

TASK-ORIENTED MATH EDUCATION

FRANK QUINN

ABSTRACT. “Learning tasks” on which students work independently with support by helpers and web materials provide an approach to math education. Experience at the Math Emporium at Virginia Tech demonstrates educational effectiveness at the college level and suggests it should work in upper grades in K-12. Implementation would be tricky so the factors involved are considered carefully and in detail. Benefits could include significant improvement in the quality and effects of high-stakes tests. Many of the educational advantages come from giving students more choices and more control over their learning.

CONTENTS

1. Goals and descriptions	2
1.1. Descriptions	2
1.2. Discussion	2
1.3. Summary	3
2. General Constraints	3
2.1. Non-terminal courses in a task-oriented subject	3
2.2. Students capable of modestly independent work	4
2.3. Computer-based	4
2.4. Helpers	4
2.5. Traditional class meetings	5
2.6. Summary	6
3. More about Tasks	6
3.1. Tasks are not Assessments	6
3.2. Learning-goals and strategies	7
3.3. Task Constraints	7
3.4. Multiple tries in assessment	8
3.5. Software generation	9
3.6. High-stakes tests	9
3.7. Summary	10
4. Course Design	10
4.1. Not An Online Course	10
4.2. The Skeleton	11
4.3. Student behavior	14
4.4. Beyond the Skeleton	15
4.5. Summary	17
5. Resource Requirements	17
5.1. Increased resources are not an option	17
5.2. Operating expenses	18
5.3. Startup	19
5.4. Development	21
5.5. Summary	23
6. Educational Opportunities	23
6.1. A Behavioral Model	23
6.2. The Main Strategy	24
6.3. Tracked Courses	24
6.4. Summary	25
7. Conclusions	25

1. GOALS AND DESCRIPTIONS

The long-term goal is to improve math outcomes in K-12 and the first two years of college. As a professor at a university with large science and engineering programs I am particularly anxious for significant improvement in the top 5-10% of high school graduates.

The educational system is highly stressed and traditional instruction seems to have reached a limit. Better outcomes apparently require a new approach, but so far there have been as many ways to fail as there have been new approaches.

This article presents yet another new approach, with strategies for avoiding all the modes of failure I have been able to identify. Because there are so many of these modes, and because avoiding one often causes trouble with another, the description is detailed and complicated.

1.1. Descriptions. We begin with two contrasting descriptions. The rest of the article can be seen as an attempt to reconcile the two.

1.1.1. Sympathetic Description. Task-oriented courses enable students to use modern learning resources in ways that best suit their individual learning styles. Course objectives are formulated as a sequence of tasks to be mastered. Students are provided with an array of web materials, video and audio presentations, printed materials, and access to individual helpers. Other opportunities might include traditional lectures, study groups, or group projects. Students choose or combine these resources with the freedom that they have come to expect with the internet, games, television programming etc. Learning is richer and more efficient than is possible in traditional classrooms. Finally, because it is more efficient, expectations can be raised without serious rise in failure rates.

1.1.2. Critical Description. This approach amounts to having “pass the test” as the course objective, and in traditional classrooms is called “teaching to the test”. It fragments material into discrete tasks and weakens development of conceptual context and connections. The result is learning that is mechanical, disconnected, and short-term. The use of new-age materials may engage students but will not fix the underlying shortcomings, and the idea that outcomes would actually be better is a fantasy. There are other problems common to most novel programs: they are usually economically unrealistic, particularly in being seriously over-budget in demands on faculty time; and heavy dependence on computers make them ineffective for a significant number of students. Neither of these would be acceptable even if educational outcomes for most students were satisfactory.

1.2. Discussion. The objections raised in the critical description are between 99% and 100% valid. The question is whether there is even a 1% window for success, and if so whether we can design a program with enough care and sophistication to squeeze through it. Specifically, is there any way a task-oriented program could provide outcomes at least as good as traditional programs, for the same student population, and within the same time and money budgets?

Reasons vary in different communities but the general conclusion would be “no”. For instance task orientation is incompatible with basic tenets of the K-12 education community, e.g. as formulated in NCTM publications.

Not long ago I also would have dismissed the idea as nonsense. As a university professor I highly value conceptual context and connections. High-school AP calculus is a prime example of a teach-to-the-test course and I have spent a lot of time getting students out of that mode so they can be successful at the university level. But I now feel that effective task-oriented courses may be possible, and indeed may have significant advantages.

The change of heart is due to experience with a computer-tested calculus course and programs in the Math Emporium at Virginia Tech. I watched and worked with students to see how they used materials, then modified the materials to work better when used that way. In effect the students taught me how to construct an effective learning environment.

I discovered that students were using practice tests as study guides. Diagnostic aids, comments, and links to reference materials were added to make this more effective. Considerable effort went into designing problems so that abstract understanding gave a problem-solving advantage. And as the materials matured students used them differently. The description of a “task-oriented learning program” is an attempt to formulate what students are actually doing. But the fact that students want to learn this way is only useful if learning goals can be met.

Learning goals are being met in the main course involved, second-semester calculus for science and engineering. Virginia Tech has strong science programs and a large engineering school so this is a key course. Weak outcomes, higher dropout rates, reduced content, or increased cost would not be acceptable. In the last four years thousands of students have taken the course divided roughly equally between task-oriented and traditional sections and with a common final exam, so there is a lot of data. Detailed analysis will be presented in another essay.

The course was not explicitly developed to be task-oriented, and is still evolving. Traditional lectures are still provided, for instance. Nonetheless it provides good evidence that the idea is workable.

1.3. Summary. Experience with a university calculus course suggests that educationally effective task-oriented courses are possible but there are a great many ways to fail. The remainder of the article describes failure modes and attempts to chart a way through.

2. GENERAL CONSTRAINTS

Tasks as described here do not provide a general approach to education. In this section we describe some of the limitations and interpret them as constraints on topics and levels where the approach could succeed. In particular the idea shows most promise in mathematics; it might be useful in other contexts but we do not speculate on this.

The limitations described here will also appear as constraints on program design in later sections.

2.1. Non-terminal courses in a task-oriented subject. Non-terminal math courses up through calculus and differential equations are essentially task-oriented. *Non-terminal* means that the ideas and skills acquired in the course are expected to be used in a later course on mathematics, science, engineering, business, etc. The bottom line for the later course is ability to routinely and accurately solve certain types of problems. Abstract understanding can be helpful or even essential for flexible and effective problem solving. In a real sense this is the job of abstract understanding in math. Understanding that does not support problem-solving is dysfunctional from the later-course perspective. Therefore it makes sense to approach even abstract understanding through tasks in these courses.

Terminal courses (not intended to be used later) typically aim for cultural exposure and a softer understanding that does not have to support problem-solving. A

task approach is less appropriate for these courses. It may work anyway: most of our task-oriented courses are actually more-or-less terminal. Lower-level and possibly terminal courses may work well as tracks in a task-based course, see §6.3 Tracked Courses. However for simplicity we focus on non-terminal courses.

2.2. Students capable of modestly independent work. In a task-oriented course students take the initiative in selecting tools and developing and implementing learning strategies, at least on a small (single-problem) scale. This requires some maturity and purposefulness. We emphasize that these are not on-line courses and do not require nearly as much independence as on-line courses, see §4.1.

We have made no attempt to adapt the approach to very young students and have no guess as to what might be needed or what the limits might be.

2.3. Computer-based. Large numbers of practice tasks are needed. Web links and interactivity are required to make them an effective learning environment. As a result tasks must be provided in electronic format and much of the work done on computers. We return to this in §3 More About Tasks.

2.4. Helpers. Human helpers are essential for most students. Helpers do not teach in the traditional information-delivery sense: their role is to help students who get stuck. Students develop skill at locating and correcting minor errors, and diagnostic aids are provided to help with this. But any student will occasionally get stuck in a way he or she cannot unravel. The helper diagnoses the specific problem and shows the student how to repair it.

2.4.1. Constraints. When appropriately offered, help sessions are short and the average total time required per student is less than in traditional classroom instruction. In other words helper time and expertise are leveraged. This is not a sure thing and making it work seems to require the following:

- Help should be in person. We have tried a number of schemes for on-line help and found them unsatisfactory. A problem that requires a helper is by definition one that the student cannot locate or articulate, and in these situations direct interaction and observation of body language are often essential.
- Help should be quickly available when it is needed. In schools this means opportunities to work in a single location (computer lab) with helpers available to respond to help requests.
- Helpers should circulate in the work area and go to students when they need help. This results in short, targeted interventions. If students have to relocate they tend to spend more time stuck and often collect a list of problems to make relocation worthwhile. They then want to settle in for an extended tutoring session to work through the problems and reconstruct the specific difficulty in each one. This is less efficient for both students and helpers and often has the effect of making helpers unavailable for other students.

Some students do need extended tutoring and we provide this as a separate resource to keep them from tying up the helpers.

2.4.2. Opportunities. Helpers can be effective with considerably less background and preparation than would be required to teach the course. In fact most of our helpers are advanced students. This provides a number of opportunities.

- There is a severe shortage of fully-qualified math teachers. Use of less-qualified helpers, e.g. advanced students, teachers with expertise in other areas, or even parent volunteers gives a way to leverage the skills of the teachers available.
- Having older students work with young ones benefits both groups. There have been proposals to incorporate this into school curricula either as a “highly encouraged” volunteer activity or as a required component of a course. Helping would incorporate it as a paid part-time job.

Paying student helpers is feasible because efficiencies elsewhere make it possible without increasing the overall budget, see §5.2 Operating Expenses. It is a good idea because it would be important to attract the best older students; developing help skills takes time and effort; and the system depends on reliable participation for most of the school term.

It would be interesting to see the effect if excellent performance in math courses guaranteed a part-time job in the senior year. Help experience might also make teaching more attractive as a profession.

- Providing high-quality math instruction is one of the biggest challenges of home schooling. But tasks as described here are computer-based so they would be available anywhere, and background sufficient for helping (rather than teaching) would enable parents to use them successfully.

2.4.3. *Proctors.* Proctors are needed to supervise computer-based tests:

- Check ID and sign in students;
- ensure disallowed materials are not brought into the test area; and
- activate for-credit tests on the machine.

Since tests are multiple-try and not tightly scheduled, they must be available for extended periods and demand is unpredictable. The way we handle this is to use one end of the lab for testing. The actual area reserved for testing expands or contracts according to need.

The number of proctors needed also varies unpredictably. This is handled the same way: when the test area expands helpers are reassigned to proctoring, and when testing contracts they are released back to helping. As a result we consider proctoring as part of the help process rather than as a separate job.

2.5. **Traditional class meetings.** Our course with a strong task orientation still provides traditional lectures as a resource. Most students find that with all the other resources the lectures are not necessary. Some student attend faithfully even though they don’t get additional credit. Attendance has—amazingly—essentially no correlation with outcomes. This needs additional study but it may reflect learning styles: students who learn best in a class come to class, and those who can efficiently use other resources don’t come. It does result in a closer and more interested class atmosphere. In any case it seems likely that success for all students will require some sort of lecture-style component, but it probably should not be compulsory.

2.6. **Summary.** A task-oriented approach may be appropriate for non-terminal math courses from approximately fifth grade through university calculus and differential equations. Materials are primarily computer-based, and opportunities must

be provided to work in an area with qualified helpers available. Some students will probably need a class or lecture component to be fully successful.

In following sections we discuss additional requirements for success in these contexts.

3. MORE ABOUT TASKS

In practice a task is presented as a collection of practice tests.

3.1. Tasks are not Assessments. There is a vital distinction that must be emphasized immediately. Traditional tests are assessments not intended to directly influence instruction. To the extent that they do, the influence is bad. “Teach-to-the-test” has a bad reputation for good reasons, and an attempt to base task-oriented learning on a traditional assessment test can be confidently expected to fail.

Differences between learning tasks and assessment tests include:

- Learning tasks are harder. Assessments frequently use simple special cases or spot-check to avoid excessive time or computation requirements. But if this guides learning then students only learn simple cases and will skip things missed by the spot-checks. Effective learning tasks must be in some way comprehensive and represent the full complexity of problems that arise in later study. Below we describe how to accomplish this.
- Learning tasks are frequently more abstract. For instance a test question on area formulas might be “What is the side length of a square with the same area as a circle of radius 6?” The numerical formulation gives students an opportunity to use calculator skills, and for test designers it has the advantage that they can get a whole family of apparently different questions by changing the number.

Questions like this are bad learning goals. We really want students to be able to do it for a circle of symbolic radius r : set the area formulas equal, $\pi r^2 = s^2$, and solve for s to get $s = r\sqrt{\pi}$. Different number versions become “Plug $r = 6$ into $r\sqrt{\pi}$ ”, “Plug $r = 7$ into $r\sqrt{\pi}$ ” etc. The numerical aspect is completely mechanical and we really don’t *want* students to see different numbers as giving different problems. A focus on numerical versions actually inhibits development of symbolic skills. Consequently learning tasks should be, for the most part, not numerical.

- Learning tasks must incorporate conceptual material by making it directly useful in problem-solving. Assessment tests tend to be formula-oriented and the role of conceptual understanding is essentially to help students choose the right formulas. Students trying to learn from them will see only the formulas. Making concepts directly useful is difficult but usually possible, and the effort often leads to deeper understanding on the part of the course developer! See the Preparation for Technical Careers web site <http://amstechnicalcareers.wikidot.com> sponsored by the American Mathematical Society for examples.
- Learning tasks must support learning. When a student looks at a problem and thinks “how do I do this?” or “I thought I knew how to do this but I can’t get the right answer” there must be some way to make progress. Typically this includes links to reference material with concise descriptions

of principles and worked-out examples. Current textbooks work poorly for targeted references: wikipedia might be a better model. Problem-specific diagnostic aids can help locate errors. Complete solutions are not so helpful: some students confuse “see how it is done” and “learn how to do it”.

3.2. Learning-goals and strategies. The student view of the process is:

- there is a test that has to be taken for a grade;
- the test is computer-generated so is actually a huge number of essentially equivalent instances rather than a single static thing. It (more precisely, different instances of it) can be taken multiple times with the highest score being the final grade;
- there is a time window during which the test can be taken for credit, with a very firm deadline;
- students can get an unlimited number of practice versions generated in *exactly the same way* as the for-credit versions; and
- there are various resources available to help with figuring it out.

The intent is that students will look at several practice tests to get an idea of what needs to be done. This will vary widely: a few will be able to do most of it immediately while a few will have a long way to go. But if they have kept up and previous courses have done their jobs then students should be able to identify their individual problem areas fairly quickly. In other words, students should be able to be able to formulate learning goals on the basis of four or five practice tests, and should be able to develop a strategy for dealing with difficulties they encounter.

3.3. Task Constraints. Our objective is to make the student view work, not fight it. This presents some serious challenges.

3.3.1. No shortcuts. Most students know that traditional tests have weaknesses:

- tests focus on simple cases and conceptual material is usually not tested;
- problems on computer tests (or human-written ones for that matter) are usually drawn from a limited database and enough practice versions will show essentially all of them; and
- most tests have structural weaknesses. Knowing how a test is constructed and scored can give a student a statistical advantage independent of content knowledge, and many students have taken courses in test-taking strategies that exploit this.

We have had students download fifty practice tests, presumably looking for repetition, systematic weaknesses or omissions. Judging by outcomes they were not successful. This is already a difficult accomplishment but it is not good enough: the goal is not just to make this a waste of time, but to make it quickly clear (after seeing four or five instances) that it *will* be a waste of time. This does not mean that serious problems on every topic must appear on every test, but they must appear often enough to convince students that learning the material will be the simplest way—and the only reliable way—to be consistently successful.

3.3.2. Consistency. Students should see different test instances as essentially similar in several ways.

- Layout: If learning goals are formulated on the basis of four practice tests then a fifth should fit into the framework. In practice this means that if the

first two problems on one test concern topic B then the first two problems on any other should also concern topic B. They might be easy on one and tough on another, and there are exceptions, but as a rule topics should be consistent.

A common and cheap way to make assessment tests look different is to scramble the questions. This is inappropriate for learning tasks because it interferes with goal and strategy formulation. Depending on scrambling is also an easily-discovered structural weakness.

- **Difficulty:** Tests should be consistent in overall difficulty. First, a realistic test must omit or simplify something, so a tough problem on topic B might be balanced by easy questions on topic C. Balancing difficulty does not undercut learning as long as students know they have to be prepared for the balance to go the other way on the next instance. Second, any real-life system will produce some instances that are genuinely harder than others. Students seem to expect this and are not bothered by it *provided* it doesn't happen often, the worst instances are never truly horrible, and they can take the for-credit versions multiple times.
- **Not adaptive:** Adaptive tests are also multiple-try, but the test system tracks results and when a student demonstrates success with one topic it is omitted from later tests and the focus shifts to other topics. One drawback of this is described in *Layout* above; here we give another.

Suppose a test has ten problems and a student wants a score of 80%. To be reasonably sure of getting this he needs enough mastery of the topics of six or seven problems to be sure he can get them right, and a good enough grasp of the remaining ones to have a 50-50 chance on each. Such a student will finish the course with a good mastery of most of the material. Recall that the courses under consideration are non-terminal, i.e. needed for further study in science, math, or some other technical subject, so mastery is an important objective. An adaptive approach would allow students to relax after achieving success (or having good luck) but before achieving mastery.

3.4. Multiple tries in assessment. In previous discussions we have assumed or asserted that assessment should be done with multiple-try tests. Here we explain why.

First, it does not pose additional difficulty in task design and development. We have emphasized that task assessment, motivation, etc. are maximized when instances of the same "test" are used for both practice and assessment. This means it must provide many equivalent instances, etc. and therefore be *suitable* for multiple-try use whether it is used that way or not.

The reason multiple tries are necessary is that a tight practice-assessment linkage has drawbacks and allowing multiple tries addresses or compensates for these drawbacks.

- Learning tasks must be harder than traditional assessments and computer grading makes partial credit impossible. As tests they often strike traditional teachers as seriously unrealistic. Students do better than might be expected because goals and standards are clear and there are no surprises. Nonetheless there is a lot of exposure to minor errors, and being able to retake the test compensates for this.

- Some instances are a bit harder than others, or a student might find one variation particularly challenging. The recourse is to retake the test.
- Multiple tries provide opportunities and incentives for improvement. It often happens that after a test a student realizes that he could do better with relatively little effort or more care (“made a dumb error”). Having another try makes this an opportunity rather than just a frustration. If it is reasonably easy to retake then some students will do it even if they already have a satisfactory grade.
- Some students cannot resist peeking at answers and hints available in practice tests, and use them as crutches rather than learning aids. Roughly speaking they confuse “being able to do it” and “seeing how it is done”. For these students proctored-for-credit tests can provide enforced-discipline practice. Fortunately this mostly effects students new to the system and they grow out of it.

3.5. Software generation. Our tasks are generated by software that writes problems directly rather than taking them from a database. Problem-generating modules take parameters that determine problem type, difficulty, etc. so a single module can provide a number of problem types, and a very large number of instances of each type. This has substantial advantages over the database approach:

- higher quality;
- greater flexibility and variety;
- much better control for balancing problem types and difficulty and ensuring full coverage;
- better quality control and problem-specific diagnostic hints;
- straightforward upgrades; and
- very low maintenance costs after development. Currently our calculus task-generating software has run for two years with almost no modification.

Software that writes problems encodes subject knowledge and educational wisdom in a way that individual problems cannot. This is a big factor in making tasks effective as learning guides and I do not believe this could be done satisfactorily with a database-oriented system.

Encoded knowledge and wisdom accounts for low maintenance costs: once the software is mature most adjustments can be made by modifying input parameters rather than modifying generator code. A less welcome consequence is that development cost are likely to be high. This is discussed in §5.4 Development.

3.6. High-stakes tests. There are now state-level K-12 math tests and some movement toward regional or national tests. The way these tests are used to grade schools has forced a teach-to-the-test response in many systems. However the tests currently in use are poor as assessment instruments and *very* poor as learning guides, so no good can come of this.

Bad high-stakes tests will undermine a task-based system. They both encourage a teach-to-the-test approach, even if for very different reasons, and if there is a disconnect between the two the high-stakes test will win.

An ideal solution is to use the same system to generate course tasks and high-stakes tests. This would be straightforward with the software problem-generators described in the previous section. Benefits are:

- high-stakes tests of high quality and designed to support learning;

- customization for state or local needs accomplished by customizing parameter settings rather than the entire test;
- synergy: motivation provided by course grades and high-stakes tests reinforce rather than conflict; and
- enormous savings in construction of high-stakes tests. Currently these are expensive and have to be redone every year. Yearly costs with the software system would be negligible by comparison.

Details and other benefits are discussed in §5.4 Development.

3.7. Summary. Students see a typical learning task as a multiple-try test with practice versions and various bells and whistles. But the objective is genuinely effective learning when used in ways that seem natural to students, not just assessment. To be successful the materials must meet different and much more demanding standards than needed for pure assessment. This section described requirements imposed by the format and the ways students use the materials. The next section describes requirements imposed by the way they are used in a course.

4. COURSE DESIGN

This section discusses how tasks can be used in a course. After clarifying the goals we begin with a stripped-down version. Possible enhancements are then described. Much of the design is shaped by student psychology and behavior. Resource constraints are discussed in the next section.

4.1. Not An Online Course. Our tasks are available online and some students use them as an online course. Our goals are quite different from those of online courses, however.

Online courses do not have to count failures. An online can be considered an outstanding success and make a lot of money even if it cannot be used by 50% of the target population.

Public schools do have to count failures. A school that “left behind” 50% of its students would be considered a catastrophic failure.

Public colleges and universities are not so different. We may have selective admission but only the most elite could restrict admission to students capable of taking ambitious online courses. If the Math Emporium at Virginia Tech discontinued help support I am sure we would have catastrophic failure rates.

The point is that helpers, supplementary lectures, and other features that make this proposal complicated and difficult are consequences of our determination to make quality education accessible to all students, not just an elite.

4.2. The Skeleton. Tasks, with their supporting resources, are the backbone of a task-oriented course. A skeletal course needs some connective tissue but little else. In this subsection we expand on this and explain why some traditional features can be omitted without causing problems while others require adjustment. Topics are course segments, grading, and high-level guides.

4.2.1. Segments. The course is divided into segments with a task (test) to be mastered in each segment. The task can be taken for a grade during the segment and becomes unavailable when the segment ends. There should be a makeup policy for tests missed for legitimate reasons but it must be restrictive enough that students

will not use it to postpone work. Our calculus course is divided into six two-week segments and a final exam. Considerations are:

- serious deadlines are necessary to keep students moving through the material;
- there must be sufficiently many segments so that students can handle the material covered in each; and
- there must be sufficiently few that each one is significant. Students cannot afford to skip one, and can work up the motivation to tackle them.

For university students and our course, two-week segments seem to be a good balance. At least a week is needed for the learning mechanisms and multiple tries to work and too many tasks may overwhelm interest and motivation.

Two-week segments may be appropriate for grades 5–12 also. In practice students relax in the first week and get down to work in the second. Task orientation is more efficient in use of student time because they focus on their own needs and choose the most effective resources. Traditional courses have uniform assignments that everyone is supposed to do whether they need it or not. The tradeoff is that tasks require more focus and active participation and therefore more motivation. The consequence is that the second week in a segment produces at least as much learning as two weeks in a traditional program, but it also requires as much focus and motivation as two traditional weeks. Trying to reduce the “inert” periods is likely to reduce engagement.

Another advantage of an easy week–hard week rhythm is that the relaxed periods enable some sort of mental digestion or long-term memory formation. This varies from person to person but there seem to be limits on how fast humans can effectively absorb material like mathematics. Learning in individual segments can be faster, but these need to be paced to avoid cumulative overload.

Class meetings could also be adapted to a two-week rhythm. Meetings could be held in the first week, and the second week reserved for work in the lab and testing. Two sections could then be run concurrently with a one-week shift: one would be in class while the other is in the lab. Other advantages of this idea are discussed in §5.2.2 Classroom Teachers and §6.3 Tracked Courses.

Finally, there will be some students who need the whole two-week segment to get the work done.

4.2.2. *Grades.* In a skeletal course the tasks and final exam are computer-graded and these grades are the sole assessments in the course. There is no homework per se, no quizzes, no extra credit, no dropped grades, and grades are not curved. These will be discussed individually but there are two general points.

First, most of these practices are artifacts of the constraints of traditional classrooms and in other settings their objectives are better achieved in other ways.

Second, these practices reduce the connection between performance and grades. But our context is a non-terminal course covering material needed later, and test performance is a bottom-line measure of preparation for later use. Disconnecting performance and grades undercuts course objectives.

In detail:

- Homework: Repetitive practice is vital for learning mathematics and this is the traditional role of homework. Equally traditionally, students see it more as busywork than mission-critical. They have to get credit to be

willing to do it. Standards tend to be relaxed so good grades are easy and students see them as a buffer against bad test scores. Low standards may be misleading: why should something that is acceptable on homework be wrong on a test? Finally corrections to homework errors have minimal impact. The student was not really engaged in the first place, genuine difficulties are hidden among sloppy errors, and there is too much of it for either the student or teacher to review carefully.

In a task-oriented course practice tests provide repetitive practice. Students see work on practice tests as directly mission-related so they do it voluntarily without credit and take it seriously. Standards are uniform. Students make an effort to avoid or correct sloppy errors, and are engaged so that when they do make an error they actually want to know how to fix it.

The difference between “homework” and “practice test” is partly psychological, and reinforcing this is another reason practice tasks should be authoritative guides to the assessment versions.

Most sections of our task-oriented calculus course do not require homework. Some teachers have been nervous about this and did require homework. It seems to have no effect on outcomes.

- Quizzes: The function of quizzes is to force students to stay engaged and compel class attendance. In a task-oriented course students are supposed to engage on their own schedule. There are problems with this, see e.g. “procrastination” below, but there are ways to address them that are more consistent with the course design. Class meetings should be considered resources rather than the main show, and some students will not need them. Rather than forcing attendance, students should be lured by making classes efficient and useful as resources.
- Curves, extra credit, dropped grades: These practices undercut learning in several ways. First, do teachers give grades, or do students earn grades? Grade curves etc. are at the teacher’s discretion so when they are used the answer is “give”. If enough students go limp the teacher will rescue them with a curve. Scores can be appealed and grades negotiated so the focus is often on the teacher as a potential patron, not on the material. I have had students whose negotiation skills were far stronger than their study skills.

The second problem with these practices is that they disconnect grades from performance. We are discussing non-terminal courses so content and standards are designed to support later work. Extra credit and dropping low scores essentially enable students to skip part of the material, often the most significant part. Grade curves lower performance standards on the remaining material. The result is poorly prepared students. Generous use can turn a non-terminal course into a terminal one.

To summarize: in the skeletal task-oriented course all assessments come from computer-graded tests and expectations are made clear by practice versions. If adjustments are impossible—grades are earned, not given—then students accept these expectations and get to work. Even a hint that adjustments are possible can damage motivation: as things get tough, when does negotiation become a better bet than further work? This is particularly an issue with learning tasks since they must be more difficult than traditional tests.

A final benefit of computer-assigned grades is that it improves student-teacher relationships. In traditional classes there is a tension between the teacher's roles as evaluator and as mentor; here the teacher is completely on the student's side. Pure mentoring is also a more consistently positive and enjoyable experience.

4.2.3. *Reference Texts.* Tasks guide learning at the segment level. Careful design can provide connections but for the most part high-level coherence must be provided other ways. The most important of these is a hierarchical web-based document along the lines of Wikipedia. This is the least well explored aspect of the proposal so details are uncertain and will depend on level, but some general principles are clear.

- Individual entries should be relatively self-contained and dependencies made explicit with links. The reason is that they will be used at unpredictable times to aid recall or sharpen understanding, not be read linearly like a classical textbook.
- Reference entries should not be designed for first-exposure learning because this would reduce usefulness for reference. Expanded presentations—generally viewed only once—should be provided for that, though in fact a great many students will be able to learn directly from the reference text.
- Entries are short, precise, and functional. In mathematical terms they should be more like definitions than explanations.
- Examples, alternate viewpoints, etc. should be given but details should be provided through links to avoid bloating and obscuring the main point.
- There should be *no distractors*: sidebars, cute graphics, video clips or animations. These are doubtful in ordinary single-use texts and irritating and counterproductive in reference texts.
- Graphic illustrations should be clearly relevant and carefully explained.
- The most detailed entries—twigs in the graph structure—typically relate to task problems and are targets of links in diagnostic aids.

Current textbooks are inappropriate in nearly every way.

4.2.4. *Presentations.* Lectures or presentations should be provided. These will probably be videos at higher levels and live in elementary grades. Videos should be linked to the reference text and may also fit together in a linear sequence like a traditional course.

- Presentations are considered part of the skeleton rather an enrichment because there are a significant number of students whose primary learning modes are best addressed this way. At higher levels most of them can survive without this support but the benefits far outweigh costs.
- Presentations generally will be viewed only once; after that most students will use the reference text. Consequently the text should be developed first and presentations coordinated with it.
- Presentations duplicate some low-level material available through tasks so students who find tasks more efficient will tend to skip them.

Introducing tasks as the evaluation component of a traditional course seems to be the best way to modify an existing curriculum. Our most-ambitious task-oriented course is evolving this way and still offers traditional lectures. We have other courses that work satisfactorily with online texts and presentations instead

of class meetings but most of these are terminal or near-terminal, and some have content compromises, so their materials may not be good models.

4.3. Student behavior. Some behaviors are addressed differently in a task-oriented course. Here we discuss procrastination and disruption.

4.3.1. Procrastination. The need to combat procrastination has driven development of the main features of standard courses: homework, quizzes, and periodic major tests. It is apparently one of the key problems in education. Procrastination is difficult to measure so it rarely figures in modern data-oriented studies, but it should be a major concern for any proposal that involves changing course structure.

I was led to this realization by the data rather than being clever enough to figure it out for myself. Multiple-try tests do provide an indicator of procrastination: waiting until the very end of the segment to take the test for credit. Students who did this had importantly lower scores than either students in general or the same students on tests started earlier. There was enough data to reveal many statistically significant correlations but this was by far the most important and the only one that clearly required action.

The intent in a task-oriented course is that work should be organized and initiated by students, and standard anti-procrastination measures would work against this. Instead we use psychological countermeasures. They work for us but we have no great confidence that they will be sufficient in other contexts.

- **Impending Doom:** In this approach the number of times a task can be taken for credit goes down as the deadline approaches. There is a maximum of two tries a day, and only one on the last day. Thus someone starting two days before the last could take the test five times, starting one day before allows three tries, and this goes down to one at the end. In practice many students take the test only once and very few take it more than three times. Nonetheless the steady evaporation of opportunity does provide enough motivation to greatly reduce the grade disparity.

There is a subtle point here. The Impending Doom strategy reduces the grade disparity more than it reduces the number of students waiting until the end. This and other factors suggest (but don't prove) that there is a sub-population—perhaps 10%—who either work effectively under pressure or already know the material, and waiting until the end has no disadvantage for them. This illustrates the importance of identifying the real problem (grade disparity) rather than focusing on an easily-measured correlate (waiting until the end). Countermeasures focused on the correlate might actually be counterproductive for some students.

- **Preemptive Strike:** This strategy requires the test to be taken for credit in the first few days of the segment. The penalty for missing it might be a 10% reduction in whatever score is finally earned. One objective is to ensure an early start on task assessment, at least at the subconscious level.

Another objective is to provide a default grade. Officially no one cares about the score because in the end only the best score counts. Generally scores will be bad. The psychological difference is that as the end approaches students have to think about fixing a bad grade, rather than the more abstract idea that they should allow time to fix a grade if it turns out bad.

We have not used the Preemptive Strike strategy, but plan to try it in the near future.

4.3.2. *Disruption.* Attentive students are easy to teach. A few obviously inattentive students in a class can noticeably pollute the learning environment. One actively disruptive student can degrade the environment enough to make real learning very difficult. Practicing teachers know this and disruption is easy to measure but—incredibly—it goes almost unmentioned in the educational research literature¹. Some educational approaches, the Discovery method for example, are quite vulnerable to disruption and descriptions really should include warnings about this.

A skeletal task-oriented course is relatively insensitive to disruption. Group activities such as lectures are optional so disinterested students generally don't come, and there no reason not to ask a disruptive student to leave. The most important point, however, is that computer-side help is one-on-one and initiated by the student. Even students who would be tempted to disrupt a group activity will be attentive in a help situation.

I have worked with students who were very reluctant to ask for help and were incredulous that they could get genuinely interested help without being scolded or put down in some way. I expect most of their interactions with teachers had involved behavior control, and posturing and one-upsmanship may have played a large role in peer interactions. I believe that the complete separation of help and evaluation was also important. In any case watching these students bloom in private one-on-one help sessions is very rewarding.

4.4. **Beyond the Skeleton.** The skeletal course is the minimum needed to get satisfactory results and major features are described in the previous section. Some issues are unclear, needing more experience and probably depending on level and what is considered “satisfactory”. Here we touch on these and some possibilities that are beneficial but not part of the skeleton. It is important that additions be efficient in use of student time, or optional; see §6 Educational Opportunities.

4.4.1. *Class meetings.* Traditional class meetings have been discussed in several places in this essay and their role remains unclear. A full course of traditional class meetings has to be considered beyond the skeleton for economic reasons. Abbreviated versions are feasible, see [link to economics]. It is hard to imagine traditional classes persisting long into the twenty-first century something along these lines is probably necessary.

Some of our computer-based courses began with optional class meetings that were later discontinued. A few students attended regularly but the benefits did not seem to justify the expense and dropping them did not cause significant problems. Our most ambitious task-oriented course has lectures but this is partly because we are not willing to run the risk of lower outcomes if they are dropped.

4.4.2. *Group activities.* The context is small groups of students working together with little or no supervision. Topics are benefits, organization, and credit. Benefits include:

¹There was a study reporting that disruptive students have essentially the same long-term outcomes as well-behaved students. In other words they don't disrupt their own learning any more than they disrupt the learning of others. The more important question of how much they disrupt others' learning was not addressed!

- Communication skills: communication has to be practiced to be learned. Computer-based courses currently do not support this. Traditional courses don't do much better. Teachers know what is to be communicated so frequently accept incoherent clues rather than requiring precision. Peer-to-peer communication requires precision to be successful.
- Conceptual skills: asking for help with a problem requires isolating and articulating the difficulty, and providing an answer requires isolating and articulating the solution. Greater care is needed when the exchange is between peers, and both parties benefit.
- Peer help: this is another way to describe the previous point.
- Social support: social interactions are very important to most students and this reinforces almost anything done in groups. We want to take advantage of this.

Our own evidence for the benefits of group work is mostly negative: students who have serious trouble are almost never part of a study group. Similarly when group projects are assigned there are almost always students who, for one reason or another, end up working alone. They seem to be significantly less successful, and consistently enough that it seems reasonable to attribute this to lack of group support rather than lack of individual ability.

Key questions are: how to get students to participate in group work; and what to expect from it. The two main approaches differ in grade credit.

- For credit: Participation is forced so is almost universal. Outcomes are occasionally impressive but vary widely and effective assessment is expensive. Groups that are not homogeneous tend to be dominated by the student who is most capable, best prepared, or most ambitious. In other words tension connected with getting a grade tends to overwhelm the beneficial mechanisms.

We developed assessment methods that ensured group projects were a learning experience for the non-dominant students, but these were so expensive (in faculty time) that they could not be sustained. Further, we could not require performance at a level that would enable us to rely on learning in these activities. As a result any significant content had to be duplicated elsewhere, and again this was too expensive to sustain.

- Without credit: Voluntary study groups are probably more effective than for-credit and cost very little, so they win cost/benefit comparisons hands down. The problem is getting students to participate.

The strategy is to make it as convenient as possible and hope that benefits and social factors sustain it. Providing convenient places and times for group work is important. Having faculty available for brief help interventions (*not* tutoring) would be valuable. Internet-dating or Facebook-type software designed to connect people with common interests might help form compatible groups. Making it a standard part of a curriculum would probably lead to high participation because students who once find it helpful are likely to continue.

4.5. Summary. In previous sections we saw that careful task design, and supporting resources including helpers and linked web materials, are necessary for the approach to work. The point here is that these seem to be sufficient for a workable

skeletal course. There are issues that need to be further explored and valuable additions that would cost little, but the basic plan seems to be in place.

5. RESOURCE REQUIREMENTS

Inadequate resources are a grim reality in education and a potential killer for new programs. Ongoing costs are discussed in concrete, immediate terms:

- **Teacher time:** Demands on teacher time are often already high enough to make the profession unattractive and promote burnout. Time must be counted as a limited and valuable resource.
- **Teacher expertise:** Expertise of the current math teacher corps is limited and uneven, partly because many were not trained as math teachers. Teacher training programs are not replacing losses in K–12 and economic pressures are forcing wide use of adjuncts and graduate students for undergraduate teaching. This is not going to change anytime soon and a realistic plan must accept this.
- **Personnel budgets:** These are essentially fixed, and—because people with more expertise are more expensive—enforce a tradeoff between time and expertise. Teacher time can be maximized at the expense of expertise by more, but less expert, teachers, or vice versa.
- **Student time:** This must be considered a valuable resource. Students resent things they perceive to be a waste of time, and as they grow older they become more consciously resentful and less tolerant. Conversely, it is easier to engage students in time-efficient learning and more can be accomplished. See §6 Educational Opportunities for discussion.
- **Facilities and equipment:** task-based learning requires a large computer lab.

The question is: can a task-oriented program stay within current budgets for these resources and get good results? Our task-oriented course actually costs less than a traditional course so the answer is probably “yes”, but getting it to work may be tricky.

The next section explains why worrying about budgets is important. The following sections discuss costs of operation, startup, and development.

5.1. Increased resources are not an option. Educational cost accounting is not required by educational grants and is almost never mentioned in research papers. New approaches tend to be generously subsidized during development and would be far over-budget in any real-life setting. Two justifications are offered for this: first, if something can really be proved to be better then people will pay more for it. Second, the objective of this kind of research is proof-of-concept and cost-effective implementation is someone else’s job.

I believe it is vital to consider costs from the beginning. An education plan that depends on additional resources is like a business plan that depends on winning a lottery: it might happen but no serious proposal should count on it. The current K–12 situation is actually worse because the No Child Left Behind strategy forces concentration of resources on failing students and subjects. If a method works well enough that most students pass then it becomes a target for resource reduction.

Dodging the resource issue often leads to concepts that cannot be made cost-effective. The New Math of the 1960s was a great concept and worked fine when

taught by professional mathematicians. The expertise requirements were far over budget and the program crashed and burned when it collided with reality. Some of the proposals for Discovery learning also depend on high expertise. Are they re-inventing the flat tire?

In other cases costs forced out novelty and implementations shared only a name and some materials with the research methodology. Persistence of the name gives a way to save face and avoid admitting failure, but it should be dishonest to claim success.

The most insidious problems come from compromises needed to stay within budget. For example “enriching” a course means adding something. If nothing is taken out then the result will always be more expensive than before and usually over-budget. Computer-enhancing a course places significant demands on time and expertise. To stay within budget some of the earlier content is typically replaced with “learning to use computers” as a course goal. But the lost content may be needed later and the computer proficiency gained is usually poor.

The big challenge in educational innovation is to do better with *the same or fewer* resources. Ignoring this leads to failure in one way or another.

5.2. Operating expenses. Primary operating expenses for a task-oriented program are helpers and classroom teachers. There are facility and equipment needs but these may be shared with other programs and may come from different budgets.

5.2.1. Helpers. Helpers are the major new expense. It is important to have enough helpers to provide real-time help to students working at computers. The tradeoff is that good helping requires far less expertise than traditional teaching. Most of our helpers are undergraduate or graduate students, or instructors. Regular faculty are simply too expensive. Faculty can be used to oversee and provide backup for helpers because this leverages their expertise enough to justify the expense.

In K-12 qualified junior and senior students should make excellent helpers and will themselves benefit from the experience. However this should be a paid position because it really is a job. Some training and experience are needed, and success of the program depends on them showing up reliably for a whole semester or year. See §2.4.2 Help Opportunities.

5.2.2. Classroom Teachers. Costs in this category must be reduced to balance the cost of helpers.

There are immediate savings in teacher time because the task system provides assessment and class administration. No more grading. This does not translate into systems savings unless the student/teacher ratio is increased, either by increasing class size or class numbers.

Task-based sections in our university course usually have three to five times as many students as traditional sections and this alone pays for helpers and leaves a tidy net savings. Teachers don't mind because there is no grading. It works better for students than usual monster courses because students who use the tasks as online courses don't come, and attendance drops back toward traditional numbers.

In a school situation it might be better to increase class numbers than class sizes. For instance when two-week segments are used, see §4.2.1 Segments, class meetings could be held in the first week and the second used for independent work in the lab and taking tests. If the schedule of another such class is shifted by a week then whenever one is meeting in a classroom the other is working independently. A single

teacher could handle both classes, again reasonable because there is no grading. This effectively doubles the acceptable student/teacher ratio, or equivalently halves the number of fully-qualified teachers needed. This would not really make half the personnel budget available for helpers, but it should suggest that the idea is workable.

5.2.3. *Facilities and Equipment.* The main facility requirement is a large computer lab where students can work with the support of helpers, and take proctored tests. In most instances space and computers will both be available and the main issue will be configuration. In particular, a single large area is significantly more efficient than several areas with the same number of machines due to the way help effectiveness scales with size.

We have also found that having the area comfortable, attractive, and free of distractions is helpful. An investment in decor and the presence of helpers sends a strong message about expectations and the importance of learning. Folding tables in a gymnasium might send the opposite message.

5.2.4. *Student time.* For reasons explained in other sections this approach should yield significant savings in student time. Student time is not usually valued or measured but this is the key to better outcomes. This is explained in §6 Educational Opportunities.

5.3. **Startup.** Startup expenses are costs incurred in each system when the program is first introduced. We have not participated in a startup other than our own so much of the following is extracted from our experience minus the false starts and groping in the dark. College-level startups are discussed in [Economics of Computer-Based Education](#) so we concentrate on K-12 here.

5.3.1. *Begin with tests.* The ideal changeover begins with use of task-generating software to produce high-stakes tests, and making related tasks available as study guides. The tasks would quickly and naturally become important course materials.

The next step is to use tasks as course assessments. Students and teachers should be comfortable with this: the tasks are obviously mission-related in a teach-to-the-test way because they have the same source as high-stakes tests and are designed to support it. Courses would not depend on them functioning as learning environments and all the usual practices (homework etc.) could continue. In particular they would not be supported by helpers. Many students will find the learning features useful, however, and teachers are likely to find themselves doing a fair amount of what amounts to helping.

The final step is to change over to task-oriented courses with computer labs, helpers, modified class schedules, maybe tracks, etc. If tasks have already been in use as assessments for a year or so then the new plan should more-or-less make sense to students and teachers and educational dislocations should be minimized. This would allow focus on organizational and institutional dislocations, which is good because they will be plentiful.

5.3.2. *All at once.* Our experience, and my best advice to a school planning a changeover in their math offerings, is that it is very important to do as much as possible all at once. There will be a chaotic period but it will settle down and work. An attempt to phase it in over time will significantly increase difficulty and aggravation in the long run and greatly increase risk of failure.

- A phased change will be thought of as an experiment that might be cancelled. People opposed to the idea will attack vigorously, trying to kill it before it gets established. There will be instances where this can't be resisted.
- The people directly involved won't be fully committed: why knock yourself out if it might get cancelled?
- An experimental program is a lightning rod for complaints from students and parents even if they aren't relevant to the program.

An obvious full commitment from the beginning minimizes these problems.

Another problem is that parts of the program, particularly help, depend on economies of scale.

- A small-scale pilot program is likely to be over-budget, or unsatisfactory because it is under-funded, even if a full-scale program would succeed.
- There will be a great temptation to support a small-scale startup with a small computer lab, and add additional labs as the program grows. This can be a killer. Testing and computer-side help work best if everything takes place in a single large lab. Using several smaller labs significantly increases cost, multiplies problems, and increases the risk of breakdown and failure.

See the essay [Economics of Computer-Based Math Education](#) for a discussion of scale-dependence.

5.3.3. *Preparation and support.* The first startups will be breaking new ground. After that there should be resources to make program conversions easier if not routine:

- Training videos, manuals, instructions, and specific data on lab size and help staffing requirements;
- Seminars and summer programs; and
- opportunities to spend time in functioning facilities.

We argue in §5.4.1 Not Commercial that software development should not be a commercial undertaking. This argument does not apply here: a business could offer a range of assistance including consulting, products like those described above, and computer-lab setups. They might also offer computer services such as test and course administration, as long as they don't try to commercialize content software.

5.4. **Development.** Initial development involves development of task-generating software and supporting materials and refining them with feedback from field testing. Reasons for using software rather than a problem-database approach are discussed in §3.5 Software Generation, but one is that full development need only be done once. Maintenance and refinement should continue indefinitely but are relatively inexpensive.

5.4.1. *Not Commercial.* High-stakes state tests are usually contracted out for commercial implementation, the SAT is a commercial test, and while the College Board is nominally nonprofit their tests are either developed by commercial subcontractors or internally in the same closely-held way. There are widespread and fully justified concerns about counterproductive effects of these tests. There is much more at stake with learning tasks than with assessment and no basis for thinking that this approach would be any more successful.

In short, it would be inappropriate to outsource a key part of our educational system. Development must be driven by concern for outcomes rather than profits, and everyone in the mathematical and educational communities must be able to participate in feedback and refinement.

Work on the first draft could be organized by a non-profit group, educational institution or professional society. Large open software projects such as linux, wikipedia and tex are useful models for subsequent maintenance and development.

5.4.2. *Develop for the top.* Task-generating software must be designed to work for the highest-level version of the course that might be offered.

- Difficulty and coverage can always be reduced by changing parameter settings, including, for instance, multiple-choice answers instead of free-response.
- Designing for the highest level requires the deepest understanding of learning and mathematical structure. In particular it requires finding ways to make abstract understanding directly useful in problem-solving, as it is for professional mathematicians.
- I believe we will find that highest-quality task design will enable all students to go further than we might currently imagine.

Recall that for non-terminal courses “high quality” and “high level” are largely defined in terms of preparation for later courses. Consequences are:

- “highest” quality requires understanding how material will be used at least through the second year of college calculus; and
- for best results the whole development from at least fifth grade through the second year of college calculus should be thought of as a unit and outlined before specifications for any level are finalized. Ideally it would be developed as a unit without grade levels hard-wired in the program. Local school systems could then decide where to place divisions to best meet their needs and there would still be general coherence in overall programs.

Finally *really* high quality would make “Profoundly Gifted” threads possible in tracked courses, see §6.3 Tracked Courses.

5.4.3. *Expertise required.* High-quality task design requires profound subject mastery, analytical ability, and educational wisdom.

- Database-oriented test developers often recruit students or math BAs to write or check problems. We have tried graduate students, instructors and others but only a few senior professors with records of original mathematical research and extensive programming experience have been really successful with task design.
- One of the hardest lessons has been that classroom-oriented educational expertise is almost irrelevant. Knowing how to teach, it turns out, is very different from knowing how students learn in a student-directed environment. Experience with such an environment, for instance as a computer-side helper, may well be necessary.

To expand on the first point, this is not just a matter of skills. Single problems can, at best, encode wisdom and expertise at the undergraduate or BA level. Software that generates problems can encode wisdom and expertise at any level. In a real sense students are being taught by the people who develop the task-generating

software. Their contributions are incredibly highly leveraged, so it is vital to do the absolute best possible job.

To expand on the second point, I had been teaching for about 25 years when I started working with computer-based learning in the Math Emporium. I had lots of ideas, plans and expectations based on my classroom experience. They were all wrong and many of them were counter-productive. Watching and working with students slowly disabused me of many preconceptions and I doubt this process is finished. Outstanding teachers heavily invested in classroom expertise have been—so far—unable to make this transition.

5.4.4. *Support for development.* First we consider resources needed. For perspective consider that this undertaking would be comparable to development of a web browser, search engine, or high-performance database system. How would this be approached professionally? How would a major software company organize such an undertaking, and what resources would they consider necessary to ensure success?

This program would require at least a few experts whose regular salaries are over \$100,000 and a specialized support staff. Careful recruiting and help from volunteers should keep the total well below the usual cost of a major commercial software development program, but it will still be a lot of money for an education project.

I have argued that software development must be undertaken as a not-for-profit activity. These are usually supported by grants from private foundations or government agencies and some of these grants are in the multi-million dollar range. However this is not likely to help with a task-generation development project.

- Grant applications are reviewed by education professionals with expertise grounded in classroom instruction. These experts tend to find the ideas advanced here counterintuitive and unconvincing if not actually repulsive, and are unlikely to support funding.
- The funding needs of this project do not fit the standard mold. Large education grants are multi-year, expected to involve many partners and collaborators, and require elaborate, costly, and for us irrelevant, assessment. The pie is so divided that it provides encouragement rather than full support.

In principle state departments of education could be a source of support for the K-12 portion. A software system that generates high-stakes math tests could save tens if not hundreds of millions of dollars each year. This is independent of any educational benefits so they would not have to believe outcomes would improve for the investment to make sense. If the learning tasks etc. provided as study guides did improved outcomes it would be pure gravy.

Unfortunately state departments of education usually have to scrape to get the next round of tests ready and are not in a position to invest in the future. Further, innovation tends to be punished. If they do things in the same old way and something goes wrong then they can't be blamed. If they are at all adventurous and something goes wrong, e.g. scores don't go up enough to avoid sanctions, they get the blame even if the traditional approach would have done worse.

Finally, attempts at collaboration among states tend to founder on questions of local control. State departments would have to be convinced that tinkering with

input parameters would give them adequate control before they could give up control over software design.

5.5. Summary. If it is done well then initial development of software for generating tasks only needs to be done once to enable long-term nationwide (and international) use. Costs would be large for a single education project but negligible compared to long-term savings on high-stakes alone, and truly trivial compared to potential benefits of improved math education. Even so there seems to be no straightforward way to get it started.

If the development gets done, and if the system is used for high-stakes testing, then in K–12 the rest of the program can develop through relatively small steps. The largest of these steps is starting up computer labs with help programs. This has immediate benefits in terms of teacher expertise and involving students in education, so once a good model is established this should also become routine.

A key point is that operational expenses are no greater than traditional programs. Better outcomes would be a consequence of high quality of the initial development and reorganization of resources. They would not require additional resources or sacrifices in other parts of the curriculum.

6. EDUCATIONAL OPPORTUNITIES

Our goal is to improve outcomes at all performance levels. This is tricky: most approaches trade off improvement at one level for losses at another. To explain why, and how to avoid it, we need an understanding of student behavior.

6.1. A Behavioral Model. The best first-approximation description I have found is: students have time budgets and grade targets, and work until one or the other is met. If they run out of time they accept a lower grade. If they reach the target grade, they quit and take more free time. I wish it were otherwise but this explains the data.

This model explains the usual achievement/failure tradeoff. If standards are raised then students who are not over-budget in time will learn more to achieve their target grades. But students who are at or over their time budgets will accept lower grades. Learning by stronger students rises but grades fall. Reducing standards has the opposite effect: students under-budget in time work less to get their target grade and enjoy more free time, while previously over-budget students may get higher grades. Lowering standards reduces the spread in learning and increases grades.

This achievement/failure analysis assumes a fixed educational method. Now suppose standards are held fixed and methods are changed. A more efficient method raises grades only of students who would have been slightly over their time budgets; others take a payoff in free time. Less-efficient methods cause a hit in free time but changes grades only for students who now go over their time budgets. Real life is more complicated but this leads us to expect—to a first approximation—that methodology will have only marginal effect on outcomes. This explains the “no-significant-difference” phenomenon often seen in education research.

Two important conclusions:

- The only sure way to improve outcomes, particularly for the best students, is to raise expectations. The challenge for educators is therefore “how can we raise expectations without unacceptable increases in failure rates?”

- The main benefit of a more effective method is likely to be reduced demands on student time. Time has to be measured or inferred to effectively compare methods; outcomes alone won't do it.

The second conclusion suggests a solution to the first.

6.2. The Main Strategy. In a nutshell the idea is to switch to more efficient learning methods and more-or-less simultaneously raise expectations, with the goal of holding demands on student time constant. When time demands are unchanged grades should also be largely unchanged, but learning outcomes will improve.

Efficiency has been a recurrent theme in our description of the task-oriented approach. For instance students choose among, or combine, resources to fit their individual learning styles. Real-time help with difficulties is an enormous time-saver. The main savings, however, come from letting students skip what they don't need. Uniform homework assignments require more than most students need, and for these students the excess is busywork. For some students many class meetings are a waste of time.

Note that the strongest students will see the greatest time savings in this approach. This means expectations for top grades can be raised quite a lot without reducing grade outcomes.

6.3. Tracked Courses. Tracked courses offer another strategy for improving outcomes. We outline the idea; see the essay [Tracks in a math course](#) for more detail.

The context is a pair of courses that cover similar material but at different levels: say "standard" and "advanced". The usual approach is to sort students by interest, ability, and preparation, for placement in the two courses. But at least 10% and frequently 20% will be in the wrong course. Students who get D or F in Advanced course should have been in Standard, and many Standard students who get an A should have been in Advanced. This is unavoidable, and in particular better placement tests will not fix it.

The idea is to combine the courses and let students choose their own level. Students who do well on Advanced tests stay in that track. Students who take Advanced tests and don't do well are offered the choice of retaking them and doing better or going into the Standard track. Students who begin in the Standard track but find the material more accessible than they expected have a risk-free upgrade path. The course for which they receive credit is not determined until the end of the term.

If class meetings are offered then courses would start with one-size-fits-all presentations. As students settle into tracks different sections could specialize to one or the other track and students could switch sections to get appropriate lectures. If the alternate-week schedule (§4.2.1 Segments) is used, and the two parallel sections specialize to different levels, then students could switch sections simply by moving to the parallel section. Times, teachers, and classrooms would be the same.

There could even be choices offered at the end of the course: a C in the Advanced track could be converted to an A in the Standard track. Is a higher GPA more important than getting a prerequisite for a technical career? The student decides. In any case no one would get a D or F in the Advanced track, so expectations could be kept high without forcing up failure rates.

This scheme is too time-intensive for use in traditional classes with current student/teacher ratios. It would be easy to implement in a task-oriented course:

- grades and course administration are managed by computer so choices and transitions could be managed automatically; and
- the same software could generate tasks for several tracks by appropriately adjusting input parameters.

Finally, it would give quite a boost to the development of first-class scientists and engineers if a “Profoundly Gifted” track could be offered to the very best students.

6.4. Summary. A task-oriented program offers several ways to raise expectations and improve learning without increasing failures. One exploits the student time made available by efficiency of the method. Another exploits computer management rather than any virtue of the method to provide separate levels. It is significant that both rely on providing students more choice and control over their learning. In one case this enables them to optimize the process for their individual preferences and needs, in the other case it gives them more input into choice of level.

7. CONCLUSIONS

The proposal is to exploit natural tendencies of students, and practices widely forced on teachers by high-stakes tests, by making “teach-to-the-test” really work. Experience with college-level courses indicates that test-like “learning tasks” with appropriate support could provide better outcomes without drawbacks such as higher failure rates.

The questions considered in detail are: how would such a system work in real practice; can we get there from here; and can we afford it? There are plenty of pitfalls, most of them beyond the ken of usual educational studies, and the way through them is a bit torturous, but there does seem to be one. In particular it should require no more resources than traditional classroom instruction and in large systems may actually reduce costs.

Putting everything together gives a best-case scenario for K-12:

- development of task-generating software and reference texts as a source of high-quality high-stakes state math tests;
- tasks and supporting material provided as study guides for the tests;
- teachers find tasks to be effective learning guides and, over time and attracted by a reduction in grading, use them as course assessments; and then
- school systems realize that by going to a teacher/helper system they can save money and leverage the effectiveness of fully-qualified teachers.

The first step is the most problematic. If that can be overcome then the others provide a way to make the change in reasonable, well-motivated and individually sensible steps.