Drawing conclusions II:
Using student drawings to inform teaching and learning in a university setting

Larry Ludlow, Damian Bebell, Katie Trong

Boston College
Lynch School of Education
Educational Research, Measurement and Evaluation

Paper presented at AERA, April 2004, San Diego CA

Understanding what makes a great teacher a great teacher is a long debated problem for those working in educational evaluation, policy, and measurement. The literature shows that from the very beginnings of public schools there have been measures in place to assess the quality of the teacher (Haney, Madaus, and Kreitzer, (1987). The earliest efforts were largely subjective observations and interviews carried out by untrained individuals. Surprisingly, many of these methods still thrive throughout education (Haney, et al. 1987).

A common theme of education reformers is that teachers within schools must become reflective practitioners if they are to become more successful in meeting the needs of increasingly diverse student populations. As Sternberg & Horvath (1995) wrote, "The current popularity of 'reflective practice' as a touchstone for teacher excellence suggests that, in the minds of many, the disposition toward reflection is central to expert teaching" (p. 15). Thus it is not surprising that there is currently great interest in promoting reflective practice among teachers.

These calls for reform have not been focused solely upon the traditional K-12 educational system. Recent reform efforts have been initiated for the teaching of college level statistics, particularly with respect to: cooperative learning (Garfield, 1993; Giraud, 1997); incorporating the ideas of context-relevant material (Sowey, 1995); devising alternative forms of testing and grading (Garfield, 1994); and team teaching the statistical

software component of courses (Rumsey, 1998). However, documenting and gauging the extent to which such pedagogical changes have been effective in classroom practice is a major hurdle that remains to be addressed. Traditionally, end-of-course summative evaluation forms (required at most universities) are the "end-all, be-all" of measuring student experiences and teaching pedagogy of a course.

Student ratings of instruction

These evaluations are usually summarized to show how each instructor fared relative to his/her academic unit. These crude comparisons are often considered by many faculty to be worthless or mildly interesting, at best. In particular, there is usually no university effort to track an individual's teaching record over time. In addition, the typical instructor does no systematic analysis of his/her course evaluations beyond that provided by the university. It is, however, possible to extract extremely useful and valuable information from one's own evaluations. Specifically, the faculty evaluation research exemplified in the case-study longitudinal analyses of Ludlow (1996, 2002) and Ludlow & Alvarez-Salvat (2001) clearly show relatively simple ways to statistically model and analyze evaluation results over time.

An example is provided in Figure 1. The percent excellent ratings (Y axis) for each class are plotted from the first class to the last class taught (X axis). The two vertical lines split the ratings into three different marital periods. The term "spillover-effect" is usually applied to the spillover of pressures from work to the home. Here it is used to refer to spillover from home to work. Specifically, during the early phase of marriage (M) and work at BC the ratings show an upward trend. The ratings, however, fall off prior to and continuing into the period of separation and divorce (S/D). During this period they again change direction and begin to recover prior to and continuing into the current

remarried stage (RM). This type of analysis is very effective for looking at ratings over time, by tenure status, as a function of class size, the degree to which students understood principles and concepts, and nearly any other classroom, administrative, or any other personal variable an instructor can think of that might be related to teaching effectiveness as measured by student evaluations (Ludlow, 2004).

The standard university-generated evaluation dataset that produced Figure 1 now contains the ratings for 99 classes taught since 1983. The data on those classes include the following variables:

- Year the course was taught
- Semester the course was taught
- Course catalog number (indicative of student level and difficulty of the course)
- Type of course (statistics, research methods, etc.)
- Number of times the course had been taught
- An indicator variable for pre-post tenure status
- An indicator variable for pre-post medical leave
- An indicator variable for department chair status (pre-current)
- An indicator variable for professor's marital status (divorced, married, remarried)
- Class enrollment
- Percent of students who Strongly Agreed "the course helped me to acquire factual information"
- Percent of students who Strongly Agreed "the course helped me to understand principles and concepts"
- Percent of students who Strongly Agreed "the course helped me to acquire academic skills"
- Percent of students who Strongly Agreed "the instructor was available for help outside of class"
- Percent of students who Strongly Agreed "regular class attendance was necessary for learning the required content"
- Percent of students who Strongly Agreed "the percent of time required for the course was more than others have the same credit hours"
- Instructor overall rating: percent who indicated either Excellent, Very good, Good, Acceptable, or Poor

Those analyses, unfortunately, still yield little information about *student affective experiences*—experiences that go beyond what was taught, how it was taught, and how well it was taught. Specifically, how did the student see, literally, the instructor standing

in front of the class; the clarity or confusion of material on the board; fellow students interacting with one another; their own personal progress over the term; the type or style of interaction between students and the instructor; the salient teaching tools; the level of anxiety, boredom, and attention of others? How would the student describe the class to a friend? The standard evaluation form provides little opportunity to express these experiences—experiences that arguably contribute to the establishment of an effective learning environment.

Student drawings of instruction

Beginning in 1995, the first author adapted the classroom drawing technique of Haney, et al (1998) in his classes as a means of obtaining additional course evaluation feedback. The instructions to the students consisted of the following written prompt:

> (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?

It was quickly realized that this type of exercise offered an extraordinary opportunity to understand classes from the perspective of the students. The drawings are both crude and detailed, insulting and complimentary, factual and metaphorical, provocative and disturbing, amusing and reflective. They provide a fascinating glimpse of what success or failure felt like; how the instructor was supportive or threatening; and, they provide an opportunity to understand how students perceived their peers as engaged and excited or bored and stressed. In comparison to the standard course ratings, the

"voice" of the student seemed nearly absent from the traditional university course evaluation the instructor had been using for nearly 20 years.

An example is presented in Figure 2. The powerful negative affect expressed by this student's experience in a statistics class could never be adequately expressed in a filled-in bubble on a rating scale form. To facilitate interpreting and explaining these drawings, an objective coding rubric was constructed that allowed quantitative comparisons and offered generalizability to other faculty who might choose to adopt this course evaluation tool. Such a 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?

Two coders independently reviewed two separate samples of forty drawings and recorded the various features present in the drawings. Features were coded either present or absent. In addition, the raters took notes of features that existed in the drawings but were absent from the coding sheet. The coders then compared their findings and condensed the list of features into a draft coding sheet. For features that had high levels of agreement, formal descriptions (operational definitions) of each feature were developed. For features that had low levels of agreement, the coders worked together to examine drawings for which there were discrepancies to identify reasons for discrepancies and to develop an operational definition of the feature. If a common definition of the feature could not be developed, the feature was removed from the list and the coding system.

At the 2003 AERA conference we addressed the strengths and weaknesses, and reliability and validity of this new methodology for gathering course evaluation information (Ludlow & Bebell, 2003). Specifically, that paper presented techniques for

coding large sets of university student classroom drawings to illustrate how pedagogical practices can be measured across time and across different types of courses. The purpose of the present paper is to present the results of the first set of analyses of an integrated dataset consisting of the standard university generated quantitative evaluations combined with the qualitative codes of the drawings associated with those course evaluations.

The final coding protocol for the drawings was:

OPERATIONAL DEFINITIONS FOR X's CLASSROOM DRAWINGS

Instructor Presence/Affect:

- Instructor is present: "X" is somehow, somewhere depicted in the drawing.
- 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).
- Depicted Neutrally: "X" face and speech (if any) are not positive or negative but visible.
- Can't discern affect: "X" is present but affect is not visible.

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).
- With student(s): Instructor is drawn EITHER physically with student(s) or metaphorically with student(s)

Instructor Interaction

- Instructor Speaking: Words/phrases are depicted coming from "X".
- Supportive: "X" is speaking (see above) words of support or encouragement.
- Asking Question: "X" is posing a question. (student's my be present or not present).
- Speaking Statement: "X" is lecturing/instructing
- Instructing single student: "X" is addressing one student. (words or phrases need not be present).
- Instructing class/lecturing: "X" is addressing(?) more than one student (words or phrases need not be present).

Student(s) Present: Figures are present in the drawing other than instructor or TA.

- 1 depicted: Only 1 student is depicted anywhere in the drawing.
- 2 or more: 2 or more students are depicted anywhere in the drawing.
- 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).

- Asking a question: 1 or more student is actively speaking a question, thinking a question.
- Answering a question: It is obvious that 1 or more students are actively speaking a response to a question or thinking a response to a question.
- Student to student interaction: There is some evidence of communication and discussion between (amongst) students (may include arrows connecting students or thoughts) (Not just sitting in groups).

Student Depiction

- Depicted Positively: At least 1 student's facial expression is positive (smiling), or positive speech.
- Enthusiasm/excited: 1 or more student is obviously depicted as being enthusiastic or excited.
- Depicted Negatively: At least 1 student's facial expression is negative (frowning), or negative speech.
- Confusion, frustration or stress: 1 or more student is obvious depicted as confused, frustrated or stressed.
- Depicted Neutrally: Face and speech of students is visible but not positive or negative.
- Can't discern affect: Students are present but affect is not visible.

Course Experience(s)

- AHA/light bulb/lightning bolt: At least 1 student depicts a light bulb, or AHA!, or lightning bolt
- Other sudden insight: Some other evidence of the sudden understanding of a concept or idea.
- Understanding over time: Some depiction of before/after learning, gradual growth of learning or understanding
- Enthusiasm/excited: The student(s) is obvious depicted as being enthusiastic or excited.
- Sleeping/bored: Student(s) is depicted sleeping or obviously bored.
- Daydreaming: Student(s) is obviously daydreaming (off topic thought bubbles, etc.)
- Crying: 1 or more students are depicted as crying as evidenced by text or tears.
- Angry: 1 or more student is expressing hostility, "pissed off"
- Sleeping/bored: 1 or more student is depicted as sleeping or obviously bored

About classroom

- Computer depicted: A computer is present somewhere in the drawing
- Overhead projector depicted: An overhead projector is present in the drawing
- Laser pointer: A laser pointer (or beam) is present in the drawing
- Clock: Clock or representation of a clock is depicted somewhere in the drawing.

Other

- Readable text: There is readable instructor related text somewhere in the drawing (words or sentences, NOT SPEECH—typically board work).
- Unreadable text: There is unreadable instructor related text somewhere in the drawing (scribbles, etc. NOT SPEECH—typically board work).

- Statistical symbols/formula: Stats or formula are located somewhere in the drawing.
- Graphical representation of the data: Statistical graphs or figures are located somewhere in the drawing.
- Assistants present: One of Dr. "X"'s graduate assistants are present (Chris, Julie, Jere, Camelia).
- Students thought(s) depicted: Words, thoughts or ideas are presented in the drawing representing the students thoughts.
- Metaphorical: The drawing uses a metaphorical representation of the classroom or experience, rather than a pictorial depiction of the actual classroom environment.

Data sets:

Each drawing was coded using the above list and was recorded into a Microsoft Excel 98 spreadsheet. These codes marked the presence or absence of 46 variables. Additionally, some limited information about the student artist (course, semester, year, level of study, etc.) was recorded.

The courses for which drawings have been gathered include: Research Methods, Interpreting & Evaluating Research, Statistics I, Statistics II, Multivariate I, Multivariate II, Psychometrics, and Seminar in Educational Research. The courses differ in subject matter, degree of difficulty, enrollment, level of student, and format. All of them have been taught more than once; most are taught yearly. The dataset now represents 587 student drawings collected from 45 classes taught since 1995. These codes were then imported into SPSS and were aggregated to obtain the mean drawing code proportions for each individual class. This file was then merged with the instructor's evaluations data file (consisting of 99 classes).

At present there are two data files: 1) one containing the class-level aggregated course evaluation ratings and drawing codes (99 classes with student mean ratings, 45 of which also have the proportion of times each drawing code was present in the class); and 2) one containing the 587 student-level drawing codes and their original individual course evaluation ratings.

Creation of New Variables

The following variables were created to better capture the essence of the drawings or when it seemed as though they could be combined into a useful category. Many of the variables were computed in various ways in order to observe the difference that would be created using various methods, and because each yields different information.

- Total Insight – Combining "AHA" and "other insight" to capture all student revelations. Computed in 3 ways -
  - Total Insight Combined = "AHA" + "other insight" = 2
  - Total Insight Either = "AHA" or "other insight" = 1
  - Total Insight All = "AHA" + "other insight" = 1

- Total Understanding – Combining "simple understanding" and "understanding over time" to capture all forms of student comprehension. Computed in 3 ways –
  - Total Understanding Combined = "simple" + "over time" = 2
  - Total Understanding Either = "simple" or "over time" = 1
  - Total Understanding All = "simple" + "over time" = 1

- Total Statistics – Combining "graphical representations" and "statistical symbols or formulas" to capture total presence of statistical representations. Computed in 3 ways –
  - Total Statistics Combined = "graphs" + "symbols" = 2
  - Total Statistics Either = "graphs" or "symbols" = 1
  - Total Statistics All = "graphs" + "symbols" = 1

- Ideal Drawing – Combining variables that instructor would hope to find in an ideal drawing, included "instructor positive", "student positive", "aha", "other insight", and "excitement". Computed in 3 ways –
  - Ideal Drawing Combined = sum of all drawings for a possible score of 5
  - Ideal Drawing Either = presence of any/all of the components yields a score of 1
  - Ideal Drawing All = presence of all of the components yields a score of 1

- Unsatisfactory Drawing – Combining variables that instructor considered as the opposite of an ideal drawing, included "instructor negative", "student negative", and "confusion". Computed in 3 ways –
  - Unsatisfactory Combined = "instneg" + "studneg" + "confusion" = 3
  - Unsatisfactory Either = "instneg" or "studneg" or "confusion" = 1
  - Unsatisfactory All = "instneg" + "studneg" + "confusion" = 1

- Total Confusion – Combining "total insight" and "confusion". This variable was created because it was noticed that students often seemed to include both aspects in their drawing, usually showing confusion before insight. Computed in a single way -
  - Total Confusion All = "total insight" + "confusion" = 1

Analyses:

These two 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.

At present, our long-term investigation into how faculty can learn more about their teaching practice and effectiveness includes the following components:

1. Quantitative course-level summary evaluations tracked over time (as illustrated through Figure 1)
2. Qualitative perceptions of the course experience (as illustrated through Figure 2).
3. Quantitative relationships among the individual-level drawing codes (new results from this project as illustrated through Tables 1-3 and Figures 3-5).

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.

That is, even though the instructor is not necessarily depicted positively and 35% of the drawings depict a level of confusion that is greater than the other courses, there are also more depictions of "aha" experiences (12%) than overall (6%) and 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 3 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). As discussed above, this pattern was a surprise because the course numbered 468 (Introductory Statistics) was predicted to be the one with the greatest confusion. It is the one the instructor admits has been his hardest to teach at a level that was grasped by all students. So, it was a surprise to see that a course that he considers to be more interesting and engaging is the one associated with the greatest level of confusion (469—Intermediate Statistics).

The relationship between confusion and insight is further investigated in Tables 2 and 3 and Figure 4 and 5. 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 4 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.

Table 3 contains similar results for the statistically significant relationship between presence of confusion and computers. This result was somewhat surprising too because computer instruction is provided in the class with usually at least 2 assistants teaching and roaming the desktops as exercises are conducted. All students have hands-on experience and assistance in-class. The interpretation of this finding did not become apparent until the drawings for the 25 students who provided both features were looked at. Figure 5 shows a typical series of scenes where the in-class instruction is relatively clear (even "exciting") but the at-home scene is clearly confusing and frustrating.

4. Quantitative relationships between the drawing codes summarized at the course-level and the course evaluation data file (new results from this project as illustrated through Figures 6 to 10).

These analyses take the form of investigating relationships (through simple correlations and scatterplots) between the class-level student ratings of instruction and the drawing features represented in each of those classes. This was done in the expectation

that different aspects of student drawings would reveal insight into what students think about and re-experience when they evaluate professors in various ways. The points in each of the scatterplots are labeled by sequence number (from 1-99) with higher numbers indicating more recent classes.

Figure 6 contains the relationship between the percent of students in each class who rated the instructor as 'excellent" and who depicted a student positively in their drawing of that class. The positive relation between the two variables (r = .326), showing that the emotional state of the student is important to their classroom experience, suggests a simple instructional principle: happy students make for happy raters.

Figure 7 represents the relationship between "total understanding" and percent excellent ratings. While both types of understanding (simple and over time) showed positive relationships with excellence ratings (r = .439 and .281, respectively), this correlation increased when these variables were combined into total understanding (r = .487). The various ways of combining these variables also revealed that the two types of understanding are different and suggests that students depict themselves learning in various ways.

Figure 8 contains an initially surprising result. There is a negative relationship between the extent to which students strongly agreed that "principles and concepts" had been taught and where the instructor was located in the drawing—in this case, at the board (r = -.342). When these drawings were looked at more closely it was observed that many showed the instructor facing or writing indecipherable text on the board, with his back to the students. This frequent depiction suggests a style of impersonal instruction associated with statistical minutiae.

Figure 8 suggests that the instructor's ratings would likely be low when there are large proportions of drawings with the instructor located at the board. In fact, that is what

Figure 9 shows. The relationship between instructor being located at the board and the percent of unsatisfactory ratings was positive (r = .3). That is, the more frequently the instructor was depicted at the board, the higher the unsatisfactory ratings.

Figure 10 is interesting because it addresses a situation that is frequently debated and investigated. This plots shows that as class size increases, the proportion of students depicting simple understanding in their drawings decreases (r = -.376). The literature is extensive on the relationship between class size and ratings. From the student's perspective, however, smaller classes are associated with more clarification and educational benefit for each student.

In these selected analyses we have explored the relationship between drawing characteristics and course evaluation ratings at the class-level. That is, the relationships have been between averages computed across the students in each class. Simply saying that as the proportion of drawings where a student is depicted positively increase, so does the percent excellent rating does not mean that those who depicted students positively were the ones who submitted the excellent ratings. The next analyses, however, address the data at the student-level.

5. Quantitative relationships between the student-level course ratings and the qualitative drawing codes attached to their personal drawings (new results from this project as illustrated through Figures 11-13)

With the recent linking of individual student drawings to individual student course evaluations we can begin looking at the relationship between individual student drawings and individual course evaluation ratings. For example, we can now explore how specific students who depict "Aha experiences" in their drawings actually rate the professor and

class using the traditional course evaluation. This is particularly important since many of the more interesting drawing characteristics occur infrequently (such as Aha's). When an infrequently occurring characteristic is examined at the aggregate (course) level the relationship of that characteristic is obscured by the large number of drawings for that course that do not depict that feature. Thus, it is plausible that the examination between and among these infrequently occurring drawing characteristics is not well served by analyses at the (averaged) course level, but rather at the student level. Examining the relationship between the individual student drawings and the individual course evaluation allows a much closer examination of the validity of the student drawings.

Recall that Figure 6 shows a positive relationship between excellence ratings and the proportion of drawings with students depicted positively. Did students who drew those positive depictions actually give higher ratings in classes than those who did not draw such student depictions? Figure 11 shows the relationship between the actual ratings provided by students and whether or not they drew students depicted positively. Only two classes are represented here, one section of 469 (Intermediate Statistics) and one section of 216 (Research Methods). Although there are not as many students in this graph as there will eventually be (since now all 587 student evaluation sheets must be pulled and the student-generated ID linking the evaluation form with the drawing has to be matched and the data then entered) it can be seen that those who did depict students positively did in fact rate the instructor higher than those whose drawing did not have such a feature.

The one final way these data need to be understood is represented in Figures 12 and 13. Figure 12 contains the same variables but for the 469 Intermediate Statistics class only. Figure 13 contains the corresponding graph for the 216 class. It is apparent that the overall pattern seen in Figure 11 across the classes does not hold at the specific class level. In 469, the students who depicted students positively did not submit ratings higher

than those who did not draw students positively—basically, both groups rated the class low. In 216, however, the students who depicted students positively did submit the higher ratings.

In addition to these standard type analyses of numeric codes there is one final aspect of this project that still has not started. That is the scanning and analysis through qualitative means of the drawings.

6. Development and analysis of a HyperResearch qualitative data file (e.g. which courses tend to show more expressions of negative experiences—"sharks in the water"—versus those courses with more positive experiences—"successfully climbing a mountain"?)

Conclusion

As educational researchers, we are well aware of the negative attitudes and belief systems that many students bring as extraneous baggage to applied statistics courses. It is clear that these attitudes and beliefs may interfere with the learning of the material. These attitudes and beliefs are not necessarily well communicated on standard course evaluation forms. For example, was the class rated low because it was poorly taught or because the student never overcame a dread of the material. The least informative evaluation occurs when the student gives a low rating with no explanation. Even more of a problem, from the instructor's standpoint, is the impossibility of students using Likert responses to articulate surprising changes in their negative attitudes and beliefs.

The systematic analysis of these course evaluation drawings, combined with the opportunity to link them to the standardized ratings from the same students, has created a unique approach to the assessment, interpretation, and evaluation of instructor and course effectiveness. Drawings depicting scenes with "ah-ha's", dead-fish expressions,

confusion), light bulbs turning on, gibberish on the blackboard, celebrations on a mountain top, students sleeping or day dreaming, and tear-drops on an anguished face effectively communicate what students feel in classes. Those expressions lead to an intensity of self-reflection about practice that is impossible to ignore and is virtually impossible to experience with any standardized course evaluation now in use.

The overall objective of this long-term, continually evolving project is to better understand instruction from multiple perspectives. The more information an instructor can gain from students about teaching effectiveness, the more opportunities the instructor may have to improve instruction.

How might one think about excellence in teaching? Through students who depict themselves, others, and the instructor positively; frequent scenes of "aha" and insight; learning occurring over time, frequent small-group interactions; explanations involving principles and concepts; and small class sizes.

REFERENCES

Garfield, J. (1993), Teaching Statistics Using Small-group Cooperative Learning, *Journal of Statistics Education* [Online], 1(1).

Garfield, J.B. (1994), Beyond Testing and Grading: Using Assessment to Improve Student Learning, *Journal of Statistics Education* [Online], 2(1).

Giraud, G. (1997), Cooperative Learning and Statistics Instruction*, Journal of Statistics Education* [Online], 5(3).

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.

Haney, W., Madaus, G. and Kreitzer, A. (1987), Charms talismanic: Testing teachers for the improvement of American education. In E. Rothkopf (Ed.) *Review of Research in Education*. Volume 14, pp. 169-238.

Jackson, P. (1968). *Life in classrooms*. New York: Holt, Rinehart & Winston.

Lipsky, M. (1980). *Street-level bureaucracy : Dilemmas of the individual in public services*. New York : Russell Sage Foundation.

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

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

Ludlow LH (**in press**). Making sense of course evaluations: A longitudinal analysis of teaching. *Practical Assessment Research and Evaluation*.

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. & 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, 2003.

Rumsey, D.J. (1998), A Cooperative Teaching Approach to Introductory Statistics, *Journal of Statistics Education* [Online], 6(1).

Sowey, E.R. (1995), Teaching Statistics: Making it Memorable, *Journal of Statistics Education* [Online], 3(2).

Sternberg, R. J & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher* (24:6), pp. 9-17.

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

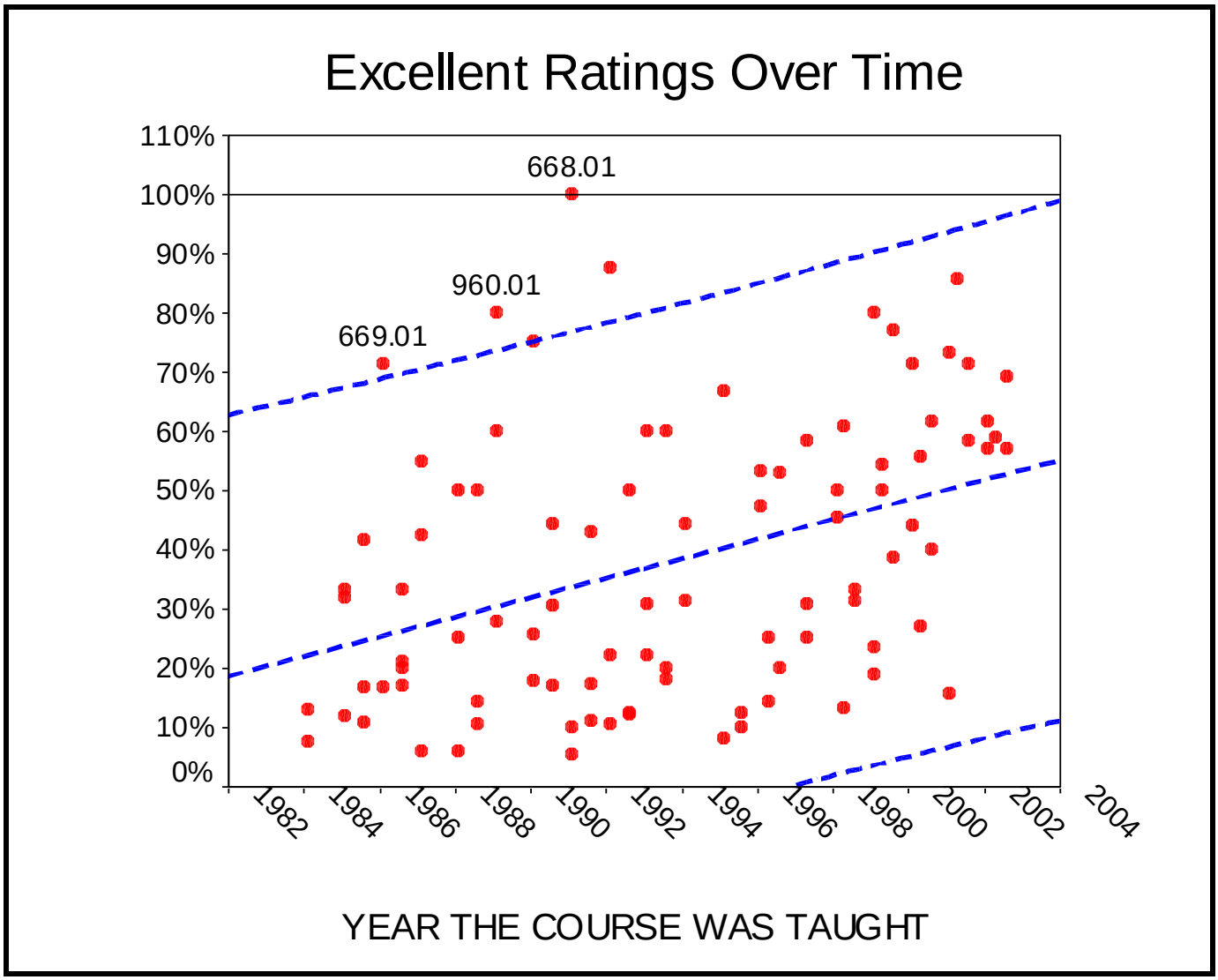
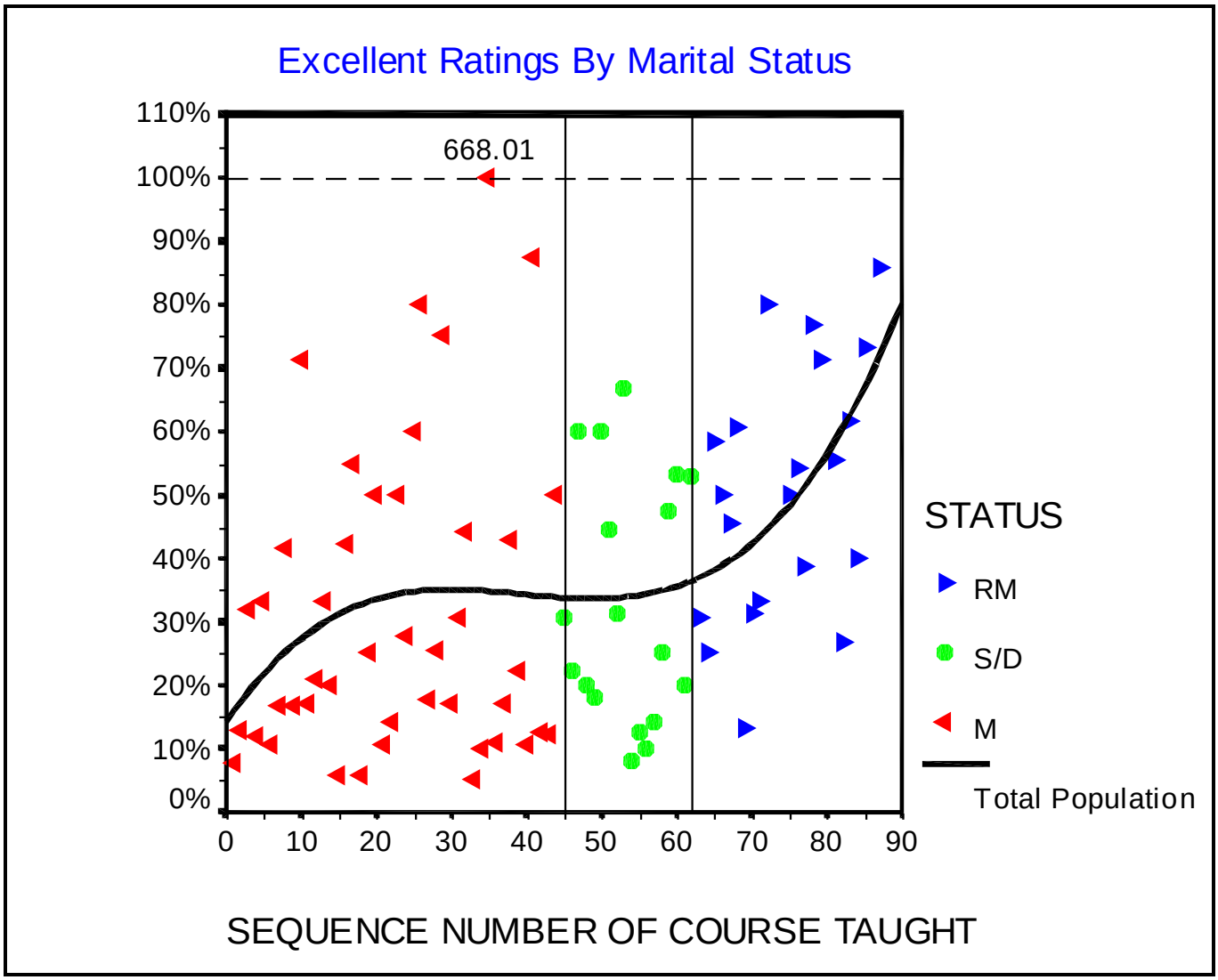

Excellent Ratings Over Time
110%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
668.01
960.01
669.01
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
YEAR THE COURSE WAS TAUGHT

Figure 1b. How do ratings look across a significant personal factor?



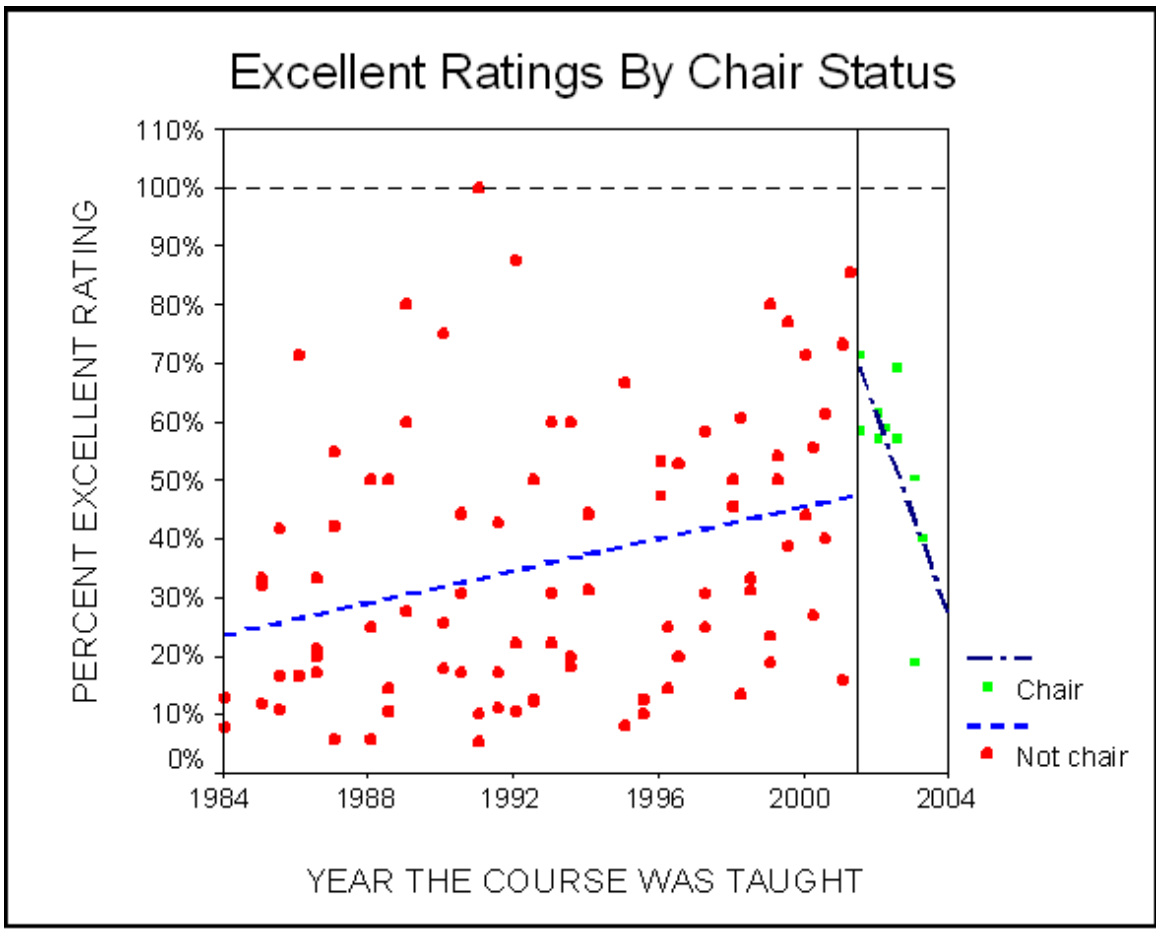
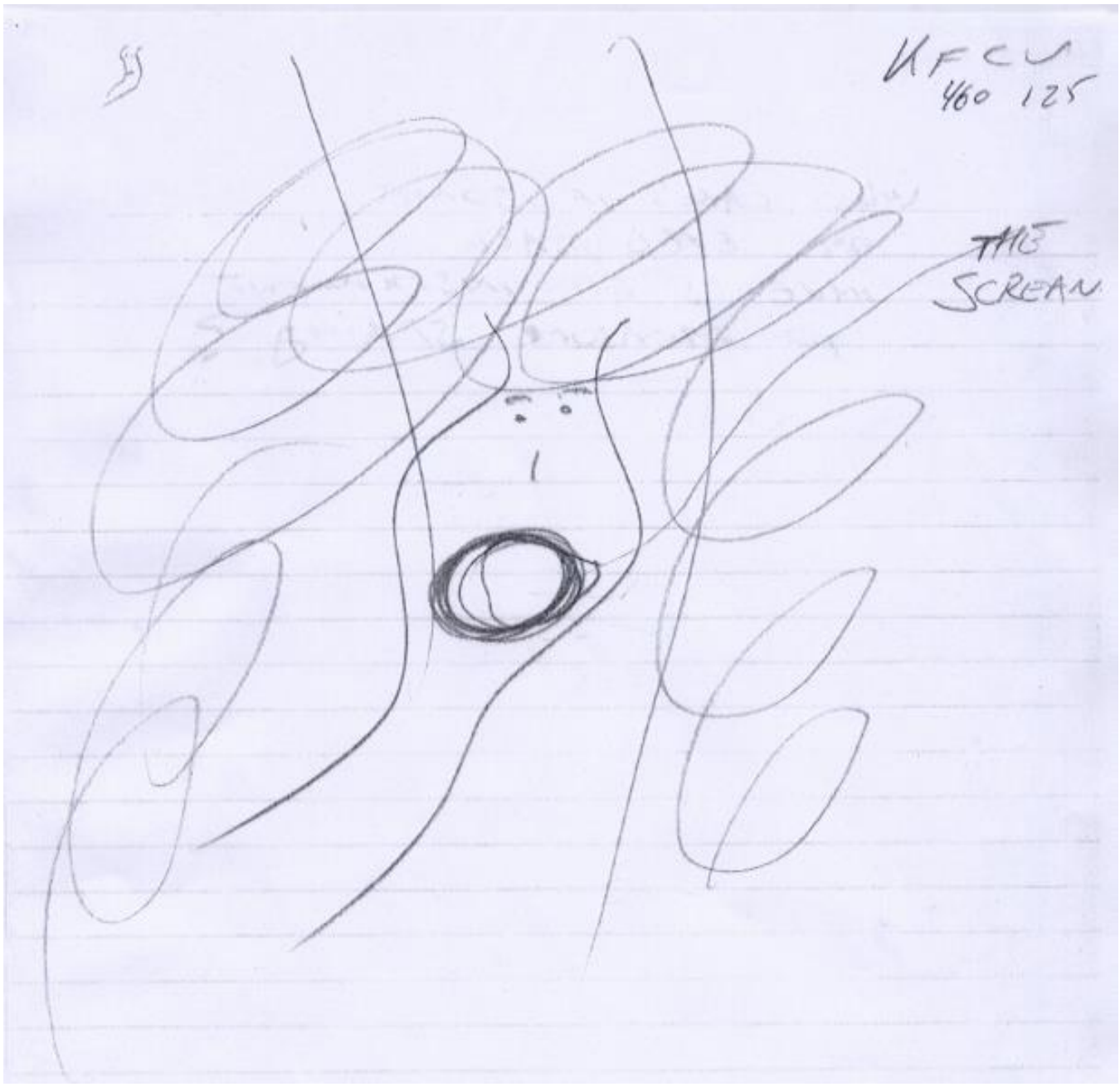
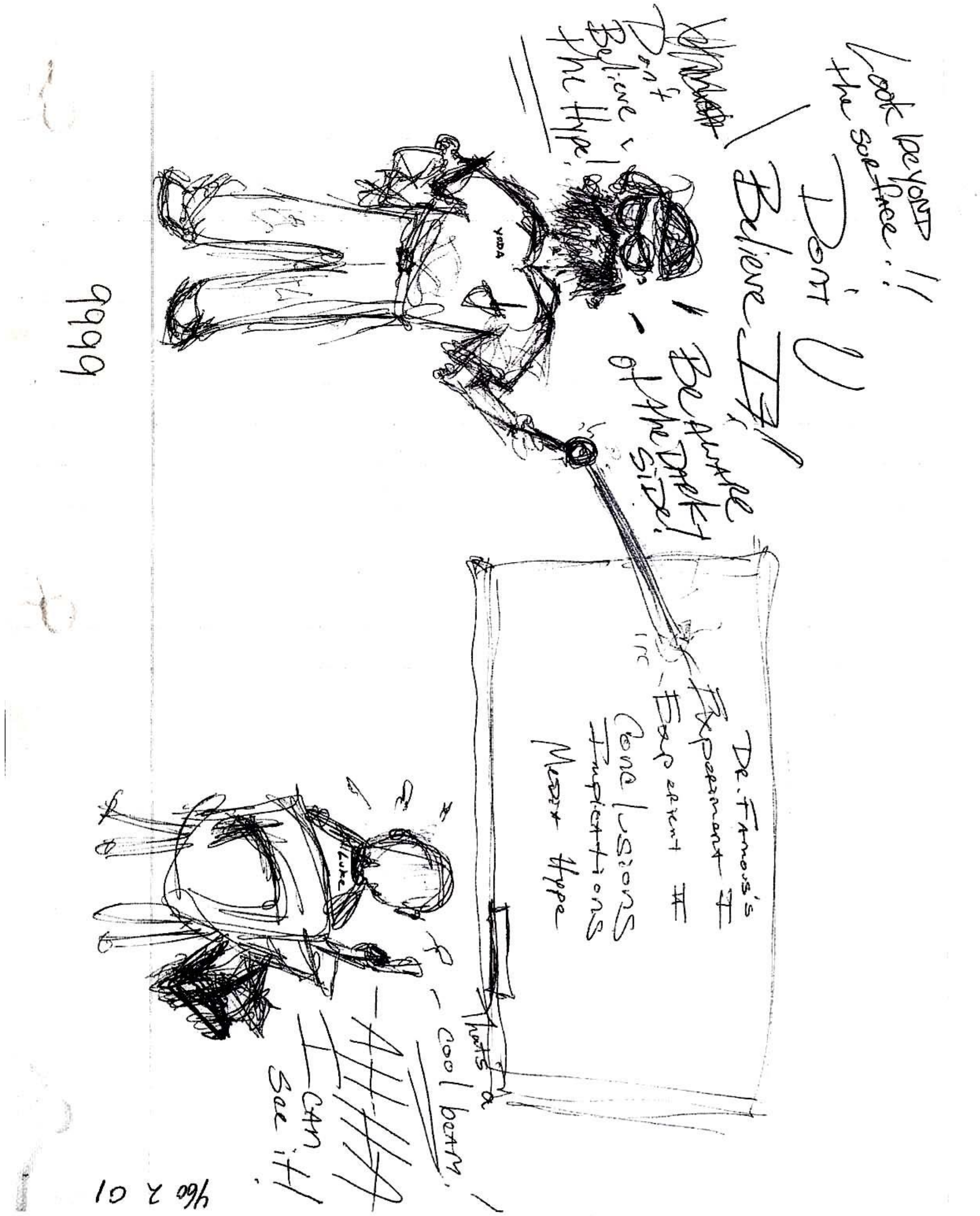

Excellent Ratings By Chair Status
PERCENT EXCELLENT RATING
110%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1984
1988
1992
1996
2000
2004
YEAR THE COURSE WAS TAUGHT
Chair
Not chair

Figure 2: One student's experience in a research methods class



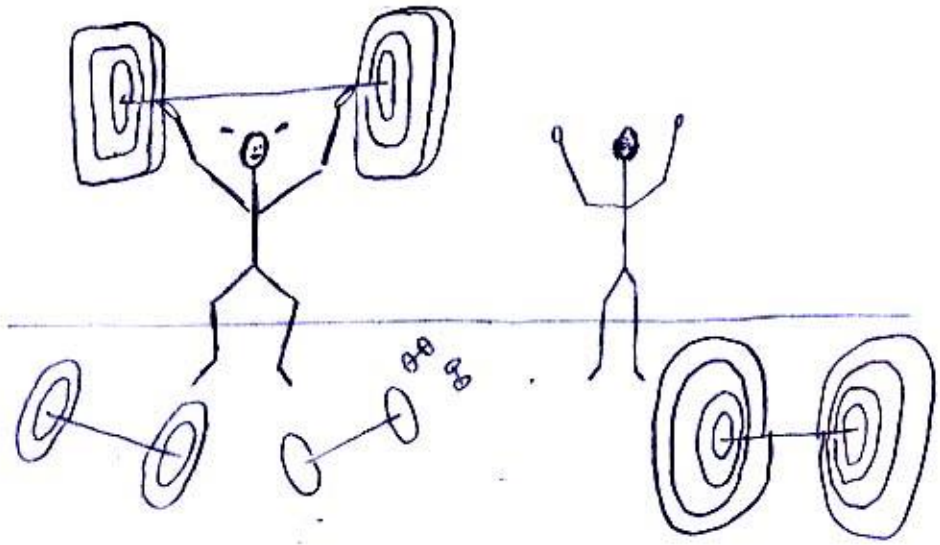

Look beyond!! the surface!!
Don't U Believe IT!
Be aware of the Dark Side!
Don't Believe the Hype!
YODA
Dr. Famous's
Conclusions
Implications
Media Hype
Luke
That's a cool beam.
AHHH I can see it!

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
L

66901021

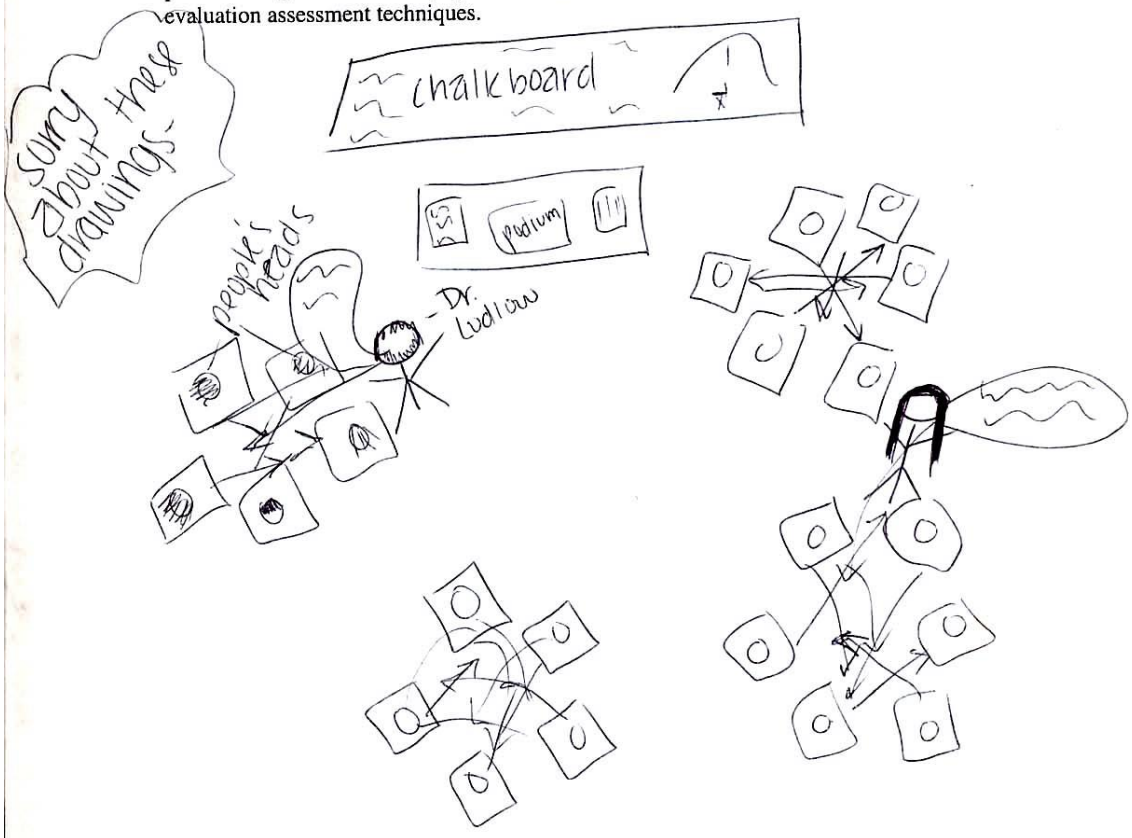
Table 1: Percent of drawings representing various classroom experiences

	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

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.

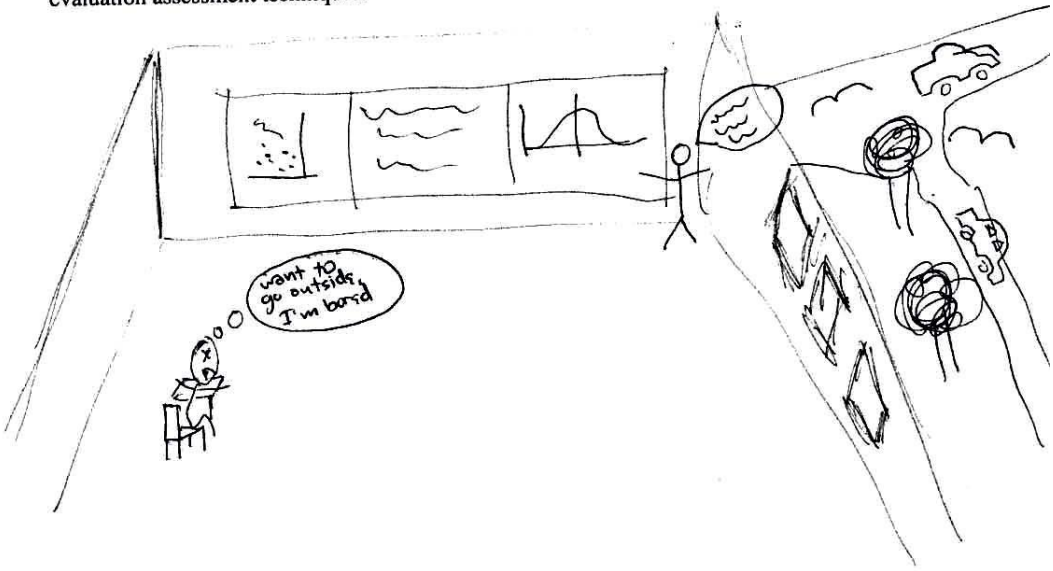


2/6/2000

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: PY216: Spring 2001: BC: Ludlow

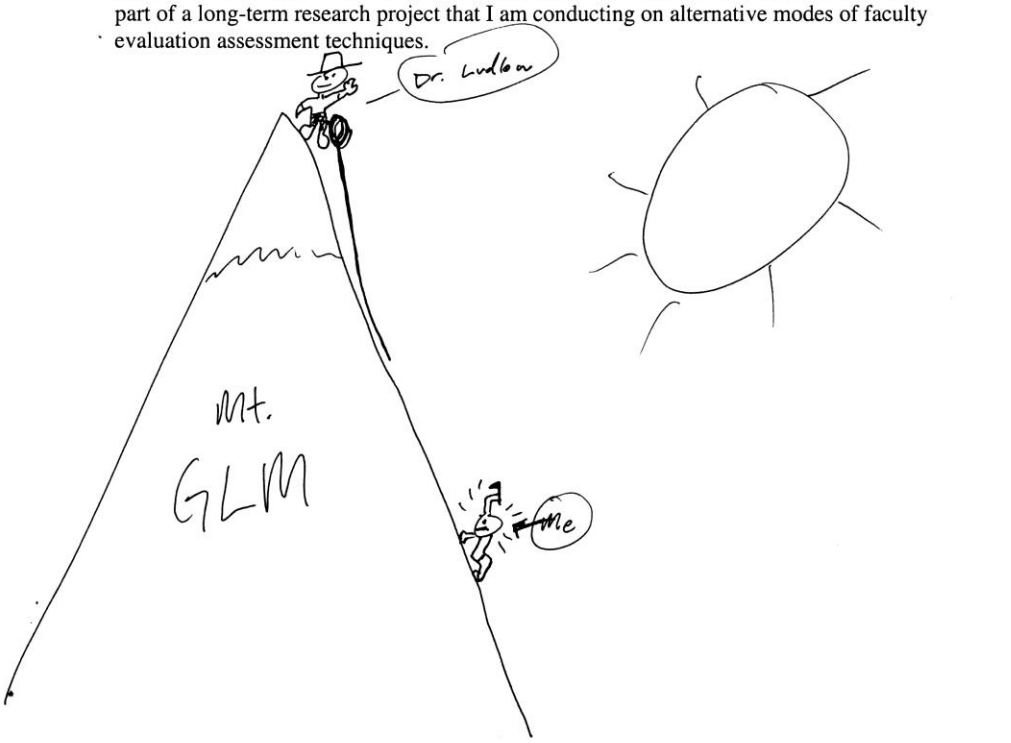
18
2160101H

4673

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.



GLM: ED/PY667: FALL 2000: BC: Ludlow

66702011

Figure 3: Relationship between level of confusion and course

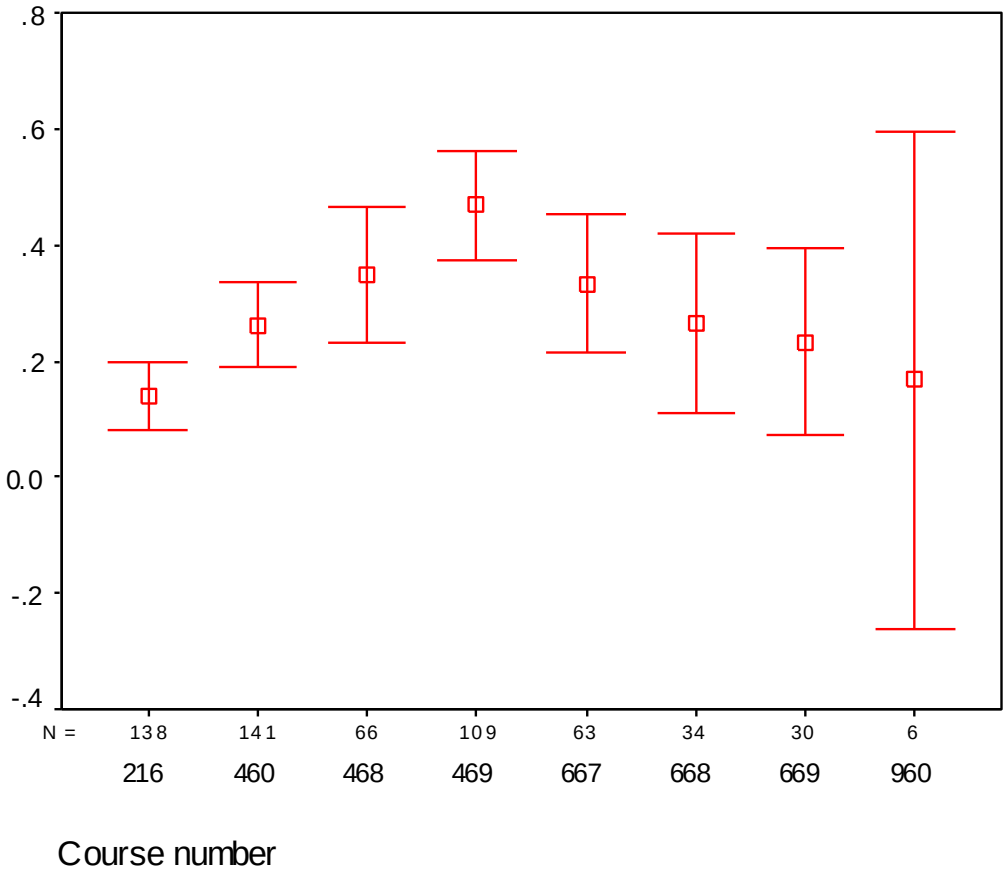


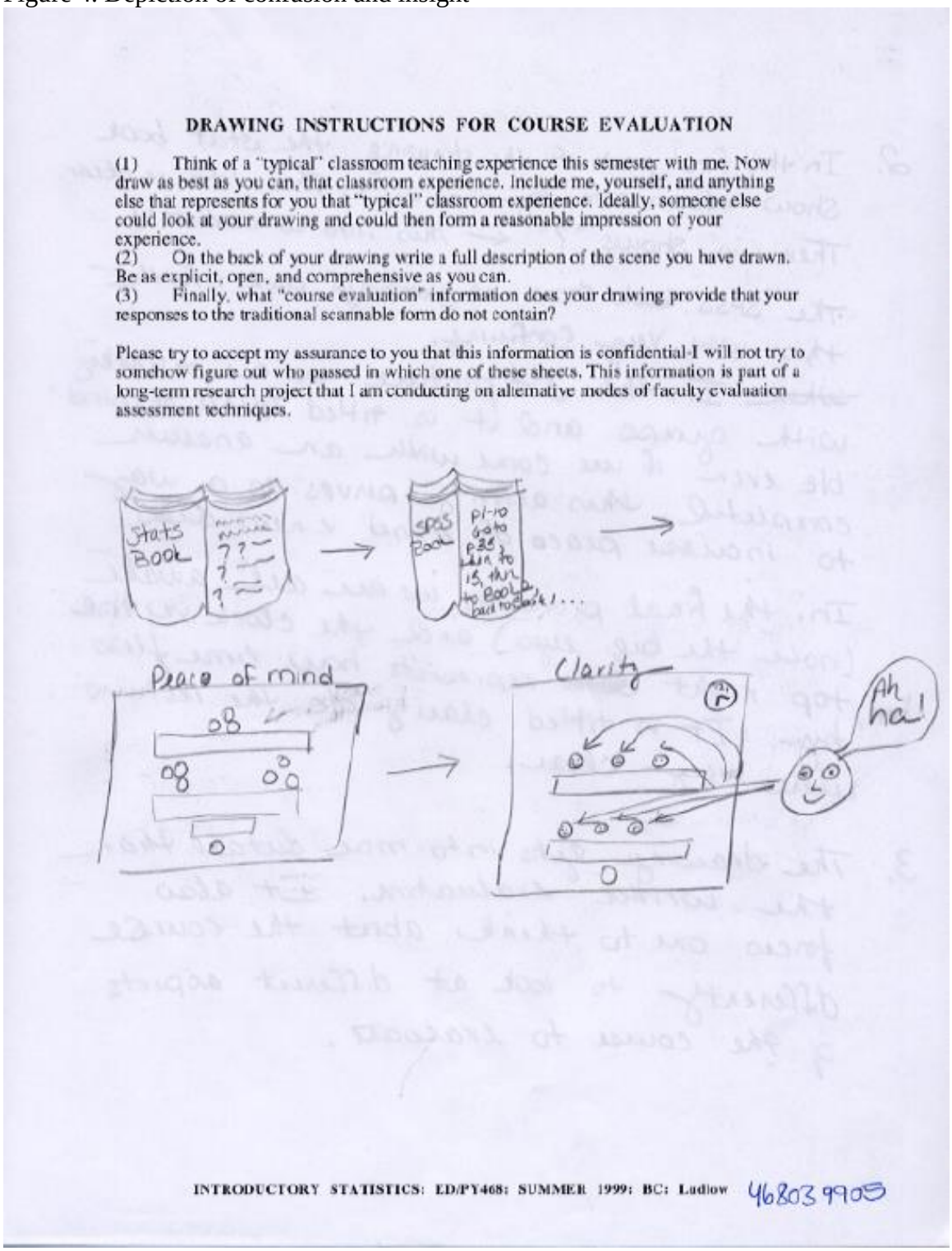
Table 2: Relationship between presence of confusion and sudden insight

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

Figure 4. Depiction of confusion and 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 468039905

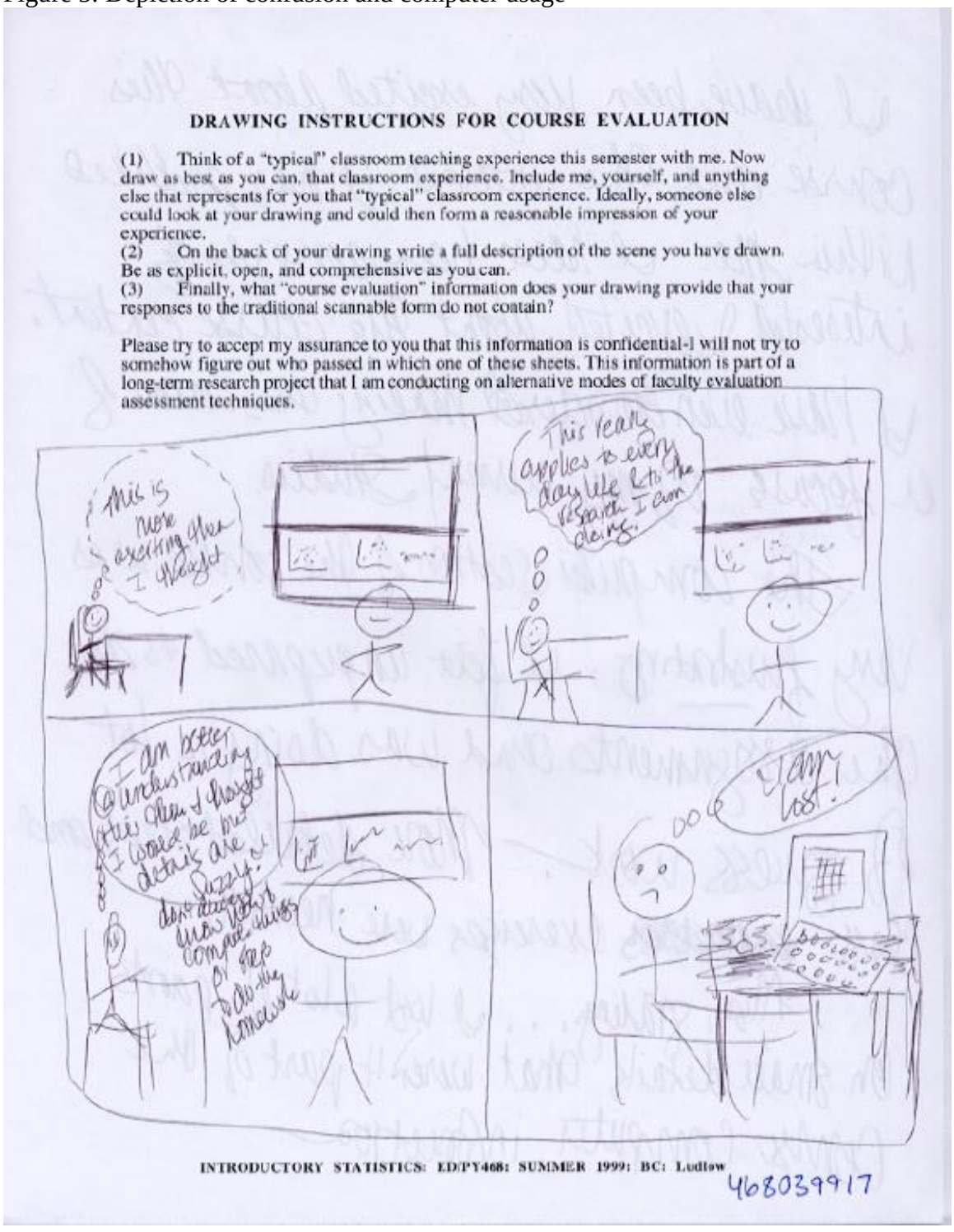
Table 3: Relationship between presence of confusion and computer depiction

Computer depicted * confused/overwhelmed/lost Crosstabulation

			confused/overwhelmed/lost		
			absent	present	Total
Computer depicted	absent	Count	392	143	535
		Expected Count	381.9	153.1	535.0
		Std. Residual	.5	-.8	
	present	Count	27	25	52
		Expected Count	37.1	14.9	52.0
		Std. Residual	-1.7	2.6	
Total		Count	419	168	587
		Expected Count	419.0	168.0	587.0

chi-square=10.6, p=.001

Figure 5: Depiction of confusion and computer usage



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

468039917

Figure 6: Relationship between ratings and affect of students

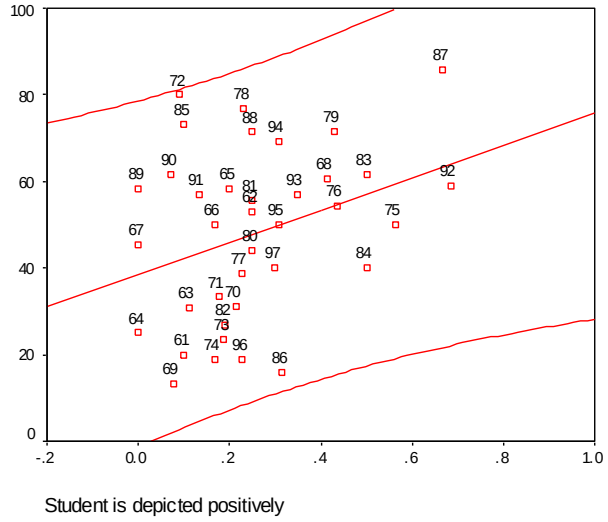
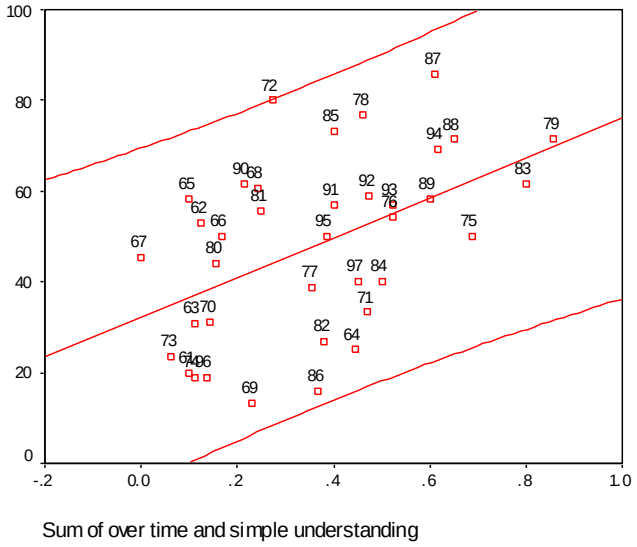


Figure 7: Relationship between ratings and "total understanding"

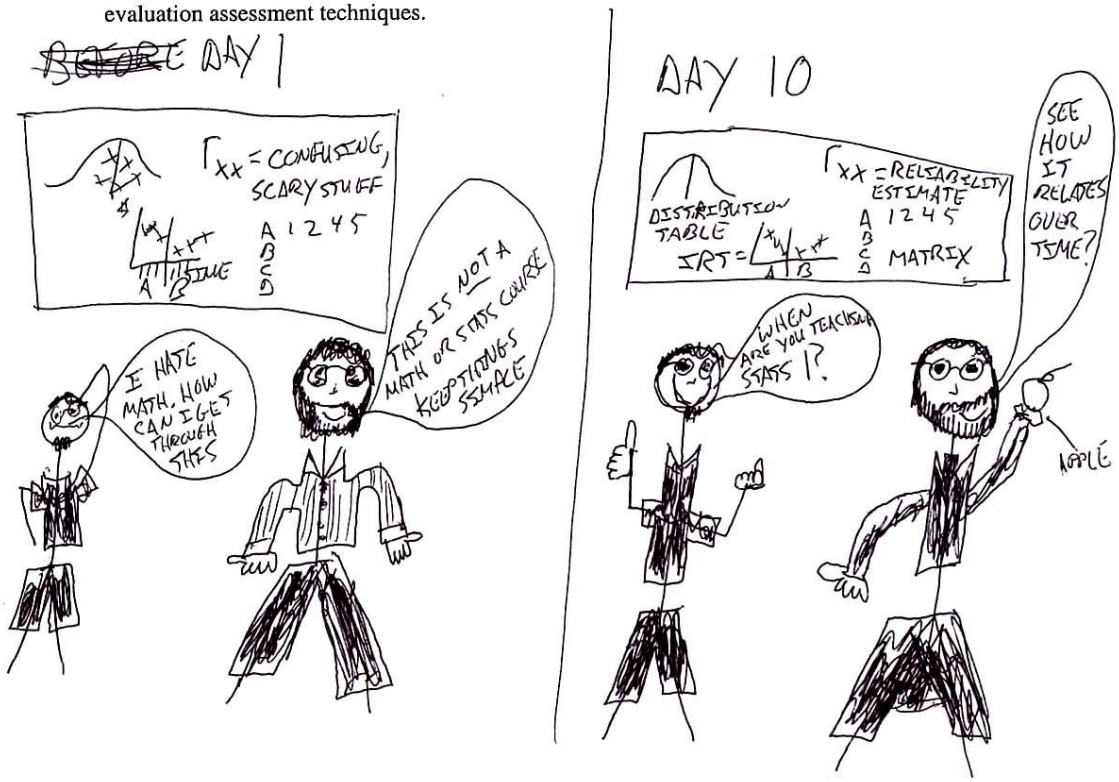


NEON

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

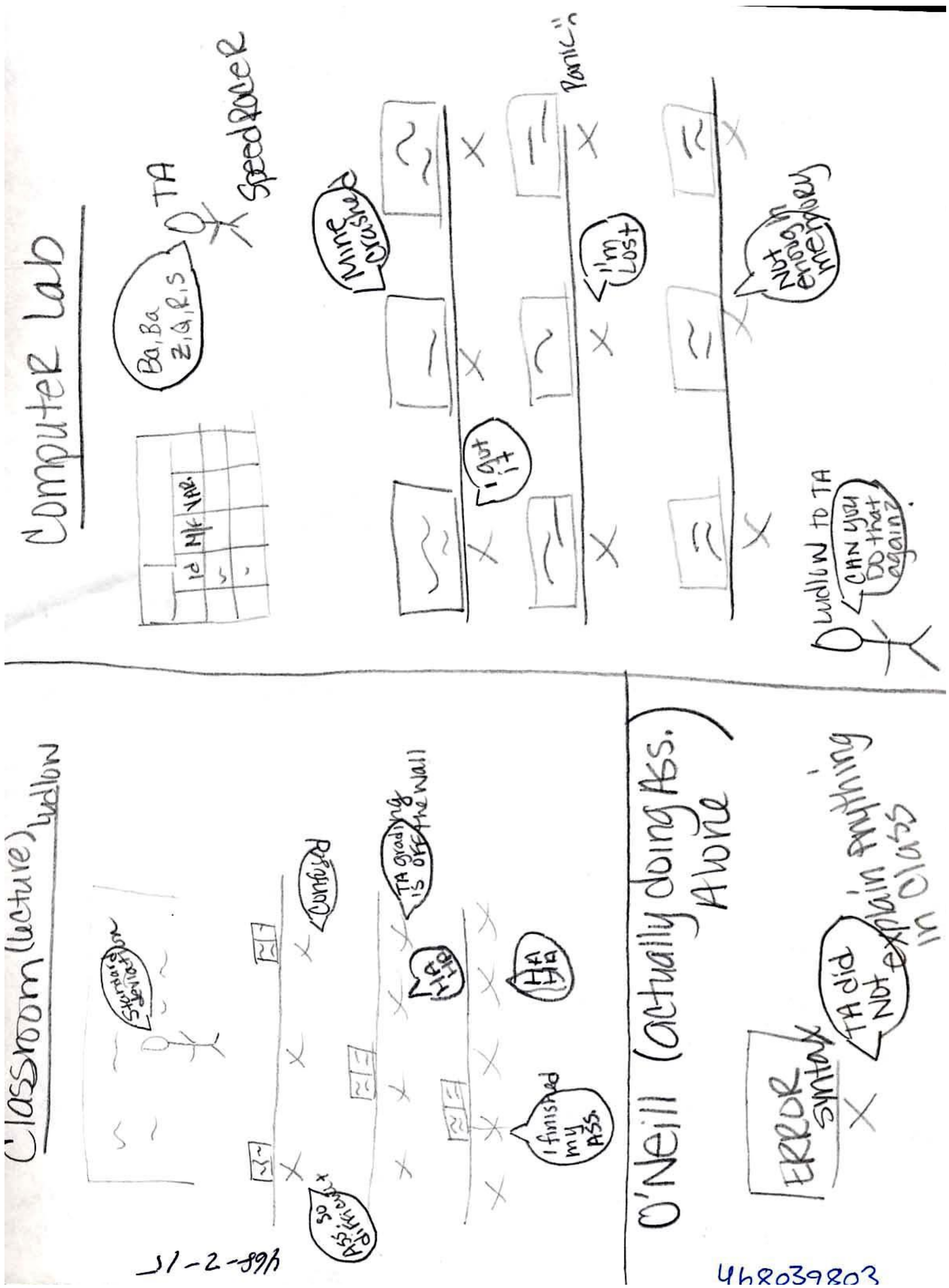

Computer Lab
Ba, Ba Z, Q, R, S
TA
Mine crashed
Panic!!
I'm Lost
Not enough memory
I got it
id M/F VAR.
Ludlow to TA
Can you Do that again?
Classroom (lecture) Ludlow
Standard deviation
Confused
TA grading is off the wall
HA HA
HA HA
I finished my ASS.
So difficult ASS.
O'Neill (actually doing Ass. Alone)
ERROR syntax
TA did Not explain anything in class

Figure 8: Relationship between principles and concepts and instructor location

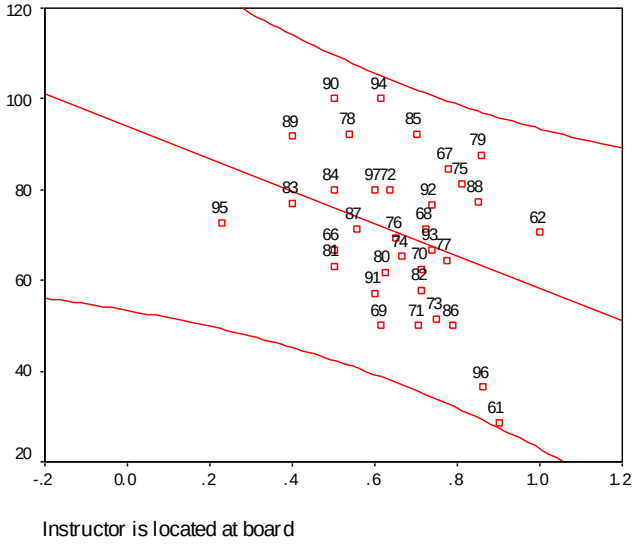
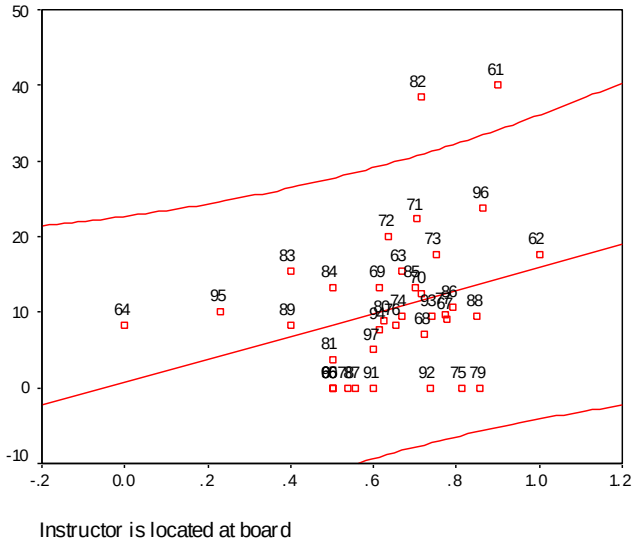
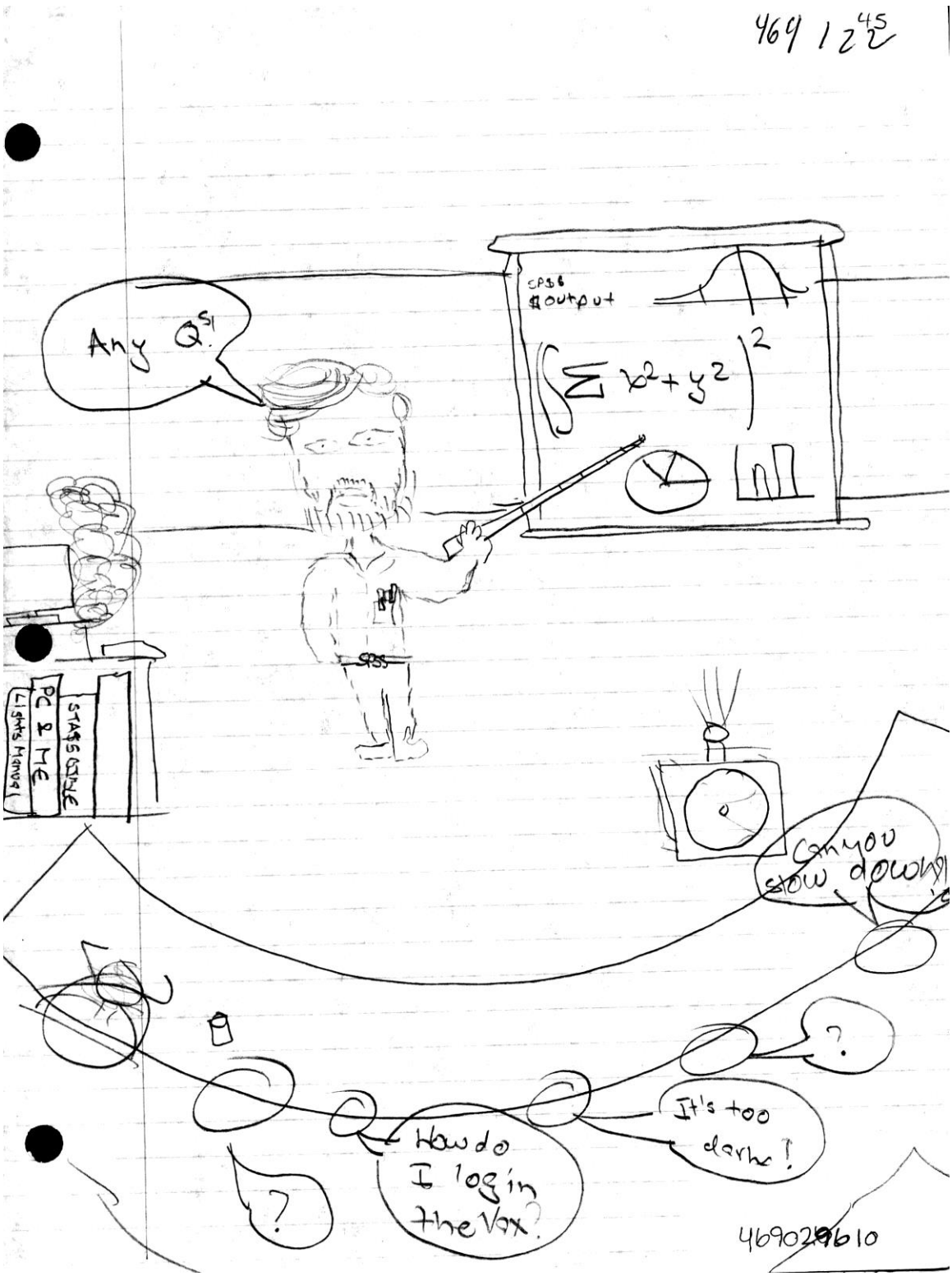


Figure 9: Relationship between unsatisfactory ratings and instructor located at the board

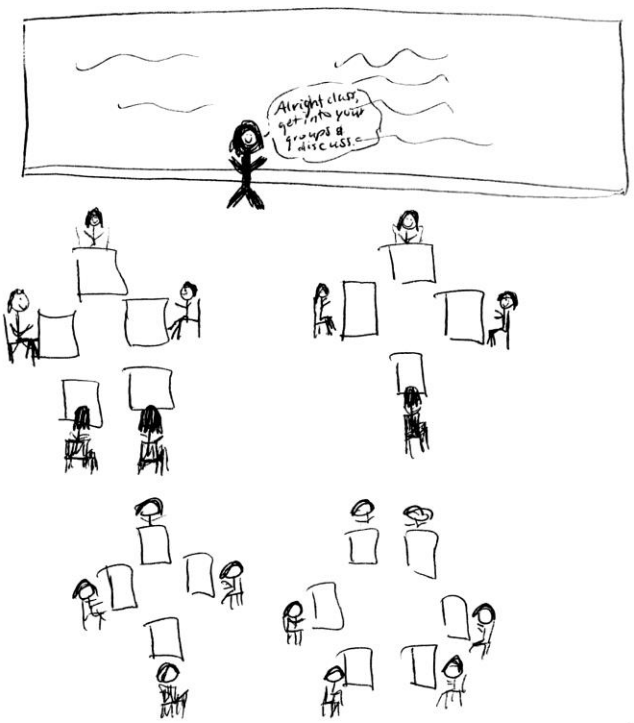



469 12 45
Any Q's!
SPSS Output
$\left(\sum x^2+y^2\right)^2$
SPSS
PC & ME
Lights Manual
STATS CLASS
Can you slow down!
?
It's too dark!
How do I log in the Vax?
?
4690296 10

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?

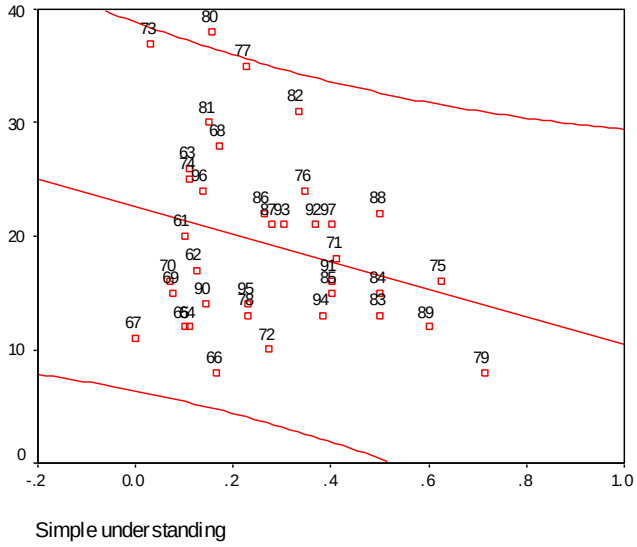
Please try to accept my assurance to you that this information is confidential--I will not try to somehow figure out who passed in which one of these sheets. This information is part of a long-term research project that I am conducting on alternative modes of faculty evaluation assessment techniques.



Research Methods: PY216: Spring 2001: BC: Ludlow

16
26010107

Figure 10: Relationship between class size and understanding of the material

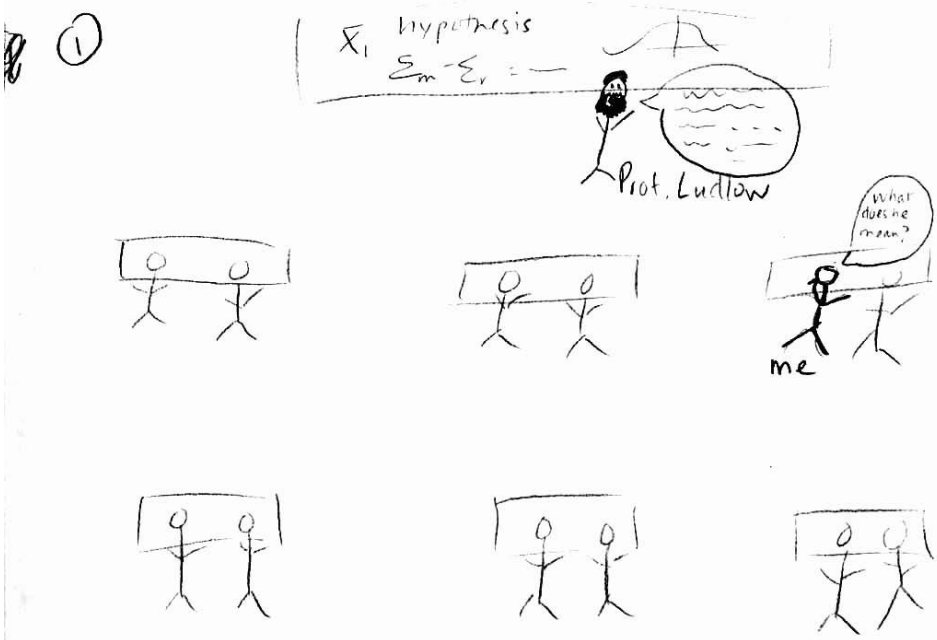


216-1-31

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.



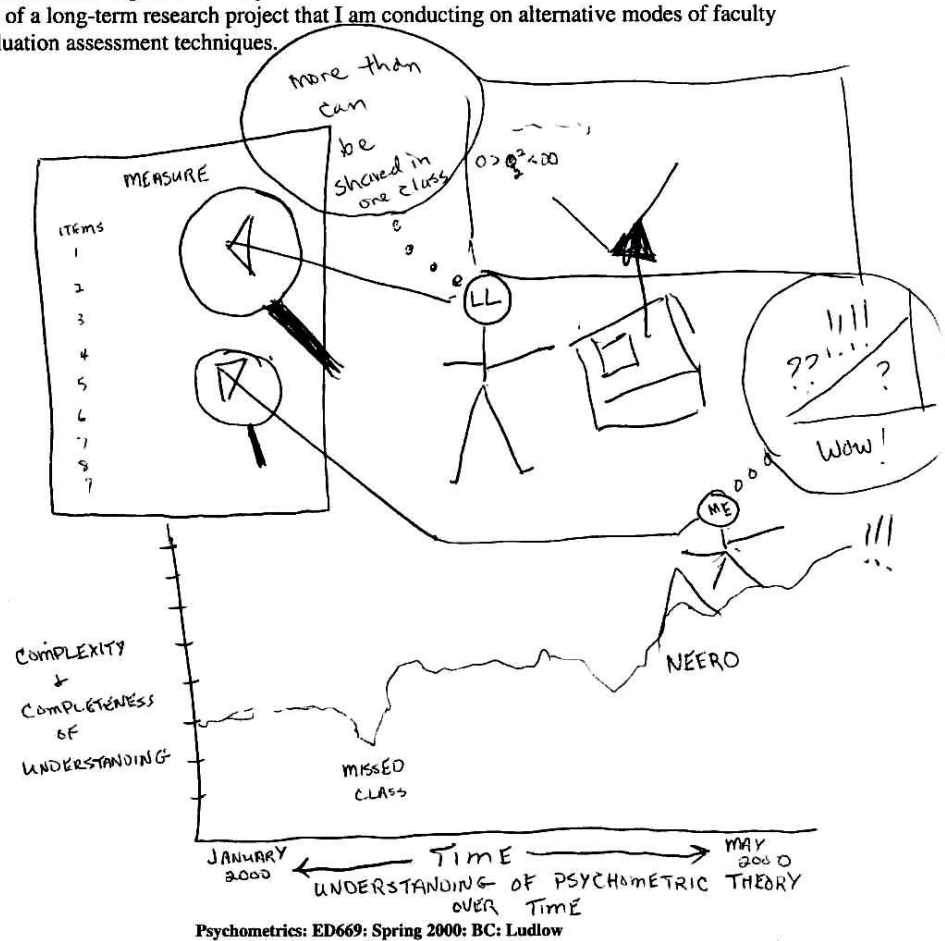
Research Methods: ED/PY216-TTH: Sprin 1999: BC: Ludlow

216019922

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 2000: BC: Ludlow

669010007

Figure 11: Relationship between student depiction and instructor rating (469 & 216 combined)

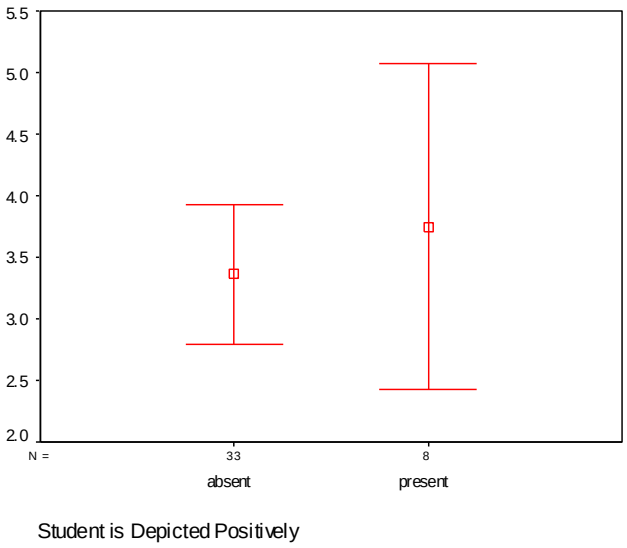


Figure 12: Relationship between student depiction and instructor rating (469 only)

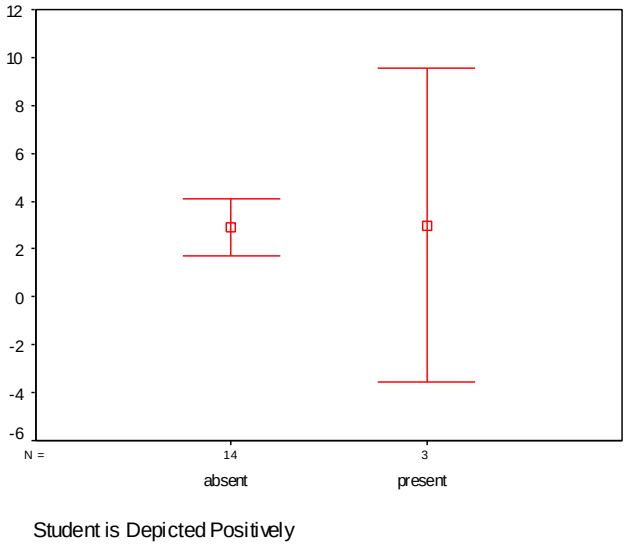
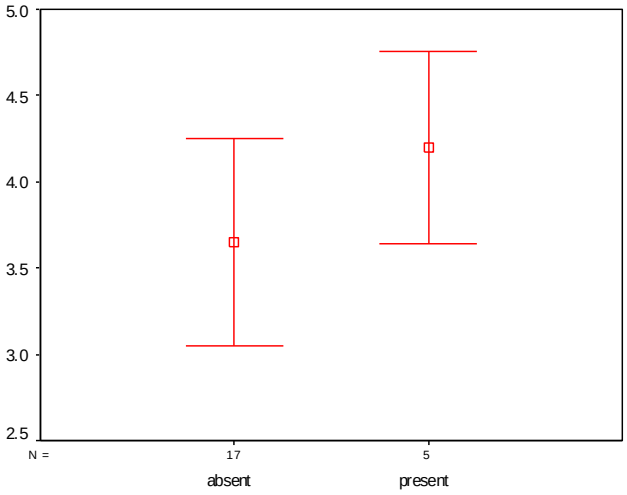


Figure 13: Relationship between student depiction and instructor rating (216 only)



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.

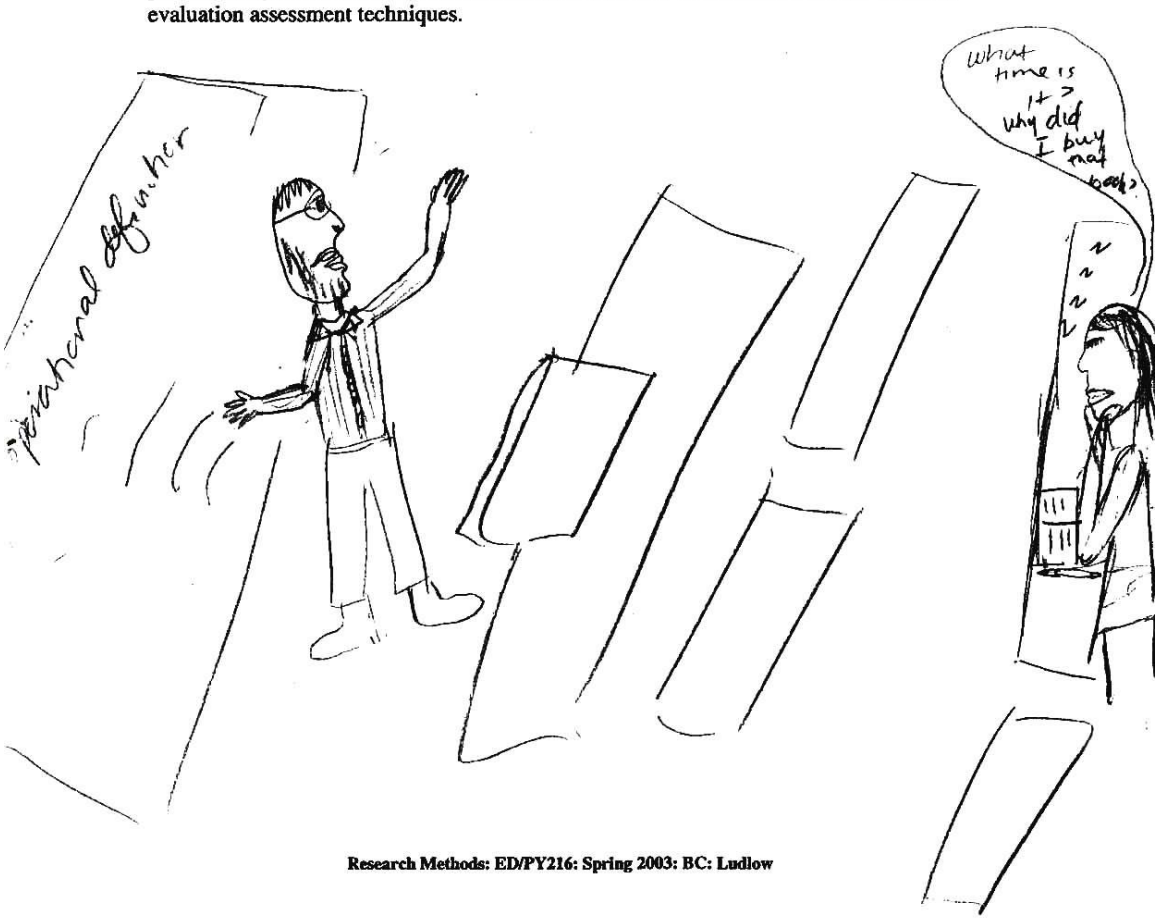
Research Methods: ED/PY216: Spring 2003: BC: Ludlow

2442

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.



21601030?