6 Conducting and Reporting Original Scientific Research: Anderson's Biology Class

Virginia Johnson Anderson
Towson State University

Barbara E. Walvoord
Loyola College in Maryland

As in the other chapters, here we present and analyze students' difficulties, but more importantly we describe and document what happened when a teacher-researcher, after identifying these difficulties through systematic observation of her classroom, changed her teaching methods.

Like the other classroom chapters, this one begins by describing the teacher's expectations. We note particularly the similarities and differences between Anderson's and the other three teachers' expectations for the professional-in-training role and for good/better/best reasoning. Then we describe Anderson's teaching methods for 1983—the first of the two sections of her class that we studied. Next, we discuss the difficulties that arose, how Anderson changed her teaching methods to address those difficulties, and the improvement we found in the research and the papers of students in 1986—the second section we studied.

Though we used outside raters to establish that the papers of the second class had improved, this chapter is not a report of a scientific experiment to prove the efficacy of Anderson's procedures. It is, like our others, a naturalistic study of events in a particular classroom, and it takes place within the theoretical framework and research approaches outlined in Chapter 2. This chapter is a story, actually—the story of how a biology teacher and a collaborating colleague observed her class, identified what she took to be the difficulties that had arisen, and then shaped teaching methods to address those difficulties. Particularly Anderson tried to provide *concrete experiences* through which her students could learn to use the scientific method—in other words, to teach procedural knowledge procedurally. It is the story of Anderson's attempt to see whether the improvement she thought had resulted from her changes could be recognized by others.

Anderson: Background

For Virginia Johnson Anderson, the process of change began in the fall of 1981 in a writing-across-the-curriculum workshop for college teachers. She recalls in her own words:

I would love to tell you that it was great insight on my part or great recruitment by the writing-across-the-curriculum (WAC) movement that brought me to study how students write in science, but it wasn't. I was pregnant. Not just pregnant, but over 40, with two years left to finish my doctoral dissertation on scanning electron microscopy, juggling a full-time, tenured, biology assistant professorship and two children 12 and 15, thrilled to death—pregnant! My motive for taking the WAC workshop was to get the "release time and/or other recognition" that my university was offering to lure faculty into the workshop.

Well, my darling son was born in February and bundled off to WAC workshops once a week from March through May. Also, he slept, almost unnoticed, in a carrier on my back as I delivered the last ten Biology 101 lectures for Spring 1982. Although I never got the released time and never figured out what the "other recognition" was, I knew that what I had learned about the writing process had profoundly changed my professional life.

Before the WAC workshop, I had told myself that students wrote poorly in their biology courses because they didn't spend enough time doing it and/or they had not been adequately trained to write in English 101. Once the WAC workshop had dispelled these myths, I wanted to know more about how students wrote in science. I shared this interest with Barbara Walvoord, one of the coleaders of the workshop. We decided to collaborate.

ANDERSON'S EXPECTATIONS

THE SCIENTIST-IN-TRAINING ROLE

We (Anderson and Walvoord) selected one assignment in Biological Literature as the focal point of our research. Since the 1983 class had only 13 students enrolled, we used the entire class for all analyses, rather than choosing a focus group as in Sherman's, Breihan's, and Robison's classes.

Biological Literature was a one-semester, three-credit course offered by the biology department at Towson State University (TSU). It enrolled juniors and seniors. Although it did not count for credit for the biology major or minor, Biological Literature did fulfill the university requirement for an advanced writing course, usually taken in the student's major discipline. To enroll in Biological Literature, students had to have earned a grade of "C" or better in the freshman composition course and completed ten or more semester hours in biology. Characteristics of the class are on p. 18.

The professional-in-training role (pp. 8–9) that Anderson expected of her students was the scientist. Biological Literature was designed to include many types of writing that scientists do (see Pechenik, 1987, for a survey). Thus Anderson's assignments in both 1983 and 1986 included:

- Paragraphs and short papers summarizing laboratory results, procedures and equipment descriptions; defining and/or describing specimens; comparing and/or contrasting taxonomic groups
- Short written exercises on *BioAbstracts, Science Citation Index,* ERIC, and/or *Index Medicus,* all of which are indexes to science literature
- Informative abstracts of scientific journal articles
- Written text to accompany graphs, illustrations, micrographs, etc.
- Short evaluations of biology seminars, lectures, or texts to simulate short position papers by scientists
- Letter to the Editor for a scientific journal
- A short library research paper designed to give the student experience in researching scientific literature, specifically *BioAbstracts* and *Science Citation Index,* or a student grant proposal (1986 innovation)
- An original scientific research report designed to give the student an opportunity to conduct and report original scientific research

Our data are related to a single assignment—the last one listed above. The longest and most demanding of them all, this assignment spanned ten weeks.

In constructing the original research assignment, Anderson was influenced by her perception of what her students would need if they were to succeed in routine research and development laboratories

(R&D labs in scientific jargon). This is the type of job that many biology graduates experience during their first years of employment. In the Baltimore area, TSU biology graduates are often employed by Noxell (manufacturers of Noxzema facial cream), Doxsee Food Corporation, and McCormick (spices). As entry-level scientists, TSU graduates might work on research questions such as which purple eye makeup pigment is easiest to remove, which milk product is most stable, or which grind consistency of pepper is most aromatic.

Similarly, Anderson's assignment required students to conduct original scientific research in which they compared two commercially available products to discover which was "better." Students were expected to prepare five pages of text in the scientific report format and to include a minimum of three appropriately labeled graphics. They were to address an audience of their classmates, to whom they were also to give oral reports of their findings.

We have noted in the other three classes how the language of the classroom helped to shape roles and students' reasoning. Strikingly characteristic of Anderson's classroom, as Walvoord observed it, were collaboration and scientific problem solving. The class of 13 students met in a small science laboratory. Seated around the lab tables in groups of four, they easily formed small working groups, and Anderson frequently broke them into groups for interactive work during the class period. They resembled scientific teams working in a scientific environment. The tasks Anderson gave them were to solve scientific problems and/or to question one another's scientific methods or ideas, not in the spirit of confrontation so much as in the spirit of helping one another. We will see some of these interactive, small-group activities later in this chapter, and we will see how, after our analysis of the 1983 data, Anderson changed the nature of some of these small-group activities to make them more effective.

Anderson parked her purse and auxiliary bags and boxes of equipment up front at the instructor's lab table, where she conducted demonstrations if she needed the equipment. Otherwise, her style was to move fluidly among the groups in the manner of a project director or senior scientist directing scientific teams. Her participation in the groups struck Walvoord as quite different from what she herself might experience in a composition class, where students are revealing their own perceptions and stories about which they are the only experts, and into which the teacher may hesitate to intrude. Anderson dipped into one group after the other with energy and direction. These were groups whose procedures were open to public scrutiny and accountable to the scientific community. Often a group would motion Anderson to their table to consult with her about a procedure or problem. Anderson's expectations and the dynamics of her classroom, then, were oriented toward helping students to become scientists—Anderson's version of the professional-in-training role.

ANDERSON'S EXPECTATIONS FOR GOOD/BETTER/BEST REASONING

The good/better/best question that Anderson's assignment addressed was "Which of two consumer products is 'better'?" As in the other disciplines, Anderson's students had to perform the five tasks of good/better/best reasoning (p. 12) in order to answer it. Figure 6.1 shows how the models for good/better/best reasoning differed in each class.

Anderson saw herself as enforcing the expectations for scientific experimentation and scientific writing that were common to the scientific community. We discuss her expectations here under two headings: Expectations for using the scientific method and expectations for organizing the research within the scientific format.

Anderson's Expectations for Using the Scientific Method

To arrive at their conclusion about which product was better, students had to use the scientific method; i.e., they had to formulate a hypothesis,

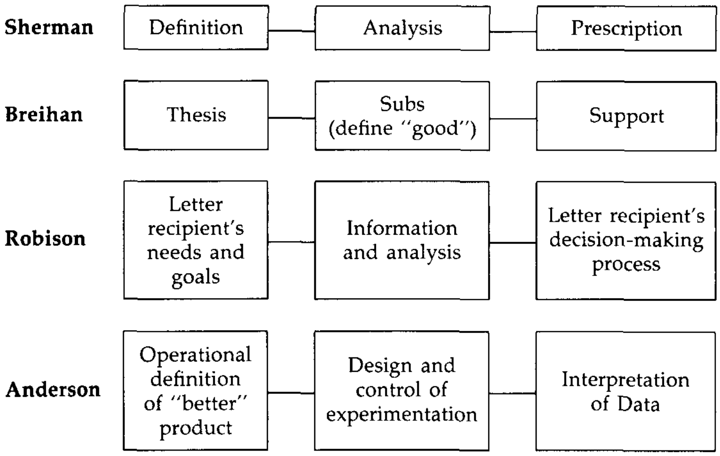

Figure 6.1. Models for good/better/best reasoning.

construct operational definitions, design an experiment, control the variables, and interpret the data.

Formulating a Hypothesis

Students could not test a statement like "Jumpy tennis balls are wonderful." As novice scientists, they had to learn to structure ideas into testable statements such as, "Tennis players will express a preference for Jumpy tennis balls over Bumpy tennis balls," or "Jumpy tennis balls will bounce higher than Bumpy tennis balls." Students were further encouraged to construct a null hypothesis: "There is no difference between Jumpy and Bumpy tennis balls." Anderson considered null hypotheses easier to accept, reject and/or interpret than directional hypotheses such as "Jumpy tennis balls are better than Bumpy tennis balls."

Defining "Better" Operationally

Anderson required the definition of "better product" to include at least four experimental factors plus cost (cost is not an experimental factor because one does not need to conduct an experiment to find it—just read the price tag). In other words, students could not decide that pickle *A* was better than pickle *B* merely because it tasted better at room temperature. They had to consider three other factors, such as taste under refrigeration, shrinkage, and pH (a measure of acidity) over time. This information had to be integrated with the nonexperimental factor of cost, as well as with any other nonexperimental factors the student chose. Further, the assignment called for students to weigh the factors, using values they chose, much as in factor rating in Sherman's class (pp. 74–76). Would shrinkage, for example, count as heavily as taste under refrigeration in defining "better pickle?" As in the other classes, the choice of factors in the definition of *good* or *better* implied that the product might be better for some people or situations than for others. Students therefore had to target the various subgroups for the product's users. By defining "better" product in this way, the student performed good/better/best reasoning Task 1 (defining "good") and also Task 4 (integrating values with evidence).

The student also had to operationally define each of the factors that comprised the definition of "better." For example, if a student was comparing Utz's[1] with Herr's potato chips, what does "better chip" mean, and if one factor in "better chip" is crunchiness, how is "crunchier chip" defined? One experimenter might define crunchier chip as the chip that 75 percent or more of persons sampling the potato chips rank higher on a 1–4 test scale. Another experimenter might define crunchier chip as the chip that breaks into the greatest number of pieces when hit with a mallet. Either definition is satisfactory, but to achieve what Anderson considered the cardinal rule of scientific inquiry—that one's research can be replicated—the researcher had to state that definition.

Designing an Experiment

Having constructed a hypothesis and defined terms operationally, students then had to determine how to answer the question, "Which is the better product?" by designing an experiment. An appropriate experimental design included operational definitions for "better" and for each factor to be tested. Then the student had to determine how to test the factors. If the student were testing shampoo, for example, what factors (bounce, cleanliness, odor, durability, sheen, growth of new hair in three months, number of split ends) were important to identify? How could quantifiable information on these factors be obtained? If some factors were to be judged by people, how many people should be included? How many times should hair be checked for "shampoo durability"—once a day? once a week? Designing an experiment, then, was Anderson's form of Task 2—analysis of the various qualities of the product—and Task 3—bringing the information about the product's traits into disciplined relationship with the definition of "good" so that a single judgment can result.

Controlling Variables

The students had to restrict the variability that entered into the experiment in order to attribute results to the proper cause. Variables in science, Anderson taught, could be controlled in three ways: manipulation, randomization, or writing the variable out of the design.

Manipulation: If the student were examining the frying ability of Crisco and Wesson Oil, the material being fried had to be the same; students could not fry chicken in one oil and potatoes in the other.

Randomization: If a student were testing the size of cereal flakes, he or she would randomly choose, say, 20 flakes, measure these, and get the average. A random selection insured that the researcher did not pick out large flakes in one cereal and small ones in another.

Writing variables out of the design: In testing for shampoo durability, the student could simply state that for this experiment the environment of the shampoo evaluators was assumed to be comparable and that the effects of relative humidity were not considered.

Interpreting Data

After the experiment was complete, conclusions and implications had to be drawn within the framework of scientific logic constructed by the researcher. Students had to use the information as evidence to support their positions, and the conclusions had to be limited by the nature of the designs. If a student designed soap comparisons and found that 40 male soap users wash with Dial for 3.2 minutes and Ivory for 4.1 minutes, the student could not conclude that *people* use Ivory soap longer than Dial soap, but would have to make the conclusion gender specific.

In interpreting the data, the student had to address the question of "better for whom?" by designating various subgroups of users. For example, a more absorbent paper towel costing twice as much might be better *for those who could afford the extra cost.*

Within Task 5 (balancing rationale-building with solution-searching), Anderson placed heavier emphasis on solution-searching than any of the other teachers. Once the definition of "good" was determined, feelings and values should not enter the process of decision making; rather, the student was expected to adopt the scientific stance, attempting to reach results through objective, quantifiable experimentation.

Anderson's Expectations for Organizing the Research Report with a Scientific Format

As research scientists, students had to report their findings in the traditional research report. Format in Anderson's class was therefore more convention-driven and more specific than in the other disciplines, where students were told to "write a letter" or "write an essay," with specific sections left to choice.

Anderson expected the students to use the standard scientific journal article format and taught them to organize their papers according to the following sections: Title, Abstract, Introduction, Review of the Literature, Methods and Materials, Results, Discussion, and Literature Cited. As we describe Anderson's format expectations, we will give excerpts from the paper of Jim Wilkerson, a high-success student from the 1983 class who wrote what Anderson considered a good report on his experiment to compare two types of erasable pens. ("Success" is defined on p. 36.)

Title

The title of an original science research paper should be explicit because the major science reference sources such as *BioAbstracts*, *ChemAbstracts*, and *Science Citation Index* use the title's key words (descriptors) to index the article. Anderson explained that titles were restricted to 25 words or less and expected students to adopt the tone of the many research article titles they had read earlier in the course. Wilkerson's title, "Comparable Research on Papermate Erasermate$_{TM}$ and the Scripto Erasable$_{TM}$ Pen," was appropriate in length and tone, but it lacked any descriptors of the qualities for which the pens were being tested. The title, "A Comparison of the Writing Ability, Erasability, and Ease of Use of Papermate Erasermate$_{TM}$ and Scripto Erasable$_{TM}$ Pen" would have been more effective in Anderson's judgment.

Abstract

Anderson expected an informative abstract in which the first sentence explained what the researcher did. Subsequent sentences were to describe how the researcher did it, what he or she found, and what implications could be drawn from the study. Never to exceed 5 percent of the original work, most abstracts, Anderson taught, are limited to about 250 words (Biddle and Bean, 1987, 41–46). Here is Wilkerson's abstract:

> The Papermate Erasermate$_{TM}$ and the Scripto Erasable$_{TM}$ Pens were evaluated on the basis of smoothness of writing, tendency not to skip, tendency not to smear, erasability, overall appearance and writing comfort. The Erasable$_{TM}$ Pen was found to be better. Seven volunteers (four males and three females) from ages 16 to 58 made up the study group. Three of the seven volunteers were left-handed. The length of time required before the erasable inks became permanent was studied inconclusively. Both pens photocopy about the same.

Introduction

Anderson told her students in class that the Introduction "attracts the reader's attention and states the purpose of the research." It was therefore important to set up a framework for the research and to identify interested audiences. In addition to this, some recent texts on scientific writing such as Day's (1979) *How to Write and Publish a*

Scientific Paper (23–25) advocate including a brief statement of findings in the Introduction. In *Writing Papers in the Biological Sciences,* McMillan (1988) explains, "Other writers, along with some journal editors, criticize this practice, arguing that results are already covered in their own section and in the Discussion and Abstract. Ask your instructor what he or she prefers" (15). Anderson preferred the more traditional form and did not instruct or expect the students to include results. Wilkerson's Introduction met Anderson's expectations:

> If you are a perfect writer who never makes mistakes, you will probably not be interested in this paper. But if you have to spell a word three different times before you get it right, you may benefit from this research. Erasable pens can make your notes or even your final drafts look much better. Instead of crossing out mistakes, you can simply erase and correct them. The frustration of ruining a birthday card by misspelling your best friend's name can be cured with an erasable pen. The embarrassment of asking for two job applications because you know you will mess up one can be forgotten. Erasable pens can give the writer freedom from mistakes and the power to write neatly with ink.
>
> In this paper the Papermate Erasermate$_{TM}$ and the Scripto Erasable$_{TM}$ Pen are evaluated. The aim of the paper is to demonstrate that one pen is better than the other or that both pens are of similar quality. Various aspects of the performance of both pens will be used in the evaluation.

Review of the Literature

Scientific writers ordinarily prepare a review of the literature to draw in previous information and techniques or to associate their research with that of others. In this assignment the review of the literature was omitted because the emphasis was on demonstrating the scientific process skills.

Methods and Materials

This section reveals how the scientific experiment was conducted. Anderson taught that it must be thorough, once again to insure replicability. The organization of the Methods and Materials section hinges on the nature of the tasks involved; it is not merely a chronological narrative. In *The Craft of Scientific Writing,* Alley (1987) states, "In scientific writing, logical sequences may be based on time, space, or any number of variables. The variable that you choose depends on your research and your audience" (156). Wilkerson's Methods and Materials section begins with the heading "Evaluation of Six Aspects of the Erasermate$_{TM}$ and the Erasable Pen$_{TM}$." Here is his explanation of how one of those tests was conducted:

> Seven volunteers were asked to write the first three paragraphs of the Declaration of Independence two times in order to evaluate the performance of each pen, the first time using the Erasermate$_{TM}$ and the second time using the Erasable$_{TM}$ Pen. The volunteers included four males and three females. They ranged from 16 to 58 years of age. Three of the volunteers (two male and one female) were left-handed. The volunteers wrote in a spiral notebook at their leisure. They were instructed to correct immediately any mistakes they made. In case they did not make any mistakes, they were told to erase four words at random and rewrite them.
>
> The volunteers were asked to rate the pens on six criteria. The rating scale ranged from one (very poor) to five (very good). The six criteria are defined as:
>
> 1. Smoothness: Smooth writing pens have minimum drag on the paper. A smooth writing pen glides easily across the paper.
> 2. Tendency not to skip: The ink flows evenly and regularly. [continues with the other criteria]

Results

The Results, Anderson taught, should include both written text and graphic information. Quantitative data should be introduced by appropriate text. Further, as McMillan (1988) explains, "The Results section should be a straightforward report of the data. Do not compare your findings with those of other researchers, and do not discuss why your results were or were not consistent with your predictions. Avoid speculating about the causes of particular findings or about their significance. Save such comments for the Discussion" (21).

Anderson was pleased that Wilkerson did just as McMillan advocated, and that he presented the results to his readers in both text and graphics. Here is an excerpt from the text of his Results section:

> Evaluation of six aspects of the Erasermate$_{TM}$ and the Erasable$_{TM}$ Pen by volunteers (see tables 1, 2 and 3)
>
> Smoothness: The Erasable$_{TM}$ Pen was found to be a much smoother writing instrument than the Erasermate$_{TM}$ for the left-handers. The right-handers rated the Erasermate$_{TM}$ as the smoother pen. By including both groups together, the Erasable$_{TM}$ Pen is rated smoother.
>
> Tendency not to skip: The left-handers rated the Erasable$_{TM}$ Pen as better by two points on the rating scale while the right-handers preferred the Erasermate$_{TM}$. Summing both groups gives the Erasable$_{TM}$ Pen a clear advantage in this category.
> [continues with rest of results]

The organization of Wilkerson's results followed the same sequence of topics in his Methods and Materials section. (See Table 6.1).

Discussion

The Discussion section is also called "Conclusions" or "Implications." In this section, students should summarize major findings, support or reject the hypothesis, and provide explanations of the significance of the data and relevant nonexperimental information. The statements in this section are interpretive. McMillan (1988) suggests that the prose in this section, unlike the Introduction which moves from the general to the specific, should move from the specific to the general and "convey confidence and authority" (26–27). Here is the first part of Wilkerson's Discussion:

> Left-handed writers clearly preferred the Erasable$_{TM}$ Pen. The tendency not to smear was an important category. Left-handed writers tend to drag their hands through their writing as they move across the paper. This causes smearing with both pens, but the Erasable$_{TM}$ Pen smeared less.

Literature Cited

This section was necessary only if outside sources were used. In Anderson's research assignment, library sources were not required; however, many students did cite advertising claims or commercial publications.

Table 6.1 *Jim Wilkerson's table with his title*:
Averaged Responses of Right- and Left-Handed Volunteers to the Six Criteria Rated.*

Volunteers were asked to rate each pen on six different criteria and their responses were averaged. The scale is from one (very poor) to five (very good). Overall averages are also included.

Criterion	Erasermate$_{TM}$	Erasable$_{TM}$ Pen
Smoothness	3.29	3.43
Tendency not to Skip	3.00	3.71
Tendency not to Smear	2.43	2.86
Erasability	3.29	2.86
Overall Appearance	2.57	3.14
Comfort	3.71	3.28
Overall Average	3.05	3.21

*$\underline{n}$ = 7

Graphics

Anderson required that each student include at least three appropriately labeled graphics in the report. Wilkerson constructed three data tables (one of which is reproduced as Table 6.1) as well as a bar graph that visually illustrated the discrepancies between preferences of right-handers and left-handers.

ANDERSON'S 1983 TEACHING METHODS

When we studied her class in 1983, Anderson had already instituted some of the teaching methods she had been led to consider through the writing-across-the-curriculum workshop. Particularly, she depended heavily on peer response. Further, she had begun to work out the philosophy of teaching that would guide her throughout our study: she believed that she was a "facilitator" of learning. But she realized after our study that she was not using many concrete experiences to guide the students in using the scientific method. That was the main ingredient she added after our analysis of her class in 1983, as will become clear in later sections. In 1983, however, she used three types of teaching methods: lecture/demonstration/response to questions, peer and teacher response, and auxiliary activities. We discuss each in turn.

Lecture/Demonstration/Response to Questions

To help her students meet her expectations, Anderson introduced the original research paper early in the semester. She identified it in the class syllabus and, after Walvoord visited to explain data collection procedures, she explained the assignment more fully in class. Classroom activities that related to the assignment included:

1. A 50-minute lecture and discussion on the scientific method using a comparison of Crisco and Wesson oils as a model.
2. A 50-minute lecture on the do's and don't's of the scientific format.
3. Two 15-minute "warning and review" sessions, where Anderson reminded students to decide on their two commercial products by the date she had set for that decision.

4. A 30-minute session in class when students announced their topic decisions.
5. Occasional class time answering questions students asked about the assignment.

Peer and Teacher Response

Anderson also used several techniques she had learned in the writing-across-the-curriculum workshop, particularly peer and teacher response:

1. A 50-minute planning and focusing activity where students, in groups of four, discussed "How I am going to test ____."
2. A 50-minute class session during which students responded to one another's drafts of the Introduction and Methods and Materials sections.
3. Individual 15-minute conferences with each student to return previous writing assignments and answer questions about experimental design.
4. An interactive 45-minute class session to review the scientific method, encourage revision, advocate peer review, and urge students to edit their papers meticulously.
5. A 50-minute session just before the final papers were due, in which students worked in pairs interviewing their peers and asking such teacher-supplied questions as "I had trouble writing the ____. What part of the paper did you find hardest?"
6. Class sessions after the reports were completed, in which students gave 7- to 10-minute speeches to their classmates, reporting their research.

Auxiliary Activities

Three other activities conducted at various times throughout the semester were also important. Anderson assumed (wrongly, it turned out) that the students would recognize the relevance of these activities to their original science research. These activities were:

- Two periods (150 minutes) in the library learning how to retrieve scientific journal articles from *BioAbstracts* and *Science Citation Index*. In separate practice assignments for each resource, the students used a generic topic descriptor (i.e., *Felix rex* for the lion or *Drosophila melanagaster* for the fruit fly) to locate a title and an author of a relevant work. After the students located, read, and cited an appropriate abstract, they used a computerized locator system to determine whether the journal article was in the TSU library.

- Five reading assignments for students to read and abstract a minimum of five original research reports taken from scientific journals. All the reports followed the format that students were to use for their own reports.
- Two class periods (150 minutes) learning how to select, construct, and label graphics.

OUR METHODS OF DATA COLLECTION AND ANALYSIS

EARLY DATA COLLECTION AND ANALYSIS

We collected our first set of data in the spring 1983 semester. The data and our analytical procedures were those described in Chapter 2. Later, we expanded those procedures, as we explain.

We had promised Anderson's students that Anderson would not look at any of their process data until after course grades had been turned in. That was optimistic. Anderson was busy (remember the 12-year-old, the 15-year-old, the new baby, the full-time teaching position, and the doctoral dissertation). Anderson completed her dissertation in May, 1984, and that summer we began to examine the data. As we listened to the students' tapes and studied their notes and drafts, Anderson kept saying, "Oh, if only I'd known they were doing that, I would have . . ." Anderson knew *that* the students had fallen short of her expectations when she graded the papers in 1983, but now she was constructing explanations for *how* and *why* some of their difficulties had occurred.

We were intrigued with the difficulties. Although we were still "knee deep in data" as the spring 1985 semester started, we wanted to know more. We collected similar data—logs, tapes, rough drafts, and so forth, to help broaden our understanding of the difficulties. Although our basic analysis of student difficulties was formulated solely from the 1983 students, in this chapter we have occasionally augmented the descriptions of those difficulties with some particularly cogent examples from 1985 students.

COMPARISON OF THE 1983 AND 1986 CLASSES

By the following year, 1986, Anderson had made significant changes in her teaching methods, and we wondered whether students' thinking and writing would appear different as a result. Thus we studied the 1986 class and collected the same kinds of data, then used outside raters to compare the quality of the final products of the 1983 and 1986 classes.

Comparing the quality of the final products of two classes and using outside raters may appear outside the naturalistic paradigm and the theoretical assumptions we explain in Chapter 2, so we want to clarify what we think the comparison study portion of our research does and does not do.

First, we do *not* view the two classes as control and treatment groups. Although the classes were remarkably similar in some ways (Table 2.1, p. 18), we could not meaningfully compare SAT scores because so many of the students were transfers for whom none were recorded by the university. Further, the classes were small (each 13 students, with 11 students submitting data). More broadly, students' performance, in our theoretical paradigm, is viewed as socially constructed, shaped by multiple interacting factors within each classroom, many of which we did not investigate or try to measure. Thus we do not claim that the improvement the raters found in the 1986 class is due to Anderson's changes in teaching. Rather, that part of our investigation was simply another tool we used to get a handle on some of the ways in which teaching, thinking, and writing might be interacting in Anderson's classroom. We *do* consider Anderson's changed teaching methods as likely candidates to have influenced the improvements in 1986, and we look at the process data for explanations of how that influence worked.

The Primary Trait Analysis

Following the principle that we would keep evaluation as close to the classroom context as possible, we tied the scoring of the paper by outside raters as closely as possible to the expectations that Anderson had held during the course. In 1983 we had drafted a crude primary trait scoring scale primarily to help us articulate Anderson's expectations and to serve as a check on in-class grades to determine students' success, as we explain on pp. 35–36. In the fall of 1986, Anderson refined the primary trait analysis and constructed a primary performance scale of 1 to 5 to serve as an instrument for evaluating students' papers. (See Appendix A for the complete instrument.) Anderson then trained Walvoord, after which the scale was refined and tested again with McCarthy. We eliminated the abstract from the rating because we had found, in developing and validating the primary trait scale, that the abstract had a "halo effect" on the raters' marking of the paper.

In January 1987, Anderson trained the two outside raters. Both were experienced, tenured, college biology teachers who had not previously been involved in the project. In a one-hour training session, Anderson answered questions about the scoring scale as the biologists evaluated components and examples from the 1985 research papers. Next, the biologists read and scored two papers from the 22 papers of the 1983 and 1986 classes. Then they were asked to compare their marks in each category and to resolve discrepancies of more than 1 point by consensus; no such discrepancies occurred. Subsequently, they evaluated the remaining 20 papers independently in the order of their choice. They believed that they were ranking a single set of 22 papers; they were not told that the purpose of the research was to compare 1983 and 1986 achievement.

In compiling the data, we gave each student a score for each primary trait from each evaluator; we averaged the scores. The highest score for each primary trait was 5. We will present pertinent data excerpts from these primary trait analyses as we discuss each student difficulty; the complete set of ratings are in Appendix A.

In the following pages, we discuss all six areas of difficulty. Under each area of difficulty, our discussion is organized under two headings:

1. The nature of the difficulties
2. Teacher's methods, student performance, and implications

DIFFICULTIES WITH CONSTRUCTING THE AUDIENCE AND THE SELF

THE NATURE OF THE DIFFICULTIES

Anderson, like Sherman and Breihan, expected her students to address their audience both as classmates and as *fellow* professionals-in-training. For example, Jim Wilkerson achieved an appropriate role and tone in his Introduction by appealing to his classmates' everyday experiences with pens while stressing that they, as scientists, would also be interested in the results of his experimentation.

A Variety of Roles

In contrast, some students constructed themselves and their readers in roles taken from other settings. In her research on paper towels, for example, Susan Bell concludes:

> Since people usually tear more than one off the roll no matter what the job is, it is wiser and economical to buy the A&P brand. It is *silly* to pay for Bounty's quality if they will not use it *properly*. [Italics ours]

Bell appropriately wants to address the issue "best for whom?" but "silly" and "properly" reflect the voice of a moralizing parent. Doug Cipes's title, "A Quality Comparison Between Two Commercial Electrophoresis Units: The BioRad DNA Sub Cell Versus the BioRad DNA Mini Cell Unit" was incongruous. "Versus" is fine for sports fans, but inappropriate, in Anderson's judgment, for the scientific reader. Kitty Cahn seemed to be writing to the Speech 101 class throughout her paper entitled "Would You Eat Machine-Made or Homemade Cookies?" Sometimes students constructed the reader as the "generic teacher": perhaps the author of "Research to Determine the Better Paper Towel" recalled succeeding with "Book Report on *Silas Marner*."

Occasionally, low-success students addressed the audience with an exaggerated or stereotyped view of scientists. On her think-aloud tape, Amy Olds read aloud from her notes Anderson's instruction that "The Introduction should get the audience's interest and state the purpose." She immediately looked at the container and wrote this dullest of first sentences: "Ivory Liquid Detergent and Lemon Fresh Joy are both manufactured and distributed by Proctor and Gamble."

Olds's difficulty—reading the instructions and then immediately taking a step that contradicts them—is another example of how hard it can be for students to use procedural instructions that they merely read or hear. We saw this same difficulty in Sherman's class, as Carla Stokes read the steps for making a location decision, and then began her task by skipping the first two steps (p. 79).

A possible influence behind Olds's opening with Procter and Gamble is students' common way of writing a paper in school: reading first, then taking the paper from written sources. We have noted the ubiquity of the text-processor role students adopted in all the classes. Olds's final paper was still haunted by the ghosts of that early dependence on reading the labels: she never separated her own findings about the products from the claims that the manufacturers made on the labels and in TV ads.

Students had difficulty constructing not only the reader but also the self. Penny Reno, who compared two men's fragrances—Polo$_{TM}$ and Timberline$_{TM}$—began her introduction with a beautiful quotation—fine, Anderson thought, for the literary essay, but not for the scientific report. Further, Reno adopted the role of persuader in a way that violated Anderson's expectations for the balanced, objective voice of the scientist. Reno's log says: "I first convinced myself that this was a good thing to write about and then I convinced my reader in my introduction."

Scientists do "convince" their readers in certain senses, but Anderson did not consider it appropriate for a scientific research report to exhibit the persuasive tone that Reno adopted on the basis of her notion of "convince." Anderson believed that "an objective tone," similar to Wilkerson's, was "convincing" to the scientific community.

Another inappropriate model for the writer's ethos was a chatty, "stream of consciousness" voice that violated Anderson's expectations for objectivity and conciseness. Compare this excerpt from Mike Siliato's low-success final paper to Jim Wilkerson's earlier statement of the purpose of his research. Siliato's paper says:

> At the start of the original research I have no evidence of which cleaning product is superior. Comet and 409 are just two names for household cleaning products. As far as I am concerned, there exists absolutely no difference between the products. The research carried out was to identify any superiority between the products. Both products are considered to be the same at the start of the research, but when I am true [*sic*] I will pick one as better. The null hypothesis prevails in this study.

The vocabulary of the passage, particularly the last sentence, shows how Siliato attempts to see himself and the readers as scientists, but his ethos is also partly that of the storyteller.

Sharon Tissinger, who grappled with many aspects of doing science, also had a hard time writing about it. She selected a personal narrative approach for the Results section:

> When I first began to prepare homemade french fries, I found the most difficult part to be cutting the potatoes into exact sizes and shapes of the frozen french fries. After measuring 10 frozen french fries, the sizes were recorded on a table. (Table l) As I began to prepare the potatoes using the various instruments, I observed that more pieces of equipment were needed for the homemade

> french fries. In this respect the frozen french fries were more accessible than the homemade french fries. The basis for this conclusion is based on the fact that it took four instruments to prepare homemade french fries and only two instruments to prepare the frozen french fries (these include the drying spoon).

Tissinger's strategy seemed doubly inappropriate to Anderson as a biologist: results should not be presented in narrative form and results should not ring with the personal voice.

TEACHER'S METHODS, STUDENT PERFORMANCE, AND IMPLICATIONS

Titles

Difficulties in constructing the audience and the self showed up in many sections of the report format; they were most easily apparent in the report's title. In Anderson's 1983 lecture, she instructed students to "choose a title of less than 25 words with appropriate descriptors." Descriptors had been clearly defined in the library sessions, and earlier in the course the students had read and abstracted a minimum of five research articles. Anderson had assumed they would use these as models.

The 1983 class did not very well meet Anderson's expectations for titles: on the outside raters' primary trait scoring, the group mean was below 3.0 (Figure 6.2). As a matter of fact, the only title receiving a perfect (5.0) primary trait score—"Comparison of the Stain-Removing Qualities of Shout$_{TM}$ and Spray and Wash$_{TM}$"—was written by Ben Blount, who told us on his think-aloud tape that he had made up the research the night before. (He really trusted our promise that Anderson would not look at the data until after the final grades were in—or perhaps he half-wanted to be found out. When students are going to invent titles for bogus papers, they probably model very carefully.) Despite his good start in deception, Blount received a "D" on his two-page paper because it failed to meet so many other requirements

												$\bar{x}$	P Value
1983	1.5	1.5	2	2.5	2.5	3	3	3.5	4	4	5	2.95	.24
1986	2	3	3	3	3	3	3	3	3.5	4.5	4.5	3.22	

Figure 6.2. Comparison of 1983 and 1986 students' primary trait scores for "Title." Shaded areas show students receiving scores of 3.0 ("adequate") or better. $\bar{x}$ = mean score. P = the probability under the null hypothesis that improvement is due to chance is 24 in 100.

(1200–1500 words, 3 graphics, etc.). Later he retook the course from Anderson and compared two car waxes, bringing to class for his oral report an actual waxed car fender to bolster his credibility.

The 1983 logs and tapes did not produce a single other piece of evidence that suggested that students made any connection between composing their own titles and the titles of the research reports they had read a few weeks previously. Students' lecture notes about the project indicated that Anderson had not made that connection explicit for students, either.

In 1986, however, Anderson supplied a concrete, teacher-directed experience designed both to help students to notice the ethos implied by titles in the articles they read and to apply that lesson to composing their own titles. The 1986 students participated in the same library tour, prepared the same number of abstracts, and received the same instructions in class as in 1983, but after they had abstracted a scientific journal article, "Relative Climbing Tendencies of Gray (*Elaphe obsoleta soiloides*) and Black Rat Snakes (*E. o. obsoleta*)" by Jerome Jackson (*Herpetologica* Vol. 32–4), Anderson asked the students how they would have felt about the author if the article were entitled "Do Snakes Get High?" In a five-minute discussion, the students were encouraged to see how what they had learned about "audience" in English was relevant to scientific writing. The discussion also reviewed the importance of adequate descriptors and communicated to students that the articles they abstracted had titles that could be modeled.

In contrast to the 1983 class, 10 of the 11 students in the 1986 class composed adequate (3.0 or above) titles (see Figure 6.2). Although not many students constructed superior (4.0 or above) titles, they were able to avoid titles that were modeled inappropriately from other disciplines and settings.

Introductions

Students' Introductions also reflected their difficulties in constructing the reader and the self. In a 1983 lecture, as we have said, Anderson told them that the Introduction section in scientific writing "gets the audience's interest and states the purpose of the paper," and she gave them several opportunities to discuss their Introductions with classmates. During one class session, pairs of students talked for about 15 minutes about their Introduction and their Methods and Materials sections. In a later session, pairs responded to each others' drafts of the Introduction. In addition, some students used parts of "open" peer conferences to discuss their Introductions.

Despite all of this activity, many students still wrote inadequate Introductions. Apparently, peers did not effectively help each other write as scientists-in-training to other scientists-in-training.

The data reveal two strategies used by the successful students in 1983:

1. Drafting or revising their Introductions *after* beginning their data collection.
2. Consulting people they considered "experts" for help with their drafts.

But these successful strategies were not common. Logs, tapes, and rough drafts document that 8 of the 11 students had conducted no research at all before writing their Introductions. Three successful students who had tried to write theirs before any experimentation with their products, succeeded only after three or more revisions. Kay Price wrote her Introduction four times before designing the experiment. Hilary Nearing ended up with a one-half page, typed paragraph Introduction, but her rough draft contains ten handwritten pages from her first effort; this includes a tedious chart in which she copied the ingredients listed on each soap powder box before using the soaps (the text-processor role again?). It seemed that when students had not *acted* as researchers, they had difficulty adopting the *ethos* of researchers. Again, as in our other classes, the creation of ethos seemed closely connected to the roles that students adopted for other aspects of the thinking and writing process. Kathy Carr seemed to have found the key: she did not draft the Introduction until two days *after* she conducted her first tests and then needed to make only surface revisions.

The second possible reason why Carr wrote a high-success Introduction is that she asked for reader response not only from peers, but from others whom she saw as experts. For her project—comparing diet colas—Carr documents in her log that she talked with both a Coca-Cola spokesperson and an avid diet cola drinker. Likewise, in the 1985 class, Doug Cipes sought help from someone he felt could give expert advice.

Early in the semester, Cipes, a senior who worked in a genetics lab, jokingly(?) remarked that he thought the assignment was "dumb." Anderson countered by suggesting that he compare some product he used in the genetics lab where he worked. He decided to compare two brands of electrophoresis units. In the first peer conference, he discovered that his classmates knew nothing about these units. After one unsuccessful draft, Cipes asked a biology graduate student for help. On a tape that Cipes marked "4-20-85, A friend gives some comments on paper," we hear the graduate student make some candid comments about constructing the audience and the self:

> I hate the way you introduced this because somebody could read this and have never seen your hypothesis. Proven, never say proven. Nothing is ever proven. [Instead, say] this can be demonstrated. Then I get to down into this other thing which I hate. You are writing it almost as if you are explaining it to a graduate student who already has a good idea about it and you don't expect him to remember much of what you're saying. It's just like a general tour. It's like: This is the Empire State Building, it weighs 30 billion tons and took three million people—anyway nobody expects anybody to remember, and that's bad news in a written paper. Your written explanation has to be perfectly clear.

After his friend's response, Cipes significantly improved his Introduction. Evidently, students' strategies of conducting some experimentation before they wrote the Introduction and consulting experts for draft response appeared to help them succeed in creating an appropriate audience and self in that section of the report.

To capitalize on the insights about how successful students had worked, Anderson made two changes in her 1986 teaching methods. Using a principle that often guided her teaching changes, she guided *all* students through the processes she found had worked for successful students. In 1986, Anderson said the same things about the Introduction as in 1983, but she required all students to bring to class a pilot report first—"two or three paragraphs on what you have learned in experiments with your two products so far." By placing initial pilot experimentation *before* they began their reports, she hoped to engage the students as scientific investigators before they drafted their Introductions. Anderson also reduced the amount of peer conference time focused on the Introduction. She retained one 15-minute session on the Introduction, but scheduled it *after* the pilot report, so that students were responding to one another as fellow scientists who had each conducted some experimentation. The second change Anderson made was to encourage 1986 students to ask "fellow scientists" in other classes to critique their Introductions.

In 1986, 10 of the 11 students wrote an adequate Introduction (3.0 or above). Over half the class performed at or above the 4.0 level (Figure 6.3).

												$\bar{x}$	P Value
1983	1.5	2	2.5	2.5	2.5	3.5	3.5	4	4	4	5	3.18	.14
1986	2	3	3	3	3.5	4	4	4	4.5	4.5	4.5	3.64	

Figure 6.3. Comparison of 1983 and 1986 students' primary trait scores for "Introduction." Shaded areas show students receiving scores of 3.0 ("adequate") or better. $\bar{x}$ = mean score. P = the probability under the null hypothesis that improvement is due to chance is 14 in 100.

FOUR INTERRELATED DIFFICULTIES: STATING A POSITION; USING DISCIPLINE-BASED METHODS TO ARRIVE AT (AND SUPPORT) A POSITION; MANAGING COMPLEXITY; GATHERING SUFFICIENT SPECIFIC INFORMATION

In Anderson's class, stating a position meant stating which product was "better." The discipline-based method was the scientific method. The scientific method manages complexity through experimental design, operational definitions, and control of variables. It also defines ways of gathering sufficient specific information.

Anderson expected students to design an experiment, construct operational definitions, control variables, gather sufficient specific data, present data in graphic form, and interpret their data in the Results section of the report. We take up these aspects one at a time in this section; however, we do not treat presenting data in graphic form. As in earlier sections of this chapter, under each aspect we first discuss the nature of the difficulties, then Anderson's teaching methods of 1983 and 1986, the student performance for each year as measured by the raters, and the implications of those findings.

DESIGNING AN EXPERIMENT

The Nature of the Difficulties

Designing an experiment was the most difficult task the students faced. It seemed difficult partly because it precluded the text-processing role we have found so common in the other classes (though, as we have seen, some of Anderson's students began with the only print available—the product container). As one student put it, "It's your baby all the way. You have to do the research and you have to write it. You can't go to a library and read about it and summarize."

Three 1983 students—Mike Siliato, Jeremy Lucas, and Sharon Tissinger—had poorly designed experiments. We will use their data to illustrate six ways in which some students had trouble in this area:

1. *Students considered few topics seriously.* The logs and transcripts revealed that most 1983 students considered and/or expanded very few topics. Siliato, for example, revealed on his think-aloud tape on February 5 that he thought of one food product "but since I'm not much on cooking myself, it is out of the question. I do feel I would perhaps start my research comparing 409 and Comet." He never considered another topic. Jeremy Lucas, an international student, began on Feb. 25 with the think-aloud statement, "There is so many house products flashed in my mind that it does make me choice (unintelligible two words) is comparing between Palmolive and Joy dish-washing liquids." These were the only products he recorded having examined on a trip to the store on February 27, although he did not purchase his Joy and Palmolive bottles until March 3.

Sharon Tissinger, on the other hand, appeared to think about possible products because she listed several words in her log, but she did not expand these ideas. When her father suggested french fries on the evening of March 14, Tissinger, like Lucas, seriously considered various factors in experimental design for only one topic. She said on the tape:

> Now for some reason I really like this idea. It seems to me differences could include the cooking time, the storage, you know, even a taste test, you know... even the stipulation that there is, are, different instruments.... This idea has definitely topped my list because I feel like it is something that could easily be compared whereas something like soap or shampoo—it's very hard to tell the difference of clean, you know, you have to define clean....

Tissinger did not work out the details of experimental design for any other products.

2. *Students concentrated on peripheral issues rather than on the critical task of designing the experiment.* Mike Siliato's data suggest he was preoccupied with making a good grade. For example, on his tape dated 2/26/83, he notes that "I just hope that this will be a success." Jeremy Lucas recorded on his March 4 tape:

> How am I going to start this paper? It is surely difficult to start the paper. Yet once I start it everything will be easy. I am thinking of a thesis statement, of all of them. I hope I come up with something good. [Long pause] Yes, I've got what I wanted—a good thesis statement.

Throughout his log it is evident that the actual writing was Lucas's main concern. On March 29, he gave the paper to a friend, as he says in his log, to "proofread," not to ask for feedback on the experimental design or content. Lucas was focused on writing the paper correctly; designing the experiment was a side issue for him.

To write the paper, Lucas, like a number of the students in Sherman's class, tried to use the "thesis" model, but interpreted forming a thesis as the writer's first act, rather than the position a scientist would reach after experimentation.

Siliato, Lucas, and Tissinger all demonstrated a third possible reason for poorly designed experiments:

3. *Students did not conduct preliminary investigations (pilots) to aid them in planning the experiment.* Only 2 students out of the 11 in the 1983 group conducted a pilot. Since this seemed so counterproductive and naive for science majors who had taken at least ten semester hours of college science, Anderson wondered whether the students had conducted pilots without recording them. However, analysis of students' experimental procedures as recorded in their logs and tapes made it clear that there had been no pilots. For example, Tissinger exposed the spontaneity of her investigation as she recorded, while conducting her experiment,

> It seems the frozen french fries are exceeding the fresh cut french fries in cooking time amazingly, immensely! In fact they are almost done and it has only been two minutes. It is now—the total cooking time is 2 minutes and 20 seconds! The frozen french fries are definitely done—brown, very crispy. I have now turned off the flame. I am immediately taking them out and putting them on a towel (pause, laughter from assistant)—there seems to be a problem with the towel.

Siliato's tapes reveal that he did not conduct a pilot project either. He records on the tape:

> Right now I am starting the project. I'll clean one half of the bathroom, rather the bathtub, with 409 and the other half with Comet. . . . What I plan to do is, one day I'll clean the toilet with Comet, the next day I'll clean it with 409 or better yet how about if I do it this way—the first 15 days of March I'll clean it with Comet and the last with 409. . . . As time goes on, I will probably think of new ways to test both products.

Thinking of new ways to test the product as one goes along is good strategy for a pilot, but a disaster for conducting the actual experiment.

As we examined the students' experimental designs in their final papers, we found three more reasons that students had poorly designed experiments. These reasons were very closely related to difficulties in gathering sufficient specific information.

4. *Students failed to locate* sufficient *information because they designed experiments that had inappropriate sample sizes.* For example, Jeremy Lucas recorded in his Methods and Materials section:

> I measured 2 mls of cooking oil in one plate and 2 mls in another. 5 drops of Joy were placed in one of the plates and 5 drops of Palmolive in the other plate. The effects were noted.

In this case, Lucas selected a sample size of N = 1 for the plates. He did not collect sufficient data. If he had tried this procedure with five plates and taken an average, his data would have been more credible. Sharon Tissinger asked evaluators to compare homemade and prepared frozen french fries. Sample size didn't seem to worry her as she recorded her results in a table—two people for homemade and two for frozen french fries.

5. *Students failed to locate sufficient* specific *information because they failed to design ways to quantify information.* In contrast to Lucas and Tissinger, Mike Siliato collected a lot of data. In his Methods and Materials section, he describes how five evaluators compared the two cleansers on five different occasions:

> When I began my research on Comet and 409 I spent 10 different days just cleaning the bathtub and toilet. On five different days I washed out the toilet with Comet and on the other 5 days likewise with the 409. The purpose of this was to see among members of my family if there was any opinion in regard to odor or looks. . . .

Table 6.2 shows Siliato's chart.

Notice that Siliato designed an experiment that records no *specific* information on "the odor or looks," the two factors he said he wanted to examine. Siliato had a sufficient amount of data for Anderson's expectations in this project, but data were not specific enough to be useful.

Compare Siliato's chart to Wilkerson's chart on page 188. Wilkerson, like Siliato, asked judges to compare two products, but he clearly defined the criteria for comparing the two brands of pens and constructed a rating scale of 1 through 5, thus allowing him to collect specific data.

In his chart (Table 6.2), Siliato reveals that he did not know what to do with his own opinion, and so he simply listed it with the opinions of his other rankers. Siliato's difficulties arose because he failed to distinguish himself as scientific researcher from himself as observer. His difficulty may be related to the difficulties of students in Breihan's class who could not distinguish between "feelings" and

Table 6.2 *Mike Siliato's Chart Comparing Comet$_{TM}$ and 409$_{TM}$*
Chart #1. ("X" marks superior, "O" same)

	Date	Rankers	#1	#2	#3	#4	#5	my view
1st	3/1/83	Comet	X	X	X	X	X	X
comparison	3/2/83	409						
2nd	3/3/83	comet	O	X	X	X	X	X
comparison	3/4/83	409						
3rd	3/5/83	Comet	O	X	X	X	X	X
comparison	3/6/83	409						
4th	3/8/83	Comet	X	X	X	X	X	X
comparison	3/11/83	409						
5th	3/12/83	Comet	X	X	X	X	X	X
comparison	3/15/83	409						

10 days set aside, 5 comparisons made. Of my view 100% of the time I saw Comet as a superior cleaner and 0% of the time for 409. Of the 5 Rankers, 92% of the time they saw comet superior, 0% for 409 and 8% no difference.

"evidence" (p. 134). Clearly, issues of roles and of the construction of the self are also involved in Siliato's difficulty with his chart.

6. *Students failed to include four experimental criteria, and they did not know what to do with nonexperimental data.* The "Factors Tested" column of Table 6.3 shows the criteria that each 1983 student chose as experimental criteria, to be reported in the Results section. The I as a factor type identifies nonexperimental criteria inappropriate for Results. Siliato, for example, selected price. Price is a nonexperimental factor that should be handled in the Discussion section, because to find it one merely reads the price tag. Tissinger counted the utensils needed for homemade and frozen french fries—useful but not experimental information; Kathy Carr surveyed can color.

Table 6.3 Experimental Designs, 1983 Students

Name	Product	Factor Type	Factors Tested	Subgroups for Interpretations
Jim Wilkerson	erasable pens	J	erasability	Righthanders & lefthanders
		J	xeroxability	
		J	tendency to skip	
		J	smudgeability	
Hilary Nearing	laundry soap	J	7 stains	Users at 3 temperatures; ecologists
		M	suds durability	
Kay Price	diet sodas	M	flat time (cup)	Servers of diet drinks; heavy users & novices
		M	flat time (bottle)	
		M	flat time (ice)	
		J	taste	
Mike Siliato	tile cleaners	J	on tubs	—
		J	on tiles	—
		J	on stains	
		I	price	
Susan Bell	paper towels	M	total absorbency	Users of 1 or more paper towels
		J	window cleaning	
		M	absorbency rate	
		J	hand drying ability	
Kitty Cahn	chocolate chip cookies	J	visual appeal	—
		M	stretch of chip	
		J	combined tastes	
Sharon Tissinger	french fries	J	taste/appearance	—
		J	greasiness	
		J	crispness	
		I	no. of utensils	
Ben Blount	pre-washes	J	4 stains	—
Amy Olds	dish soaps	M	spots on glass	—
		M	suds life	
		M	baked-on foods	
		J	hand preference	
Kathy Carr	diet sodas	J	blind taste test	—
		J	known taste test	
		J	can preference	
		I	can color	
Jeremy Lucas	dish soaps	J	spot removal	—
		J	grease removal	
		J	combined preferences	

J = judged factor M = measured factor I = nonexperimental factor inappropriate to design

N = 11 students

Teacher's Methods, Student Performance, and Implications

Aware of these indications about how and why the 1983 students were having difficulty designing experiments, Anderson reexamined her teaching techniques. Her first assignment in 1983 had asked students to submit a topic to her by a certain date. She had done this to help the students start early in the semester, but our data revealed that since the students thought seriously about very few topics, Anderson's assignment caused many to close prematurely on a poor topic. The next exercise was for students, in groups of four, to discuss "How am I going to test [my product]?" In 1983 they were not, at that point, able to help each other with experimental design. We saw in Robisons' class, also, how peer response could fail when students did not know enough, or did not have sufficiently specific guidelines, to appropriately evaluate others.

On February 14, 1985, Anderson experimented with a new set of beginning assignments designed to encourage students to consider more topics in greater depth, and to help them to help each other with experimental design. She asked the students to bring in ten topics that might be used for the original research paper. She put students in groups of four and they shared their lists for about 10 minutes. Their next assignment was "Do *not* decide on a topic; decide on at least four possible topics. Write a paragraph about how you would design an experiment to test each of these." The next assignment was to list four criteria each for two kinds of products. This was followed by a pilot report.

The logs and tapes of the 1985 class convinced Anderson she was on the right track. For example, Matt Brady, raised on Maryland's Eastern Shore renowned for duck hunting, recorded his quest for a topic in his log:

Feb. 15.	Thought about testing two brands of Beer. Heniken vs Molsen, or Molsen vs Moosehead.
Feb. 16.	Talked to my girlfriend's father who is a surgeon hoping to gain some insights. He suggested testing trash bags. Considered trash bags, remembered that someone in class suggested them so I did not want to do it.
Feb. 23.	Talked to my dad—he suggested shot gun shells.
March 17.	Called my old high school Science Teacher talked to him about research paper he suggested that I talk to a pharmacist he knows because the pharmacist had said that brand name drugs were a rip-off compared to generics. I did not feel qualified to test drugs on people.

March 23. Talked to my dad and decided to test the difference between steel and lead shot in shotguns—lead shot is supposed to be hazardous to diving ducks who eat the shot and die of lead poisoning.

Unlike the 1983 students, Brady took his time, considered a number of different topics in a serious way, talked to other people, and worked through each possible topic far enough to decide specifically why he rejected it.

In 1986, Anderson again used these new activities. In addition, on the day the students brought their rough drafts of the pilot reports, she asked them to share these reports in groups of three or four and to focus on helping each other develop a list of "4 to 6 testable, quantifiable" criteria by which they would judge which product was "better." Immediately after that 30-minute, in-class session, Anderson lectured for about 10 minutes on the importance of sample size and the difference between experimental and nonexperimental information. Table 6.4 describes the 1986 experiments. By comparing it with the 1983 experiments (Table 6.3) in terms of experimental design, we identified several telling differences between the two classes (Table 6.5).

These differences between the 1983 and the 1986 classes were reflected in the judgment of the raters (Figure 6.4).

We concluded that in 1983 Anderson used peer conferences prematurely to accomplish a task too complex for students to handle at that time. The new activities helped students focus on the many aspects of designing an experiment before limiting their options. In the other three disciplines, as well, teachers came to the same conclusion—put more time into guiding the *beginning* of the thinking/planning/writing process. In Anderson's class, as in Robison's, peer-group success seemed to depend on giving the groups specific, structured tasks and enough teacher guidance so that they knew what to look for and how to help each other.

CONSTRUCTING OPERATIONAL DEFINITIONS

The Nature of the Difficulties

Once students had designed experiments, they were expected to demonstrate two less complex but essential scientific skills: to define operationally and to control variables. Each skill was independent but integral to the student's success on the final paper. We will discuss each separately.

Table 6.4 Experimental Designs, 1986 Students

Name	Products	Factor Type	Factors Tested	Subgroups for Interpretations
Betty Farr	beer	J	taste	Drinkers of cold or warming beer
		M	bitterness	
		M	foam (amount)	
		M	foam (duration)	
Tia Stoffer	pickles	J	flavor	Smokers and nonsmokers; those who do and do not refrigerate
		M	texture	
		J	appearance	
		J	aroma	
Ken Johnson	trashbags	M	puncturability	—
		M	dragability	
		M	stretchability	
		M	tie performance	
Valery Hobbs	laundry soaps	J	5 stains	—
		J	softness	
		M	static cling	
Duncan Solski	paper towels	M	strength	Users of wet & dry paper towels
		M	absorption rate	
		M	pull test	
		M	total absorption	
		J	scrubability	
		J	softness	
Donna Conner	raisin brans	M	crumbliness	Vitamin users; fiber users
		M	crispness in milk	
		M	no. of raisins	
		J	taste	
Gary Galvez	typing papers	M	durability	—
		M	strength (wet/dry)	
		M	ink retainability	
		J	photocopiability	
Molly Sutton	peanut butters	M	texture/oiliness	—
		M	spreadability	
		J	combined tests	
Roy Dodd	popcorns	M	volume	Users hot/cold; smokers & non-smokers
		M	percent waste	
		J	combined tastes	
Kara Pettit	breakfast beverages	J	taste	Adults & children; dorm students
		J	shelf life	
		M	dissolvability	
		I	storage	
Mary Hart	horsehair polishes	J	shine	Routine users; horse show users
		J	durability	
		M	mane/tail tangles	
		M	preparation time	

J = judged factor M = measured factor I = nonexperimental factor inappropriate to design

N = 11 students

Table 6.5 Experimental Design Strategies (1983 and 1986 Classes)

Strategy	1983	1986
Students seriously considering four or more topics	4	11
Students choosing four or more appropriate experimental criteria	4	7
Students adequately distinguishing/using nonexperimental data	6	9
Students designing superior experiments (4.0 or above on primary trait score)	1	6
Number of quantified, measured criteria included by all students	10	26

N = 11 students (1983); N = 11 students (1986)

Difficulties in constructing operational definitions cut across all levels of achievement in designing an experiment: Mike Siliato designed a poor experiment, Susan Bell designed an average one, and Karen Price designed an above-average experiment for the 1983 group, but they all had difficulties in defining operationally. We use their data to illustrate three aspects of these difficulties: (1) Constructing no operational definition, (2) confusing operational with vocabulary definition, and (3) including no operational definition for a "better" product.

Constructing No Operational Definition

Some students did not make any operational definitions. Choosing the simplest approach, Siliato makes a heading "Operational Definitions" and writes, "When I compared the Comet and the 409 I looked for such qualities like abrasiveness, smoothness of certain areas cleaned, cost ratio and scent." Later, he writes, "Several variables were employed during the course of the research project. One was a test removing bacon grease, bathtub rings, dirt (common ground dirt) and some food stains. The same number of sponge strokes were used in the removal of the filthy substances."

If Siliato had understood operational definition, he could have defined "better grease cutting agent" as, for example, "the cleanser that required the fewest strokes with a sponge to be returned to clean in the judges' opinion after equal quantities of bacon grease, bathtub rings, and dirt have been applied."

												$\bar{x}$	P Value
1983	1	1.5	1.5	2.5	3	3	3	3	3.5	3.5	4	2.68	.07
1986	1.5	1.5	3	3	3.5	4	4	4	4	4	4	3.32	

Figure 6.4. Comparison of 1983 and 1986 students' primary trait scores for "Designing the Experiment." Shaded areas show students receiving scores of 3.0 ("adequate") or better. $\bar{x}$ = mean score. P = probability that improvement is due to chance is 7 in 100.

Confusing Operational with Vocabulary Definitions

Some students confused constructing operational definitions with locating or composing vocabulary definitions. Susan Bell records in her log on 3/24/83, "Towel—an absorbent cloth or paper for wiping or drying." Next she copies the definition of "paper" from the *Concord Desk Encyclopedia*. Bell is trying to operationally define, but she doesn't know what to do. She is only describing the words. She is not constructing an operational definition of "paper towel" for this experiment (e.g., one perforated section from a 2-ply roll made by Bounty or A&P). We noted in Sherman's chapter, also, the students' tendency to use dictionary definitions instead of constructing definitions for the purpose of their arguments (pp. 88–89).

Having worked as a technician in a scientific lab, Kay Price was familiar with what research was like. She compared Diet 7-Up and Tab. She had an adequate design, an excellent N = 50, and the potential for interesting results in comparing the taste preferences of 25 men and 25 women. Her paper seems very scientific; she even subtitles her "Operational Definitions." Here are her first three entries:

> Flat test—a test to determine the amount of time it takes for Tab and Diet 7-Up to go flat.
>
> Flat time—the amount of time for Tab and Diet 7-Up to go flat.
>
> Flat—loss of appealing taste, no longer possessing refreshing qualities.

After reading these definitions, could a reader replicate Price's "flat time"? What have her definitions told the reader to do to obtain the same results? Nothing. She has given the description of the term, but not the operations to be performed.

Compare Price's operational definition for "flat time" to Matt Brady's definition of shot pattern:

> The shot pattern can be studied by firing the shotgun into an open sheet of paper. What is most desirable is to have a shot pattern which is concentrated. A greater concentration of shot hitting the target means that more individual pellets will hit the target. And the more pellets that hit the target (in this case a duck or a goose), the greater the likely-hood of a successful kill . . . The shotgun shell's pellet concentration (Shot Pattern Concentration) was determined by counting the greatest number of shot holes found in a circle with a diameter of 20 inches and then dividing the number of holes within the circle by the number of pellets originally in the shotgun shell.

Including No Operational Definition for a "Better" Product

Although most students made operational definitions of at least some specific characteristics or tests, many students did not operationally define the "better product." Although Kay Price addressed the issue of operational definitions by making the list quoted above, she did not put "better diet drink" on the list. Neither she nor Susan Bell had an operational definition of "better." Bell explains to her readers, "The purpose of this research is to describe the experiments performed on paper towels and to present the conclusions that have been reached as to which is the better paper towel." Her paper stays in the descriptive mode; her conclusions describe which is the better paper towel for each of her four tests, not overall. In terms of the five tasks of good/better/best reasoning (p. 12), she does not complete Task 3—to bring the information about the options (in this case the four tests) into a disciplined relationship with her definition of "good" so that a *single* judgment can be made.

Teacher's Methods, Student Performance, and Implications

As we listened to the 1983 tapes, we heard several students repeat to themselves the phrase, "must operationally define the better product." The tone in their voices, as well as their final papers, revealed that this was a skill that they knew they were expected to demonstrate, but they really didn't know what it was.

Anderson recognized the phrase from her one-hour lecture in which she used Crisco and Wesson Oil as model products. She recalled asking questions like "How would you define 'crispy'?" or "What is the 'better' cooking oil?" She remembered reminding the students in both the 45-minute peer conference on their drafts of the Methods and Materials section and in their last 45-minute peer session to "Be sure that all your operational definitions are perfectly clear."

We realized that many of these upper-level biology students, like the students in Sherman's and Breihan's classes, could not easily move from merely reading or hearing a description of a complex intellectual operation to using it on their own. Since Anderson had assumed that upper-level biology majors knew how to construct operational definitions, she assumed that group work would help students focus on improving their definitions. As Anderson listened to the tapes, she realized, "no wonder peer groups hadn't helped—the blind were leading the blind."

Anderson decided to teach students concretely how to define terms operationally. She used several oral "If we were doing an experiment on . . ." exercises with the 1985 class; they got much better at defining a specific characteristic such as "shot pattern" (see Brady's operational definition, pp. 210–11). But when the students still had trouble defining the better product, we realized that two kinds of operational definitions were critical to success. We called these *specific operational definitions* (e.g., "shot pattern") and *comprehensive operational definitions* ("better" product).

Based on this insight, Anderson developed for her 1986 class a writing activity to help students learn how to construct both kinds of operational definitions. Figure 6.5 is a copy of her Operational Definitions Worksheet. The handwritten comments are the student's and the comments in brackets are ours. Anderson used a simple question as basis for this exercise: "Under which conditions—very wet, moist, or dry—does mold grow best?"

In addition to the worksheet, Anderson asked each student to "write out your comprehensive operational definition of better product" before the second peer conference session. This time, the students knew enough to help each other.

Students in the 1986 class received a higher mean score from the raters on defining operationally, and a larger percentage of students had scores of 3.0 or above (Figure 6.6).

The difference in the average achievement for operational definitions, as measured by the primary trait score for Anderson's 1983 students (2.68) and her 1986 students (3.50), has a $\underline{P}$ value of .01, indicating that the null hypothesis is 1 in 100.

CONTROLLING VARIABLES

The Nature of the Difficulties

In Anderson's experience, controlling variables is a skill that college biology students often use in their laboratory courses. Many lab manuals ask questions such as "Which variables have been controlled in this experiment?" or "What additional factors must be controlled in this experiment?" Teacher-constructed tests and standardized tests commonly use multiple-choice questions to assess this skill. Anderson was confident that her junior and senior science majors, all of whom had taken at least ten semester hours of college science, knew how to control variables in experiments designed by others. In her assignment, however, the students had to control variables within their own designs.

Operational Definitions Worksheet

I. Explain in your own words what an operational definition is. Include an example in your discussion.

A definition which explains a specific process or how a certain task is to be performed.
* An explanation for a procedure to use something
- data matches outcome
- the procedure can be repeated

[In the three or four minutes allotted for them to answer, many of the students, like this respondent, did not come up with an example. However, these few minutes made them aware that they couldn't think of an example.]

II. In the described mold experiment, state which terms must be operationally defined and define one of them.

wet → 5 ml H_2O
moist } Bread → 2 ml H_2O
dry → 0 ml H_2O

Lighting – 12 hr light
12 hr dark

Enviornment – moisture will be contained ~~aerobic~~ in a limited envirinment

Mold growth – excessive = 7 cm +
moderate = 2-7 cm
minimum = 0-2 cm

[added later] grow best – most ~~excessive~~ growth during a 14 day period. (~~success~~ consistant) (Do A pilot To get a feel for time and)

[All students listed very wet," "moist," and "dry." Many recognized that "mold" had to be identified. These were *specific* operational definitions. Only a few students realized that they had to define "grows best"—the *comprehensive* operational definition which would integrate the values of the specific tests. When students had filled out the sheet, Anderson explained the meanings of specific and comprehensive operational definitions. This student then added "grow best" to her list.]

continued

Figure 6.5. Operational Definitions Worksheet.

Figure 6.5 (cont.)

III. Was the term that you defined a specific operational definition or a comprehensive operational definition? As you explain the difference, define the type of term you did not use in II.

C-O def. of Research topic: <u>Best</u> <u>beverage</u> - which product resulted with the most positive responses ~~to experiment on~~ ~~for~~ on a taste-test, had a minimum preparation time, + maximum shelf life, ~~cost~~.

[This student's mind went right to the original research topic that she had chosen and for some reason she wrote that down—an indication of the success of this in-class exercise in helping students to apply the principles directly to their own research projects.]

IV. Examine the terms in the article that you abstracted and identify an example of a specific operational definition and of a comprehensive definition. Explain if these were stated or had to be inferred by the reader.

[At this point groups of students examined a journal article handed out the day before. They worked on this task orally with peers for about 7 minutes until the end of class. Most groups quickly moved to talking about their own research—another indication of their ability to transfer this in-class lesson to their own research projects.]

As we reviewed the 1983 papers, we found that when students developed an adequate experimental design, they controlled variables adequately. When they had poorly designed experiments, they had difficulties in controlling variables as well. Finding this expected high correlation between designing experiments and controlling variables, we really did not try to document how the low-success students went wrong. Quite frankly, to have helped Mike Siliato or Jeremy Lucas control the variables in their designs would have been like "arranging deck chairs on the Titanic." On the other hand, successful students' logs made very clear what they found helpful:

												$\bar{x}$	P Value
1983	1	1.5	1.5	2	3	3	3	3	3.5	4	4	2.68	.01
1986	2.5	3	3	3.5	3.5	3.5	3.5	4	4	4	4	3.50	

Figure 6.6. Comparison of 1983 and 1986 students' primary trait scores for "Defining Operationally." Shaded areas show students receiving scores of 3.0 ("adequate") or better. $\bar{x}$ = mean score. P = the null hypothesis is 1 in 100.

1. *Peer conferences were an effective way to help students control variables.* Although this was documented in the 1983 logs, we found two particularly cogent examples in the 1985 logs. On 2/27/85 Ivan Ford records,

> I was eating breakfast and decided to do a comparison test on two types of cornflakes. The idea of a taste test popped immediately into my mind and when my flakes got soggy I decided to do an absorption rate test first thinking of soaking and weighing the flakes at one minute intervals. We broke into groups in class (me, Kathy, Eric, and Mark) and I told them my plans. I thought of having some physical tests to determine crunchiness of dry flakes, a 20 person taste test a nutritional comparison, and some way of telling how many times you can drop the box without getting all those scummy crumbs. They all liked the ideas.

Lisa Land, a student who designed a good experiment to test microwave popcorns, recorded in her log,

> April 4—sat around in class and discussed the paper. Once you see the other ideas it becomes easier—your ideas are on the right TRACK. As I describe what I'm doing + how I'll control the experiment, it's not so bad.

2. *Manipulating the products and/or conducting a pilot helped students do a better job of controlling variables.* For example, Jim Wilkerson did not identify the age of the pen as a variable to control until he wrote with the pens himself. Two days after she wrote her log entry above, Lisa Land wrote in her log,

> April 6—I did a simple chart for my *variables.* Popped the popcorn like the pilot—put into unmarked bags—I took them over to my mother-in-law's first—had the family do my taste tests. They hem-hawed, but they did it. . . . took the rest of the popcorn to my mother's & did the same thing. I put a special mark next to the smoker—? *new variable.* [Italics ours]

Teacher's Methods, Student Performance, and Implications

In 1983, Anderson instructed students in the scientific method in a one-hour lecture, using Crisco and Wesson Oil as examples, and reminding students that it was important to control variables. She broke them into various groups for three 45-minute sessions to work on a variety of topics, including controlling variables.

In 1986, Anderson reviewed the term *control variables* in a 15-minute overview of the scientific method. All students conducted at least one pilot experiment and turned in a rough draft of a pilot report. Students focused on controlling variables in peer groups when they worked on the Methods and Materials section of their papers. The 1983 class average for controlling variables was 2.73; it increased to 3.18 in 1986 (Figure 6.7).

Anderson was most interested in the scores for controlling variables earned by students who had received 3.5 and better for designing good experiments, because those were students who had variables *worth* controlling. Those averages increased from 3.33 in 1983 to 3.64 in 1986. We concluded that controlling variables was closely linked to experimental design. Further, Anderson's assumption that students did know how to control variables was probably correct. These averages, 3.33 and 3.64, are higher than the average scores for any other scientific skill category.

PRESENTING DATA IN GRAPHIC FORM

The Nature of the Difficulties

Anderson encouraged both the 1983 and the 1986 students to use graphics to communicate scientific information, but to be on the safe side, she also required three graphics. In 1983 Anderson spent one 75-minute period lecturing on the basic functions of pie graphs, bar graphs, line graphs, diagrams, flow charts, tables, and photographs of organisms and/or their representations, i.e. X-rays, EKG's, photomicrographs, and so forth. In a second period, she went over the five different types of graphics the students constructed from some fish data she had given them as a homework sheet, and she helped students

												$\bar{x}$	P Value
1983	1.5	2	2	2	2.5	2.5	3	3	3.5	4	4	2.73	.10
1986	2	2	2.5	3	3	3	3.5	3.5	4	4	4.5	3.18	

Figure 6.7. Comparison of 1983 and 1986 students' primary trait scores for "Controlling Variables." Shaded areas show students receiving scores of 3.0 ("adequate") or better. $\bar{x}$ = mean score. P = the probability under the null hypothesis that improvement is due to chance is 10 in 100.

compose self-contained headings (consistent with APA style guidelines) for their fish graphics.

Looking at the 1983 graphics, we concluded that the two-day "mini-unit" was not a smashing success. Although some graphics were good (see Jim Wilkerson's table on p. 188) and some were poor (see Mike Siliato's table on p. 204), most of them were just mediocre. The students' rough drafts document that they did not experiment with data in different graphic forms. Their tapes and logs also indicate that they picked the graphic first such as, "I'll put a pie graph in my paper" or "I will put some stuff in a table," rather than taking the information and deciding which graphic would most effectively illustrate their findings. Less successful students seemed to reason that any three graphics would do in much the same manner that less successful writers seem to view writing as coming up with a specified number of words. For example, Jeremy Logan records that he "prepared graphics"; we found out that meant he soaked off the Joy label and put it in his paper as "Figure 3."

Teacher's Methods, Student Performance, and Implications

In 1986 Anderson gave the same lecture on graphics. She handed out the same fish data as a homework assignment. But to make explicit the connection between the way they illustrated fish data and the graphics students should use in their household products experiment, she spent the last 20 minutes of each class having students discuss in groups what kinds of household product data would be appropriate to each type of graphic. Anderson emphasized that scientific writers focus on *selecting*, not constructing, the most appropriate graphic in the rough draft stage. The 1986 students used more graphics and they used a greater variety of graphics, as Table 6.6 indicates.

Clearly, providing students with in-class, teacher-directed time to transfer learning principles was quantitatively productive. The primary trait scores for collecting and interpreting data indicates that the 1986 students constructed more effective graphics as well (see Table A.1, Appendix A, p. 247).

INTERPRETING DATA

The Nature of the Difficulties

The other essential skill in using evidence to support a position was interpreting data. Although students collected and communicated data

Table 6.6 Graphics Strategies (1983 and 1986 Classes)

Strategy	Totals 1983	1986
Students having 3 required graphics	9	11
Students having more than 3 graphics	4	7
Students having diagrams	2	7
Students having line graphs	1	4
Students having rating scales	3	6
Number of Graphics included in all papers	38	55

N = 11 students (1983); N = 11 students (1986)

in the Results section of the scientific journal format, they had to make sense of that data in the Discussion. In this section, the students had to "put it all together," as Wilkerson did in his Discussion of the better erasable pen.

Teacher's Methods, Student Performance, and Implications

In both 1983 and 1986, Anderson told both groups that the Discussion, Conclusions, and Implications section should "summarize the research, accept or reject the hypothesis, and explain the significance of the research in terms of price and quality." Figure 6.8 compares 1983 and 1986 students' primary trait scores for interpreting data.

Since Anderson's teaching methods for interpreting data were virtually identical, we wondered *how* and *why* 1986 students had been more successful than the 1983 students. First, we focused on quantitative differences between what each class included within the Discussion section. But as we examined the data, we found that differences between the two groups in summarizing, accepting, or rejecting the hypothesis, and discussing price adequately, were small and did not involve the same students. The striking difference was that the 1986 students designed experiments that contained more evidence to support a position (see Table 6.7). Further, they refined both their data collection

												$\bar{x}$	P Value
1983	**1**	**1.5**	**2.5**	**3**	**3**	**3**	**3**	**3**	**3.5**	**4**	**4.5**	**2.90**	**.03**
1986	**2.5**	**3**	**3**	**3.5**	**3.5**	**3.5**	**4**	**4**	**4**	**4**	**4.5**	**3.59**	

Figure 6.8. Comparison of 1983 and 1986 students' primary trait scores for "Interpreting Data." Shaded areas show students receiving scores of 3.0 ("adequate") or better. $\bar{x}$ = mean score. P = the probability under the null hypothesis that improvement is due to chance is 3 in 100.

Table 6.7 Use of Evidence (1983 and 1986 Classes)

Strategy	1983	1986
Students using special subgroups for interpreting data (see Tables 6.3 and 6.4)	4	7
Students achieving an average of 3.0 or above on the primary trait checklist for other skills	4	8*
Students interpreting data more than adequately (3.5 or above on the primary trait checklist)	3	8*

N = 11 students (1983); N = 11 students (1986) *Same population

and their conclusions by designating subgroups for whom one product might be better (see Tables 6.3 and 6.4).

We highlight with an asterisk two groups of 8 students because they were *identical* groups. We conclude, therefore, that the differences in the 1983 and 1986 students' achievement in interpreting data can be attributed to the fact that the 1986 students designed better experiments, made better operational definitions, controlled more variables, and collected better data.

We have discussed difficulties that concerned Anderson's expectations for use of the scientific method. Now we discuss the difficulties that arose as students tried to meet Anderson's expectations for the scientific report format.

DIFFICULTIES WITH ORGANIZING THE PAPER

Because the research paper had a prescribed format, students faced two types of organizational problems. They had to organize their information in order to fit it into the appropriate format sections—Title, Introduction, Methods and Materials, Results, and Discussion—and they also had to organize the appropriate information logically within a section.

FORMAT ORGANIZATION

The Nature of the Difficulties

Although practicing scientists have traditionally regarded a standardized, research-report format as an asset, it was the students' nemesis. An interview between 1983 students Hilary Nearing and Susan Bell reveals how frustrating students found this organizational task. For the interview with each other, Nearing and Bell are using a peer review sheet that Anderson has provided, the first question of which is, "I had trouble writing the _____. What part of the paper did you find hardest?" Nearing, who plans to compare a high-phosphate and a low-phosphate detergent to find out which is better, replies:

Nearing: Well, see, I had percent of phosphates in one cup of detergent was 6.3 in one and .3 in the other. Now I don't know whether to put that in my results because it *is*—

Bell: Well . . .

Nearing: Because my thing is—

Bell: Well, see, usually it's, well . . .

Nearing: Or would it go in the Conclusion [Discussion section]. See, that, that's particularly what I had trouble with.

Bell: I don't know. Results of data, maybe you could put? Under your ta— under your Data and Results? And then put that as your first table with the main characteristics of your, of each detergent, you know, if it's broken up into that?

Nearing: Oh. Well, I'm going to have a bar graph representing that, just to show people.

Bell: In your Conclusion? And where's that going to go?

Nearing: In my, in my—

Bell: In your Results?

Nearing: My Results—

Bell: 'Cause then that would be OK, I mean, that would be telling that information then. I mean if you were going to have a graph for it in your Results and then, then you really don't have to expl—

Nearing: [unintelligible syllable] OK go ahead.

Bell: Then you really don't have to explain it.

Nearing: Price is kind of relative, because it's going to be varied with the amount of phosphate.

Bell: Is that what your main factor is?

Nearing: Yeah, that's like a variable. Well, I'm going to do phosphate-free and then Cheer which has lots of phosphate.

Bell: Oh, OK, oh.

Nearing: So I put the amount of phosphate in my Results. It's pretty important for them to know it before they read the Discussion. How about looking at it that way?

Bell: Well, if you're going to have a graph. Are you planning on putting price in the Results? Is that what you're—

Nearing: Yeah.

Bell: How are you going to list that? Are you going to write it out?

Nearing: Yeah, write it out.

Bell: Then I would write the other factors in there, too. In the Results, not the Conclusion. If you're going to put price in the Results,

Nearing: No, I put the price in the Conclusion.

Bell: Oh. Let's see.

This "Who's on first?" routine went on in a peer conference, at the end of which Nearing asked Anderson if she should include the price in the Results. The answer, of course, was no, and Anderson reminded her that, as a nonexperimental factor, it should be stated in the Methods and Materials section first, then discussed at the end.

Students who did not ask for guidance often constructed inappropriate guidelines for themselves. Sharon Tissinger, on her tape, reads a price and says, "It has numbers in it; it goes in Results." Quantification is a characteristic of all scientific writing, not just the Results portion of the format. By the same token, Nearing's percent of phosphate per cup of detergent does not belong in the Results just because it is expressed as a number. It belongs in the Introduction in order to clarify that Nearing has chosen to compare the cleaning abilities of a high-phosphate and a low-phosphate detergent.

In essence, 1983 students placed materials in the wrong sections. They omitted sections. They invented sections. They even put some things into two sections because they couldn't decide which section they belonged in.

When we read the students' logs, we discovered a major reason why they had difficulties: they believed that the order of the format dictated the order of composing. Jeremy Lucas's log reveals this notion:

March 3rd.	Bought 32 oz. each (Joy and Palmolive)
March 4th.	Used about 15 minutes to think of how to start the paper.
March 5th.	Spent 1 hour in the library to write the first page of the paper. (Introduction)
March 10th.	Compared stain removal by these two products. Also compared physical differences.

March 14th.	Wrote 1 and 1/2 pages on the materials and methods. I spent 1 hour on this.
March 17th.	Tested spotlessness on two cups. Still thinking of data for the graphs.
March 18th.	At 2 p.m. I wrote the remaining part of the methods and materials.
	[Lucas has 12 more log entries running through April 9th, 1983. He never mentions the Introduction again]

Believing, as Lucas did, that the order of the format dictated the order of composing, students tried to write their original research papers in sections from the beginning, rather than writing rough drafts of all related information and revising the drafts to meet format demands at a later date. Many 1983 students were never able to resolve the conflict between the order of the format and the order of the composing process—a difficulty that also plagued Sherman's students (p. 80). When Lucas began writing the Introduction, he had not conducted a pilot and had not experimented with the two dish detergents in any way. He seems to have begun with writing the Introduction because, like nine of his ten classmates in 1983, he thought that was where everybody started. Jane Chance, a 1985 student who made a "C" on the original research paper but a "B" in the course, thought so too. When she was asked by a peer which part of the paper was the hardest part to write, she said:

> The Introduction, no question. I didn't have too much problem writing the Methods and Materials and the Results other than phraseology, I suppose. The Introduction I had the most problem with—how to lead into it . . . I went at it the wrong way. I tried to sit down and hammer it out, the Introduction, before I did anything else. I finally had to switch and just write the Methods and Materials and then fill in the Introduction from there.

After struggling with the Introduction, Lucas moved on to the next section of the format—the Methods and Materials section. On March 10, he tested the spot removal qualities of the dish soap. The only writing he did was to record data. He wrote nothing about his procedures, what he found out, or what it meant, because he was too busy writing the first section of Methods and Materials. He finally wrote about his spot test four days after his experimentation. It is not surprising that he left out important information and organized his data poorly. Mike Siliato separated by a number of weeks the experimentation and writing in his "scrub now, write later" plan; he too was unsuccessful.

In addition to the major problem—letting the order of the format dictate the order of composing—some students also thought there should be "transitions" from section to section. In Anderson's view, one function of the scientific format divisions, such as Introduction, Methods and Materials, etc., is to eliminate the need for transitions.

Teacher's Methods, Student Performance, and Implications

In 1983, Anderson lectured on the parts of the scientific journal article and their functions. At that time the students had already read five research articles arranged in that format. She asked the students to "bring in three pages of your Methods and Materials section." This was the basis of a 40-minute peer conference session in which students were free to address other concerns about the paper as well.

Anderson's conclusion? "What a mistake!" By asking the students for three pages from the Methods and Materials section, she forced them to do exactly the activity they needed help with *before* they got to class. She also perpetuated, if not created, for the students the myth that scientists compose their reports in format sections. One student even scratched out text that would have been good for her Conclusions section because she decided it didn't fit in Methods and Materials, and she never brought it back.

In 1986, to counter these problems, Anderson, in a 15-minute lecture on format, clarified that successful scientists write before, during, and after research, and then rearrange their written text to meet format demands. She stressed that the order of the sections in a scientific research paper does not determine the order of composition, and encouraged students to link experimenting and writing together as closely as possible. She listed each unit of the scientific article and explained its function, then gave the students a short research article that she had cut into chunks. In groups of three or four, the students put the "format puzzle" together, the cut-and-paste activity replacing her longer 1983 lecture. All of this helped students to focus on the issue of format in reading a scientific article, and gave them hands-on experience in manipulating material within that format.

Also in 1986, Anderson gave the students a reprint of one of her own published articles to serve as a format model. It is clear from notes, tapes, and rough drafts that the students referred to it for organizing the entire report.

In the next assignment, the students were asked to "bring in three pages of your research." Notice the big change: Anderson did not ask for a specific section of the paper, she asked only for pages of text, thus reinforcing the notion that format did not dictate the order of composition. In class, peer groups discussed where the text they had written would fit within the format. Toward the end of this session, Anderson explained to the students *why* transitions were not necessary between sections because the format itself resolved that writing problem.

But one problem arose with the students' in-class discussion of their pages of text. Although it was Anderson's intent that the three-page assignment would be a "hands on" time for students to put information into all five format sections, as it turned out, students' three pages usually contained information that belonged only in Methods and Materials. Thus the groups worked almost exclusively on Methods and Materials because those were the only pages they had. The primary trait scores for "format organization" measured the way students placed information in all sections. For that reason, we were not surprised that "Scientific Format" was the writing category that showed the smallest margin of increase (Figure 6.9).

SECTION ORGANIZATION

The Nature of the Difficulties

Anderson's students not only struggled to determine which section their information should go into, they also had difficulty in organizing information within the sections. It was a common problem; for example, 45 percent of students in 1983 had a primary trait score below 3.0 for the Methods and Materials section—a score partly dependent upon organizing that section.

Logs and tapes made clear that students thought chronological order was always important. Hilary Nearing and Susan Bell agreed that the Methods and Materials section was easier than the other sections, able to be written, as they said in their peer-response session, "zoom, zoom, zoom." However, they often included extraneous material arranged in

												$\bar{x}$	$\underline{P}$ Value
1983	2	2	2	2	2.5	2.5	4	4	4	4.5	4.5	3.09	.31
1986	2	2	2	2.5	3	3	4	4.5	4.5	4.5	4.5	3.32	

Figure 6.9. Comparison of 1983 and 1986 students' primary trait scores for "Scientific Format." Shaded areas show students receiving scores of 3.0 ("adequate") or better. $\bar{x}$ = mean score. $\underline{P}$ = the probability under the null hypothesis that improvement is due to chance is 31 in 100.

merely chronological order. Bell wrote in her Methods and Materials section about conducting the "first, second, third, and fourth tests." Since these were not a sequence of time-related tests, she should have referred to them as "maximum absorbency test," "hand dryability test," and so forth. Further, the order of the tests in Materials and Methods should have corresponded to the order in the Results section.

In a short, informal assignment asking students to reevaluate this paper and explain what they would do differently next time, Duncan Solski, a 1986 student, replied,

> The overall feeling of the paper was written in chronological order, with words like *then, next,* and *after* to show sequence of events. I think leaving these words out would make the paper sound more like gathered research information and less like a story.

Teacher's Methods, Student Performance, and Implications

In addition to the activities designed to help students with format organization as a whole, Anderson in 1986 instituted an activity to help students move away from chronological organization, particularly in their Methods and Materials sections. In the last peer conferencing session, when students were working on revisions, Anderson asked them to exchange papers and "Circle all the words on one page of the Methods and Materials that imply chronology." After this, they were to reread this section and determine whether the chronology was significant or simply the result of an inappropriate narrative approach. The primary trait scores on the Methods and Materials section went from 3.00 in 1983 to 3.55 in 1986 (Figure 6.10).

ANDERSON'S AND WALVOORD'S CONCLUSIONS

This chapter has explored the nature of students' difficulties in the same six areas of difficulty we constructed for all four disciplines.

												$\bar{x}$	P Value
1983	1.5	1.5	2	2	2.5	3	3.5	3.5	4	4.5	5	3.00	.14
1986	1.5	2.5	3	3	3	3.5	4	4.5	4.5	4.5	5	3.55	

Figure 6.10. Comparison of 1983 and 1986 students' primary trait scores for "Methods and Materials." Shaded areas show students receiving scores of 3.0 ("adequate") or better. $\bar{x}$ = mean score. P = the probability under the null hypothesis that improvement is due to chance is 14 in 100.

Though students in Anderson's class had to address the good/better/best questions within the scientific framework, and though some of their difficulties arose from the scientific requirement to quantify data, nonetheless, we found that their difficulties fell into the same general categories as those of students in the other classes.

We found similarities, too, in students' strategies. As in the other three disciplines, difficulties arose with roles. Anderson wanted the professional-in-training role of scientist. Many students adopted this role, but we also saw traces of lay roles (the advice-giving parent, the storyteller) and text-processor roles (students relying in inappropriate ways on the product labels). Students experienced difficulties not only in adopting the role of scientist, but also in performing it appropriately—that is, using the scientific method and writing their reports in the appropriate format.

Anderson's teaching methods in 1983 were already using peer and teacher response. Our investigation of the 1983 class showed Anderson where students were still having difficulties, and how her methods were either working in contradiction to the processes she wanted them to use, failing to offer appropriate guidance when it was needed, or placing too much reliance on the advice of peers before they were able to help one another. To direct her changes, Anderson was guided by the information we had gathered and by her own strong conviction that students trying to engage in complex reasoning and methodology need concrete experiences under the guidance of their teacher. In our terms, she taught procedural knowledge procedurally.

After we examined the data from 1983 and Anderson implemented changes, the 1986 section of students performed better, according to the judgment of outside raters, in those areas that Anderson had addressed. Given the small sample (11 students in each class) and our lack of full information about other factors, such as SAT scores (lacking for the many Towson State students who were transfers), we certainly do not have a scientific basis for proving that the improvement resulted from the changed teaching methods. Nonetheless, students improved in every category (Appendix A). In eight of the eleven primary trait categories, the probability that any difference was due to chance was less than 15 percent. In eight of the eleven categories, the probability was less than 15 percent. We think that the changes in teaching methods are likely candidates for helping to explain the improvement. Moreover, our findings about the nature of students' difficulties may provide useful clues about students in other settings trying to learn scientific processes and scientific writing.

But this chapter is not just a report of an "experiment" on our part.

It is the story of a living classroom, a story of teacher and students working together in order that learning could take place for all of them, a story of a teacher's growth and change. Anderson's classroom will never be the same again. So the only way to "replicate" our "experiment" is for other teachers to do what we did: systematically and collaboratively observe students, and then, guided by the best theories and intuitions at their command, try to shape teaching methods that address the difficulties that observation has revealed.

Note

1. Brand names used by the students have been retained for authenticity. However this is student work. In no way do we imply judgments about the relative merits of any product named in this chapter.