews typically consider **just the two semesters** in a given academic year—which **ignores the trajectory and pattern of evaluations** for that instructor over the course of contiguous years. This means that significant contextual information about an individual is missing when teaching reviews ignore past performance and circumstances.

Frustration with these various problems provoked a line of research that focuses on the **individual instructor’s teaching over time** and it is in direct contrast to administrative snap-shot comparisons that offer little evidence and understanding of factors that contribute to systematic growth, maintenance, or deterioration in teaching. A key characteristic of this approach is that the analysis of one's ratings is intended to inform and guide individuals—**regardless of what the administration's aggregated summaries may suggest**.

Ideally, one's institution provides the ratings in an electronic format that is easily imported into a spreadsheet (EXCEL) or a statistical package (SPSS). In the event that evaluation results are only returned in hardcopy, it is relatively simple to build a data file through hand-entering the results. **See Figure 7**

SHOW DATA FILE SHEET

In such a file each row of data corresponds to a separate class. Each column corresponds to a different aspect of the SRI summary results.

Once the file is created it is **simple to add new** information over time. This includes adding records for successive classes but it also includes adding new variables. For example, an indicator variable was added when the instructor was appointed department chair. This opportunity to add variables retrospectively makes it possible to test hypotheses about a wide variety of potentially influential variables, e.g. tenure, rank, or marital status at the time the class was taught.

The data file consists of **101** separate records summarizing the course evaluations received by one instructor for all classes taught from fall 1984 through **summer 2004**. The data were extracted from the end-of-semester rating summaries I just showed you. The evaluation questions remained the same across the 20 years covered by this dataset.

The **file is updated each fall** with the ratings from the previous fall, spring, and summer classes. Although the specific courses are largely irrelevant for the purposes of this paper, they range from entry-level freshman "**Child Development" to a capstone third-year doctoral "Seminar in Statistical Methods**." Most of the evaluations are for graduate courses in applied statistics. The data file consists of four relatively distinct categories of variables for each class taught:

- administrative characteristics (e.g. year taught, class size, course code, level of students),

- student-level perceptions (e.g. percent of time spent on the course, extent to which they acquired factual information),
- instructor-specific variables (e.g. tenure status and marital status at the time the class was taught), and
- overall evaluation ratings (percent who marked excellent, very good, good, acceptable, or poor).

There are a total of 27 variables associated with each class. Overall, the dataset summarizes the evaluations submitted by 2174 students.

Analysis

Now **what can we do with all** this information? One of the first things many faculty are interested in is: "What do my ratings look like over time?" **Figure 8** shows the percent of students in each class who rated this instructor "excellent." The center dashed line is the regression line—it shows the predicted excellence rating for each class taught in any given year. Although the ratings show a positive upward trend over time across all classes.

The other two dashed lines represent the 95% confidence interval around the regression line—the region within which most of the class ratings should lie (given this particular simple statistical model). There are four points corresponding to four classes that fall above the upper confidence interval. Those classes received excellence ratings that were much higher than expected. The class identifier reveals that three of these high ratings occurred the first time that particular course was taught by this instructor (indicated by the ".01" designation).

This is an **interesting findin**g because faculty often believe that the first time they teach a course they work harder to get it "right" than for subsequent offerings. Hence, they believe ratings tend to be **lower** for first-time classes.

This is a **useful graph for one's tenure and promotion portfolio**. It is also useful when negotiating work area priorities and load considerations for the coming year and for documenting unexpectedly high ratings in the past year.

What do the ratings look like for the **different types** of courses? **Figure 9** plots the percent of excellence ratings for each class against the institutional course code assigned to that class—the course code is a proxy for the **level of complexity** of the material and **sophistication of the student**. The column of low ratings in the left region of the plot corresponds to courses 030 and 031—freshman child development courses taught early in my career. The next vertical column of slightly higher ratings corresponds to 216—undergraduate research methods. The next ratings correspond to 460, 468 and 469, required graduate research methods, introductory and intermediate statistics, respectively.

The vertical column containing the highest ratings corresponds to doctoral specialty courses (667, 668, 669—general linear models, multivariate statistics, psychometrics). **Overall, there is a general upward trend** in ratings as the course code increases—undergraduate courses have lower ratings than graduate courses, required statistics courses have lower ratings than the specialty courses.

This is an **extremely useful type** of analysis because it shows that **ratings differ for the type and level** of course taught. In particular, if an administrator must compare an individual's ratings against some aggregate, let the summary be constructed from similar relevant courses and students.

One of the variables that faculty typically think has a **negative effect on ratings is the size** of the class. We generally think that we do better in small classes and, hence, receive better ratings than when we teach large classes. **Figure 10** tests this hypothesis by plotting the excellence ratings by class size. There is a **clear, unmistakable negative relationship** between the ratings and class size—as enrollment increases, the ratings tend to drop. Statistically, for each additional student added to a class there is a decrease of about 1% in the excellence ratings.

There are three classes with unexpectedly high ratings. The **particularly unusual class** with the highest enrollment and relatively high rating is 468.01—this was the **first time** introductory statistics was taught by me, it **attracted a crowd** of curious students, and its **curriculum and format differed** substantially from the way it had previously been taught (it changed from equation-based lectures to lectures followed by applications and statistical software instruction). This particular graph has been used in numerous annual reviews to **support arguments for reducing class sizes**.

I was appointed department chair in fall 2000. I quickly discovered that my new administrative duties were interfering with my class preparation, holding of office hours, and critiquing of assignments. I thought these problems might be reflected in his ratings.

Figure 11 show the same ratings presented in **Figure 8** but now the two periods of pre-chair and chair status are represented. It is apparent that the ratings in the two different periods **reflect different trends**. Although the overall trend seen in Figure 8 is positive, the ratings during the period as chair are **dropping**.

This type of event history graph is helpful when trying to understand the effects that critical experiences may have had, or are currently having, upon one's teaching effectiveness. Is there a **drop in ratings from pre to post tenure**? Is there a **rise in ratings after a sabbatical or medical leave**? Do ratings **reflect personal changes such as marital status**?

(Figure 12)

By their very nature these types of **self-reflective questions and analyses are unique** for each individual.

Positive **educational effects** that may be attributed to **deliberate pedagogical change** are often hard to detect and document. One way of detecting such an effect is illustrated in **Figure 13**. I felt I was **not adequately meeting the needs of students** in this required graduate introduction to research methods course. The material was taught in a **traditional lecture format** that was **boring** me and the students—as reflected by their low ratings. In 1997 (indicated by the vertical line) I changed the format to part lecture and part small-group interaction. The books, handouts, examples, and assignments essentially stayed the same. The primary change was a period of time during each class session that **required students to interact** with one another on practical exercises. I wander from group to group and serve as a facilitator aiding and guiding their discussions.

The effect of this **relatively simple change** is shown by the direction of the ratings before and after the change—the **slope of the ratings quickly changed** from negative to positive. In fact, this course is now one of my favorites. This particular analysis recently convinced me to add small-group interactions and in-class exercises to my specialty classes. This **type of analysis may be a useful tool** as faculty debate the various pros and cons associated with moving from **traditional chalk-and-talk** formats to the various evolving **point-and-click** technology based formats.

Statistics faculty, like all faculty, must make choices about how a given topic is presented in class. For example,

- are equations presented as mathematical expressions to be memorized,
- is the emphasis placed on how to run and interpret statistical software,
- is the emphasis placed on linking the techniques in such a way that they are understood as logical ways to ask and answer increasingly sophisticated questions about one's data?

Figure 14 shows the percent excellent ratings plotted against the percent of students who strongly agreed that they understood principles and concepts. Note the extreme contrast between the 600-level courses (specialized statistics) with high principles and concepts ratings and the 30-level courses (child development) with low principles and concepts ratings.

Faculty are often **curious about the extent to which the workload** required of students has an impact on the ratings. While some faculty may believe that a **heavy workload is desirable** regardless of what the student thinks, others wonder if there is a **negative relationship between the workload and ratings**. **Figure 15** shows the ratings plotted against the percent of students who stated that they spent "much more" time on the present class than other classes that semester.

There are **four distinct clusters** of courses in this graph.

- There are child development courses (30, 31) with trivial amounts of time commitments and the lowest excellence ratings. These were taught when I first started teaching—and these were courses outside my training.
- There are the research methods classes (460) with slightly higher time requirements and some of the highest excellence ratings. These consist of frequent small-group interactions (as seen in Figure 5), are taught in the summer, and do not require extensive take-home assignments.
- There is a cluster of required statistics courses (468, 469) with a heavy time commitment and very low excellence ratings. These also define the mid-range cluster of points in Figure 6—students in these courses did not tend to strongly agree that they understood principles and concepts.
- Finally, there are the higher level specialty statistics courses (664, 667, 668, 669) with heavy time commitments and high excellence ratings. These courses tend to be the ones with small class sizes and the highest ratings on understanding principles and concepts.

Apparently, heavy workloads and time commitments are valued by students if they understand why they are doing the work.

In summary, one of the key features of these graphs is that they show direction and magnitude of relationships, patterns over time, clusters of similar classes, and individual instances of surprising ratings.

Now, as **soon as I receive the latest summary** of evaluation results I enter them into the data file and start looking at how those results add to the overall picture. The first question I ask is: “Have the new ratings **strengthened or weakened** previous patterns?”

Lessons learned from these analyses?

These are a few ways that I've made use of these graphs and results:

- **stress principles and concepts (instead of tedious calculations),**
- **form small-groups to facilitate interactions that reveal areas of confusion (instead of constant lecturing),**
- **incorporate real-world examples in all statistical applications (instead of artificial textbook examples),**
- **encourage email communications and hallway interactions outside the classroom (instead of sending everyone to the assistant),**
- **try to balance coursework between detailed thoroughness and unnecessary burden (instead of expecting every lecture point to be reflected in the assignments), and**
- **acknowledge to students when personal variables outside the classroom may affect day-to-day teaching effectiveness (instead of perpetuating the ivory-tower myth).**
- **And, *<u>try</u>* to keep a harmonious marital relationship.**

How has this longitudinal approach been of value to other faculty and administrators?

- **It has helped faculty build their own data bases for their own personal interests.**
- **It has helped others prepare their teaching portfolio for promotion and tenure, and annual review considerations.**
- **And this work has led to LSOE and University discussions of systematic use and review of evaluations—we are now going online this coming fall with the collection, storage, and analysis of University course evaluations!**

But, over time, I became **dissatisfied** with just the analysis of the quantitative summaries. They do **not provide a rich enough** source of information about what students experience in classes. These graphs are very **effective** for showing long term trends in ratings over one's career—also for comparing different types of courses and levels of students, and various other creative professional and personal characteristics. Ultimately, however, those graphs are all **dependent upon one very simple piece of data**—a filled-in circle on a standardized form. The question, then, is

> "**what was driving** the student to make the mark they did—what were they **re-experiencing** about the course, what **images** came to mind to guide their mark, what **lies behind the 1000's of numbers** that produced those graphs?

My broader goal was to somehow **get more personal information** about what was going on in class. I believe this type of information can yield insights in how to **construct learning opportunities** to enrich their statistics experience. Statistics is a **tough topic** for many students and anything that can address not only the professional presentation of it but their personal reaction to it, seems like a good thing to do. This desire led to an adaptation of the elementary and middle school classroom drawing project that Walt Haney was conducting in the mid-90's.

Which now brings us back to where we started. The instructions for the drawings now state:

(1) What visual image of a classroom experience comes to mind when you think of this course? Now draw as best as you can, that classroom experience. Include me, yourself, and anything else that represents for you that classroom experience. Ideally, someone else could look at your drawing and could then form a reasonable impression of your experience.

(2) On the back of your drawing write a full description of the scene you have drawn. Be as explicit, open, and comprehensive as you can.

(3) Finally, what "course evaluation" information does your drawing provide that your responses to the traditional scannable form do not contain?

Please try to accept my assurance to you that this information is confidential--I will not try to somehow figure out who passed in which one of these sheets. This information is part of a long-term research project that I am conducting on alternative modes of faculty evaluation assessment techniques.

I now have over 800 student drawings. These cover nine years of teaching undergraduate and graduate classes in measurement, evaluation, and statistical analysis. Some courses follow a traditional lecture format, others include a cooperative learning component. Most of the students are from the school of education. The lower level courses are **required**. Students in the higher level courses specialize in measurement, evaluation, and statistical analysis.

The “drawing evaluation” immediately **follows their completion of the standard** evaluation form. They are told that the drawing evaluation is part of a long-term research project and I encourage them to take the exercise seriously. They are also asked to **write an arbitrary four-digit code** of their choice on their scannable evaluation forms and their drawing.

VI. What do the drawings look like?

To my amazement and delight, the drawings are **rich beyond anything I expected**! As we look at the drawings we can ask from a **broad perspective**:

- what is **important** in these drawings?
- what are students **trying to convey** about a particular course and instructor?
- what is **unique and different** about the courses?
- which **patterns** are similar across courses? and
- how can these drawings be **systematically analyzed**?

The drawings are presented to you in sets of **common ideas and themes as I see them**. My presentation of these drawings is **entirely qualitative** in that I simply went from one drawing to the next , reacting to each as I looked at them, and started placing them in common piles.

(Show drawings)

VII. Analysis of the drawings

Now, we’ve had some fun but is there anything serious here? The **broader research problem**, of course, is how to analyze, interpret and explain not only these but other faculty initiated drawings in a way that is **not self-serving, idiosyncratic, or arbitrary**.

To facilitate interpreting and explaining these drawings, an objective coding rubric was constructed that allows quantitative comparisons and offers generalizability to other faculty who might choose to adopt this course evaluation tool. This coding method indicates whether individual drawings exhibit particular features. For instance, is the instructor depicted alone or with students; is he or she verbally addressing the class or writing on the blackboard; were computers, books, or projectors shown in use?

The final coding protocol for the drawings was:

SELECTED OPERATIONAL DEFINITIONS FOR “X’s” CLASSROOM DRAWINGS

Instructor Presence/Affect:

- Depicted Positively: “X”’s facial expression is positive (smiling), or depiction of positive speech (praise, support).
- Depicted Negatively: “X”’s facial expression is negative (frowning), or depiction of negative speech (confusion, malice).

Instructional location of instructor

- At board: Instructor is drawn located at or near the board (not necessary actively using it).
- At overhead: Instructor is located at/or near overhead projector (not necessary actively using it).

Instructor Interaction

- Asking Question: “X” is posing a question. (student’s my be present or not present).
- Speaking Statement: “X” is lecturing/instructing

Student(s) Present:

- Sitting in groups: More than 1 student is located in a cluster of desks or with other students.
- Sitting in rows: Students are arranged in rows or columns (either in desks or not).

Student Depiction

- Depicted Positively: At least 1 student’s facial expression is positive (smiling), or positive speech.
- Depicted Negatively: At least 1 student’s facial expression is negative (frowning), or negative speech.

Course Experience(s)

- AHA/light bulb/lightning bolt: At least 1 student depicts a light bulb, or AHA!, or lightning bolt
- Understanding over time: Some depiction of before/after learning, gradual growth of learning or understanding

About classroom

- Computer depicted: A computer is present somewhere in the drawing
- Laser pointer: A laser pointer (or beam) is present in the drawing

Other

- Unreadable text: There is unreadable instructor related text somewhere in the drawing (scribbles, etc. NOT SPEECH—typically board work).
- Metaphorical: The drawing uses a metaphorical representation of the classroom or experience, rather than a pictorial depiction of the actual classroom environment.

These codes marked the presence or absence of 46 variables. These codes were entered into SPSS and were aggregated to obtain the mean drawing code proportions for each individual class. This file was then merged with the course evaluations data file (consisting of 99 classes).

Analyses:

The quantitative and qualitative data sets allow for multiple forms of analysis of how drawings and course evaluations depict systematic changes across courses and time. Taken together, they provide a remarkable opportunity to analysis multiple forms of evidence regarding an individual instructor's teaching "gestalt". For example,

- do highly rated classes tend to have greater proportions of "aha" incidents than lower rated classes?
- Do introductory statistics classes tend to have greater proportions of "confused" expressions?
- Are frequent depictions of small-group interactions associated with higher ratings than drawings of traditional rows of seats?
- Are depictions of statistical symbols associated with higher or lower ratings?
- Across all classes is there a higher proportion of drawings depicting the instructor with a positive or negative affect?

These data provide a unique opportunity to support the validity of both quantitative student ratings of instruction and qualitative depictions of context.

Table 1 contains the proportion of times a particular feature was present across all the drawings in a particular type of course. For example, there are 138 drawings across six sections of Research Methods that have been taught. Across those 138 drawings, 57% of them depicted the instructor positively. The instructional opportunity from such a table is seen when we concentrate on the Introductory Statistics course. It is not surprising that across the 66 students in four sections of this course that the proportion of times the instructor was depicted positively was only 26% compared to the non-statistics based methods course and the specialty course of General Linear Models (40%). In fact, the standard evaluations using the standard university rating system always have this course rated below the other courses taught by this instructor. But the story becomes more interesting when the other bolded values in the table are interpreted.

Note the depictions of aha" experiences (12%) compared to the other classes. Interestingly, these students also tended to create more metaphorical images of their experiences than other students (21%), e.g. depictions of sharks on the attack, successfully lifting heavy boulders, deer facing oncoming headlights.

Figure 16 is a useful graph for depicting the proportion of times a characteristic was present across all the drawings for a specific course. It shows the proportion of students across all classes who depicted some level of confusion in their drawing (along with a standard error region around the estimate that corresponds to the number of students in the calculation).

The relationship between confusion and insight is further investigated in Table 2 and **Figure 17**. Table 2 presents a statistically significant relationship between the simultaneous presence of confusion and insight. When this was first seen it was a surprise—it **did not seem** reasonable. When we looked at the drawings for the 18 students who portrayed both situations we understood the table. **Figure 17** represents a typical drawing showing the student starting off confused with the statistics book and then arriving at insight ("aha"). This information is invaluable to an instructor because on the first day of class students can be told about (and shown through these drawings) the experiences they are likely to share over the course of the semester, e.g. persistence, patience, and effort will eventually counter the initial confusion.

What about relationships the student ratings of instruction and the drawings they created in each of those classes? Figure 18 contains the relationship between the percent of students in each class who rated the instructor as 'excellent" and the percent who depicted a student positively in their drawing of that class. The positive relation suggests a simple instructional principle: happy students make for happy raters. **Figure 19** is a drawing form a student who depicted students positively and who also gave an "excellent" rating on the form

Figure 20 represents the relationship between "understanding occurring over time" and percent excellent ratings. Like the previous figure, this graph clearly shows that as the percent of drawings representing understanding occurring over time increases, so does the percent of excellent ratings for those classes increase. **Figure 21** is a drawing from a student who depicted "understanding occurring over time" and who also gave an "excellent" rating on the form.

Figure 22, in contrast, contains a negative relationship between the extent to which students strongly agreed that "principles and concepts" had been taught and where I was located in the drawing—in this case, at the board. When these drawings were looked at more closely it was observed that many showed me facing or writing indecipherable text on the board, with my back to the students. This frequent depiction suggests a style of impersonal instruction associated with statistical minutiae. **Figure 23** is a drawing from a student who gave a "disagree" response to the question about understanding principles and concepts in the class.

Summary:

I've been **trying to understand how students experience classes** and how those **experiences affect learning** ever since I first started teaching. For example:

- What **techniques** of engagement or interaction work well
- What **delivery systems** (overhead, handouts, board work) work well
- What do students **pay attention to** during class
- What **do they think** about the overall class environment

- How do they **experience the class** and describe it to others.

Basically, I’ve questioned **how different professional and personal characteristics affect one’s teaching** and the extent to which those **variables affect their learning** and the extent to which those variables and their **affects are alterable**. To answer these kinds of questions, I’ve been working on ways to extract more useful personal classroom experience information from students.

One solution has been to develop an **intensive statistical investigation** into the understanding and modeling of student response to **standard course evaluations**. The systematic, longitudinal analysis of evaluations presented here has been a useful means of offering insight into what factors underlie the ratings students provide. Such an analysis can be particularly powerful and contextually relevant when would instructor specific professional and personal variables are included.

In addition to this relatively straightforward approach, the classroom drawings can also be useful at extracting more emotional, affective student experiences—both in terms of the structural aspects of instruction and how that instruction had an impact on the student.

What are you going to take away from this presentation?

- You will remember that **graphs can be useful** for showing trends over time and various other creative aspects of one’s teaching career but you aren’t going to remember the details of those graphs and you may be hard pressed to build your own data file and conduct analyses of it.
- You will, however, remember some of these drawings and you will wonder about the kinds of **drawings you would get** if you did this in one of your classes.

These drawings are extremely personal, emotional, and powerful. On one hand, they can bring us **pleasure and satisfaction**—and those are good feelings for teachers to experience—pay certainly isn’t the only reason we do what we do. On the other hand, they can be **profoundly disturbing and haunting**.

Which ever is the case, they will affect you, and potentially your practice, in ways that will never be possible from looking at a summary rating on a computer printout.

Through the work described in this presentation, it has been possible to detect variables that influence my teaching quality and effectiveness. These modifications were all prompted by patterns found through these efforts to extract meaningful information from students about their course experiences.

REFERENCES

Haney, W., Russell, M., Gulek, C. & Fierros, E. (Jan/Feb. 1998). Drawings on education: Using student drawings to promote middle school improvement. *Schools in the Middle: Theory into Practice*, 38-43.

Ludlow, L.H. (1996). Instructor evaluation ratings: A longitudinal analysis. *Journal of Personnel Evaluation in Education, 10,* 83-92.

Ludlow,L.H. (1999). Putting the psych in psychometrics. *Popular Measurement*, *2*, 50 51.

Ludlow,L.H. & Alvarez-Salvat,R. (2001). Spillover in the academy: Marriage stability and faculty evaluations. *Journal of Personnel Evaluation in Education*, *15:2*, 111-119.

Ludlow,L.H. (2002a). A structural model for understanding faculty evaluations. Paper presented at the American Educational Research Association annual meeting. New Orleans, LA. April 5.

Ludlow, L. H. (2002b). Rethinking practice: Using faculty evaluations to teach statistics. *Journal of Statistics Education*, 10(3). www.amstat.org/publications/jse/v10n3/ludlow.html.

Ludlow,L.H. (2003). Faculty course evaluations: Is a picture worth a 1000 numbers? Michigan State University. Center for the Scholarship of Teaching. March 26.

Ludlow,L.H. & Bebell,D. (2003). Drawing conclusions: Using student drawings to inform teaching and learning in a university setting. Paper presented at the American Educational Research Association annual meeting. Chicago, IL. April 23.

Ludlow,L.H., Bebell,B. & Trong,K. (2004). Drawings conclusions II: Using student drawings to inform teaching and learning in a university setting. Paper presented at the American Educational Research Association annual meeting. San Diego, CA. April 14.

Ludlow LH (2005). Making sense of course evaluations: A longitudinal analysis of teaching. *Practical Assessment Research and Evaluation*, http://pareonline.net/pdf/v10n1.pdf.

Ludlow,L.H., Enterline,S. & Famularo,L. (2005). Student drawings and student ratings of instruction: When numbers bow to figures. Paper presented at the American Educational Research Association annual meeting. Montreal, Canada. April 14.

Ludlow,L.H., Famularo,L, & Enterline,S. (2005). Drawing conclusions III: Student drawings to inform teaching and learning in a university setting. Roundtable presentation at the American Educational Research Association annual meeting. Montreal, Canada. April 13.

Ludlow,L.H., Diaconu,D., Erberber,E. & Muncaster,K. (2005). Does instructor availability matter: The size, time, outclass model by degree. Paper presented at the American Association of Colleges for Teacher Education annual meeting. Washington, DC. February 21.

Weber, S. & Mitchell, C. (1995), *That's Funny, You Don't Look Like a Teacher*, Washington, D.C.: Falmer.

Wheelock, A., Bebell, D., & Haney, W. (2000). "What Can Student Drawings Tell Us About High-Stakes Testing in Massachusetts?" *TCRecord Online* www.tcrecord.org.


Look beyond the surface!!
Don't U Believe It!
Don't Believe the Hype!
Be aware of the Dark Side!
YODA
Dr. Famous's
Experiment I
Experiment II
Conclusions
Implications
Media Hype
That's a cool beam!
AHHA I can see it!
Luke
9999
460 2 01

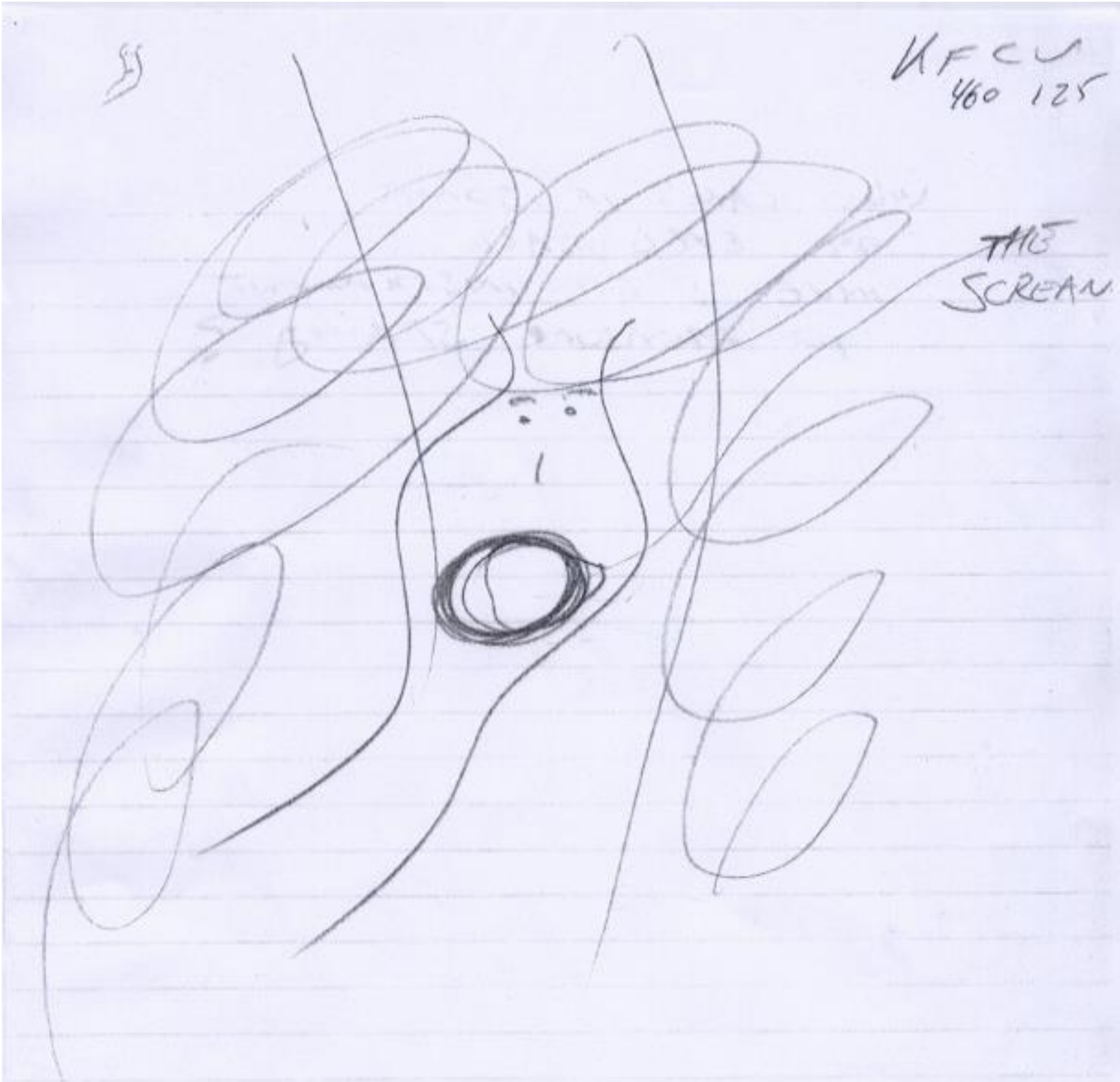
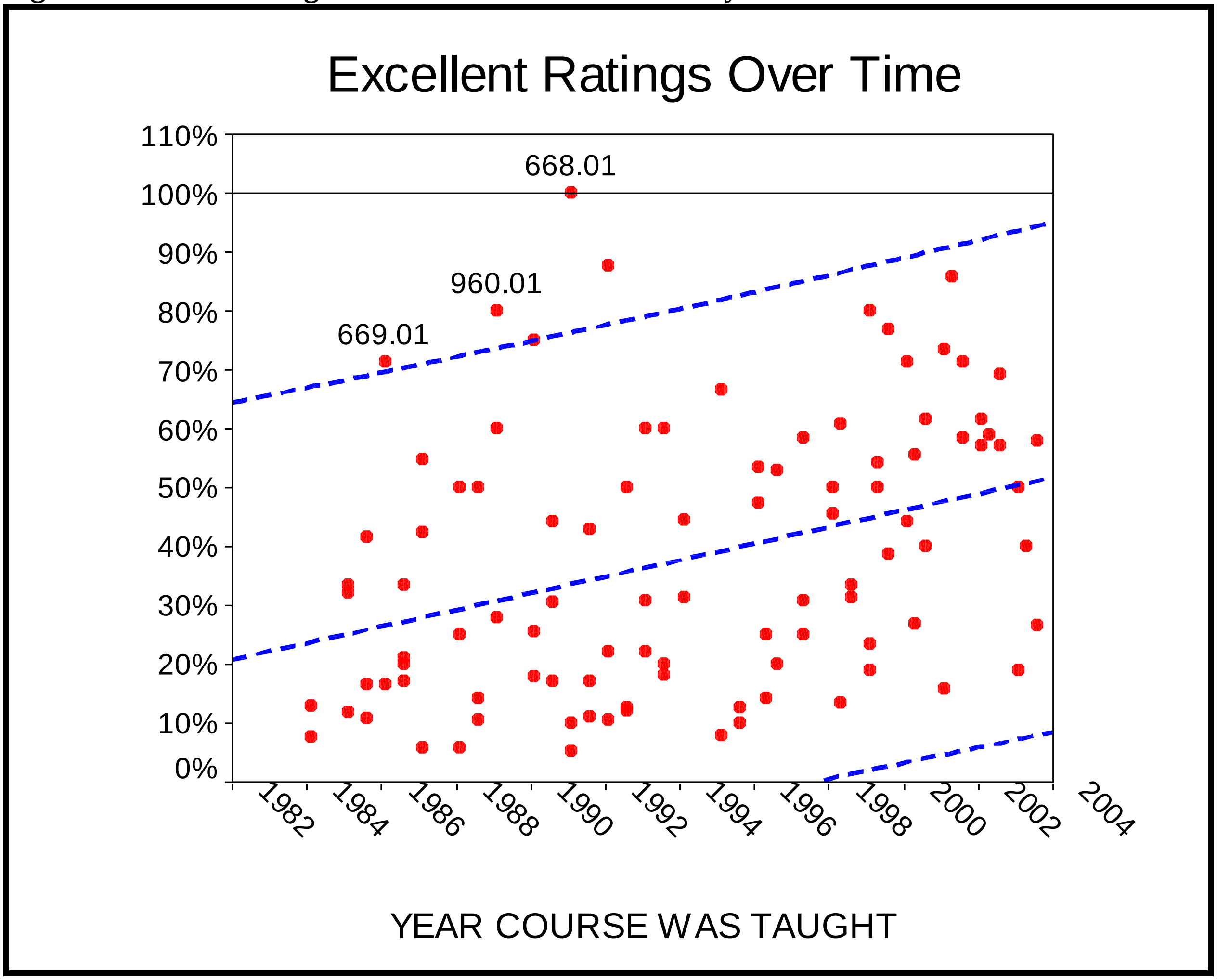
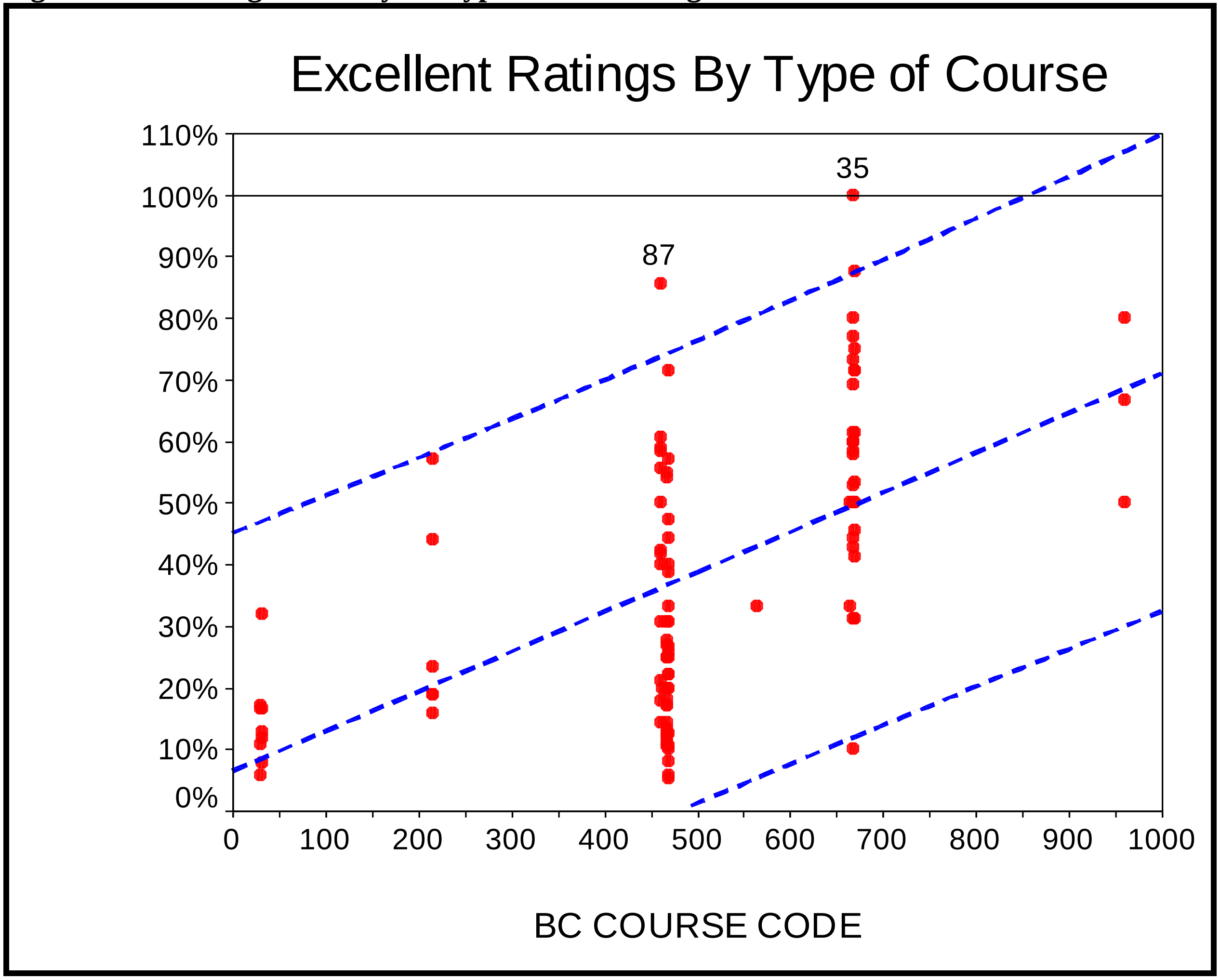

KFCU
460 125
THE
SCREAM

Figure 8. How do ratings look across all classes and 20 years?



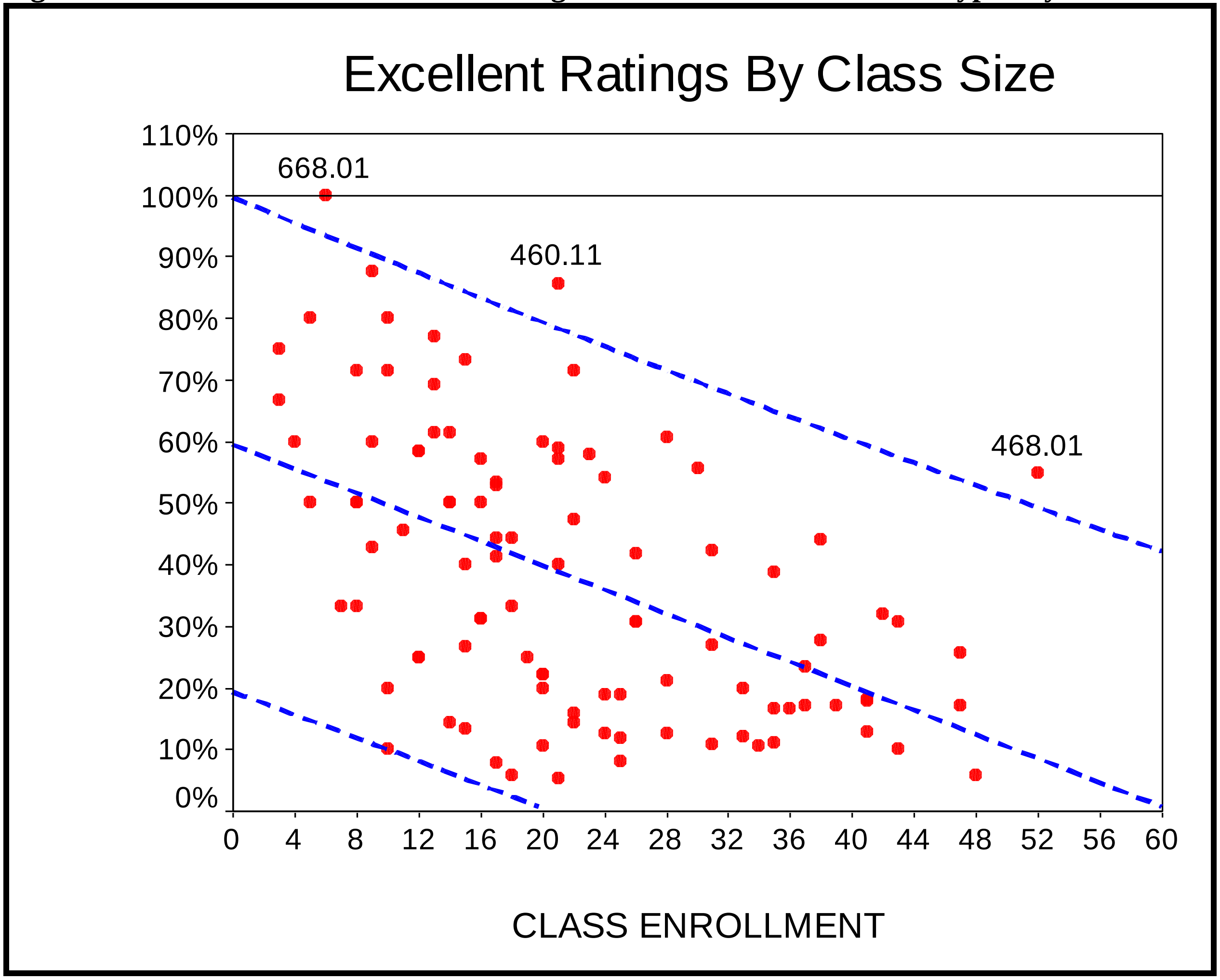
The ratings follow a general upward trend over time across all classes. The center dashed line is the regression line—the predicted excellence rating for a given year the course was taught. The other two dashed lines represent the 95% confidence interval around the regression—the region within which we most of the ratings to lie. Note that at the time that 668, 669, and 960 were taught for the first time (.01), their ratings were much higher than expected based on other classes up to that point.

Figure 9. Do ratings differ by the type of course taught?



We see a general upward trend in ratings as the course code increases. If we assume the course code is a proxy for level of complexity of the material and sophistication of the student (30 and 31 were freshman child development courses while the 600+ courses are doctoral specialty statistics courses), then this makes sense. Now an individual course (say, 469—Stat II) could be selected and plotted to see how it looks over time.

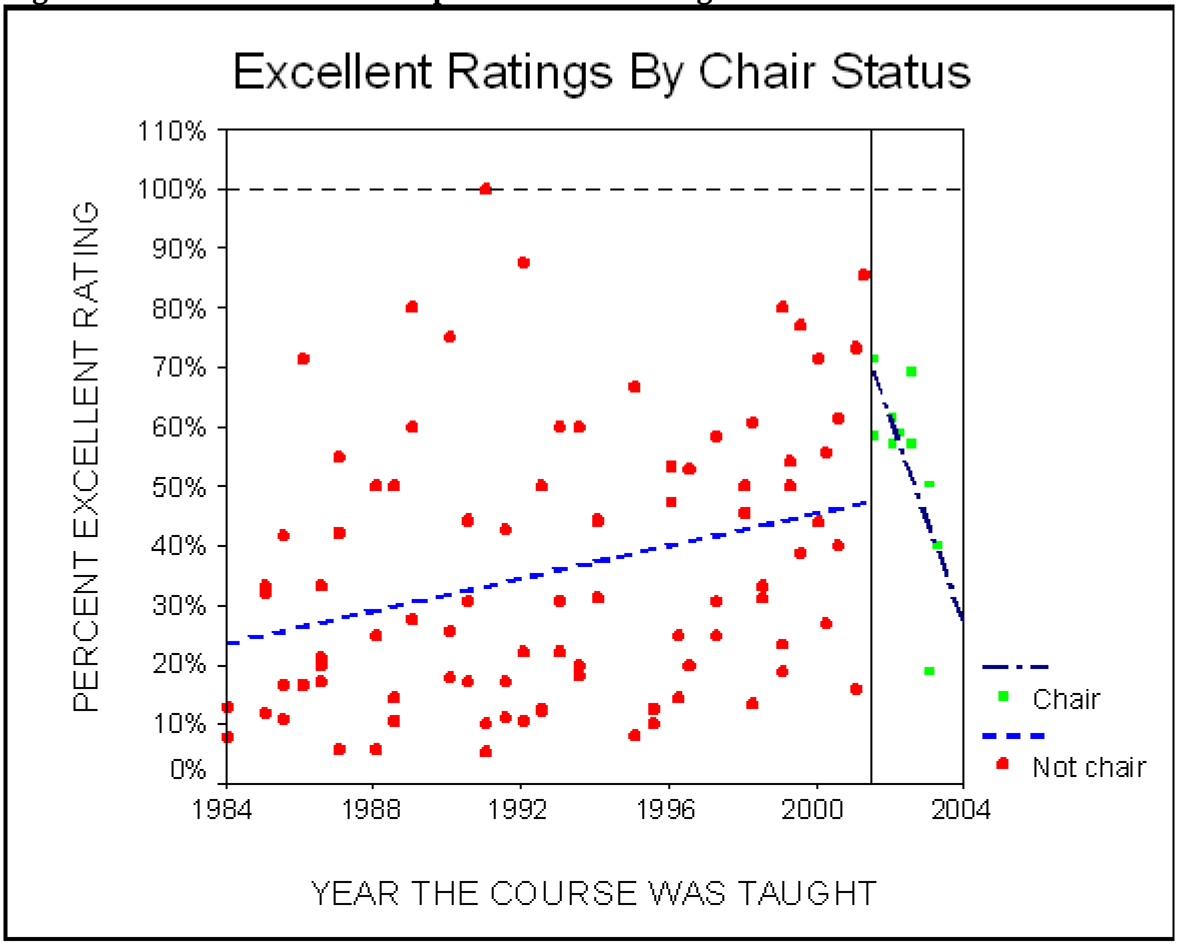
Figure 10. What is the evidence for the negative class size effect that we typically assume exists?



There is a clear, unmistakable relationship between the ratings and class size—as enrollment increases, the percent of excellent ratings tends to drop. The drop is actually at the rate of a decrease of 1% in excellence rating for each additional student added to a class. Note that 468.01 is an outlier—it was the first time it was taught, it attracted a crowd, and it differently substantially from the way it had been previously taught.

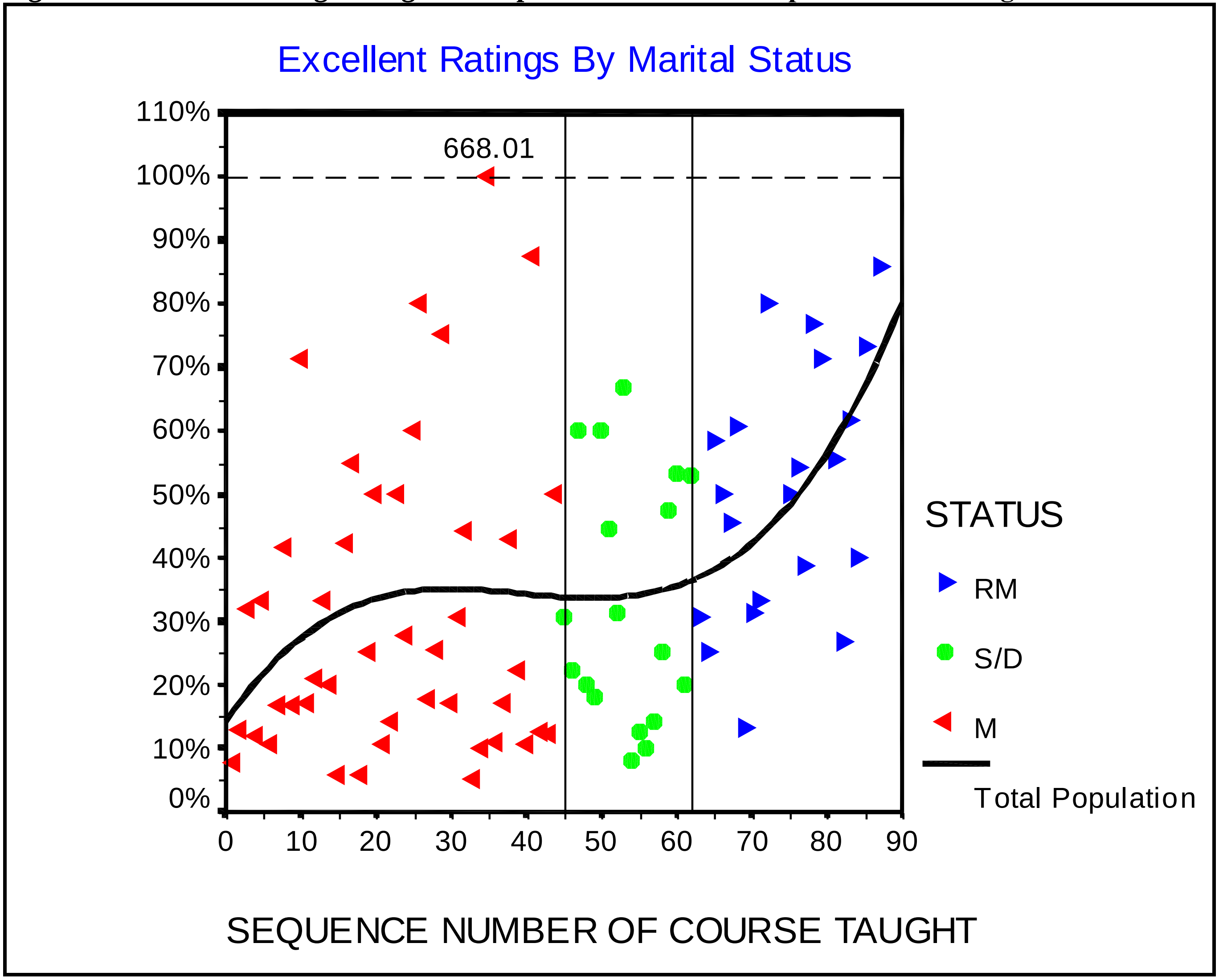
This is what we take to the Dean's office to argue for smaller classes!

Figure 11. Is there a relationship between the ratings and administrative status?



Although the overall long-term trend of percent excellent ratings is positive, as seen in Figure 8, the specific ratings while serving as department chair show a downward trend! This is consistent with my feelings of rushing into a class and not being as fully prepared as I would like at times. It is also consistent with abbreviated office hours and comments on assignments.

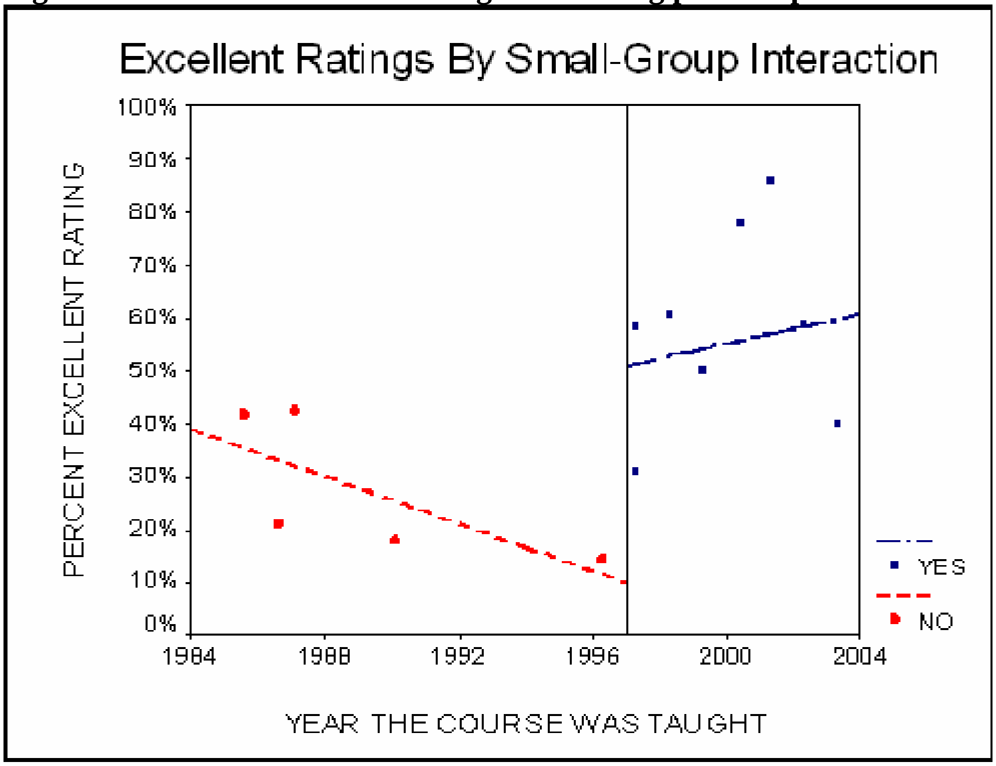
Figure 12. What effect might a significant personal factor have upon one's teaching?



The term "spillover-effect" is usually applied to the spillover of pressures from work to home. Here, I use it to refer to spillover from home to work. Specifically, during the early phase of marriage and work at BC the ratings show an upward trend. The ratings, however, start to fall off prior to and continuing into a period of separation and divorce. During this period they again change direction and begin to recover prior to and continuing into my remarried status.

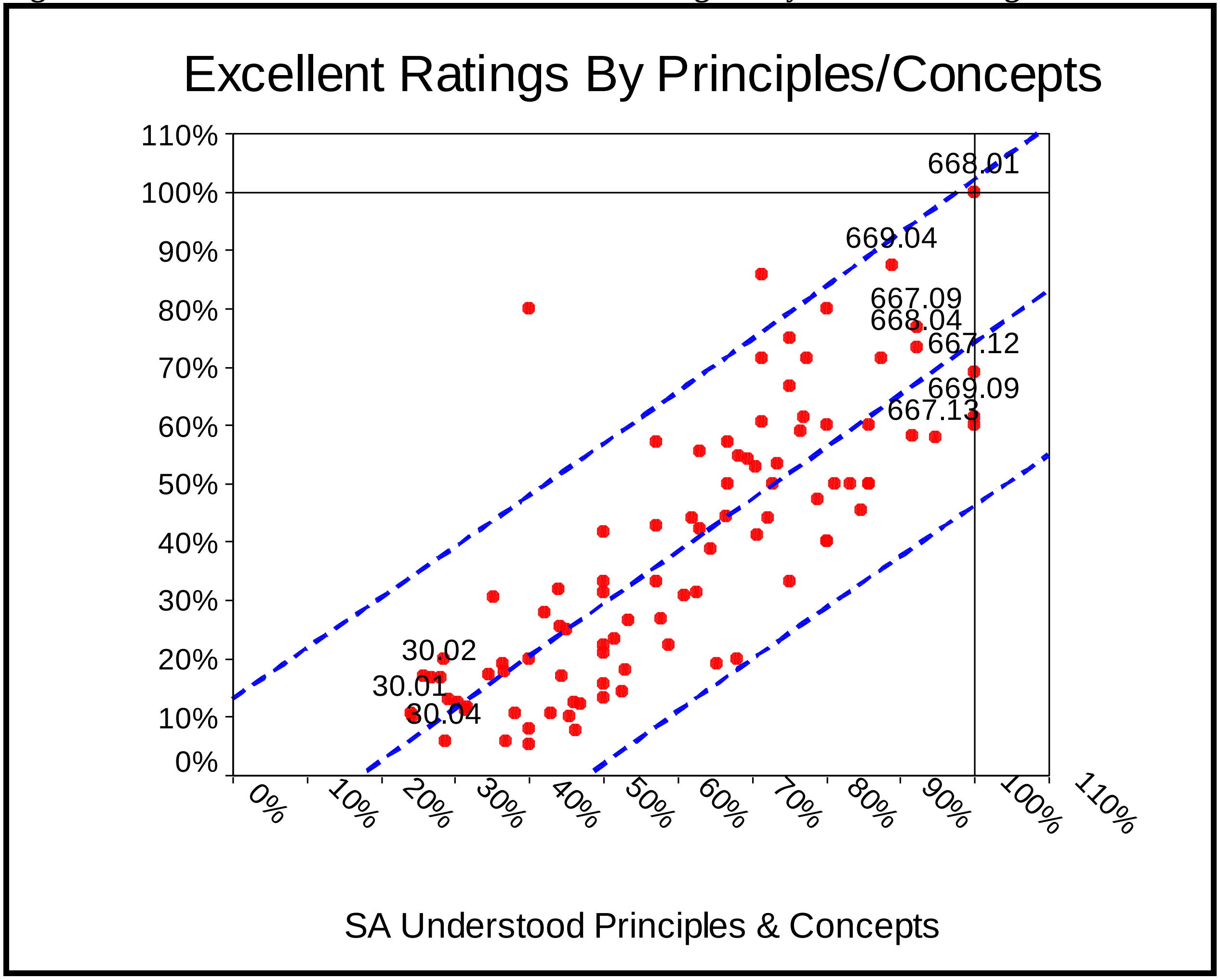
Technically, this is a statistically significant cubic relationship—non-technically, this is the "*Marilyn Effect*".

Figure 13. What effect does a change in teaching practice produce?



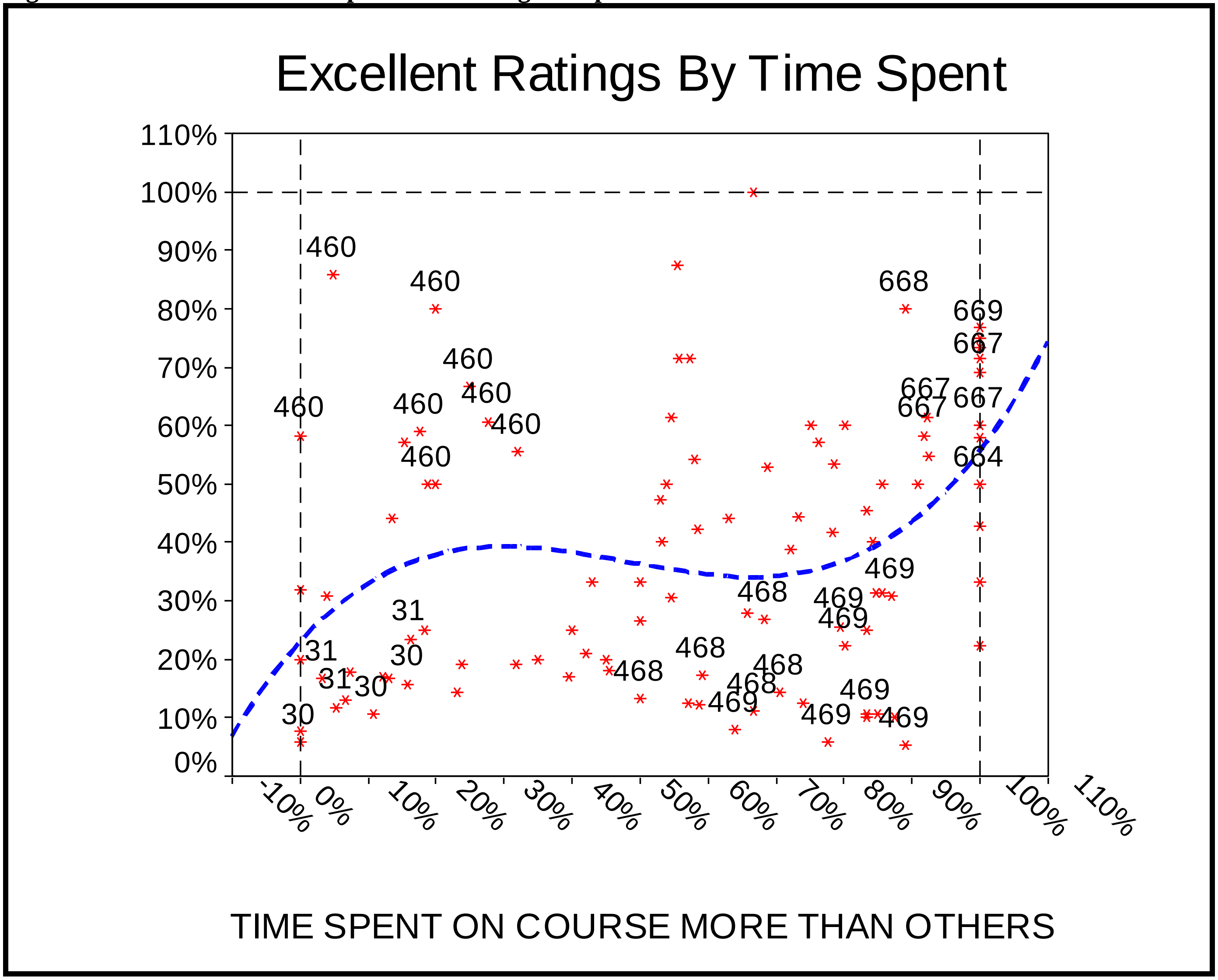
The ratings show a clear difference between when this research methods course was lecture-based ("NO" small-group interaction) versus its present format ("YES" small-group interactions with lectures). This type of graph is extremely useful for assessing the impact of any major change in course structure.

Figure 14. What factors are controllable and how might they affect the ratings?



For these types of courses there is an unmistakable positive relationship between the extent to which students perceive they have been taught principles and concepts and the percent of excellent ratings they gave. And we see again the difference in percent excellent ratings for the lower-level undergraduate courses versus the doctoral-level specialty courses (30 indicates freshman child development courses while the 600+ courses are doctoral statistics courses).

Figure 15. Is there a relationship between ratings and perceived workload?

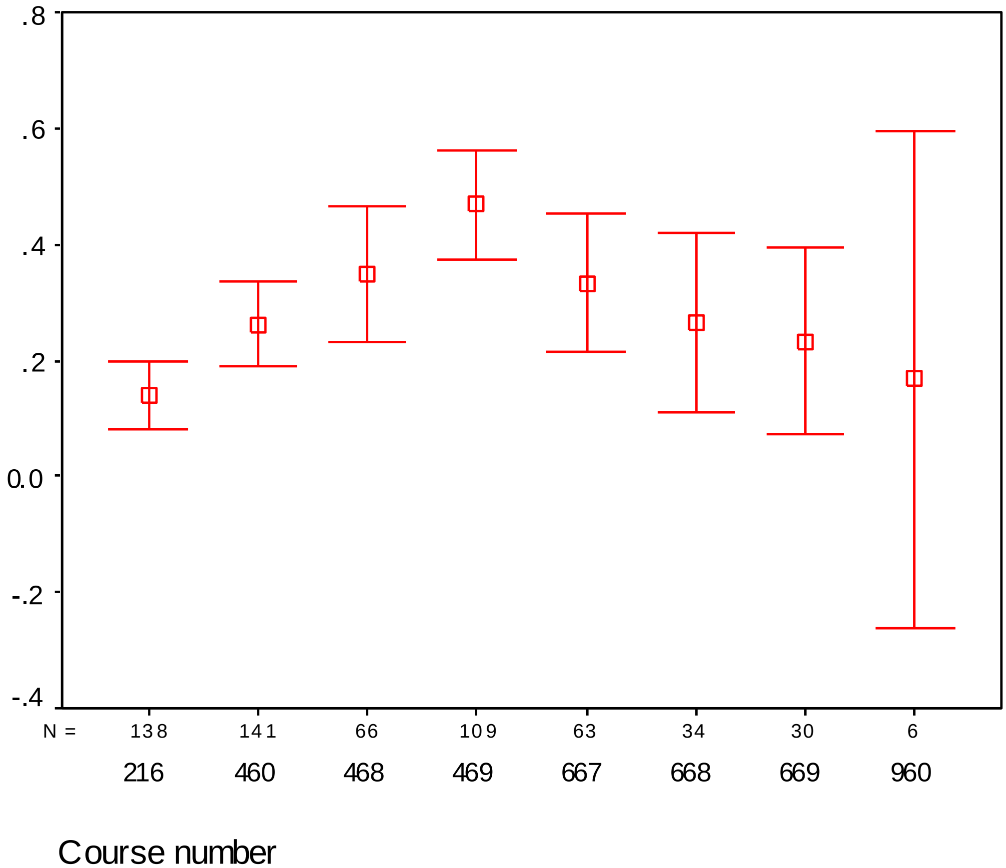


The child development classes (30, 31) did not require much time and were not highly rated; the research methods classes (460) did not require much time but were highly rated; the required statistics classes (468, 469) did require a great deal of time but were not highly rated; the specialty statistics classes (664, 667, 668, 669) did require a great deal of time and were highly rated.

Table 1: Proportion of drawings representing various classroom experiences

	Research Methods (UG)	Introductory Statistics	General Linear Models	Overall Drawings
Instructor Presence/Affect:				
Depicted Positively	0.57	0.**26**	0.40	0.41
Depicted Negatively	0.00	0.00	0.02	0.01
Location of Instructor:				
At board	0.72	0.58	0.62	0.66
At overhead	0.13	0.11	0.13	0.13
With student	0.27	0.18	0.24	0.24
Instructor Interaction:				
Instructor Speaking	0.33	0.14	0.40	0.31
Supportive	0.05	0.08	0.08	0.07
Instructing class/lecturing	0.74	0.50	0.68	0.68
Student(s) Present:				
1 depicted	0.14	0.24	0.37	0.24
2 or more	0.**82**	0.61	0.49	0.69
Sitting in groups	0.25	0.14	0.08	0.18
Sitting in rows	0.62	0.52	0.40	0.53
Asking a question	0.12	0.09	0.11	0.12
Student to student interaction	0.20	0.12	0.10	0.16
Student Depiction:				
Depicted Positively	0.22	0.23	0.27	0.27
Depicted Negatively	0.05	0.17	0.14	0.13
Depicted Neutrally	0.11	0.14	0.10	0.12
Course Experience(s):				
AHA/light bulb/lightning bolt	0.02	0.**12**	0.03	0.06
Other sudden insight	0.01	0.02	0.10	0.06
Understanding over time	0.02	0.15	0.16	0.09
Enthusiasm/excited	0.05	0.08	0.08	0.08
simple understanding	0.16	0.26	0.29	0.26
daydreaming	0.09	0.03	0.06	0.06
confused/overwhelmed/lost	0.14	0.**35**	0.33	0.29
About Classroom:				
Computer depicted	0.07	0.23	0.11	0.09
Overhead projector	0.17	0.20	0.21	0.18
Laser pointer	0.12	0.09	0.14	0.12
Clock	0.04	0.05	0.05	0.04
Other:				
Readable text	0.28	0.36	0.24	0.29
Unreadable text	0.39	0.27	0.19	0.32
Statistical symbols/formula/tables	0.07	0.45	0.**54**	0.28
Graphical representation of the data	0.25	0.44	0.32	0.30
Students thought(s) depicted	0.24	0.27	0.37	0.30
Metaphorical	0.02	0.**21**	0.19	0.12
n=	138	66	63	587

Figure 16: Relationship between level of confusion and course



The courses with the most drawings depicting confused students are the required statistics courses. The general research methods courses and the advanced statistics courses have the fewest depictions of confusion. This is not surprising and is consistent with the evaluation ratings.

Table 2: Relationship between presence of confusion and sudden insight

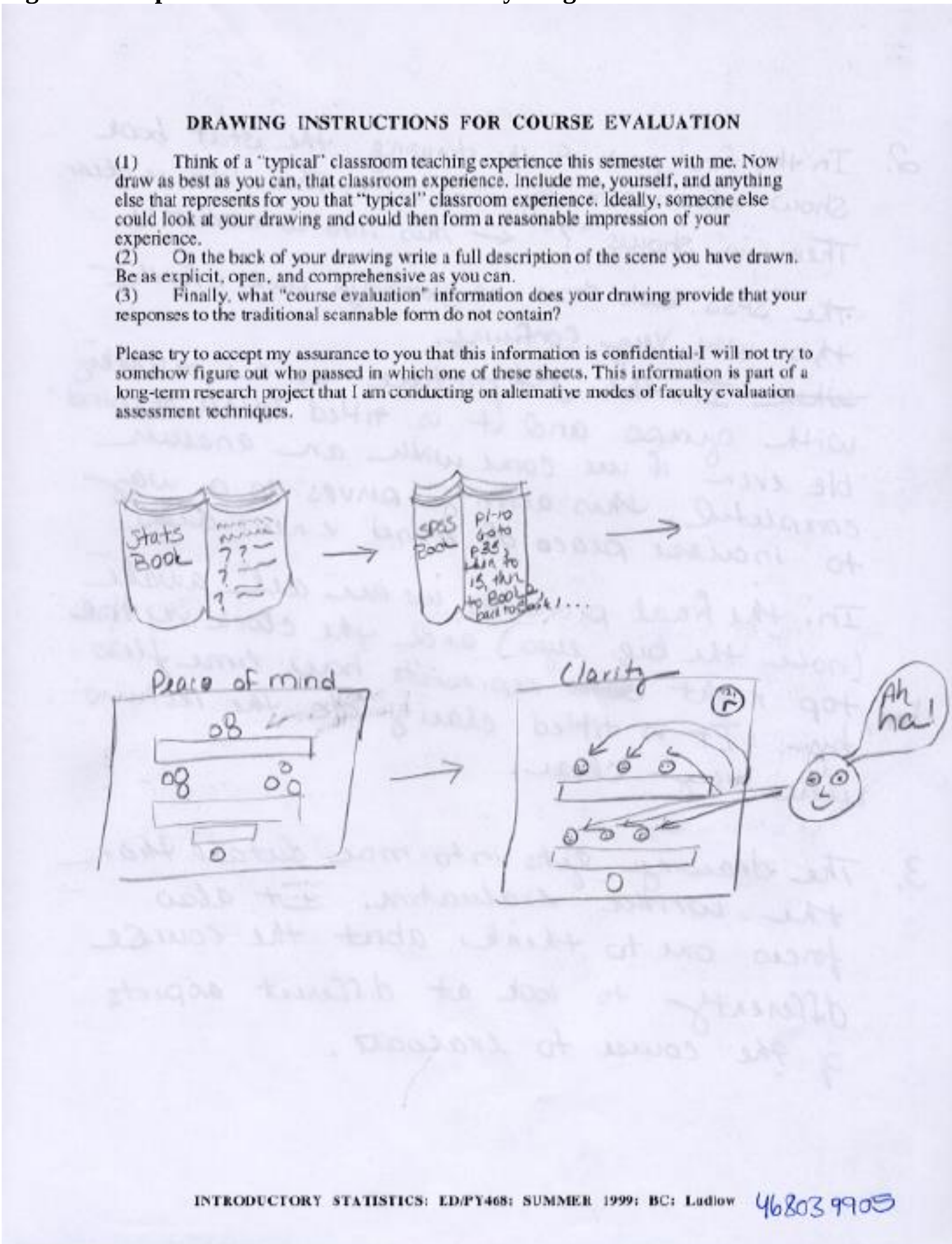
Other sudden insight * confused/overwhelmed/lost Crosstabulation

			confused/overwhelmed/lost		
			absent	present	Total
Other sudden insight	absent	Count	401	150	551
		Expected Count	393.3	157.7	551.0
		Std. Residual	.4	-.6	
	present	Count	18	18	36
		Expected Count	25.7	10.3	36.0
		Std. Residual	-1.5	2.4	
Total		Count	419	168	587
		Expected Count	419.0	168.0	587.0

chi-square=8.6, p=.003

The particularly interesting cells in this table are in the “confused/overwhelmed/lost—present” column. There were 150 drawings where confusion was depicted with nothing suggesting “sudden insight”. This finding was not statistically surprising. But, in the “confused--present” and “other sudden insight--present” cell, there were 18 drawings depicting both situations—confusion followed by insight. Statistically, it was expected that there would be 10 such drawings. This is helpful information to pass along to students who are feeling particularly lost.

Figure 17. Depiction of confusion followed by insight



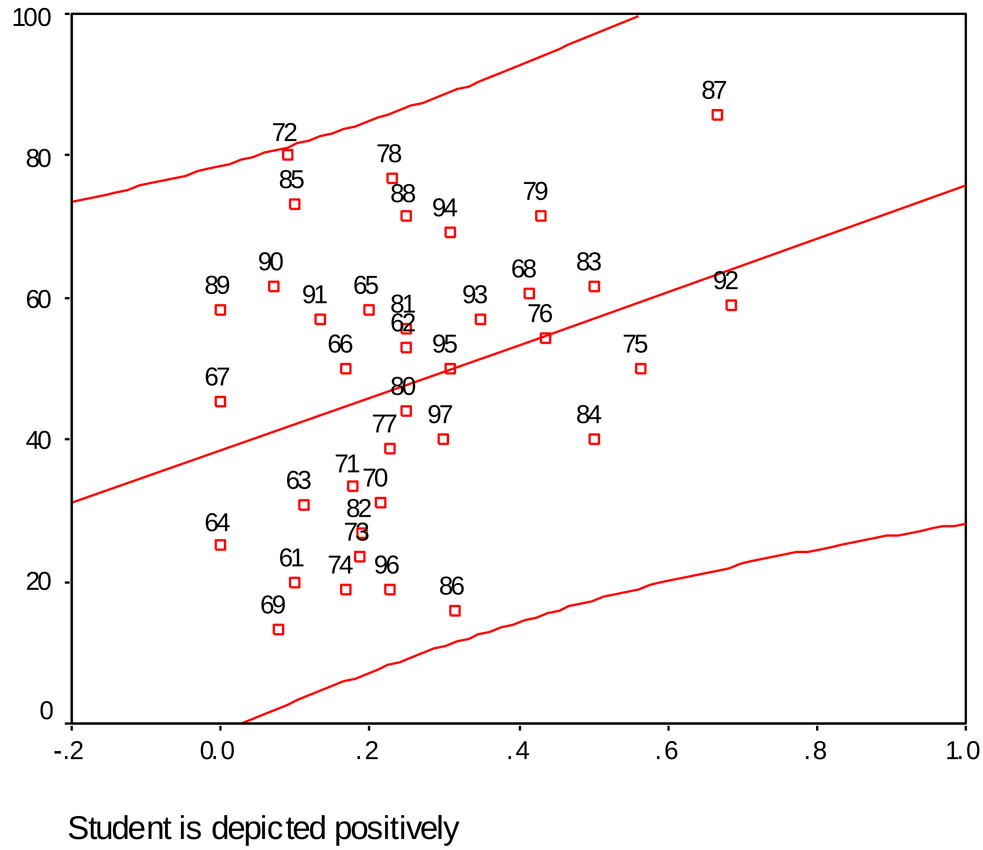
DRAWING INSTRUCTIONS FOR COURSE EVALUATION

(1) Think of a "typical" classroom teaching experience this semester with me. Now draw as best as you can, that classroom experience. Include me, yourself, and anything else that represents for you that "typical" classroom experience. Ideally, someone else could look at your drawing and could then form a reasonable impression of your experience.
(2) On the back of your drawing write a full description of the scene you have drawn. Be as explicit, open, and comprehensive as you can.
(3) Finally, what "course evaluation" information does your drawing provide that your responses to the traditional scannable form do not contain?

Please try to accept my assurance to you that this information is confidential-I will not try to somehow figure out who passed in which one of these sheets. This information is part of a long-term research project that I am conducting on alternative modes of faculty evaluation assessment techniques.

INTRODUCTORY STATISTICS: ED/PY468: SUMMER 1999: BC: Ludlow 46803 9905

Figure 18: Relationship between ratings and affect of students



This plot shows that as the proportion of drawings that depicted students positively increased in classes, so did the percent of students who marked "excellent" on their course evaluations.

Figure 19. Drawing of positive students by student who gave "excellent" rating.

4244

DRAWING INSTRUCTIONS FOR COURSE EVALUATION

(1) What visual image of a classroom experience comes to mind when you think of this course? Now draw as best as you can, that classroom experience. Include me, yourself, and anything else that represents for you that classroom experience. Ideally, someone else could look at your drawing and could then form a reasonable impression of your experience.

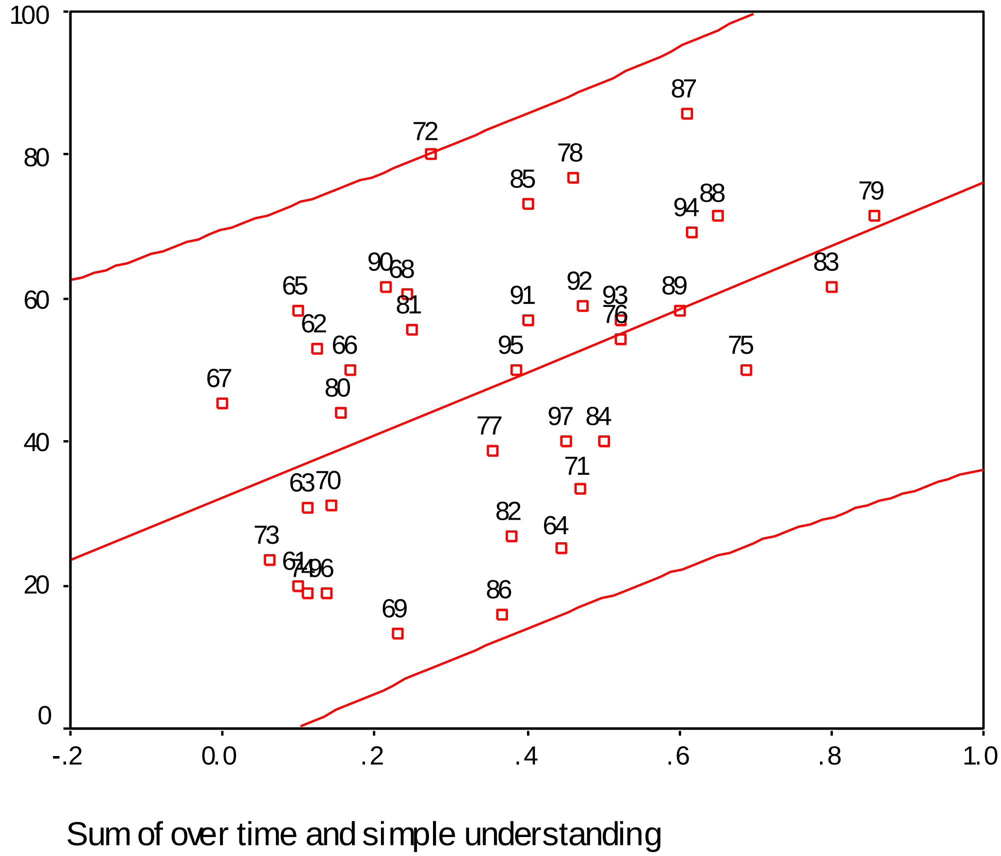
(2) On the back of your drawing write a full description of the scene you have drawn. Be as explicit, open, and comprehensive as you can.

(3) Finally, what "course evaluation" information does your drawing provide that your responses to the traditional scannable form do not contain?

Please try to accept my assurance to you that this information is confidential--I will not try to somehow figure out who passed in which one of these sheets. This information is part of a long-term research project that I am conducting on alternative modes of faculty evaluation assessment techniques.

Research Methods: ED/PY216: Spring 2003: BC: Ludlow

Figure 20: Relationship between ratings and "total understanding"



This plot shows that as the proportion of drawings depicting "simple understanding" or "understanding over time" increased in classes, so did the percent of students who marked "excellent" on their course evaluations.

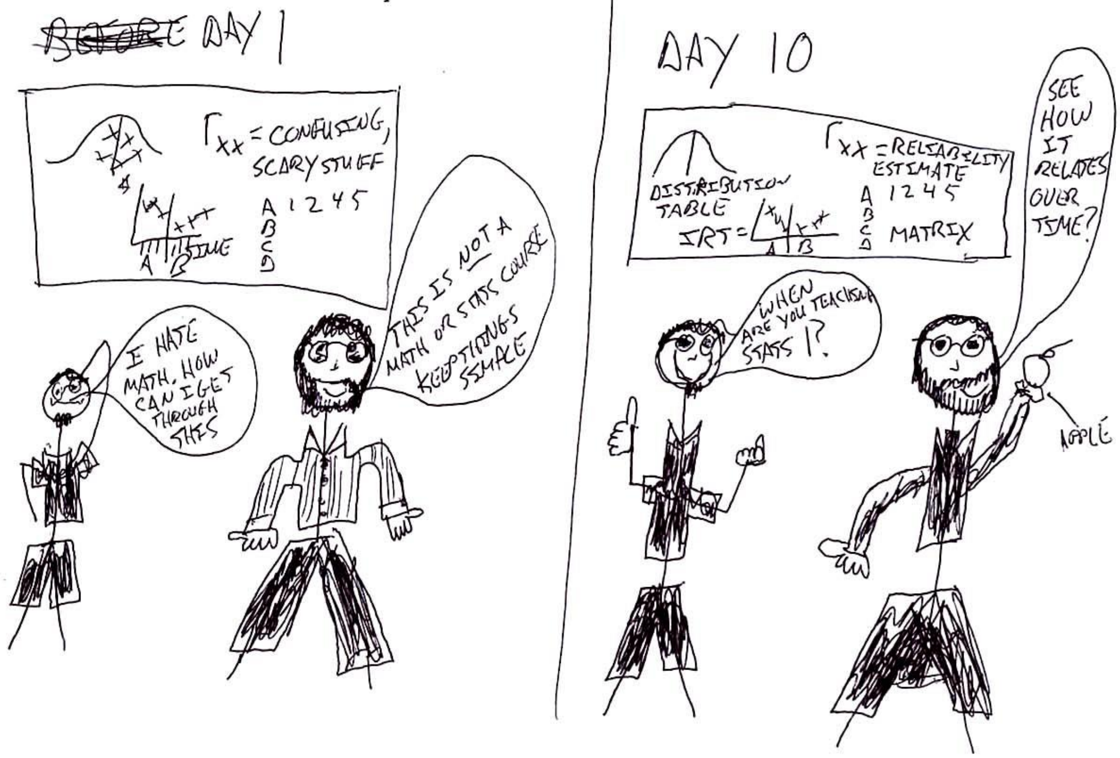
Figure 21. Drawing depicting understanding over time by student who gave "excellent" rating

NEON

DRAWING INSTRUCTIONS FOR COURSE EVALUATION

(1) What visual image of a classroom experience comes to mind when you think of this course? Now draw as best as you can, that classroom experience. Include me, yourself, and anything else that represents for you that classroom experience. Ideally, someone else could look at your drawing and could then form a reasonable impression of your experience.
(2) On the back of your drawing write a full description of the scene you have drawn. Be as explicit, open, and comprehensive as you can.
(3) Finally, what "course evaluation" information does your drawing provide that your responses to the traditional scannable form do not contain?

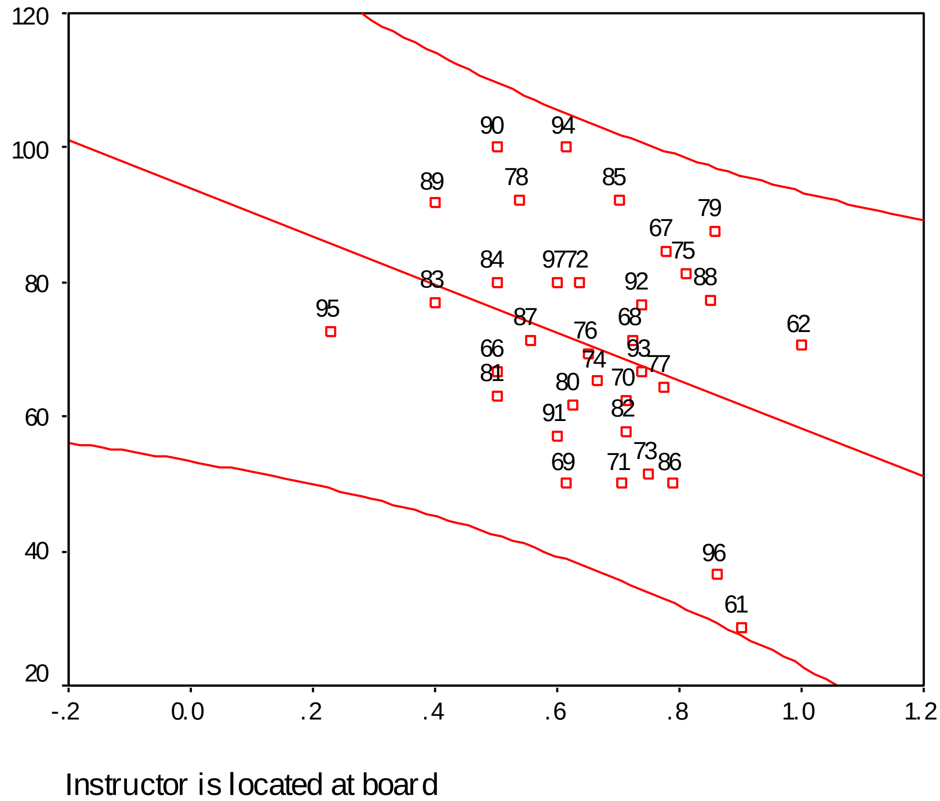
Please try to accept my assurance to you that this information is confidential--I will not try to somehow figure out who passed in which one of these sheets. This information is part of a long-term research project that I am conducting on alternative modes of faculty evaluation assessment techniques.



Research Methods: ED/PY460: Summer 2001: BC: Ludlow

460030111

Figure 22: Relationship between principles and concepts and instructor location



This plot shows that as the proportion of drawings depicting me standing at the board increased in classes (traditional lecture format), the percent of students who agreed that they understood principles and concepts decreased.

Figure 23. Drawing from a student who gave a low evaluation and a low rating on understanding principles and concepts.

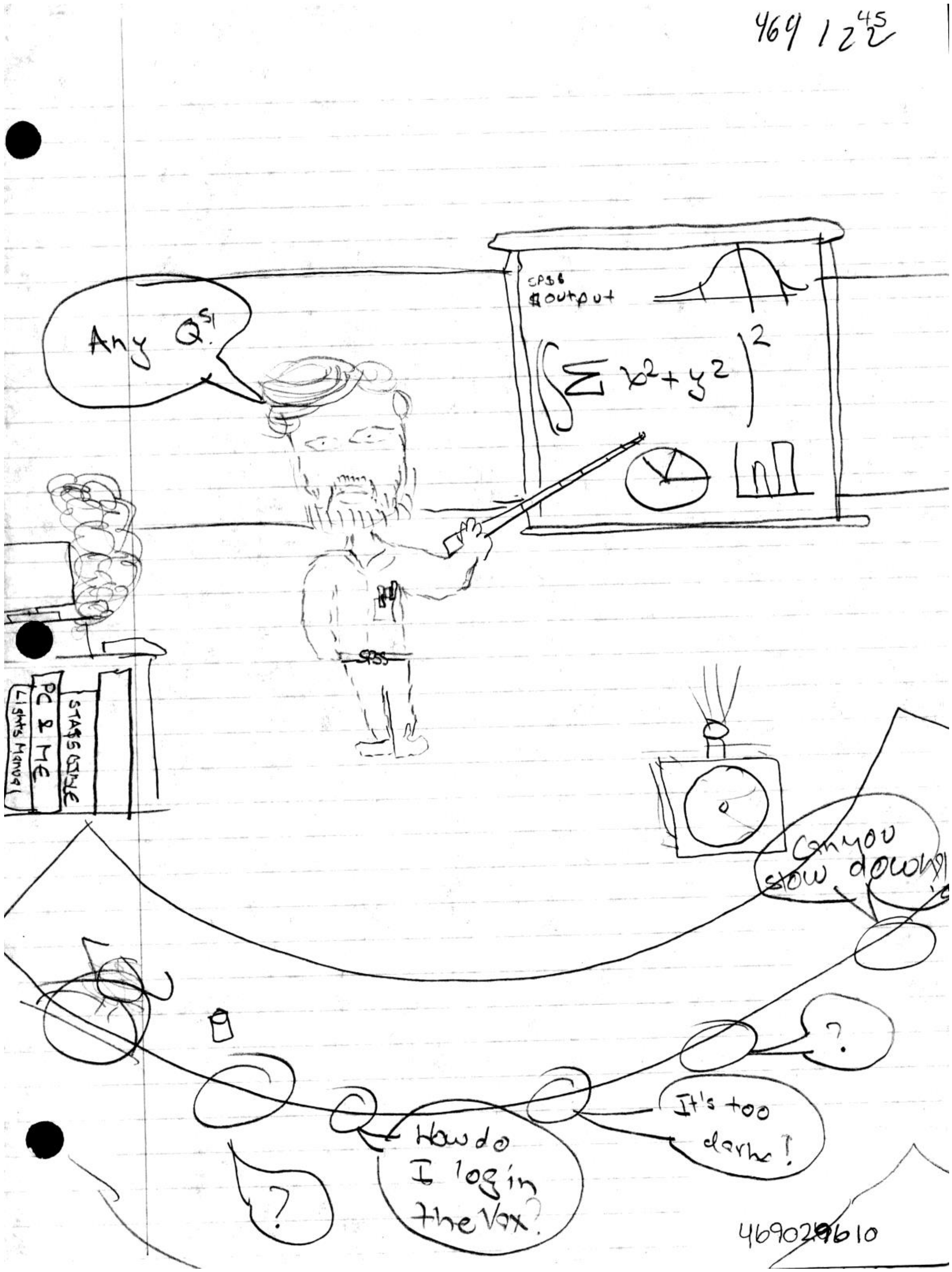


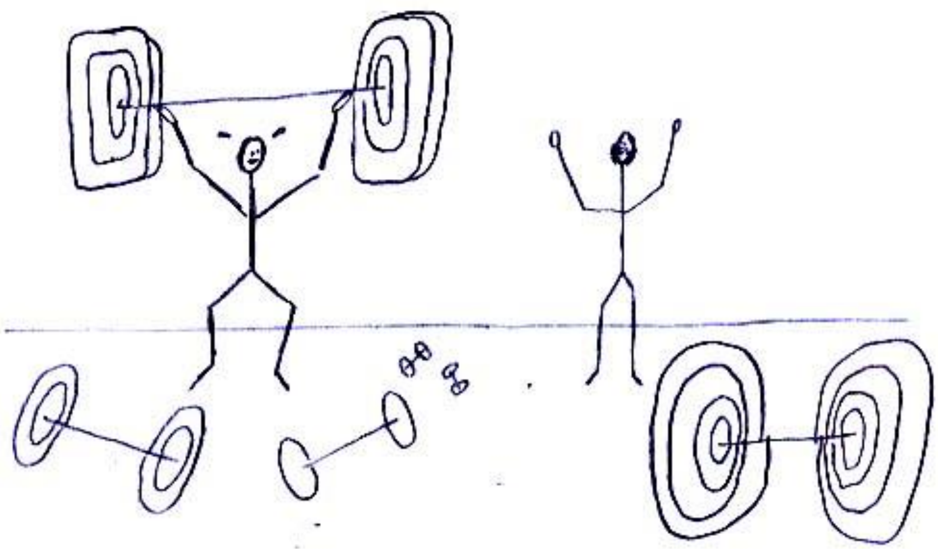
Figure 24: Finally, what does it all mean in the end?

4788

DRAWING INSTRUCTIONS FOR COURSE EVALUATION

(1) What visual image of a classroom experience comes to mind when you think of this course? Now draw as best as you can, that classroom experience. Include me, yourself, and anything else that represents for you that classroom experience. Ideally, someone else could look at your drawing and could then form a reasonable impression of your experience.
(2) On the back of your drawing write a full description of the scene you have drawn. Be as explicit, open, and comprehensive as you can.
(3) Finally, what "course evaluation" information does your drawing provide that your responses to the traditional scannable form do not contain?

Please try to accept my assurance to you that this information is confidential--I will not try to somehow figure out who passed in which one of these sheets. This information is part of a long-term research project that I am conducting on alternative modes of faculty evaluation assessment techniques.



Psychometrics: ED669: Spring 2002: BC: Ludlow

Coded
1

669010201

Hard work pays off!