4

EVALUATION

4.1 INTRODUCTION

A summary of what we have learned

Students care deeply about how they are evaluated, and small details in the way that students are graded have disproportionate effects on how they perceive the course. The other early users of *Six Ideas* and I have experimented with a variety of evaluation schemes during the past decade, and we have learned that:

1. *Exams and quizzes tell students what you value*, so it is very important to design exams and quizzes appropriately (see **sections 4.2** and **4.3**).
2. *Weekly quizzes can help students keep up* (see **section 4.4**), but *daily assignments* (of some kind) *do this better* (see **section 4.8**).
3. *Standard homework problems generally build the wrong skills,* so problems should be carefully designed and chosen (**section 4.5**).
4. *Students can correct their own homework*, and this has some real pedagogical and practical advantages (see **sections 4.7** and **4.8**).
5. *Homework grades should depend mostly on effort* for this course to work successfully (see **section 4.6**)
6. *I don't know how to use computerized assessment software* in a way that is consistent with the goals of the Six Ideas course (see **section 4.9**).
7. *Evaluation contracts* may lead to more positive attitudes about the course but have other problems (see **section 4.10**).

4.2 DESIGNING EXAMS AND QUIZZES

Exams and quizzes tell students what to learn

Students pay a *lot* of attention to what exams and quizzes tell them about what is important about the course and exactly what they should be learning. If you are trying to get students to develop a deep and flexible understanding of the qualitative concepts of physics and an ability to use these principles to solve realistic problems, then an hour-long multiple-choice exam consisting entirely of short problems with quantitative answers sends the wrong message. Considering the impact that they have on a student's grade, your exams will probably speak louder to them about what you value than whatever you might say in class.

Things that a *Six Ideas* exam should test

Therefore, whatever you really want your students to learn needs to appear in some way on the exam. In a *Six Ideas* course, this means that in addition to short questions with quantitative answers, exams and quizzes should include questions or problems that test a student's ability to

1. reason using *qualitative* principles of physics,
2. supply and interpret *symbolic* as well as numerical results,
3. make appropriate *approximations* and *simplifications*,
4. use specific *problem-solving techniques* (such as the frameworks, interaction diagrams, and free-body diagrams), and
5. set up and solve complicated realistic problems.

There are several things that make doing this difficult. (1) Grading essay-type responses (which are almost required to test the last two items on the list) can be very time-consuming and/or impractical, (2) solving even *one* realistic problem can take more than the hour one can allot to a typical unit exam, and (3) it is difficult to design exams that examine the issues listed above. Here are some strategies for dealing with these problems.

Some suggestions for constructing good exams

(1) **"Easy to Grade" Doesn't Have to Mean "Quantitative Multiple-Choice."** The two-minute problems in the text as sell as the quizzes and exams chapter 10 illustrate a variety of questions that deeply probe students' *qualitative* understanding of physics concepts and yet ask for only relatively simple responses (multiple-choice, draw a vector, draw a force diagram, sketch a curve). Even questions that require a more complicated response (for example, an essay) can be made easier to grade using the techniques discussed below.

(2) **Divide complicated problems into parts.** It would be nice to test each student's ability to solve a realistic, complex problem from scratch, but such problems require unrealistic amounts of time and are difficult to grade consistently. I have found it helpful to divide complicated problems into shorter parts, each of which is more specific and therefore easier to grade. (I also try to make it possible to solve the parts independently so that students stuck on one part can complete other parts, sometimes by providing hypothetical answers that students can use to go on if they can't find the real answer.)

(3) **Ask students to do only a *part* of a complicated problem.** A complex realistic problem can be made both shorter and easier to grade if you only ask students to complete a part of it. For example, you could (a) ask them to set up the equations for a problem but not solve those equations, (b) fill in a missing part of an otherwise complete derivation or worksheet, (c) ask them to outline a solution, (d) ask them to correct a solution having deliberate errors.

(4) **Construct a grading chart.** With essay problems, I often set up a grading chart that lists the features that I am looking for and outlines how the score on that problem depends on these features. This saves an enormous amount of decision time (see the example exams).

(5) **Bite the bullet.** Some of these techniques make it *easier* to grade complicated-response questions, but nothing is going to be as easy as a purely multiple-choice exam. You will have to weigh for yourself the importance of testing students in these ways with the practicalities of your own situation.

(6) **Use the provided exams.** It is not easy to create exams that probe deeply, test a range of skills, and yet can be completed in an hour and are easy to grade. See chapter 10 for exams that you might use instead of making up your own. This could save you more time than you lose in grading essay questions.

4.3 IMPROVING ATTITUDES ABOUT EXAMS

Suggestions for lowering student anxiety

Exams cause students a lot of anxiety and stress, much of which is unnecessary. Here are some ways that students' attitudes about exams can be improved (without coddling them).

(1) **Keep exams short.** Too much time pressure strongly affects students' attitudes and performance (and often leads to sloppy responses that are hard to grade). The rule that I have developed over time is that if I can write satisfactory responses to the exam questions in *less* than 15 minutes, it is short enough for students to take in 50 minutes.

(2) **OR let them take their time.** I have experimented with scheduling an evening time to take a exam so that students could take as much time as they needed to do the exam. Students *generally* really like this (and will vote for it, given the choice), but *some* will complain that it means committing yet more time outside of class to the course. The logistics of setting up such a session become exponentially more difficult as class size increases.

(3) **Keep the mean score relatively high.** If the average hardworking student can only get about half of the exam right, such a student will feel stupid and discouraged, no matter what you might say about making the exam deliberately difficult to spread the grades, etc. We have found that students feel much better about themselves and their work if exams are designed so that the likely mean score is somewhere above 70%.

(4) **Set absolute standards, but adjust grades if needed.** Students graded on a curve feel that they are competing with each other. A pre-announced grading scale (for example, ≥ 90% = A, ≥ 80% = B and so on) has the advantage of setting standards and encouraging students to meet them *without* competing with each other. The disadvantage is that it is easy to create a exam that is too hard and thus discouraging to hardworking students. What we do is announce a grading scale ahead of time (promising that if everyone scores above 90%, everyone will get A's), but then adjust students' letter grades (or scores) upward if *necessary* so that the average class grade is at least on the B–/C+ boundary.

(5) **Possibly allow students to make up exam points.** Another way to encourage students (and motivate them to study) is to have a section on the final where they can make up points on a previous exam. If done properly, this can help students feel like they have some power to rectify past mistakes and motivate them to review material that caused them problems before.

(6) **Let them use a cheat sheet.** We have been allowing students to use various-sized sheets or cards as "cheat sheets". This makes it clear that the exam will *not* be about regurgitation of formulas and physical constants, but rather about *using* these formulas and constants effectively. Even though students rarely actually use them, it helps them study and makes them feel better.

4.4 SCHEDULING EXAMS AND QUIZZES

The format of the *Six Ideas* course make it natural to schedule one 50-minute exam per unit. At Pomona, students typically take the last unit exam and a comprehensive final of similar length during their 3-hour final exam period.

We have also found it advantageous in past years to schedule regular quizzes. These quizzes consisted of a single page of simple-response, mostly qualitative questions (*deliberately* much like the two-minute problems in the text) and are designed to be done in 10 minutes. We give the quiz at the end of the class period (partly so that students are not distracted by their quiz performance during class and partly so that students can grab a little time from the interclass period if they need to). These quizzes take about 1 minute per student to grade. Examples of such quizzes are found in chapter 10.

Advantages of regular quizzes

The advantages of regular quizzes are as follows. (1) They help students keep up. (2) They emphasize the importance and relevance of qualitative understanding in general and two-minute problems in particular. (3) They give students an opportunity to get feedback about how they are doing on a regular basis.

Surprisingly, we also find that students actually *like* having these quizzes: when we have asked students on evaluations whether they would prefer doing weekly quizzes, biweekly quizzes, or no quizzes, most respond that they prefer the weekly quizzes or biweekly quizzes, and hardly anybody prefers no quizzes. Their comments indicate that they like the extra motivation that the quizzes provide to keep up with the course.

Some disadvantages

The main disadvantage is that *creating* a good quiz takes quite a bit of time (it takes me about 2 hours). You can save some time by using the quizzes provided in chapter 10 of this manual. Important secondary disadvantages are that the quizzes do eat up precious class time and do take time to grade and process.

In the past, I gave *weekly* quizzes in an attempt to encourage students to prepare for class on a daily basis. Evaluations showed that the quizzes helped students keep up on a *weekly* basis, but had little effect on the degree to which they came prepared to individual class sessions. The daily homework scheme described in section 4.8 proved to be much more effective at doing this.

At Pomona, we currently offer a quiz every other week or so (two or three quizzes per unit) and a 50-minute essay exam at the end of each unit. Most of the example materials that you will find in chapter 10 of this manual reflect that scheme.

4.5 HOMEWORK PROBLEM CATEGORIES

The problem with traditional homework problems

Researchers in physics education have documented some serious problems with the standard pattern that homework problems in almost any introductory text follow. Such problems tend to tersely describe the situation, clearly detailing the exact information required to solve the problem, and then ask for a specific and well-defined quantitative result. (A reason that one might actually *care* about the result is only rarely provided.)

What is wrong with this approach? Because such problems specify exactly the information needed and require a specific quantitative answer, these problems can be solved by simply doing a random search for the formula in the text that links the knowns and unknowns (understanding of the *physics* involved is not required). Unfortunately, students quickly discover that this is also usually the *fastest* method for solving the problem. One particularly troubling study [D. Hammer, "Two approaches to learning physics," *Phys. Teacher* **27**(12), 1989] tracked the progress of two young women through an introductory physics course. One was genuinely curious about physics and made a real effort to understand deeply the physical concepts discussed. The other adopted whatever strategies would get her the highest grade with the least effort. The first found that she could not keep up with the course or get a reasonable grade *until* she adopted the strategies of the latter. Other research has shown that when standard problems are assigned for group work, the groups tend to do individual searches without discussing any of the physical principles involved in the problem.

Problem categories

The homework problems in the *Six Ideas* texts are designed to address this issue. These problems are divided into four categories according to the type of reasoning required to solve them: *basic, synthetic, rich-context,* and *advanced*.

Basic problems are most like the standard problems, though this category includes some qualitative problems as well as standard problems. These problems encourage students to apply a *single* formula or concept from the chapter to answer a question. The purpose of these problems is to have students practice using the formula in preparation for more complicated problems. I rarely assign such problems, but encourage students to work them on their own as a study aid.

Synthetic problems require that students bring together (synthesize) several concepts and/or formulas in multi-step solutions. These problems often involve more realistic situations, sometimes provide too much information (requiring students to choose the relevant information), and sometimes ask for an algebraic or qualitative answer instead of a quantitative one. Most of the homework problems I assign come from this category.

Rich-context problems are meant to be most like the problems that one might really encounter in real life: they generally describe a more or less realistic situation in narrative form. These problems generally require a deeper level of synthesis, and often provide too *little* information as well as irrelevant information, requiring students to look up or estimate quantities. These problems almost always ask a *question* instead of asking for a numerical quantity, forcing students to decide what they might calculate to answer such a question. I usually assign one such problem a week.

Advanced problems require advanced mathematics, unusually complicated algebra, or unusually sophisticated reasoning. These problems are provided to challenge the very best students in the class (and/or show you the instructor how to resolve various subtle theoretical issues that might come up).

The majority of problems assigned should be at the synthetic level

While these categories are listed in generally increasing order of difficulty, they overlap somewhat: for example, an especially challenging synthetic problem may be more difficult than the easiest rich-context problem. *Rich-context* problems are especially designed for group work (see section 3.11) and most are likely too difficult for individuals. On the other hand, *basic* and *synthetic* problems are intended to be within the capability of any hardworking student in the class. Homework assignments should consist primarily of synthetic problems.

4.6 EFFORT-BASED GRADING IS ESSENTIAL!

A traditional approach to grading homework can lead to serious problems!!!

Doing homework problems is absolutely essential for building students' physics skills: this is where they will probably do most of their learning. However, *evaluating* homework can be very time-consuming and (if not done correctly) harmful to the atmosphere of the course.

In 1994-1995, two professors new to the *Six Ideas* course at Pomona tried offering the course using a traditional weekly homework scheme (where students' homework grades were based on the correctness of their first and only effort). Students found this very daunting, received poor grades, and became very discouraged. Their dissatisfaction eventually poisoned the atmosphere of the course, preventing them from learning nearly as well as they could have.

Why traditional homework grading doesn't work

The homework problems in the *Six Ideas* text are *not* designed to be used this way. My desire to make the problems more realistic means that they are also deliberately more open-ended than traditional problems. This makes it difficult for students to ensure that they get the right answer by slavishly following some nearly identical example in the text. Most of the problems are also designed to make fruitless the strategy of randomly searching for a single formula that has variables corresponding to the knowns and unknowns in the problem. For these reasons (and others), the problems in this book are generally more challenging than problems in traditional texts. To ask students to do all of these problems correctly the first time is to ask too much.

It is important that homework grades depend *mostly* on effort

Students can really benefit from such problems, though, if they are allowed to *try* to do such problems and then learn from their mistakes without getting discouraged. This means using some kind of grading scheme that awards points *mostly* for effort. Such a scheme fits in with the philosophy of "homework as practice" that I have outlined earlier and (as the anecdote illustrates) avoids a serious possible failure mode.

After the 1994-1995 debacle, I polled users of the text about this issue. In situations where the text worked well, the instructors universally used *some* kind of homework grading scheme that rewarded students' efforts and allowed them to make mistakes without getting a bad grade.

4.7 AN EXPLICIT EFFORT-BASED SCHEME

A two-pass system for grading weekly homework

Over a number of years, I developed a scheme where students correct their *own* homework. This scheme (which I used for many years) has many of the benefits of expert grading without requiring the same investment in resources. For reasons I will discuss in the next section, this scheme has proved more practical for upper-level courses (where I continue to use it) than for the introductory course, but (1) it provides a useful starting place, and (2) it is more "tried and true" than the scheme I will describe in section 4.8, and (3) it provides a useful example of a grading scheme that *explicitly* rewards effort.

The physical resources required to make the scheme practical are a locked "mailbox" where students submit work to be graded and a set of hanging file folders (one labeled file folder per student) for returning graded work.

For the sake of discussion, let's assume that students are given weekly assignments of six to eight synthetic problems and one rich-context problem every week. Students submit initial efforts for these problems to the homework "mailbox" on Monday by class time. These initial efforts are graded as follows:

4 = solution is complete and well-written
3 = solution is missing minor parts or some important explanations
2 = solution is missing major parts and/or has few if any explanations
1 = very little coherent initial effort was expended
0 = no initial solution was submitted

The grader does *not* write comments about the *content* of the solution at this point, but may request additional details, explanations, or the like to be supplied in the next round. The grader then returns the solutions to the students' folders before class on Wednesday. Problem solutions are either handed out, posted in a display case, and/or posted on the web at this time.

Students then correct their own work using the provided solutions by tracking down their errors and correcting them and/or adding missing elements as needed. All changes are made with a colored pen. Students then return their corrected solutions to the box on by Friday.

The grader then looks at the solution again, and asks two questions: (1) Did the student correct their initial work completely? (2) To what extent did the initial solution *need* correction? The grader then assigns a score of 0 to 3 points to the correction *effort* according to the following chart:

3 = all errors were caught and corrected in a satisfactory way
(OR no correction was needed and this was correctly recognized)
2 = minor errors escaped correction
1 = major errors escaped correction
0 = the solution needed a complete rewrite and nothing was done

The grader also assigns a score of 0 to 3 depending on the extent to which the initial effort *required* correction in the first place, as follows:

3 = no or very little correction was necessary
2 = minor errors needed correcting
1 = major errors needed correcting
0 = the solution needed to be virtually rewritten

Students can thus earn a score of up to 10 points on each problem. (I have often graded the weekly rich-context problem on a 20-point scale to allow for finer distinctions within the categories.)

The scheme is explicitly effort based

Note that students can earn *most* of the points on a given problem simply by making an honest initial effort and then correcting it carefully. This is deliberate. My philosophy (following the sports metaphor developed in chapter 1) is that homework is where students *practice* problem-solving skills and thus they should be rewarded mostly for *showing up* for practice rather than just on how *well* they practice. The exams provide the arena where the skills that they have learned in practice are tested and evaluated.

Advantages of this scheme

This scheme has several important advantages:

1. **It forces students to study the solutions.** Doing this not only asks them to think critically about their work, but also exposes them to expert problem-solving styles and techniques. (Our experience is that students will usually ignore provided solutions unless forced to examine them this way.)
2. **It gets more bang for the buck.** Grading goes *much* faster, because students have (normally) found and written comments on their own errors. For multiple reasons, I usually did the grading myself when I used this system: it took me about 2 to 3 minutes per student per pass, much less time than I would need to grade papers in the traditional manner. The grading also does not require as much expertise to do well: I found that undergraduate TAs were able to do a reasonable job.
3. **It rewards effort** more than right answers, as discussed previously.
4. **It decreases the number of students in your office begging for points.** The grading is more objective than the traditional scheme (particularly when you are not doing the grading personally) and the generous effort-based scale means that students have less grounds for complaint. (I also help this process by allowing students to drop a certain number of their lowest scores. This also means that students don't have to ask as often for permission to turn in late papers.)

However, there are also some disadvantages.

Disadvantages

1. **The scheme involves quite a bit of paper shuffling.** The homework “mailbox” and student file folders previously described are essential for making the scheme practical for large classes. Even so, the probability of errors and glitches seems to rise exponentially with class size.

2. **It is possible to cheat** by making corrections appear as if they were original. One can require that students submit a xeroxed copy of the weekly problems initially and the corrected original later, but this is a hassle for the students and kills trees. I have found that just *threatening* to xerox a few random initial efforts and compare them to the final sets kept cheating on weekly problems sets under control at Pomona.

3. **Solutions must be accessible but not *too* accessible.** I handed out xeroxed solutions until a couple of years ago, when I caught a student copying his roommate’s handouts from a previous year. I now post the solutions in a display case and in non-printable form on the web, but this does limit student accessibility and involves more work for the instructor.

4. **Introductory students need instruction in making corrections.** Upper-level students know what a good solution looks like and have learned to be self-critical, but introductory students often have trouble recognizing conceptual errors and/or flaws in logic. They therefore either over correct (by essentially recopying the solution, even if their solution is fine) or under correct (by missing important errors). Either approach limits the effectiveness of doing the correction, and makes grading correction “effort” difficult.

5. **The grading rubric does not fit all situations well.** For example, consider a student who carefully and correctly does the wrong problem. Does this student all of initial effort points? How should he or she “correct” the problem? Or consider a student who makes a fundamental conceptual error that makes a problem seem simpler than it actually is. Do they get the full initial effort points, or should the penalty be only assessed in the “quality of initial effort” step, in spite of the basic nature of the error? These kinds of problems are again more acute with introductory students than they are with upper-level students, and in larger classes as opposed to smaller classes.

Because of these problems and because I have found it important to assign at least some homework for each class session (see section 4.8), I no longer use this scheme for the *Six Ideas* course. Still, it illustrates a time-tested and very reasonable effort-based grading scheme that you might find useful.

4.8 A DAILY HOMEWORK SCHEME

Daily homework is the best way we’ve found to get students to come to class prepared

It is *very* important in this course that students read the book *before* class. However, I mentioned in section 2.5 that it is *very* difficult to get students to do this. Over the years, I have tried many different kinds of carrots and sticks (including counting on student embarrassment during active-learning exercises, weekly quizzes, reading quizzes at the beginning of class, web-based reading quizzes, and so on). Only one method has proved both elegant and effective with my students: requiring that students hand in solutions for one or two problems at the beginning of the class from the chapter assigned for that class.

It is not really fair to hold students *too* accountable for correctly solving problems on material not yet discussed in class, but they are willing to do such problems if (1) the reasons for assigning the problems ahead of class are explained and (2) if grading is primarily based on effort. Still, there is a tricky balance to be achieved: basing grading *entirely* on effort prompts students to treat the assignment too casually, but grading too much on the basis of (initial) correctness generates student complaints that the assignment is unfair.

A 1.3-pass grading system for daily homework

The following is a slightly modified version of the scheme we used for the *Six Ideas* course at Pomona during the spring of 2002 (I'd like to thank David Tanenbaum for working with me to co-develop this approach). For each class day, we assign two problems from the assigned chapter (usually two of its simpler synthetic problems, though sometimes one may be a basic problem). Students hand these daily problems to an "initial effort" tray as they come into class. After class, each problem is graded on the basis of both correctness *and* effort on a 5 point scale, using the following rubric as an approximate guide:

0: no initial effort
1: a poor initial effort
2: a fair effort involving modest conceptual errors
or a good effort involving serious conceptual errors
3: a good effort with modest conceptual errors and/or math errors
or a fair effort with minor errors
4: a good effort with only a few very minor errors
5: good effort with correct results *and* reasoning

A person who does the wrong problem will receive no more than a 2. A "good effort" implies that the student supplies *some* explanation and/or diagrams with calculations and has noted if the answer is obviously incorrect. If I am grading, I do not write any comments unless I think that students will not see an important error. I return the papers the following class session and post solutions.

Students *may* then submit corrections in colored ink to a "corrections" tray at the beginning of any subsequent class session up to and including the session a week after the original due date. The submitted problems are given a final score on the same scale subject only to the following constraints:

1. the correction must explicitly locate the student's error of reasoning or algebra (i.e. it cannot be just a paraphrase of the solution), and
2. the final score cannot exceed the initial score by more than 2 points.

This is an *implicitly* effort-based grading system

Note that even though students' solutions are always explicitly graded on *quality*, this scheme *implicitly* gives credit for effort because students who have read the book, made a serious effort initially, and submitted a good correction will almost always earn either a 4 or a 5 on a problem. (The "punishment" for not doing well on the initial submission is mostly the extra work required!)

Advantages of this scheme

This scheme has the following advantages over the one in section 4.7:

1. **Grading is easier.** The single-category 5-point rubric and lack of comments makes grading easier than the 10-point multiple-category scale used in the previous scheme. The rubric is simple enough that I think that undergraduate TAs can do the grading in a reasonably objective way.

2. **The scheme is more efficient.** Students do not have to examine or resubmit correct problems, and are not *required* to resubmit anything. In practice, we found last year that only about 30% of the problems were resubmitted for a second grading pass (so this is a "1.3-pass" grading system).

3. **Incentives for cheating are very low.** Since the initial grade is recorded, students cannot cheat by passing off a correction as an initial effort, and since a pretty modest initial effort can still earn a good score, there is little incentive to cheat on the initial effort.

4. **The system is easier for students to understand.** Students are used to having problems evaluated on quality.

5. **The system provides an adequate incentive to read the book** without requiring separate goads. Students reported (on anonymous evaluations) much higher numbers for the percentage of classes for which they were prepared during semesters where we required daily homework.

This scheme still has some disadvantages. The paper shuffle must still be managed carefully so that everything flows smoothly. Solutions still have to be posted somehow. One should also be careful to develop sound policies about regarding late papers, students who forget to bring papers to class, students who have to miss class, and so on (though this is true of any scheme).

Remaining disadvantages

We supplement these daily problems with a weekly set of three problems due on Monday, one problem from each of the previous week's chapters. One of these problems is a rich-context problem that students can elect to work on during a collaborative "recitation" section held during the weekend. Because I think that learning to write coherent solutions with good style is an important skill, I grade these problems on an 8-point scale, with the extra three points devoted to quality-of-presentation issues such as the following: Does the solution provide adequate diagrams and/or explanations? Does the solution use units and vector notation correctly? Does the solution avoid doing algebra with numbers? The grading rubric looks like this:

Weekly problems round out this scheme

0: extremely poor presentation
1: major presentation problems
2: minor presentation problems
3: great presentation

Again, students can submit a correction to these weekly problems within a week subject to the two-point increase rule.

In spite of the paper shuffle, this is (I think) the best and most elegant homework scheme we have ever developed for evaluating *Six Ideas* homework at Pomona College. Your mileage, however, may vary: I present this scheme not because I think that it is *the* right one, but because it might give you some ideas for creating a scheme that works well at your institution.

Use this as a springboard for your own scheme!

4.9 ASSESSMENT SOFTWARE

Over the years, I have carefully examined computerized grading software (the premier example being *WebAssign*). I have had mixed feelings about such software. On the one hand, it is clear that there are some real advantages, particularly in for professors facing very large classes.

However, I remain concerned about a number of severe disadvantages. It is very bad, I think, to focus student attention on "right answers" (either numeric or symbolic) at this stage. A problem solution, in my opinion, is really about constructing a *model* of the situation that accurately predicts the result, and the model to me is far more important than the final result. Students need to be focused on the model-building process instead of on any final results. (I constantly have them repeat the mantra "the answer is the *model*").

The core problem with computerized homework evaluation: the answer is the *model*

But successfully *describing* a model usually requires a mix of words, symbolic equations, and/or diagrams (such as interaction diagrams, free-body diagrams, coordinate systems, circuit diagrams, and so on). I have not yet discovered a good way for students to enter an acceptable model description into a program such as *WebAssign* that would be any easier to grade than physical papers.

This is not to say that it cannot be done; I am only saying that I don't know how to do it. I would be very happy to hear creative ideas about how use such software to grade the quality of models.

4.10 EVALUATION CONTRACTS

The last edition of this manual touted *evaluation contracts* as being useful for maintaining good class morale and a good way to get students to *choose for themselves* to prepare appropriately for class. We have since abandoned such contracts, but they might still prove to be a useful tool for you. The purpose of this section is to describe the scheme and its advantages and disadvantages.

A description of a possible evaluation contract

Here is how the scheme works (in abstract). Let's say for the sake of argument that 25% of a student's grade for a semester is determined by their work for each of the three units in that semester. (The remaining 25% being based on lab work and perhaps a final comprehensive exam). At the beginning of each unit, students have the opportunity to sign a contract regarding how their grade will be calculated *for that unit*. The choices might be:

1. *A practice-intensive contract* where 60% of a student's unit grade depends on weekly homework, daily homework, and/or class participation and 40% on exams and quiz scores. The contract's text expresses the expectation that the student will come to class prepared to participate.

2. *A test-intensive contract* where 80% of a student's unit grade depends on the unit exam (and maybe quiz scores), and 20% on the weekly rich-context problem. This contract carries *no* expectation of class participation or attendance (indeed, unprepared students may even be encouraged *not* to attend).

Before students sign the contract for the first unit, we discuss the advantages and disadvantages of the two contracts carefully, emphasizing what research says about the importance of active reading and class participation for effective learning. Students do not need to sign contracts for subsequent units unless they intend to switch their grading modes. If they do plan to switch, they much do so within a certain window of time at the beginning of each unit.

Advantages of using evaluation contracts

Some students genuinely feel that they work best alone and on their own schedule, and that with a little study, they can perform well on exams: these students will choose the second contract. (Some of these students learn a valuable lesson when they get the first exam back, but there a few students who do just fine.) *Most* students will freely choose the practice-intensive contract, partly because it gives them more control over their grade. But the very fact that they can *choose* makes all of the students feel more respected as individuals, and makes those choosing the practice contract feel better about (and thus more likely to do) what they are contracted to do.

This is not by any means a complete solution by itself to the problem of student preparation; many students who sign the contract still generally need specific and concrete incentives to come to class prepared. The main advantages of the contract system are that it (1) improves class morale, (2) increases students' willingness to follow the practice-intensive regimen, (3) keeps the few students who might be especially resistant to the active-learning atmosphere of class from disrupting class-sessions and poisoning the atmosphere, and (4) provides a natural excuse to educate students about the importance of active learning.

At Pomona, we found this scheme helpful for allaying student's concerns about the transition from a traditional lecture class to an active *Six Ideas* class, and also for keeping students unalterably opposed to preparing for class and participating in active-learning exercises from disrupting the experience for others.

Why we no longer use such contracts at Pomona

However, after more than a decade of teaching *Six Ideas* at Pomona, the format of the class is well-known, and the handful of good students who are really put off by this approach have other options. We have also found that as we have evolved the course design, it has become more elegant and less bizarre to students. It is also not at all clear from the evidence that students take the contract aspect very seriously. But the most important problem with evaluation contracts, we have found that (in the last several years especially) the vast majority of people electing to sign the test-intensive contract really *needed* the practice-intensive option to succeed in the course, and so allowing them to make such a choice ultimately allowed these students to hurt themselves. For all of these reasons, we have now discontinued offering evaluation contracts.

Again, however, your mileage may vary. The system was useful at Pomona for a time, and might be at your institution as well. I offer it as one possible approach to dealing with certain kinds of problems.