

Contributions to a Science of Mathematical Learning¹

Frank Quinn

March 4, 2011

¹Incomplete draft. The current version, and many of the component essays, are available at <http://www.math.vt.edu/people/quinn/education/>.

Chapter 1

Preface

Early in this century I began to see serious discrepancies between the way my students seemed to learn and standard educational theories. The essays collected here describe these discrepancies, my attempts to find remedies, and attempts to understand how such an apparently counterproductive situation came to be. Brief summaries of the essays are given at the end of this preface. The rest of the preface provides a context for organizing and understanding the material.

In a nutshell, mainstream mathematics education is essentially philosophical, not scientific. This is an oversimplified summary of a large number of specific observations, not a philosophical assertion, and only emerged when I was trying to organize the essays for publication. I had worked extensively with specific small-scale issues of learning. Larger-scale topics, for instance design of high-stakes tests or computer-based learning environments, were heavily dependent in non-obvious ways on these details. Accordingly it made most sense to me to begin with the most-detailed essays and work up. The custom in education is the reverse: begin with big-picture theory and then fill in details. In fact this is a fundamental difference between the philosophical and scientific approaches to knowledge. The organizational confusion led me to realize I had been, without explicit intent, developing a scientific approach to learning mathematics. The traditional approach is essentially philosophical, and the discrepancies are eerily typical of the discrepancies seen in the past between philosophical and scientific accounts of the same subject.

It may be helpful to unpack the title of this volume, “Contributions to a science of mathematical learning”. ‘Science’ reflects the realization described above that the approach is more scientific than philosophical. These are ‘contributions’ because they are incomplete, but more fundamentally because they are work of a single individual and science must be the work of a community. Finally, ‘mathematical learning’ rather than ‘mathematics education’ reflects the realization that the traditional educational focus on teaching, and learning as mediated by teachers, has actually been a barrier to deeper understanding of learning itself.

Below I give brief distillations of science, philosophy and human nature,

and locate the work in these contexts. Again I note this is an after-the-fact explanation and the essays are not dependent on it. People who reject the explanation, on philosophical grounds for instance, should not claim that the essays fall with it. These must be rejected one-at-a-time on their own merits, or lack thereof.

1.1 Science

The core strategy of science is to start with small, concrete observations or questions, and work up in abstraction and down in detail. The objective is to be as effective as possible. From these seeds mighty fields of knowledge grow.

Scientific fields essentially define themselves through these basic ingredients, driven by the ambitions of researchers. First, the starting-point “concrete observations” may turn out to be composite or derivative, and more effectively described in terms of more primitive objects. In other words the objects of study are revealed as the field develops, and cannot be fixed in advance. Second, methodology evolves as researchers refine it to maximize effectiveness. Therefore, methodology is also revealed and cannot be fixed in advance. To paraphrase [Knorr-Cetina], the commonalities in science are not in specific knowledge, nor in the methods for constructing this knowledge, but at one more remove: the approaches used to develop knowledge-constructing methodologies. Actually this understates the coherence (what [Wilson] calls “consilience”) of scientific knowledge and methodologies, but it does describe the short-range impressions of participants. Finally, the whole enterprise can fail to get off the ground if nothing effective can be extracted from the starting point. We can *say* ‘science of extrasensory perception’, ‘science of homeopathic medicine’, etc., but this does not mean there is any such thing. In short, people do not control science, and attempts to do will make it unproductive. Successful scientists actually take the opposite tack: they do their best to adapt to their subject and have the subject control *them*, so it can lead them in productive directions.

Mature sciences develop levels of generality and abstraction, and these interact in complex ways. Each level develops insights, questions and concerns that inform and challenge other levels. Exploration at the research frontier is often tentative and unconsolidated. In the long run, however, the bottom-up strategy and the quest for effectiveness still shape the activity. Mature abstractions are required to account for and be consistent with well-established data, using methodologies optimized for the subject matter. If they cannot, then adjustment is required: either the abstractions could be made more effective, or they are less effective than previously believed and need to be refined, or (rarely) there is a methodological flaw that needs to be hunted down and exterminated. Sometimes these adjustments take a long time, but the need is acknowledged as soon as genuine inconsistencies are verified.

Bottom-up development makes scientific knowledge powerful, robust, and cumulative, but from a human point of view it has serious drawbacks. It is very complicated. Because methodology is to a degree subject-specific, drawing

sound conclusions requires deep subject-specific wisdom, and this is often at odds with intuition or naive expectations. Questions or objections are resolved by going back to details, and this requires time, effort, expertise and integrity. Finally, scientific knowledge develops at a glacial pace.

1.2 Philosophy

Classical philosophy, politics, religion, alternative medicine, new-age mysticism, feng shui—essentially every area outside science and engineering—approaches knowledge in a top-down way. The simplest picture is that certain abstract principles are taken as primary, and lower-level understanding and practice is either derived from these or obtained by filling in details in ways consistent with them.

In some cases the basic principles seem almost random. More often they are abstracted from experience, prior beliefs, innate inclinations, etc., but they are not driven by details as in science. Questions or objections are usually resolved by appeal to authority (top) rather than data (bottom), and actual practice is complicated by the variety of ways in which authority operates. For example, lower-level conclusions are “derived” from more abstract principles using certain rules of argument. However these rules are determined by authority and are usually designed to inhibit unwelcome conclusions rather than maximize effectiveness. Further, the rules are usually nebulous enough that when unwelcome conclusions do occur they can be rejected as misinterpretations or rule violations.

Philosophically constructed knowledge has serious drawbacks. First, it is ineffective. It usually performs poorly at nitty-gritty practical levels and lacks mechanisms for improvement. This issue is usually dodged by presenting the knowledge as a “world view” or belief system rather than a source of practical tools, and it is wrong to expect immediate results. Another drawback is that it tends to fragment rather than accumulate. Disagreements among non-authorities can be settled by appeals to authority, but there are no graceful mechanisms for settling disagreement among authorities. Instead philosophical knowledge communities tend to divide into mutually incompatible schools of thought, factions, sects, etc., each with its own beliefs, rules, and authorities.

The last point has an ugly corollary. Top-down knowledge communities are held together by the strength of their belief and commitment, not by the robust power of their knowledge. As a result, challenges are taken as attacks to be defended against, not new data to be accounted for or incorporated. Rival factions have borders defended against intruders, and either warfare or uneasy truces reign outside these borders.

1.3 Humans

Why would anyone use a method with the drawbacks of top-down knowledge construction? To be specific, the top-down Aristotelian approach to physics now seems pathetically ineffective and easily refuted. Why did it dominate the subject for more than two thousand years? The answer seems to be that it is natural for humans in two different ways.

First, and most important here, top-down construction seems to be the innate human approach to knowledge. Semi-deductive reasoning comes easily, and specific systems seem attractive and accessible because—compared to science—the approach is fast and the answers are simpler and complete. They do require some thought and study, however, and this activates another innate mechanism. Learned material becomes physically implemented in neural structure. As time goes on it becomes increasingly easy to see the world from this point of view, and increasingly difficult to see it any other way, c.f. [Barton]. Eventually it becomes visceral: this is right and everything else is wrong. People implicitly acknowledge the weakness of the top-down approach by having no qualms about rejecting beliefs of others. But the innate mechanisms prevent them from seeing the immediate logical corollary: their own beliefs are no more likely to be correct.

The other attraction of Aristotelian physics is that it embodies our naive innate physical beliefs. Our innate physics is now known to be wrong [Dunbar], but it *feels* right and was easy to believe as long as it wasn't expected to actually do anything. Details vary but the phenomenon is general. Beliefs and philosophies develop from what we already believe and are comfortable with. If disciplined study shows the facts to be otherwise then we are likely to find them as uncomfortable as children find Newtonian physics.

Among other things, these points clarify public attitudes toward science. Yes, the subject matter is complex and demanding. The real problem, however, is that the bottom-up approach to knowledge is unnatural for humans. Not only do we lack the innate receptacle we seem to have for top-down knowledge, but we have to *suppress* this innate tendency, and *suppress* naive intuitions in order to learn the methods of science. Few succeed, and for most people science remains just another belief system. They accept the golden eggs laid by science but do not see effectiveness as indicating a special status for scientific conclusions, and do not hesitate to reject these if they conflict with other beliefs.

1.4 Scientific Revolutions

Every area in contemporary science was originally approached philosophically. At some point scientific approaches got footholds and, by virtue of much greater effectiveness, replaced philosophy. The pre- and post-foothold processes are quite different. The defense mechanisms of philosophical knowledge systems make the replacement process contentious, highly visible, and deserving of the

term ‘revolution’¹. The pre-foothold period is quieter and more difficult. Scientific material accumulates very slowly because the mindsets needed for routine production are considered unnatural and wrong and are correspondingly rare. Further, there is nothing to sustain the activity until enough accumulates to provide payoffs, and until these payoffs are recognized as valuable. The pattern is that once philosophy is well-established, it takes science between two and three millennia to get to the foothold stage.

The first major scientific revolution, and the best known, occurred in physics in the seventeenth century. Chemistry followed in the eighteenth, and biology and geology mostly in the nineteenth.

It is irrelevant whether or not this work is the beginning of a ‘scientific revolution’ in mathematics education. The revolution picture is important because there was a scientific revolution in mathematics. This is virtually unknown. The conventional view is that either mathematics is not a science or it is uniquely compatible with philosophy, but in any case a heavily-philosophical approach seems to be effective and has thrived for three millennia. There was a well-known period of turmoil—the “foundational crisis”—in the early twentieth century [Gray]. During this period professional practice broke from philosophy and became incredibly more powerful. The conventional view of philosophers, historians, educators, popularizers, and most users of mathematics is that this was a sort of collective insanity that will eventually pass, not a permanent revolution, and there is no connection between seemingly unnatural modern methods and incredible modern success. However, historical, philosophical, sociological, technical, and other aspects of this event are traced out in detail in [Quinn], and it really does seem to have been an essentially typical philosophy/science transition.

The significance for the development here is that contemporary mathematics education is built on *two* sets of philosophical constructs: philosophical accounts of teaching and learning, and philosophical accounts of mathematics. A science of mathematical learning requires scientific approaches to *both* components. We see, for instance, that the recent ‘math wars’ were not the beginning of a scientific revolution in mathematics education because both sides accept standard and rather romanticized philosophical descriptions of obsolete mathematical practice. This work connects instead with the bottom-up description in [Quinn].

This account of the development of this work is mostly hindsight. Foundations for the description of contemporary mathematics had been laid a decade earlier but I did not take it seriously until I realized that some of my most effective help strategies were drawn from my professional training. When I tried to explicitly describe modern methodology, as clarified by the work with students, the result was nothing like any of the standard philosophical descriptions. Historical study (especially Gray) showed that this disconnect dates roughly from

¹‘Scientific revolution’ here refers to the *onset* of scientific methodology. This seems to have been the general meaning until T. S. Kuhn [Kuhn] and successors extended the term to include “paradigm shifts” *within* science. Kuhn, perhaps understandably for a philosopher, did not see a qualitative difference between these and philosophy/science transitions.

the period 1880–1920. Mathematicians, along with everyone else, like to think of mathematics as a seamless development over almost three thousand years. It took me quite a while to realize this view is not just wrong but *very* wrong, and that a profound transition occurred only a century ago. The biggest surprise was finding that contemporary professional practice is highly adapted for human use—much more so in fact than the philosophically-constrained nineteenth-century version. My professional training connected with my students’ needs for this reason as much as for the greater technical power.

1.5 Education, and this Work

The main input for this work is a decade spent in a computer-learning facility. I followed students around to see how they used the materials, and provided one-on-one help to diagnose and repair learning problems. I also developed courseware with ambitious goals, designed to avoid the problems and to work the way they wanted to use it. The objectives were practical and immediate so the activity was closer to “learning–engineering” than science. From the science perspective described above, concrete problems and observations were guiding me to identify both core issues and and effective methodologies.

The next component was the bottom-up description of professional mathematics practice described above.

The most recent ingredient was a study of cognitive neuroscience. As might be expected, this considerably clarified cognitive aspects of both student learning and professional practice. Unexpectedly, an observation of J. T. Bruer about scale gaps (see Chapter 3) also crystalized the macro/micro and philosophy/science perspective described here.

It now seems clear that contemporary mathematics education, as embodied in schools of education, the contemporary research community, educational associations and think-tanks, and most of the teacher corps, is a collection of top-down philosophical constructions. When they speak of “educational philosophy” they really mean it.

The philosophy/science perspective clarifies why educators tend to react negatively to this material. Philosophy depends on belief and authority rather than data, and the primary allegiance of practitioners is to their principles and methodology. It is their *duty* to oppose things not imagined in their philosophy. These essays are unconventional in many ways, but three seem to be particularly problematic. First, educational philosophy is built around teaching and teacher-mediated learning. Learning at the micro (neuroscience) level, and in computer-based programs, is not teacher-dependent, and the traditional classroom is a rather poor model. From the science-defines-itself perspective it seems that they have not correctly identified the basic subject of study. But suggesting a departure from the classroom model is a serious challenge to the system. The second problem is that goals in contemporary education are more-or-less “whatever we can achieve with these methods”. This is reenforced by tight compartmentalization into elementary, secondary, early college, and “other”, the

divisions being justified as “age-appropriate”. If micro-level learning insights enable students at a given level to work a wider variety of problems, this can be rejected as “age-inappropriate” or “not contributing to established goals”. Suggesting that it might be “better” is a challenge to the system. Finally, as explained above, the mathematical content connects with a bottom-up description of modern professional practice rather than one of the usual philosophical account of nineteenth-century practice. This is problematic for interesting historical reasons. As usual for such events, the early twentieth-century ‘revolution’ in mathematics was very contentious. Traditionalists, including such luminaries as Henri Poincaré and Felix Klein, denounced the new methods as rote formal manipulation, and soulless abstraction disconnected from reality. They had no audience in the professional community because young people voted with their feet for the new ways. But philosophers and educators were listening. Not only do these remain committed to the old ways, but the venomous invective against the new is deeply embedded in their ways of thinking.

1.6 Organization

The essays here are the beginnings of a bottom-up (scientific) approach to mathematics education. Accordingly, they are ordered by scale going from micro to macro. Rough descriptions of scale divisions, and Parts of the book, are given in Figure 1.1. As usual in science, material at larger scales is a delicately-balanced

educational theory	Commonalities of teaching, learning, and behavior that transcend content.
subject matter	Content, and its influence on teaching and learning.
course/curriculum	Learning goals and methodologies at the course and curriculum level.
cognition/activity	Neural implementation of skills and understanding, and strategies for developing these.

Figure 1.1: Educational levels from macro to micro

synthesis of lower scales. Scientific reading strategies are adapted to this: if something doesn’t make sense, follow the roots down a few levels to find out what it really means, or to see the constraints that make it necessary. Quickly rejecting things that don’t seem to make sense is a philosophical strategy that will not work here.

Finally, the goals of the essays varies with level. Lower-level essays are mostly observations, suggestions and proposals. Top-level essays are mainly concerned with deficiencies in mainstream practice and philosophical principles, as revealed by micro-level studies. It is premature to try to extract much in the way of high-level conclusions or principles.

1.7 Brief summaries

Part I: The Cognition/Activity Level

These essays are primarily concerned with learning at the individual level. Issues related to teaching and synchronous group learning, or the organization of asynchronous learning, are explored at the next level.

Chapter 2: Neuroscience Experiments for Mathematics Education

Neuroscience experiments are described that could have significant impact in mathematics education. The first group explores why students find certain problem types difficult and how to fix this. The second group explores subtle but important oddities of human learning.

A goal is to explore the kinds of understanding and experience that would enable good neuroscience/education interactions. The next essay addresses the other side of the coin.

Chapter 3: Mathematics Education versus Cognitive Neuroscience

[incomplete first draft] For quite some time mathematics education has seemed an area in which cognitive neuroscience might make important contributions. This has not happened: studies have been large in number but small in impact, and education has been influenced more by misunderstandings and oversimplifications than actual science. Are these ‘important contributions’ an illusion? If not, why have they not been realized?

This article describes difficulties and obstacles to effective use of neuroscience by the contemporary education community. Comparison with the previous essay identifies specific shortcomings that are, in a sense, ways in which contemporary education is not scientific.

Chapter 4: Contemporary Proofs for Mathematics Education

In contemporary mathematical practice the primary importance of proof is the advantage it provides to users: proofs enable very high levels of reliability. This essay illustrates how a similar approach might have similar benefits in elementary education. Written for the proceedings of the ICMI Study 19 on Proof in Elementary Mathematics (Taipei, 2009).

Chapter 5: Proof Projects for Teachers

In contemporary mathematics, concise definitions and previously-internalized structure are exploited to give rapid and effective access to new material. It will be difficult, and is certainly beyond my expertise, to incorporate these methods into school materials. To provide a first approximation this essay offers course material for teachers that treats standard topics (including fractions and area) in ways that show contemporary methodology at work.

Chapter 6: K-12 Calculator Woes

Graphing calculators have been widely adopted in some K–12 curricula. Unfortunately the way they are used seems to cause significant learning deficits that limit success at higher levels. Written for the Notices of the American Mathematical Society.

Chapter 7: Student Computing in Mathematics: Interface Design

This is the first in a series on computing environments designed to support learning in mathematics and other technical areas. They draw on many years experience with students working with computers and in computer environments.

The main point is that human learning is quite complex, and the full complexity is only beginning to unfold as we move away from the tightly-bundled package of hand calculation in traditional classrooms. There are more ways for learning to fail than most of us imagined; many are different from the things educators traditionally look for; and underlying causes are obscure.

This article concerns basic student-computer interactions. Among many other things we see that standard cut-and-paste can undercut some learning objectives and has to be modified. A planned sequel should concern computational functionality. Careful limitations are needed to avoid turning the subject into keystroke sequences.

Part II: The Course/Curriculum Level**Chapter 8: Task-oriented Math Education**

“Learning tasks” look like tests to students, but are designed as learning environments. Experience at the Math Emporium at Virginia Tech, described more detail in subsequent essays, demonstrates educational effectiveness of this approach at the college level and suggests it should work at least in upper grades in K-12. Benefits could include significant improvement in the quality and effects of high-stakes tests. Some of the educational advantages come from giving students more choices and more control over their learning.

Chapter 9: Downstream Evaluation of a Task-Oriented Calculus Course

(Not yet available) The study concerns second-semester calculus for science and engineering. Over a period of eleven semesters roughly half the students (3720) were in traditional sections, half (3987) in task-oriented (computer-tested) sections, all with common final exams. More precisely, these numbers are students who passed the course. Our data includes all students who signed up for the course, with drop-outs included for reference, for a total population of 9235. Analysis will include outcomes in subsequent courses, in most cases to graduation. Preliminary analysis indicates essentially equivalent outcomes. The final

version will be much sharper quantitatively but is not expected to be much different qualitatively.

The “equivalent–outcome” conclusion of this study seems to fall short of the announced goal of better outcomes. It is actually very encouraging:

- The task (test) materials were developed in 2004 and have not been updated. Understanding has advanced considerably in the last six years and substantial improvement should be possible.
- Task-oriented courses are considerably more efficient. Focusing on outcomes in this course overlooks the substantial benefits of resources released to the rest of the program.

Some of these issues are explored in more detail in the next essay.

Chapter 11: Economics of Computer–Based Mathematics Education

Experience at the Math Emporium at Virginia Tech indicates that computer-based education can be both efficient and effective. However it may require large-scale development and infrastructure, and educational techniques have to be adapted to economic constraints. Moreover it must be assessed as part of the system, not just in terms of outcomes.

Chapter 10: Beneficial High-Stakes Math Tests: an Example

A worked-out example is given to show how mathematical and educational insights can be incorporated in the structure of high-stakes K–12 math tests in a way that promotes better teaching practices and more effective learning. The example concerns symbolic skills deficits seen in students from calculator-oriented K–12 programs. This is a variation on the “task-oriented” design described above.

Chapter 12: Levels in a Math Course

Variation in student interest, preparation, and performance is usually accommodated by offering courses at several different levels and placing students in them at the beginning of the term. This practice has serious drawbacks that might be avoided by reversing the placement strategy.

In the multi-level structure envisioned here, students enroll in a combined course, sort themselves into levels according to performance, and determinations about the credit they receive are made at the end of the term. Resource constraints may make this approach impractical in some cases, but when it can be used it could significantly improve outcomes.

Chapter 13: Teaching versus Learning in Mathematics

Teachers seem to be far too focused on what happens on our side of the desk. It now looks as though teaching and learning were never as closely linked as

we wanted to think, and the gap will widen unless we focus on students and learning, particularly long-term learning, and *not* through the lens of teaching. Examples concern calculator arithmetic, “clickers”, computer courseware, and diagnosis of errors.

Part III: The Subject Level

Chapter 14: Professional Practice as a Resource for Mathematics Education

A detailed study of professional practice is given *Towards a Science of Contemporary Mathematics*, and an extract will be included in the final version of this collection. This work describes how professional mathematical practice changed in the early twentieth century, and why current practice is much more effective. It also traces historical reasons why mathematics education continues to be modeled on the methodology of the nineteenth century. Continued use of an obsolete model may be one reason why education has been unable to improve much on nineteenth-century outcomes.

Chapter 15: Updating Klein’s ‘Elementary Mathematics from an Advanced Viewpoint’: content only, or the viewpoint as well?

The Klein Project, organized in part by ICMI (International Commission on Mathematical Instruction), seeks to update Felix Klein’s influential 1908 book. However Klein was a strong critic of twentieth-century methodologies being developed and adopted at the time, and his ‘advanced viewpoint’ is that of the nineteenth century. Indeed, his influence has been a barrier to educational use of contemporary methodology. The goal of the Klein project is to update the work to include some topics in twentieth-century mathematics, but (by default) retain Klein’s nineteenth-century viewpoint. Shouldn’t the viewpoint be revised as well, at least to be upward-compatible if full modernization is impossible?

Part IV: The Educational-Theory Level

These essays mostly concern shortcomings and failures of contemporary education at the policy and theory level.

Chapter 16: Dysfunctional Standards Documents in Mathematics Education

Standards documents attract a great deal of attention, and reasonably so: they should provide structure and common reference points for teachers, administrators, curriculum developers, textbook writers, test developers, etc. Unfortunately current documents do a poor job with all this and it seems unlikely they will improve.

Chapter 17: Math / Math-Education Terminology Problems

Many common terms have very different meanings in the two communities, and sometimes neither is appropriate. The slogan “understanding, not rote learning or mechanical calculation”, for example, has been quite influential. Unfortunately the mathematical meaning for “understand”, adapted to support long-term learning in math, is too strong to be a realistic goal in K-12. Equally unfortunately, the weak math-ed sense is easily achieved but does not support long-term learning. Actual solutions will require us to transcend terminology problems. Possibilities are explored in the next essay.

Chapter 18: Communication Between the Math and Math-Education Communities

Communication between K-12 and college educators is sorely needed to reverse a decline in preparation for study in technical fields. Attempts have been largely unsuccessful and sometimes so unpleasant they are described as “math wars”. I analyze obstacles and particularly try to separate linguistic differences from conflicts of underlying mindsets and priorities. Annotated lists of sample problems offer a good solution, but philosophical preference for abstract high-level discourse seems to rule this out.

Chapter 19: Evaluation of Methods in Math Education

Some methodologies in education research and curriculum development, especially in the U.S., seem almost designed to generate spurious findings and discourage deeper insights. As a result the current emphasis on “research-validated” methods is more likely to block good ideas than to promote them.

Web Sites**1.7.0.1 EduT_EX Working Group on T_EX-based educational materials.**

The web site for this project is at <http://edutex-wiki.tug.org/wiki/>

The Wiki for a NSF-funded project to develop support for computer-based testing and other educational materials.

The Math Emporium at Virginia Tech

<http://www.emporium.vt.edu>

This facility supports computer-based and computer-tested math courses for more than 5,000 students each semester. The essays in this collection are largely based on experience with developing computer materials for the Emporium and many hundreds of hours of one-on-one help sessions with students working there.

Personal web page

<http://www.math.vt.edu/people/quinn/> The education link on this page leads to PDF versions of these essays, as well as reports and draft versions of other essays. Many of these concern the Math Emporium.

1.8 References

- [**Barton**] Kevin Barton; Jonathan Fugelsang; and Daniel Smilek; *Inhibiting beliefs demands attention*, *Thinking and Reasoning*, **15** (2009) pp. 250–267.
- [**Dunbar**] K. Dunbar; J. Fugelsang; and C. Stein; *Do naïve theories ever go away?* In M. Lovett, and P. Shah (Eds.), *Thinking with Data: 33rd Carnegie Symposium on Cognition*, Erlbaum pub. (2007) pp. 193–205.
- [**Gray**] Jeremy Gray; *Plato’s Ghost: The Modernist Transformation of Mathematics*, Princeton University Press 2008.
- [**Knorr-Cetina**] Karen Knorr-Cetina; *Epistemic Cultures*, Harvard University Press 1999.
- [**Kuhn**] Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 1st. ed., Univ. of Chicago Pr., 1962.
- [**Quinn**] Frank Quinn; *Towards a Science of Contemporary Mathematics*, current draft available on personal web page.
- [**Wilson**] E. O. Wilson; *Consilience: The Unity of Knowledge*, Knopf, New York 1998.

Contents

1	Preface	3
1.1	Science	4
1.2	Philosophy	5
1.3	Humans	6
1.4	Scientific Revolutions	6
1.5	Education, and this Work	8
1.6	Organization	9
1.7	Brief summaries	10
1.8	References	15
I	The Cognitive/Activity Level	25
2	Neuroscience Experiments	27
2.1	Outline	27
2.1.1	Background	27
2.1.2	Cognitive interference, outline	28
2.1.3	Subliminal learning and reenforcement, outline	31
2.2	Cognitive interference	34
2.2.1	Cognitive interference in multiplication	34
2.2.2	Cognitive interference in word problems	38
2.2.3	Interference from customary usage of parentheses	41
2.2.4	Interference from customary integer names	42
2.3	Subliminal learning and reenforcement	44
2.3.1	Subliminal algebra in integer multiplication	45
2.3.2	Subliminal learning of number facts	47
2.3.3	Kinetic reenforcement of geometric structure	50
3	Mathematics Education versus Cognitive Neuroscience	57
3.1	Background and outline	57
3.1.1	Outline	58
3.1.2	Note on style	58
3.2	The macro/micro spectrum	59
3.3	Ineffective cognitive psychology	59

3.3.1	Specific questions	59
3.4	Ineffective neuroscience	60
3.4.1	Multiplication	60
3.4.2	Solving equations	62
3.4.3	Errors	62
3.4.4	Word problems	63
3.4.5	Summary	64
3.5	Dangerous neuroscience	64
3.6	Ineffective education	67
3.6.1	Lack of scientific skills	67
3.7	Mathematics and learning	68
3.7.1	Teaching vs. diagnosis	68
3.7.2	Modern mathematics	69
3.7.3	Educational hostility	69
3.7.4	Theoretical incoherence	70
3.7.5	Summary	70
3.8	Appendix: Technical difficulty, and consequences	71
3.8.1	Technical difficulties	71
3.8.2	Needs of neuroscience	73
3.9	Fragments	74
3.9.1	Unmet needs	74
4	Contemporary Proofs	81
4.1	Proofs, Potential Proofs, and Formal Proofs	82
4.1.1	Potential Proof	82
4.1.2	Formal Potential Proof	84
4.1.3	Proof and Correctness	87
4.1.4	The Role of Diagnosis	88
4.1.5	Other Views of Proof	88
4.2	Proof Templates	89
4.2.1	Polynomial Multiplication	90
4.2.2	Solving Equations	91
4.2.3	Standardizing Quadratics	92
4.2.4	Summary	94
4.3	Long Problems	95
4.3.1	Big Multiplications	95
4.3.2	Big Additions	97
4.3.3	Digits in Big Products	98
4.3.4	Puzzles	101
4.4	Word Problems and Applications	102
4.4.1	Word Problems and Physical-World Applications	102
4.4.2	Applications	103
4.4.3	Mathematical Applications	105

5	Proof Projects for Teachers	109
5.1	Fractions	110
5.1.1	Commutative Rings	111
5.1.2	Examples of Commutative Rings	111
5.1.3	Preliminary Definition of Fractions	112
5.1.4	Inverses	113
5.1.5	Unpacking Definitions	114
5.1.6	More about Inverses	114
5.1.7	Difficulties with the Preliminary Version	115
5.1.8	Zero divisors	115
5.1.9	Final version of the Definition	115
5.1.10	Rings of fractions	117
5.1.11	Ring Homomorphisms	118
5.1.12	Grothendieck Groups	118
5.2	Area	121
5.2.1	Polygonal closed paths	121
5.2.2	The Project	122
5.2.3	A Difficulty, and a Strategy	123
5.2.4	Properties of A	124
5.2.5	Orientations and Convex Polygons	125
5.2.6	Simple Closed Polygons	127
5.2.7	Winding Numbers	128
5.3	More About Area	130
5.3.1	Area and Rings	130
5.3.2	Area and Vectors	132
5.4	Derivatives	136
5.5	Comments for Students	137
5.5.1	Route to a Formula	137
5.6	Comments for Educators	138
5.6.1	Formal Definitions and Unpacking	138
5.6.2	The Fraction Project	139
5.6.3	The Area Project	141
5.6.4	Cautions about Definitions and Internalization	142
5.6.5	Summary	143
6	K–12 Calculator Woes	147
6.1	Disconnect from Mathematical Structure	147
6.2	Disconnect from Visual and Symbolic Thinking	148
6.3	Lack of Kinetic Reinforcement	148
6.4	Lack of Verbal Reinforcement	148
6.5	Summary	149

7	Student Computing: Interface Design	151
7.0.1	Learning and the Interface	152
7.1	Guiding Principles	152
7.1.1	Learning, not Technology	152
7.1.2	Symbolization	153
7.1.3	Modes of Learning	154
7.1.4	Process, not Answers	158
7.2	Interface Design	159
7.2.1	Input Modes	159
7.2.2	Windows and Sessions	161
7.2.3	Input Formats	163
7.3	Summary	165
II	The Course/Curriculum Level	167
8	Task–Oriented Math Education	169
8.1	Goals and descriptions	169
8.1.1	Descriptions	170
8.1.2	Discussion	170
8.1.3	Summary	171
8.2	General Constraints	171
8.2.1	Non–terminal courses in a task–oriented subject	172
8.2.2	Students capable of modestly independent work	172
8.2.3	Computer–based	172
8.2.4	Helpers	172
8.2.5	Traditional class meetings	174
8.2.6	Summary	174
8.3	More about Tasks	175
8.3.1	Tasks are not Assessments	175
8.3.2	Learning–goals and strategies	176
8.3.3	Task Constraints	176
8.3.4	Multiple tries in assessment	178
8.3.5	Software generation	179
8.3.6	High–stakes tests	179
8.3.7	Summary	180
8.4	Course Design	180
8.4.1	Not An Online Course	180
8.4.2	The Skeleton	181
8.4.3	Student behavior	185
8.4.4	Beyond the Skeleton	186
8.4.5	Summary	188
8.5	Resource Requirements	188
8.5.1	Increased resources are not an option	189
8.5.2	Operating expenses	190
8.5.3	Startup	191

<i>CONTENTS</i>	21
8.5.4 Development	193
8.5.5 Summary	196
8.6 Educational Opportunities	196
8.6.1 A Behavioral Model	196
8.6.2 The Main Strategy	197
8.6.3 Tracked Courses	198
8.6.4 Summary	198
8.7 Conclusions	199
9 Downstream Evaluation of a Task-Oriented Course	201
10 Beneficial high-stakes tests	203
10.0.1 Outline	203
10.0.2 Other Issues	204
10.0.3 Web Resources	205
10.1 Analysis of the Problem	205
10.1.1 A Sample Test Question	205
10.1.2 Mathematical approach	205
10.1.3 Calculator approach	206
10.1.4 Traditional approach	206
10.2 Diagnosis and a Remedy	207
10.2.1 Remedy	207
10.2.2 Better Sample Question	207
10.3 Test Design and Implementation	209
10.3.1 Learning Tasks	209
10.3.2 Functionality	209
10.3.3 Formats and Programming	210
10.3.4 Advanced Functionality	211
10.4 Conclusions	211
11 Economics of Computer-Based Education	213
11.1 Educational Models	214
11.1.1 On-line	214
11.1.2 Gigantic Lectures	214
11.1.3 Computer-based	214
11.1.4 Computer-tested	215
11.1.5 Computer-Enriched	215
11.1.6 Computer Labs	215
11.1.7 Small Traditional Classes	215
11.2 Program Costs	215
11.2.1 Notes	216
11.2.2 Scale and Efficiency	216
11.3 Outcomes	217
11.3.1 Outcomes in the Program	217
11.3.2 Outcomes outside the Program	218
11.4 More about Economics	218

11.4.1	Goals and Evaluations	219
11.4.2	Internal and External Resources	219
11.4.3	An Example	220
11.5	Summary	220
12	Levels in a mathematics course	223
12.1	Performance Diversity	224
12.1.1	Under-performing students	224
12.1.2	Over-performing students	225
12.1.3	Summary	226
12.2	Placement tests are not the answer	226
12.2.1	False positives and negatives	226
12.2.2	Tests are Untested	227
12.2.3	An Example, and Gateway tests	227
12.3	Tracked Courses	228
12.3.1	Basic Plan	228
12.3.2	Introduction to Proofs course	229
12.3.3	Both Tracks and Levels	230
12.4	Implementation Problems	231
12.4.1	Resource Constraints	231
12.4.2	Institutional Barriers	231
12.5	Conclusion	231
13	Teaching vs Learning	233
13.1	Goals vs. Responsibilities	233
13.1.1	Generalities	233
13.1.2	Calculator arithmetic	234
13.1.3	Computer calculus?	236
13.2	Improved teaching vs. Improved learning	236
13.3	Computer teaching vs. Computer-based learning	237
13.4	Information delivery vs. Diagnosis	237
13.5	Summary	238
III	The Subject Level	239
14	Professional Practice as a Resource for Mathematics Education	241
15	Updating Klein's Viewpoint	243
15.1	The Question	243
15.2	Perspective	244
15.2.1	Historical Baggage	244
15.2.2	Ineffective Response	245
15.2.3	Is Change Needed?	245
15.2.4	Conclusions	246
15.3	Example: Concept Formation	247

15.3.1	Definitions	247
15.3.2	Arguments for the old approach	247
15.3.3	Arguments for the contemporary approach	248

IV The Educational Theory Level 251

16 Dysfunctional Standards 253

16.1	Roles for Standards Documents	254
16.1.1	Tests	254
16.1.2	Textbooks	254
16.1.3	Coordination	255
16.1.4	Process and Outcomes	255
16.1.5	Mathematical Structure	256
16.1.6	Summary	257
16.2	More About Tests	258
16.2.1	Tests as instruments of terror	258
16.2.2	Tests as defective standards	258
16.2.3	Tests as suppressors of quality and diversity	259
16.2.4	Summary	260
16.3	Conclusion	260

17 Math / Math-Education Terminology Problems 261

17.1	A Search for Meaning	261
17.2	Misunderstanding Understanding	262
17.3	Right, Wrong or Different?	262
17.4	Plea	263

18 Math/Math-Ed Communication 265

18.1	Language differences	266
18.1.1	Understanding	266
18.1.2	Remembering, recalling, knowing	267
18.1.3	Applying, evaluating	268
18.1.4	Creating, discovering	269
18.1.5	Teaching vs learning	269
18.2	Mathematics and learning	270
18.2.1	Mathematics	270
18.2.2	Human learning	270
18.2.3	Math education	271
18.3	Communicating about student preparation	272
18.3.1	Quadratic example	272
18.3.2	The general pattern	272
18.3.3	Benefits	273
18.4	Conclusion	275

19 Evaluation of Educational Methods	277
19.1 Missing Criteria	278
19.1.1 Resources	278
19.1.2 Student effort	278
19.1.3 Procrastination	280
19.2 Process as a criterion	281
19.2.1 Multimedia	281
19.2.2 Technology as a goal	281
19.2.3 Trendy methods	282
19.2.4 NCTM Standards	283
19.3 Outcome measurement	283
19.3.1 Student variation and statistics	284
19.3.2 Goal selection	287
19.4 Teaching, learning and errors	290
19.4.1 Teacher-centered education	291
19.4.2 Error correction	291
19.4.3 Repetition	292
19.5 An Obsolete Model	293
19.6 Summary	293

Part I

The Cognitive/Activity
Level

Chapter 2

Neuroscience Experiments for Mathematics Education

November 2010

2.1 Outline

These experiments contrast standard educational approaches that seem to cause problems, with alternatives adapted to mathematical and cognitive structure. Many of the alternatives have been used successfully in individual cases, so the job of neuroscience is not so much to detect the effects as to quantify them and clarify the mechanisms. Moreover, the tools of neuroscience are difficult and indirect so specific tasks are designed to maximize expectation of clear outcomes.

The specific examples might be thought of as opening moves in collaborations, addressed to neuroscientists. They are sensible of the needs and limitations of neuroscience, and informed by neuroscience studies, but for the most part are not specific about neuroscience protocols because this is the job of potential collaborators. Cognitive, mathematical, and educational issues are discussed in some detail because these are responsibilities of the mathematical-educational partner.

These studies should have significant implications for educational practice. A companion essay [16](c) explores why educationally-oriented neuroscience studies to date have had so little impact. In this context, a motivation for developing high-impact examples was to test this picture, and in particular see if there are barriers not yet identified.

2.1.1 Background

A few clarifications about background and constraints for these proposals.

2.1.1.1 Diagnostic experience

These proposals draw on extensive diagnostic work with students. The format is a session in which a student describes work in which something has gone wrong, while the helper listens and diagnoses the source of the problem. The helper then (briefly) explains the error and how to avoid it in the future. This gives a much finer-grained view of learning and its problems than does traditional teaching.

A consequence is that the concerns here are only indirectly related to teaching, and are often inconsistent with mainstream (teaching-oriented) ideas about learning.

2.1.1.2 External working memory

Mathematics uses written scratch work as external working memory, and many neural processes are adapted to this.

Suppose, for example, that one step in a procedure collects numbers to be added, and the next step is to add them. The first step is purely symbolic, unconcerned with specific number properties. The second step requires processing them as numbers, but little or no engagement of organizational or symbol-manipulation facilities. The two tasks use different neural processes. Written work is used for reformatting (different processes read symbols in different, task-appropriate ways) and inter-process communication. This is much more efficient and accurate than trying to do it internally.

Consequences for neuroscience studies are:

- Subjects must be able to do scratch work during all procedures.
- A time-stamped record of external work provides a window on internal activity.
- One goal is to understand the interactions between neural activity and external memory. Effectiveness can often be substantially improved by optimizing procedures and notation for external-memory use.

2.1.1.3 Mathematical structure

This is a cautionary note. Mathematical work is both enabled and constrained by mathematical structure, and experimental design and interpretation must be carefully adapted to this structure. The discussion here only hints at the constraints involved, and these should be fully understood before modifications or variations are undertaken.

2.1.2 Cognitive interference, outline

The four examples in §2.2 concern interference between subtasks of a task. In a nutshell, mixing or switching between subtasks can reduce success and limit

scope of application. In some cases procedures or algorithms could be reorganized to separate or eliminate such tasks, but this is constrained by mathematical structure and the need to avoid introducing new cognitive difficulties. The challenge is to identify interference severe enough to need change, in circumstances where structure permits it.

2.1.2.1 Polynomial multiplication

The example in §2.2.1 concerns conflict between the organization and calculation aspects of multiplication. The experiment uses multiplication of polynomials by high-school or beginning college students, comparing performance with standard (mixed-task) methods and a task-separated algorithm. The immediate outcome should be better methodology for polynomials. A long-term goal is better methodology for multi-digit multiplication in elementary mathematics, and this is taken up in §2.3.1. The polynomial version provides a simpler (in some cognitive senses) and slower (for imaging purposes) model.

The context for the study is task switching between two basic task types. There is an extensive literature on mechanisms and costs of switching in very simple tasks, see the review [12]. The ACT-R computer model [1], [2] has been extensively tested seems to model some elementary tasks reasonably well, c.f. [20]. The work done so far is only marginally relevant, however. It corresponds, roughly speaking, to discrete behavior of matter at very small scales, while we are concerned with statistical behavior at significantly larger scales.

A useful general conclusion from cognitive psychology [14] and clearly visible in small-task studies, is that our thinking is essentially single-track. Ample working memory and frequent switching between superficial tasks give the impression of “multitasking”, but this is mostly an illusion [24][c]. Specifically, if we switch from task A to a different task B but know we will shortly be doing A again, we usually cannot economize by keeping task- A instructions loaded but off-line. Instead we have to flush task- A material, load B , and when B is complete, reverse the procedure. Further

- “Flushing” task A may involve *inhibiting* task- A instructions, not just emptying a buffer. Some errors (e.g. adding instead of multiplying) result from incomplete inhibition.
- Residual effects of this inhibition can slow or complicate reloading for the next A task. In other words, repeated switching reduces the effectiveness of working memory [11]
- Following an A task by another instance of A usually requires less reorganization and has lower costs.

Algorithms with subtasks $ABABAB\dots$ usually cannot be reorganized as $AAABBB$ for mathematical reasons, but when it is possible it should have cognitive benefits. The multiplication proposal is of this type.

2.1.2.2 Interference from customary usage

Customary usage often interferes cognitively with mathematical work. This is usually easy to fix by changing notation, avoiding customary forms, or introducing translation as an explicit separate step. But customary usage is, essentially by definition, transparent to adults. As a result the problems caused by it are invisible, and attempts to deviate from customary usage make adults uncomfortable and are resisted. The role of neuroscience is to help locate these problems, and unequivocally identify them as problems.

For example \sqrt{A} is a customary notation for $A^{1/2}$. The exponential form fits into general patterns and is usually easier to manipulate. The customary notation alerts us to the possibility of using special properties of square roots. The best practice in this case seems to be to use the customary notation so we get the special-property alert, but teach students that in most problem types the first step should be to rewrite it in exponential form. In contrast the customary notation $\sqrt[3]{A}$ for cube roots does not have benefits that compensate for translation overhead. This should always be written as $A^{1/3}$.

§2.2.3 concerns rather severe problems with customary use of parentheses. The problem is discussed because it is important and effects several other proposals. No specific experiment is proposed, however, because it has been difficult to find one with clear and useful outcomes.

§2.2.4 concerns the translation overhead of irregular customary names for integers (e.g. ‘thirteen’ for 13). Cross-cultural and imaging studies suggest that short-term working memory is mostly verbal, even when working with numbers. For instance, the total length of names for things is often a stronger limit than the number of things. Another clue comes from the additional difficulty children have in learning to count in languages with irregular customary number names.

A conclusion is that counting and arithmetic might—in some languages at least—be simplified by the use of “math names” for numbers. The experiment explores this through its effect on iterated addition.

2.1.2.3 Word problems

The example in §2.2.2 concerns cognitive interference between the modeling and analysis aspects of word problems. The experiment compares performance and neural activity in the standard (mixed-task) approach and a task-separated modeling approach. The short-term goal is to show that the educational approach is counterproductive. The longer-term goal is to explore ways to use word problems effectively in elementary education.

Educators see word problems as essentially mathematical; a different format rather than a different activity. As a result educators encourage a gestalt approach in which students “develop strategies” to work directly with the formulation of the problem. Students find this hard, and accessible problems have either mathematical or modeling component (or both) trivial.

Mathematicians and professional users of mathematics split real-world applications into two steps: ‘modeling’ translates physical data to a self-contained

symbolic formulation called ‘the model’, and then the model is analyzed mathematically. These two steps use very different methods and, technically, the modeling step is not mathematical. Diagnostic experience with students suggests that modeling and analysis are also quite different cognitively. Mixing seems to cause interference considerably stronger than that seen in multiplication, and success in science, engineering, and related mathematics, requires use of task-separated modeling.

2.1.3 Subliminal learning and reinforcement, outline

We are concerned with learning that takes place during an educational activity such as a lesson or assignment, but that is invisible to the student, and frequently to educators as well. There are two variations: subliminal learning from the content; and learning that depends in an unrecognized way on the structure (e.g. kinetic or verbal) of the activity.

When mathematics is done by hand, a lot of activity comes as packages that are activated by simple goals. New methods—especially technology—cause these packages to come apart, and important subliminal learning may be lost. For instance “find 365×86 ” requires a lot of neural activity when done by hand and rather less when a calculator is used. Is the extra activity pointless, or are the main benefits in the package rather than the number obtained?

Diagnostic work with students suggests that there is quite a lot of subliminal learning in by-hand elementary mathematics that is lost in calculator curricula [16](d). A goal of the experiments is to fix this: understand instances well enough to design programs that use technology and also provide this learning. Curiously, this should also enable improvement of traditional programs. Subliminal learning in by-hand work is usually accidental and inefficient. Better understanding should enable more efficient approaches, either with or without technology.

The role of neuroscience is this: neural effort in well-learned skills is usually focused in a small number of regions. Early attempts usually recruit much wider activity, and development requires exercising the necessary regions and connections between them, and also requires suppression of recruitment of unneeded regions. Neural activity alone is not a definitive guide to learning, but it gives excellent clues:

- Activities that exercise appropriate regions probably contribute to skill development.
- Activities that do not engage these regions cannot contribute much to skill development.
- Too much emphasis on activities that consistently engage unnecessary regions may impede skill development.

2.1.3.1 Subliminal algebra in integer arithmetic

This experiment concerns subliminal internalization of algebraic structure from by-hand integer multiplication. The point is that the symbols we write to represent numbers are symbols, not numbers, and by-hand arithmetic involves a lot of symbol manipulation. Students seem to internalize some of the algebraic structure used in these manipulations.

The place-value notation presents integers as polynomials in powers of ten, with single-digit coefficients. For instance $438 = 4 \cdot x^2 + 3 \cdot x^1 + 8 \cdot x^0$, with $x = 10$. The standard algorithms for multi-digit multiplication essentially multiply the corresponding polynomials and then evaluate at 10.

The experiment in 2.3.1 has two parts. The first compares neural activity in 3×3 -digit multiplication by hand, and with a calculator. An objective is to see to what extent the hand work recruits neural regions used in algebra, and more specifically in polynomial multiplication.

The second part explores the use of a task-separated algorithm modeled on the polynomial algorithm of §2.2.1. The first version is for hand use. It requires more writing than the traditional algorithm but should display the structure more clearly and be easier to use accurately. The second version uses a calculator, but in a way that still requires expansion and display of algebraic structure. The objectives are to assess potential cognitive benefits by comparing neural activity with that associated to standard by-hand multiplication.

There are many studies of numerical multiplication, c.f. [9, 20]. Unfortunately conceptual and methodological weaknesses [21], [16](c) render these only marginally relevant.

2.1.3.2 Subliminal learning of number facts

Examples illustrating mathematical procedures almost always have arithmetic subtasks. If these subtasks can be done transparently then new features can be seen and, conversely, if the subtasks require serious breaks in attention then new features will be obscure. A certain amount of transparent mental arithmetic is therefore vital for learning in algebra and beyond.

For instance, the example used to illustrate the task-separated polynomial multiplication algorithm in §2.2.1.2 has coefficients contrived so that the calculation steps should be very easy, but still display the structure of the algorithm. We presume that much more complex problems could be worked once the algorithm is understood, even if coefficient calculations required the use of calculators or extensive scratch work. Transparency, however, is necessary for initial learning.

The amount of mental arithmetic needed is a compromise between what students can learn relatively easily, and how simple the examples can be contrived to be and still effectively illustrate structures. The standard compromise is:

- Fully-automatic addition and subtraction of single-digit integers.

- Addition of four or five single-digit integers, or one two-digit and one three-digit integer.
- Multiplication of a single-digit and a two-digit integer.

Further, the *structure* of arithmetic should be automatic enough that a few symbols will not cause a problem.

Slightly more complex tasks, such as multiplying a one-digit and a three-digit integer, should be a minor distraction. Multiplying two 2-digit integers is cognitively more complex (see §2.2.1.1) so it is worth going to some lengths to avoid it. It seems likely that a better algorithm would put larger addition problems within easy reach, see §2.2.1.5, but this does not seem to be a bottleneck in actual use.

A consequence is that calculators cannot substitute for automatic recall of single-digit multiplication facts. This does not mean we are stuck with rote memorization, however. The proposal in §2.3.2 section explores a subliminal approach using the algorithm described in §2.3.1.

2.1.3.3 Kinetic reinforcement in graphing

This experiment concerns reinforcement of internalization of geometric structure of function graphs, by the kinetic aspect of by-hand graphing. In non-technology programs, both assignments and testing require drawing by hand. In programs using technology, student work has visual outcomes, and testing is also usually visual (choose the correct graph among a number of alternatives).

Diagnostic experience suggests that many graphing-calculator trained students cannot either verbally describe or qualitatively sketch standard curves. When they do try to draw pictures they often reproduce a calculator display, to scale, with typical poor choices of range and microscopic features of interest. In other words they have not internalized the qualitative geometric structure. It seems that the kinetic aspect of drawing powerfully (and subliminally) reinforces learning of qualitative structure, and some students seem unable to learn without this reinforcement.

A general context is that serious learning benefits from, and often requires, active reinforcement. Recent studies ([18], [19]) report that young children do not learn from videos. To learn vocabulary, for instance, they must say the word, not just hear it. Verbal reinforcement seems to be more effective when ‘social cognition’ facilities are engaged by the presence of an attentive human, and this may be the primary mechanism in some cases. None of this should be a surprise. Children in rural areas learn their local dialects but usually not (in the US) standard English, even though they hear as much or more standard English on television. Similarly, what a child sees makes far less impression than what he draws or writes himself.

The experiment uses two versions of a brief lesson on qualitative graphs of sums of functions [[draft note: task to be changed]]. The first version uses a typical visual computer-graphic approach, and the second a hand-drawing approach. Students are quizzed using version-appropriate methods: visual multiple-choice

in the first case, drawing in the second. Finally they are tested with the opposite methodology.

The questions concern similarity and differences in neural activity in the two modes, and transfer of learning from one mode to the other. Diagnostic experience suggests that kinetically-reinforced learning should transfer, visual learning usually will not. This experiment is more complex than the others because the questions concern neural activity during learning, not just during use of a learned procedure.

This is the end of the outline.

2.2 Cognitive interference

Mixing different tasks often slows and degrades performance in both. It seems likely that such interference has a neural basis. Understanding this should enable design of algorithms and procedures better adapted to humans use, mainly by separating internal tasks and using scratch work for reformatting and high-precision interprocess communication. The proposals address two instances in which interference has been observed: multiplication and word problems. See §2.1.2 above for an outline.

2.2.1 Cognitive interference in multiplication

There are two important cases that use essentially the same algorithm: multi-digit integer multiplication in elementary school, and polynomial multiplication in high school and college. We begin with polynomials because:

- the polynomial version is actually a bit simpler because there are no overflow problems associated with converting polynomial-like outcomes into place-value integer notation;
- the separated polynomial tasks take long enough to be imaged by fMRI, and this is unlikely with integer multiplication;
- more-extensive scratch work (external working memory) can be used to correlate cognitive and neural activity;
- high-school or college students are more consistent and cooperative experimental subjects;
- arithmetic skills of older students are already well-established and stable, and should produce clearer and more consistent signals.

Another reason to begin with polynomials is that the problem seems to involve a genuine limit on human ability: experienced professors of mathematics seem to have as much trouble as students with the mixed-task algorithm, and get as much benefit from the task-separated version. This should mean that the underlying neural issues should be relatively uniform and clear. In contrast,

the interference experienced by children with multi-digit integer multiplication can be eventually be managed, so may have a developmental component. In fact there are likely to be a number of different difficulties, and before we can assess any of them we must understand the mature endpoint. Further, the proper course of action may be unclear. If the problem is *only* developmental then finding ways to speed development would probably be more useful than tinkering with algorithms.

The integer case is discussed further in §2.3.1. Detailed mapping of component functionality is discussed in [16](e).

2.2.1.1 Sample problems

These examples show escalating conflict between organization of the polynomial structure, and coefficient arithmetic. Coefficients are contrived so individual operations are easy; difficulties come from mixing rather than from individual operations.

- 1) Write $(3x^2)((2-a)x^3)$ as a polynomial in x .

Note that “simple arithmetic” in coefficients may include symbols, to emphasize that we need transparent internalization of structure (associative, distributive etc.), not just number facts. This example has one coefficient operation and one polynomial operation: they are performed separated and there is little conflict.

- 2) Write $(3x^2 - x + 5a)((2-a)x^3)$ as a polynomial in x .

The result has three terms. The standard practice is to do coefficient arithmetic as each term is generated, so there are two arithmetic interruptions of the polynomial procedure. There is relatively little interference, partly because there are few interruptions. Another reason is that the structure of first term provides a template for sequential organization of the task. Minor interference is suggested by more-frequent sign mistakes with the -1 coefficient on x in the first term, as compared to errors in isolated arithmetic tasks.

- 3) Write $(3x^2 - x + 5a)(x^3 + (2-a)x^2 - a)$ as a polynomial in x .

Simple expansion gives nine terms, with eight interruptions for coefficient arithmetic. Moreover the data is a 3×3 array so a strategy for organization as a sequential task must be devised. Finally, terms with the same coefficient have to be collected and combined. The success rate is low and errors in both organization (missed terms) and arithmetic are common. The difficulty comes from the algorithm rather than the problem itself, however, as we see next.

2.2.1.2 Task-separated algorithm

The basic plan is to separate different tasks as completely as possible. In polynomial multiplication, organizational work related to the polynomial structure should be completely separated from coefficient arithmetic, and multiplication

and addition separated in the arithmetic. This is illustrated with problem (3) above: write

$$(3x^2 - x + 5a)(x^3 + (2 - a)x^2 - a)$$

as a polynomial in x .

Step 1 A preliminary scan shows that the output will be a polynomial of degree 5. Set up a template for this:

$$x^5(\quad) + x^4(\quad) + x^3(\quad) + x^2(\quad) + x^1(\quad) + x^0(\quad)$$

Step 2 Fill in the blanks one at a time. For example, the terms with total exponent 3 are obtained as follows: the highest-order term in the first factor is x^2 ; record its coefficient (3). The complementary exponent is 1, but there is no x^1 term in the second factor so we record 0. Move to the next lower power in the first factor and the next higher in the second, and record coefficients $(-1)(2 - a)$. Continue to get $((3)(0) + (-1)(2 - a) + (5a)(1))$. Put everything in parentheses, and not do *any* arithmetic on the fly. Do not, for example, omit the 3 coefficient on x^2 because there is no complementary term in the second factor, and do not write $(5a)(1)$ as $5a$.

- This enables reading the coefficients only as strings to be copied, with no arithmetic significance. This reduces cognitive overhead.
- Even completely trivial arithmetic requires a momentary change of gears, and watching for an opportunity to do it is a distraction.

The outcome is:

$$x^5((3)(1)) + x^4((3)(2 - a) + (-1)(1)) + x^3((3)(0) + (-1)(2 - a) + (5a)(1)) + x^2((3)(-a) + (5a)(1)) + x^1((-1)(-a)) + x^0((5)(-1))$$

Step 3 Do multiplications:

$$x^5 \underbrace{((3)(1))}_3 + x^4 \underbrace{((3)(2 - a) + (-1)(1))}_{6-3a} + x^3 \underbrace{((3)(0) + (-1)(2 - a) + (5a)(1))}_0 + \dots$$

The process and notation is designed to avoid organizational activity: input for each operation is in standard position in the visual field, the underbrace specifies the input so it does not have to be reconstructed for other steps or checking, and output is put in a standard place.

Step 4 Do additions.

$$x^5 \underbrace{((3)(1))}_3 + x^4 \underbrace{((3)(2 - a) + (-1)(1))}_{5-3a} + x^3 \underbrace{((3)(0) + (-1)(2 - a) + (5a)(1))}_{-2+6a} + \dots$$

Again the process and notation minimize organizational activity that would interfere with arithmetic.

2.2.1.3 Experiment

The subjects should be high-school students who have been successful in a standard algebra curriculum, or students in first-year college calculus (i.e. not remedial). Proficiency with problems like (2) above might be used as a criterion. They should also be screened for dependence on calculators for basic arithmetic (see below). They are asked to work problems similar to the one above, using standard methods. They are then taught the task-separated version, and after enough practice to become familiar with it, they are imaged working similar problems with this methodology.

- To keep the picture clear the arithmetic should be kept minimal. Multiplication of multi-digit integers, for instance, would produce a small version of the entire process.
- Half the problems should have numerical coefficients, half have symbols in the coefficients (as in the example).
- Subjects should be told that accuracy is more important than speed. Errors due to speed or carelessness will mask significant features of more intrinsic mistakes [17, 7, 6].
- Scratch work should be videotaped and time-stamped, for correlation with imaging results.

fMRI should provide general information about the areas used and the degree of usage, c.f. [9, 20]. It would be useful if there is an easily-identified MEG or EEG signature of major task switching—a mental shifting of gears—in the task-separated versions, c.f. [22].

2.2.1.4 Calculator version

The core experiment concerns students with good manual arithmetic skills. If resources permit, it can be expanded to include students with similar proficiency but who use calculators for numerical work. Let them use calculators as they like during the trials, and record this use. Expected differences are described below.

2.2.1.5 Analysis

The basic plan is to look for neural and performance differences between the standard and task-separated versions. The expectation (based on diagnostic work with students) is that performance should be significantly better with the task-separated version, and the hypothesized reason is that the task-separated version reduces interference caused by arithmetic interruptions of the polynomial organizational task. A qualitative picture should emerge reasonably quickly. It might be possible to directly explore interruptions and their short-term neural consequences by carefully correlating imaging with scratch work.

The above is the basic plan. We now discuss potential complications and refinements.

First, there may be a sub-population with substantially better performance with the modified algorithm. The goals are algorithms that benefit everyone when used as the standard approach, but it is unlikely that everyone will benefit when they are used as a retrofit. The cognitive interference signal should be clearest in the high-performance group. Note that subjects cannot be screened in advance for quick adaptation because the control experiment (using standard techniques) becomes impossible after the modified algorithm is taught. Predictors of success found after the fact, however, would certainly be useful.

Calculator arithmetic requires a significant attention shift and input/output processing, and there are a great many discrete arithmetic tasks in these problems. It is hard to imagine that calculator use could become so transparent that this would not be a source of interference. The prediction, therefore, is that students who actually make substantial use of calculators during the trials will have low success with any form of these problems.

Next, if at all possible, individual variation in the task-separated version should be investigated. Currently the statistical techniques used to analyze data have a built-in assumption that everyone does these things in essentially the same way. Variation is treated as noise. The data showing that multiplication facts are stored in the angular gyrus using verbal memory, for instance, demonstrates that this is the dominant mode. But is it really true that no-one uses visual memory for this? Understanding variation in successful learning is essential for understanding all the barriers to success, and the separated tasks may be long and uniform enough to permit this. Again, students who use calculators are likely to have significantly different characteristics.

The number and nature of mistakes made is more significant than time required to complete the tasks. Time measurements might be useful for comparing different task instances done by one individual, however.

Finally, it will be very important to assess the effects of symbols in the coefficients. The hypothesis suggested by behavioral data is that students who have effectively internalized the symbolic structure of arithmetic should show little difference in either performance or neural activity. There is some support for this in very simple tasks [2], [23]. Conversely, students who have not internalized this structure, or who think of symbols and numbers as essentially different, will find symbolic coefficients significantly more difficult. Most calculator users are likely to be in this group.

2.2.2 Cognitive interference in word problems

The modeling and analysis components of word problems seem to interfere when mixed, and this interference is often very strong. This is explored through comparison of student work using standard (mixed-task) and modeling (task-separated) procedures. See §2.1.2.3 for discussion.

2.2.2.1 Sample problem

The following have the same mathematical core.

Food version A basket contains six loaves of bread. Half of these are put in another basket that already contains nine loaves. Then one-third of the total contents of the second basket is put in the first. How much bread ends up in the first basket?

Social version Jen and Brad have six loaves of bread. Brad takes half with him when he leaves to share *everything* with Angelia, who already has nine loaves. Jen's lawsuit against Brad and Angelia is settled by giving her one-third of Brad and Angelia's bread. How much bread does Jen end up with?

Money version A basket contains six dollars. Half of these are put in another basket that already contains nine dollars. Then one-third of the total contents of the second basket is put in the first. How much money ends up in the first basket?

These are easy to model and solve, but difficult with the gestalt approach because interpretation and calculation are mixed.

2.2.2.2 Task-separated (modeling) version

Let A denote the bread in the first basket, with subscripts 1, 2, 3 corresponding to the three times. B_i similarly denotes the bread in the second basket. Translating the data for the beginning state gives:

$$A_0 = 6, B_0 = 9.$$

Changes that give the second state translate as:

$$A_1 = A_0 - \frac{1}{2}A_0, B_1 = B_0 + \frac{1}{2}A_0.$$

Finally changes that give the third state give:

$$A_2 = A_1 + \frac{1}{3}B_1, B_2 = B_1 - \frac{1}{3}B_1.$$

This is a symbolic form (model) suitable for mathematical analysis. After doing a few of these they become immediately recognizable as short recurrence relations.

Analysis proceeds in two stages; first substitute in two steps to reduce to a numerical problem:

$$A_2 = A_1 + \frac{1}{3}B_1 = (A_0 - \frac{1}{2}A_0) + \frac{1}{3}(B_0 + \frac{1}{2}A_0) = 6 - \frac{1}{2}(6) + \frac{1}{3}(9 + \frac{1}{2}(6))$$

and finally do the the arithmetic. See §2.2.2.5 for discussion of cognitive and conceptual features.

2.2.2.3 Experiment

The subjects are high-school students who have been successful in a standard algebra curriculum. They are imaged working problems similar to the ones above. They are then taught the task-separated version, and after enough practice to become familiar with it, they are imaged working similar problems with this methodology. They should be asked to give the model as part of the solution (to ensure actual separation), and some problems should ask only for the model.

In both trials, problems to be worked should be interspersed with controls in which students are asked only to identify problem type (food, social, etc.).

Finally, subjects should be interviewed before and after the imaging trials. Pre-trial questions would concern attitudes toward word problems (enjoy, dread, etc.), neutrally probe reasons (actually interesting, easy grades because the math is trivial, believe teachers' assertion that they are important, etc.), and ask the subject's impression of his general competence and success rate. Post-trial questions would include feelings about task separation (helps, is a waste of time), and assess changes in interest and feelings of competence.

There are two points to the interviews. First, is there a correlation between reduced cognitive interference and increased interest or confidence? Second, the main justifications for word problems are motivation and relevance, because the analytic tasks are trivial. It is therefore important to assess how a procedural change might effect these.

2.2.2.4 Analysis

Unseparated work should show extensive activity, probably including prefrontal recruitment to sort out confusion from interference. Active areas will probably depend on the nature of the problem, and different types should be analyzed separately to see this. The social version, for instance, should engage neural structures devoted to interaction with others of our species. Comparison with type-identification versions should reveal activity specific to the mathematical task. Questions:

- Do some types interfere with mathematical activity more strongly than others (i.e. have lower success rates)?
- Do different types lead to differences in the mathematical components, as revealed by subtracting type-identification responses?
- Is there systematic variation, for instance sex differences in responses to social versions, or socioeconomic level effects in responses to food or money versions? If so, how do these correlate with success rates?

Subtasks in unseparated work will have irregular timing and sequencing, and will be hard to image. This is not a big problem.

Task-separated versions should show clearly-defined shifts between modeling and analysis. Questions are:

- How do the areas and degrees of activation compare to the non-separated versions? For instance, are the same areas used, just in sequence rather than simultaneously?
- Modeling has some symbolic activity, and this should be revealed by subtracting type-identification responses. Where does this take place, and is it essentially the same for all problem types?
- The symbolic aspect of modeling seems not to interfere with other parts of the process, as long as no analysis is done. Is this true on the neural level, or does it reveal interference too mild to be obvious?

2.2.2.5 Further discussion

The immediate cognitive benefit of the task-separated version is that translation and analysis are both routine and reliable, and can be extended. Adding another layer, for instance if Brad goes back to Jen and there is another redistribution of bread, could easily be done in the task-separated version but would be a serious challenge with the gestalt approach.

Modeling also has conceptual benefits. The model displays the mathematical structure as a recurrence relation rather than a sequence of arithmetic operations. Similar models describe superficially different problems, showing the underlying unity and demonstrating the power of abstraction. It can be connected to other methodologies, for instance vectors and matrices: set $C = (A, B)$ and the model becomes

$$\begin{aligned}
 C_0 &= (6, 9) \\
 C_1 &= \begin{pmatrix} 1/2 & 0 \\ 1/2 & 1 \end{pmatrix} C_0 \\
 C_2 &= \begin{pmatrix} 1 & 1/3 \\ 0 & 2/3 \end{pmatrix} C_1
 \end{aligned}$$

Multiplying the coefficient matrices gives a direct description of output from input and enables exploration of the relationship. Is there an initial distribution that leads to exactly the same final distribution? In another direction, one can also see how a large number of “players” could give a cellular automaton.

Finally, modeling can be a rich activity even when students cannot analyze the model. For instance as soon as ‘rate of change’ is introduced they could model physical systems as differential equations, and then see computer graphs of solutions. Or they might be motivated to learn relevant analytic techniques.

2.2.3 Interference from customary usage of parentheses

Grouping of sub-expressions is an essential part of the structure of most mathematical expressions. Further, parsing expressions should follow this structure from the outside in: locate outermost groups and their relationships, then find immediate subgroups of these, and so on down to indivisible components.

Customary usage interferes cognitively with mathematical work in two ways:

- The customary parsing order used in reading (left to right in English) is almost always different from mathematical parsing order.
- The customary parenthesis notation does a bad job representing grouping: the opening ‘(’ and closing ‘)’ of a group are mathematically connected, but they have to be found by preliminary parsing (usually using reading order) because they are not notationally connected.

Current practice in elementary education is to avoid the issue by avoiding the use of grouping, and largely sticking with reading parsing order. This has unfortunate consequences:

- Most expressions cannot be written without grouping notation, so scope is very limited.
- On-the-fly arithmetic is often necessary to avoid intermediate expressions that require grouping. The task-separated multiplication algorithm used in §2.2.1, for instance, requires extravagant use of parentheses. As a result the cognitive costs of task-switching cannot be avoided.
- Parenthesis-avoidance is embedded in goals: the customary meaning of “simplify” is essentially “find an equivalent expression without parentheses”. This interferes with more intelligent goals in later work.
- Students do not learn how to parse non-trivial mathematical expressions.

We suggest fixing the notation rather than avoiding it. In 2.3.1.4 the underbrace used in §2.2.1.2 to indicate outcomes of evaluation is used to connect parentheses. This is not a good general solution because the underbrace is a powerful way to indicate subexpressions being manipulated, and these subexpressions usually do not correspond to parentheses. A better approach would be to join matching parentheses with an underline:

$$A + B \left(\underline{C - D(E + F)} \right).$$

This seems to address the problems, but it would have to be extensively tested and explored before it could be promoted as a “solution”.

It is quite easy to design experiments to probe the effects described above. However no experiment is suggested here because this is a complex issue, and we have not identified a key or especially revealing special case.

2.2.4 Interference from customary integer names

The English name for 513 is “five hundred thirteen”. This might be shortened to “five thirteen”, but not to “five one three”. It seems likely this interferes with mental arithmetic in two ways: first through overhead in translating 13 to “thirteen” and back, and second because “thirteen” is a cognitive unit that has to be disassociated into two digits for arithmetic processing.

The proposal is to see if the use of “math names” for integers to reduce cognitive overhead associated with customary names improves modest mental addition. The math name is simply the sequence of names of the digits: 513 is “five one three” for example. The other novelty is use of verbal working memory to store the running total, to reduce interference with single-digit operations.

2.2.4.1 Example

To do the addition $367 + 12 + 57$ do the following:

- say “three six seven” out loud, to read it into verbal working memory;
- next add the 1 digit in 12 to the running total. Unless there is an overflow this changes only the 10^1 digit, so the new total will be “three, (new digit), seven”. Begin by saying “three”, then think ‘six is the running-total digit and $6 + 1 = 7$ ’ say “seven”, and finish with “seven” from the running total. The digits said out loud are the new running total.
- next add the 2 in 12. This usually changes only the 10^0 digit in the running total so first say “three seven” from the current total, think ‘seven is running-total digit and $7 + 2 = 9$ ’ and say “nine” out loud.
- now proceed to the 5 digit in 57. The verbal running total is “three seven nine” and we expect the seven to change. Say “three”, think ‘seven is next running-total digit and $7 + 5 = 12$ ’. There is an overflow that increments the previous digit by one so overwrite the “three” by saying “**four**”, then new-digit “two”, then “nine” from the running total.
- finally add the 7 digit in 57. Current total is “four two nine”. Begin with “four two” from the total, think ‘nine, and $9 + 7 = 16$ ’. The overflow changes the 10^1 digit so update this by saying “four **three**”, and then the new digit “six”.

In this context the interference-reducing strategies can be made more explicit. First, the digits are added one at a time, so using math names avoids conflict with common names that combine digits. Second, verbal-auditory short-term memory is distinct from the working memory used for single-digit operations. It may take practice to access it independently. For instance in the second step above, after the addition one must recall the final digit ‘seven’ that was stored before the operation. It might help to think ‘what was the final digit I heard myself say?’

Another helpful learning strategy is to refresh the running total between major steps. For instance the operation $367 + 12$ ends saying “three seven” and then “nine” while adding the 2 digit. Repeating “three seven nine” before beginning the next step helps prevent erosion during preparation for the next step.

2.2.4.2 Experiment

The plan is to compare accuracy and neural activity of mental addition using customary methodology, and with the method described above.

Subjects can be selected as in previous experiments, and should be screened for automatic facility with single-digit sums. Imaging should be preceded by a practice session to refresh use of customary skills.

In the first imaging trial subjects are asked to do tasks mentally (no external working memory) using customary methodology. Tasks (described below) are presented visually and answers are given verbally. There are no time constraints and they are asked to be as accurate as possible.

Subjects are then taught the reduced-interference procedure described above, and practice enough to become reasonably proficient. The imaging trial is then repeated with this methodology.

2.2.4.3 Outcomes, and task selection

The reduced-interference version should enable significantly higher accuracy for some problem types. For instance with practice it should be possible to start with a four-digit integer and add ten two-digit integers, a feat almost impossible to do accurately with traditional methods. This is not a particularly useful skill, however, so the experiment goal is quantitative comparison with smaller problems.

Tasks should be designed so the two methods have clear differences in outcomes and imaged activity. With high-school or beginning-college students it seems likely that adding three 2-digit integers to a 3-digit integer (e.g. $367 + 12 + 57 + 24$) would do this, but task design should be explored with preliminary trials.

The objective of neural imaging is to guide long-term applications of the approach. First, the expectation is that using ‘math names’ for integers from the very beginning would make arithmetic easier for young children. Designing and conducting a trial extensive enough to show clear behavioral advantages would be a huge undertaking. However neural correlates identified in trials with older children might be detectable before behavioral differences are clear. This would enable preliminary studies, and refinement of techniques before full-scale trials.

The method here also employs short-term verbal memory to augment internal working memory. Written external memory is the standard way to do this, and is so effective that it will be the method of choice in nearly all cases. However there are cases where an alternative is useful, and this study provides a starting point for explicit exploration of an alternative.

2.3 Subliminal learning and reinforcement

Human brains are complex, and the relative lack of integration in childrens’ brains means early learning has additional complexity. The fact is well-known

but many of the details are invisible to adults. The proposals concern subliminal learning of algebraic structure in by-hand arithmetic, and reenforcement of qualitative geometric structure in by-hand graphing of functions. Both of these are usually lost in calculator-oriented programs. The goal is to understand these well enough to design programs in which subliminal learning and technology can coexist.

2.3.1 Subliminal algebra in integer multiplication

The first part of the experiment compares multiplications done by hand and with a calculator. This is to establish bases for comparison in the second part, and to compare the by-hand activity with algebraic manipulation. The second part compares two versions of a task-separated algorithm: one by hand, and one with primitive computational support. See the discussion for explanation.

2.3.1.1 Experiment, part one

Subjects should be high school or beginning college students, with reasonable facility with both calculators and hand arithmetic.

The tasks are to find 3×3 -digit products (e.g. 946×735) either by hand using the method they were taught in school, or with a calculator, as directed. Answers should be written in either case. They should be told that accuracy is more important than speed.

2.3.1.2 Discussion, part one

The number of digits is chosen so by-hand work will fully engage the algorithmic structure, but not be overwhelmed by written intermediates.

Neural activity in the calculator case should be input/output and translation of digits to key presses. Little or no numerical or symbolic activity is expected. By-hand multiplication should show input-output, number-fact recall, and organizational activity. The interesting questions concern the organizational activity and errors; see the discussion for part two.

2.3.1.3 Experiment, part two

Subjects are taught to use a task-separated multiplication algorithm modeled on polynomial multiplication, and a final assembly (see below for notation and an example). The experiment has two versions:

- Use the algorithm to reduce 3×3 -digit products to 1×1 -digit products and additions. Carry these out by hand.
- Use the algorithm with 2-digit blocks (see [2.3.1.5](#)) to reduce 6×6 -digit products to 2×2 -digit products and additions. Carry these out with a calculator.

2.3.1.4 Single-digit algorithm

The place-value notation describes integers as polynomials in powers of ten with single-digit coefficients. For example $946 = 9x^2 + 4x^1 + 6x^0$, evaluated at $x = 10$. The plan is to multiply using the polynomial algorithm of 2.2.1.2, then evaluate at powers of ten. Some care with notation is necessary.

We can avoid writing numbers explicitly as polynomials, by writing the exponent over the digit: use $\overset{2}{9}$ as a shorthand for 9×10^2 . For instance to compute 946×735 write

$$\overset{210}{946} \times \overset{210}{735}$$

Next write a template for the organizational step:

$$\overset{4}{*} \underbrace{(\quad)} + \overset{3}{*} \underbrace{\left(\quad \right)} + \overset{2}{*} \underbrace{\left(\quad \right)} + \overset{1}{*} \underbrace{\left(\quad \right)} + \overset{0}{*} \underbrace{(\quad)}$$

Note that parentheses are connected by underbraces that will eventually be used to indicate outcomes. The polynomial model only has the parentheses at this stage, but disconnected parentheses are problematic in elementary education (see 2.2.3). As above $\overset{2}{*}$ is used as a shorthand for 10^2 , but it is not clear this is a good idea.

Collect coefficient products for each total coefficient 0–4:

$$\overset{4}{*} \underbrace{(9 \cdot 7)} + \overset{3}{*} \underbrace{\left(9 \cdot 3 + 4 \cdot 7 \right)} + \overset{2}{*} \underbrace{\left(9 \cdot 5 + 4 \cdot 3 + 6 \cdot 7 \right)} + \overset{1}{*} \underbrace{\left(4 \cdot 5 + 6 \cdot 3 \right)} + \overset{0}{*} \underbrace{(6 \cdot 5)}$$

Then do the multiplication and addition (in separate stages):

$$\overset{4}{*} \underbrace{(9 \cdot 7)}_{63} + \overset{3}{*} \underbrace{\left(\underbrace{9 \cdot 3}_{27} + \underbrace{4 \cdot 7}_{28} \right)}_{55} + \overset{2}{*} \underbrace{\left(\underbrace{9 \cdot 5}_{45} + \underbrace{4 \cdot 3}_{12} + \underbrace{6 \cdot 7}_{42} \right)}_{99} + \overset{1}{*} \underbrace{\left(\underbrace{4 \cdot 5}_{20} + \underbrace{6 \cdot 3}_{18} \right)}_{38} + \overset{0}{*} \underbrace{(6 \cdot 5)}_{30}$$

Finally, assemble the pieces by writing them in offset rows and adding:

0					3	0
1				3	8	
2			9	9		
3		5	5			
4	6	3				
sum	6	9	5	3	1	0

The left column contains the exponent, which is also the offset.

2.3.1.5 Block algorithm

Multiplication using 2-digit blocks begins by expressing integers as polynomials in 10^2 with 2-digit coefficients. For instance $638521 = 63x^2 + 85x^1 + 21x^0$, with $x = 100$. A 6×6 -digit product thus becomes a 3×3 -block product, and uses the same format as above.

Example Use 2-digit blocks to find 638521×997201 .

As above we avoid writing explicit polynomials by splitting into blocks and recording the exponent over each block:

$$\begin{array}{cccc} & 2 & 1 & 0 \\ 63 & 85 & 21 & \times 99 & 72 & 01 \\ & & & & & \end{array}$$

Next collect coefficient products for each total coefficient 0–4, and do the coefficient arithmetic with a calculator:

$$100^4 \underbrace{(63 \cdot 99)}_{6237} + 100^3 \underbrace{(63 \cdot 72 + 85 \cdot 99)}_{12951} + 100^2 \underbrace{(63 \cdot 01 + 85 \cdot 72 + 21 \cdot 99)}_{8262} + \dots$$

Note we are explicitly writing powers of 100 instead of the shorthand used in the single-digit case.

The final step is to assemble the pieces by writing them in offset rows and adding, as above.

2.3.1.6 Discussion

The traditional algorithm has been optimized for production use by experienced users, by minimizing the writing needed. Essentially any modification will be less efficient. But production arithmetic is no longer done by hand, so improved cognitive benefits may well justify some loss of efficiency. The goal of this experiment is to assess the cognitive benefits of the expanded algorithm.

In actual practice the efficiency/clarity tradeoff should mean that many fewer problems are assigned, but a success rate of 100% (after locating and correcting errors) would be expected. The presumption above is that single-digit multiplications would be done mentally, but see the next section for an alternative.

The two-digit block version would be used to lead students (subliminally) to separate the structural pattern from the blocks (i.e. not think of the algorithm as something special about digits). The result should be an effective template for multiplication of polynomials or other compound expressions in algebra.

Finally, advanced students, or group projects, can use the 4-digit block analog to multiply integers with 15 or more digits using ordinary calculators; see §3.1.1 of [16](a).

2.3.2 Subliminal learning of number facts

The goal is to have students learn single-digit products subliminally and in context rather than by explicit memorization.

The context is the task-separated algorithm described in §2.3.1.4. Students would be given a multiplication table on a card, see Figure 1, and given multiplication problems beginning with 1×1 -digits and working up to 3×3 . In multi-digit cases they would do the organizational step to reduce to single-digit products. Then they would do the batch of single-digit products, using the card for ones they do not remember. Remembering has payoffs in faster work and avoiding attention breaks. If cards and procedures are well-designed for subliminal assimilation then children would learn these fairly quickly and painlessly.

Behavioral studies can incrementally improve design of cards and procedures. The job of neuroscience is to guide improvements that educators will not reach by incremental changes. Examples illustrated in the card in Figure 1:

- Students should be instructed to read the entry out loud each time they use the card, to provide verbal reinforcement and because most people store multiplication facts in verbal memory.
- The entries on the card are complete segments to be read, not just the answer.
- Entries are designed for accurate recall. For instance $\times 7, 5; 35$ for $7 \times 5 = 35$ begins with the operation (\times) because beginning with 7 invites confusion with $7 + 5 = 12$.
- Two-digit answers should probably be read as digits rather than customary names, to avoid translation overhead (§2.2.4).
- “Equals” is omitted to shorten the entry and because it is redundant in context. Emphasis can be used as a substitute to clarify the separation between input and output, e.g. read $\times 7, 5; 35$ as “times seven, five; **three five**”.

Finally, neuroscience studies have confirmed that incorrect internalizations quickly become very difficult to correct [5] [3]. It is therefore vital that they be found and fixed as soon as possible. To accomplish this, *every assignment* be checked for correctness, and students required to locate and correct errors in their work record. (Recall that this approach would use fewer assignments than is now the custom.) *Always* having to find errors also provides consistent reinforcement for accuracy and good work habits.

2.3.2.1 Experiment

Most of the neuroscience input for this topic will be inference from other studies (e.g. put the operation first). Experiments like the one suggested here might fine-tune the ideas, but serious evaluation must wait on classroom trials.

Subjects would be children (perhaps fourth grade) who are successful with standard arithmetic. The task is to perform the organizational step of the task-separated multiplication algorithm, and use the multiplication card to carry out the multiplication step. The addition step would be omitted. There should be

2	$\times 2, 2; 4$	$\times 2, 3; 6$	$\times 2, 4; 8$	$\times 2, 5; 10$	$\times 2, 6; 12$	$\times 2, 7; 14$	$\times 2, 8; 16$	$\times 2, 9; 18$
3	$\times 3, 2; 6$	$\times 3, 3; 9$	$\times 3, 4; 12$	$\times 3, 5; 15$	$\times 3, 6; 18$	$\times 3, 7; 21$	$\times 3, 8; 24$	$\times 3, 9; 27$
4	$\times 4, 2; 8$	$\times 4, 3; 12$	$\times 4, 4; 16$	$\times 4, 5; 20$	$\times 4, 6; 24$	$\times 4, 7; 28$	$\times 4, 8; 32$	$\times 4, 9; 36$
5	$\times 5, 2; 10$	$\times 5, 3; 15$	$\times 5, 4; 20$	$\times 5, 5; 25$	$\times 5, 6; 30$	$\times 5, 7; 35$	$\times 5, 8; 40$	$\times 5, 9; 45$
6	$\times 6, 2; 12$	$\times 6, 3; 18$	$\times 6, 4; 24$	$\times 6, 5; 30$	$\times 6, 6; 36$	$\times 6, 7; 42$	$\times 6, 8; 48$	$\times 6, 9; 54$
7	$\times 7, 2; 14$	$\times 7, 3; 21$	$\times 7, 4; 28$	$\times 7, 5; 35$	$\times 7, 6; 42$	$\times 7, 7; 49$	$\times 7, 8; 56$	$\times 7, 9; 63$
8	$\times 8, 2; 16$	$\times 8, 3; 24$	$\times 8, 4; 32$	$\times 8, 5; 40$	$\times 8, 6; 48$	$\times 8, 7; 56$	$\times 8, 8; 64$	$\times 8, 9; 72$
9	$\times 9, 2; 18$	$\times 9, 3; 27$	$\times 9, 4; 36$	$\times 9, 5; 45$	$\times 9, 6; 54$	$\times 9, 7; 63$	$\times 9, 8; 72$	$\times 9, 9; 81$

Figure 2.1: Multiplication Card

enough pre-trial practice to learn the procedure but not enough to internalize the card material. Then subjects would be imaged working problems, and locating and correcting errors in incorrect problems.

The first objective is to track internalization of the table. These students will already know single-digit products in another format, but if they are instructed to use the cards (especially reading entries out loud) then they will probably internalize the new format rather than translate what they already know. Patterns in successful internalization might help refine the procedure.

The second and more important objective is to track error handling. It is well-established that internal uncertainty about correctness causes delays and unusual patterns of activity [7, 6, 4, 17]. For operational purposes we call an internalization “bad” if it is incorrect but is so firmly embedded that it does not provoke this error-related activity. It is urgent that incorrect internalizations be identified and fixed before they go bad. However, little is known about the process or the size of the window of opportunity.

- What is the repetition rate of errors during a session if error feedback is not received until the next session? How does internal error awareness change with repetition? Sessions should involve 30–40 problems for this.
- Compare this with correctness feedback and error correction after each problem.

The final question concerns durability. Durable knowledge requires practice well beyond achieving accurate performance (cognitive psychologists use the unfortunate term “overlearned” for this). It will be important to know how much reinforcement is necessary to achieve durability with this particular task. This might be addressed with followup studies, but getting reliable conclusions will be difficult: long periods of disuse will lead to serious interference from standard multiplication habits.

2.3.3 Kinetic reenforcement of geometric structure

Qualitative geometric structure is used to explore questions about functions, and to clarify the quantitative information needed for specific questions. For example the curves $y = ax^{2n}$ for a positive and n a positive integer, all have pretty much the same shape. We can see, for instance, that a straight line will intersect any of them in either two points, one point (when they are tangent), or no points.

We want to compare a purely visual approach with one that includes reenforcement. The comparison is done by cross-testing so the precise questions addressed are: how well does kinetic learning transfer to visual testing, and how well does visual learning transfer to kinetic testing. In fact actual *use* of qualitative structure requires hand drawing, so the crucial question concerns visual to kinetic transfer.

The role of neuroscience is to throw light on the mechanisms (or non-mechanisms) of transfer between domains. To what extent does learning in one mode recruit activity in regions that are used in testing the other mode? Does recruitment, or lack thereof, explain success or failure of transfer? Answering these questions requires imaging the learning activity, not just the testing.

2.3.3.1 The experiment

Subjects should be non-remedial first-year college students, as above. The study design depends on the number of subjects that can be tested.

If the number is twenty or fewer then students should be pre-tested to assess competence in the two learning modes, and assigned to the variant corresponding their strength. In other words, students from largely-visual technology programs should be in the visual track, and students from traditional programs should be in the kinetic track. There should be about the same number in each track.

If the number is significantly greater than twenty then students should still be pre-tested for reference purposes, but then assigned to tracks at random. This would allow assessment of cross-training. Do visually trained students adapt reasonably quickly to kinetic training, for instance?

Training sessions should last between 30 and 60 minutes, with at least three short quizzes to reenforce learning and familiarize students with the quiz format. It should be possible to repeat at least the first subsection if the corresponding quiz outcome is unsatisfactory. Students should be imaged during the training sessions. Students in both tracks should be able to do scratch work, and this should be recorded. See below for sample materials.

Next, students should be imaged taking quizzes, in a one or two day window at least three days after but within a week of the training session. The first quiz would be in the mode in which they were trained, to assess retention by comparison with the final quiz of the training session. The second quiz would be in the other mode, to assess transfer of learning.

Genuinely qualitative internalization should include some abstraction and

provide flexibility. The later quizzes should be slightly different from the lesson materials to probe this.

2.3.3.2 Discussion

It seems likely that there will be significant differences in learning and transfer between the two modes. Quantifying this would require much more careful controls and larger numbers, but this experiment should suggest explanatory neural mechanisms that could substantially sharpen design of followups. For example:

- When kinetic students take visual tests, to what extent is the transfer internal, or external? External transfer would use visual comparison with a scratch sketch, while internal would presumably require communication between kinetic and visual regions, probably mediated by activity in the prefrontal cortex.
- When visual students take kinetic tests (i.e. are asked to draw something), does the learning transfer, or does the output look like a reproduction of a recalled visual image? (Sketches by students trained with graphing calculators are frequently reproductions of a calculator display.) How does neural activity reflect this?

If kinetic reenforcement is important for durable qualitative learning, then a long-term goal is to find ways to incorporate kinetic reenforcement in technology-based programs. This experiment should help make a start on this.

2.3.3.3 Materials

[[draft note: replace task with multiplication of functions]] The experiment requires learning something unfamiliar but reasonably accessible. The proposal is to explore how the shape of a monomial ($y = x^n$) is modified by addition of a lower-degree polynomial. This subliminally includes the qualitative similarity of the families $y = x^n$ for n even, and for n odd.

- The visual version is illustrated (as usual) with graphics generated by computer or calculator. Quizzes are visual multiple-choice.
- The kinetic version is illustrated by videos of hand drawing. Quizzes require drawing.

The following illustrates visual lesson materials:

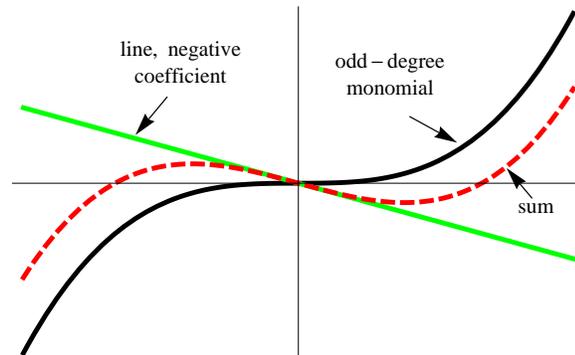


Figure 1: sum of $y = x^n$, n odd, and a line with negative coefficient.

Roughly, adding a line with negative coefficient tilts the graph a bit to the right. For very large values of x the two graphs are essentially the same.

The following illustrates a visual test item:

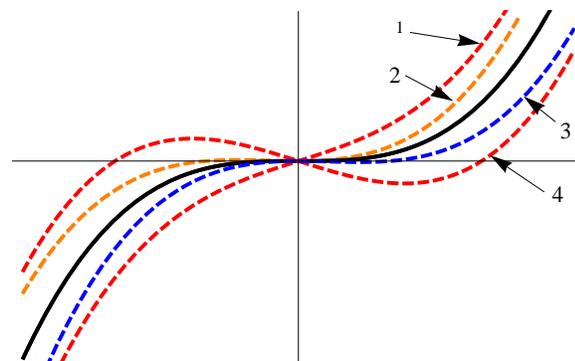


Figure 2: The solid line is the graph of a cubic monomial. Which of the functions 1–4 is the sum of this and a quadratic with negative coefficient? Which is the sum with a line with positive coefficient?

A corresponding kinetic test item would be: “sketch the graph of a cubic monomial with positive coefficient. On the same graph, sketch the sum of this and a quadratic monomial with negative coefficient.”

Bibliography

- [1] ACT-R: Theory and Architecture of Cognition, web site <http://act-r.psy.cmu.edu/>.
- [2] Anderson, J.; Qin, Yulin; Sohn, Myeong-Ho; Stenger, V. A.; Carter, C. S. *An information-processing model of the BOLD response in symbol manipulation tasks* Psychonomic Bulletin and Review, **10** (2003) 241–261.
- [3] Barton, Kevin; Fugelsang, Jonathan; Smilek, Daniel; *Inhibiting beliefs demands attention*, Thinking and Reasoning, **15** (2009) pp. 250–267.
- [4] Carter, Cameron; van Veen, Vincent; *Anterior cingulate cortex and conflict detection: An update of theory and data*, Cognitive, Affective, and Behavioral Neuroscience **7** (2007) pp. 367–379.
- [5] Dunbar, K., Fugelsang, J., and Stein, C. *Do naive theories ever go away?* In M. Lovett, and P. Shah (Eds.), *Thinking with Data: 33rd Carnegie Symposium on Cognition*, Erlbaum pub. (2007) pp. 193–205.
- [6] Fleming, S.; Weil, R.; Nagy, Z.; Dolan, R.; Rees, G.; *Relating introspective accuracy to individual differences in brain structure*, Science **329** (2010) pp. 1541–1543.
- [7] Eichele H; Juvodden H. T; Ullsperger M; Eichele T; *Mal-adaptation of event-related EEG responses preceding performance errors*, Frontiers of Human Neurosci. **4** (2010) 65.
- [8] Ioannides, Andreas A. *Dynamic functional connectivity* Current Opinion in Neurobiology **17** (2007) 161170.
- [9] Krueger F., Landgraf S., van der Meer E., Deshpande G., Hu X, *Effective connectivity of the multiplication network: A functional MRI and multivariate Granger causality mapping study*, Human Brain Mapping (Wiley), 2010.
- [10] Krueger, F.A.; Spampinato, M. V.; Pardini, M.; Pajevic, S.; Wood, J. N.; Weiss, G. H.; Landgraf, S.; Grafman, J.; *Integral calculus problem solving: An fMRI investigation*, Neuroreport **19** (2008) 1095–1099.

- [11] Liefvooghe, B.; Barrouillet, P.; Vandierendonck, A.; Camos, V. *Working Memory Costs of Task Switching* Journal of Experimental Psychology: Learning, Memory, and Cognition **34** (2008) pp. 478–494.
- [12] Monsell, Stephen *Task switching*, TRENDS in Cognitive Sciences Vol.7 No.3 March 2003
- [13] Periañez, J. A.; Maestú, F.; Barceó, F.; Fernández, A.; Amo, Carlos; Alonso, T. O. *Spatiotemporal brain dynamics during preparatory set shifting: MEG evidence* NeuroImage **21** (2004) pp. 687–695.
- [14] Pashler, Harold, *The Psychology of Attention*, MIT Press 1998.
- [15] Quinn, Frank *Towards a Science of Contemporary Mathematics*, current draft available at http://www.math.vt.edu/people/quinn/history_nature
- [16] _____ *Contributions to a Science of Mathematical Learning*, preprint at <http://www.math.vt.edu/people/quinn>
- a *Contemporary proofs for mathematics education*, to appear in the proceedings of ICMI Study 19.
 - b *Student computing in mathematics: interface design*
 - c *Mathematics Education versus Cognitive Neuroscience*
 - d *K-12 Calculator woes*, Notices of the Amer. Math. Soc. May 2009, p. 559.
 - e *Cognitive neuroscience and mathematics*
- [17] Ravizza, S. M.; Anderson, J. R.; Carter, C. S. *Errors of mathematical processing: The relationship of accuracy to neural regions associated with retrieval or representation of the problem state* Brain Research, **1238** (2008) pp. 118–126.
- [18] Richert, R.; Robb, M.; Fender, J.; Wartella, E. *Word Learning From Baby Videos*, Archives of Pediatric and Adolescent Medicine **164** (2010) p. 432.
- [19] Roseberry S; Hirsh-Pasek K; Parish-Morris J; Golinkoff RM; *Live action: can young children learn verbs from video?* Child Dev. **80** (2009) pp. 1360-1375.
- [20] Rosenberg-Lee, M., Lovett, M., Anderson, J. R. *Neural correlates of arithmetic calculation strategies* Cognitive, Affective, and Behavioral Neuroscience, **9** (2009) pp. 270–285.
- [21] Smith, S.M; Miller, K. L; Salimi-Khorshidi, G.; Webster, M; Beckmann, C. F; Nichols, T. E; Ramsey, J.D; Woolrich, M. W. *Network modelling methods for FMRI* NeuroImage **54** (2011) 875891
- [22] Sohn, Myeong-Ho; Ursu, Stefan; Anderson, John R.; V. Stenger, Andrew; Carter, Cameron S. *The role of prefrontal cortex and posterior parietal cortex in task switching*, Proc. Nat. Acad. Sci. **97** (2000) pp. 13448–13453.

- [23] Stocco, A.; Anderson, J. R. *Endogenous Control and Task Representation: An fMRI Study in Algebraic Problem Solving* *Journal of Cognitive Neuroscience* **20** (2008) pp. 1300–1314.
- [24] Willingham, Daniel, *Ask the Cognitive Scientist: Have Technology and Multitasking Rewired How Children Learn?*, *American Educator*, Summer 2010.

Chapter 3

Mathematics Education versus Cognitive Neuroscience

November 2010

3.1 Background and outline

Cognitive neuroscience is concerned with the neural mechanisms underlying human behaviour and cognition. The area has roots in medicine, psychology, sociology, and philosophy, but it was largely advances in brain imaging that led to development of a distinct discipline in the 1990s.

Mathematics education was an early theme in cognitive neuroscience. Elementary mathematical activity is more well-defined and consistently localized than most cognitive activities, and in the late 1990s Stanislas Dehaene [14, 15] exploited this in a pioneering exploration of innate number sense. Applications to education seemed a natural and valuable next step. At the same time, however, Bruer [9] pointed out that education and neuroscience are concerned with phenomena on vastly different scales, and trying to make a direct connection would be a “bridge too far”. Instead, he recommended a two-stage approach with neuroscience providing input to cognitive psychology, and psychology guiding applications to education. This seemed sensible, and the idea prevailed for almost a decade.

By 2005 there were calls for direct education-neuroscience interactions as a “two-way street” [6], [24], and the term “educational neuroscience” (with “cognitive” removed) began to be used. For later accounts see [37], [11]. The reason offered was that Bruer’s two-stage approach was not working: educators were using distorted popular accounts rather than solid science [4, 25, 21], and psychologist seemed to be ineffective as intermediaries. The ‘two-way-street’

approach is still the main theme. For instance, a major conference intended to plot a course for neuroscience and mathematics education was held in 2009; see the program [18] and position paper [17].

Unfortunately ‘brain-based’, ‘brain-friendly’, etc. educational methods are multiplying, and still based on ‘neuro-myths’ [3]. Most neuroscience is inconclusive or irrelevant. Solid neuroscience findings that conflict with educational dogma are being ignored. Nothing is working. The goal here is to analyze individual articles and specific issues to try to understand why not. Is it just ineptitude, or is there a deeper incompatibility as the ‘versus’ in the title suggests? And whatever the problem might be, is there a way around it?

3.1.1 Outline

Section 3.2 (The macro/micro spectrum) describes several levels of generality and abstraction between brain and educational theory. These flesh out Bruer’s scale mismatch observation, and have turned out to be a good way to organize not only this inquiry but the whole volume of essays (see the Preface).

Section 3.3 (Ineffective cognitive psychology) explores reasons why Bruer’s suggestion was unsuccessful. In brief, recommendations of an external agency are too easily ignored or misinterpreted, and being external means it lacks contact with important parts of the process.

Section 3.4 (Ineffective neuroscience) illustrates how neuroscience studies are often rendered ineffective by doubtful educational assumptions and lack of subject sophistication.

Section 3.5 (Dangerous neuroscience) illustrates how sophisticated neuroscience can go dangerously astray when educational philosophy and goals are accepted uncritically.

Section 3.7 (Mathematics and Learning) takes a different approach. The sections so far describe problems and some of the missing expertise that might have prevented them. This section illustrates how incorporating this expertise could lead to high-impact outcomes, again through examples. but most of the details are in other essays [?, ?]. The discussion touches on additional problems.

Section 3.8 (Appendix: Technical difficulty, and consequences) provides a brief overview of the technical challenges of neuroscience. Everything is difficult, outcomes are complex, and useful outcomes require insightful guidance on what to look for. External collaborators with appropriate expertise will be hard to find.

3.1.2 Note on style

In this article the passive third-person voice traditionally used in science is reserved for scientific material. I use the first person for personal interpretation or observation because I believe it is important to keep the distinction clear.

3.2 The macro/micro spectrum

Education is concerned with phenomena on a vast range of scales, from brain regions used by an individual student in a single activity, up through general features of teaching, learning and behavior. Figure 1 provides a summary.

educational theory	Commonalities of teaching, learning, and behavior that transcend content.
subject matter	Content, and its influence on teaching and learning.
course/curriculum	Learning goals and methodologies at the course and curriculum level.
cognition/activity	Neural implementation of skills and understanding, and strategies for developing these.

Figure 3.1: Educational levels from macro to micro

Neuroscience applies to the lowest level, while educational theory is concerned primarily with the top. Bruer's point in 1997 [9] was that the level gap alone is likely to prevent effective direct interaction. There have been many attempts since that time and Bruer was right about them being ineffective. In analysis of examples I identify levels in Figure 1 where disconnects occur. Putting these together in §?? suggests that Bruer was right about a level disconnect, but wrong about the nature and consequences of it.

3.3 Ineffective cognitive psychology

There are now quite a few qualitative conclusions from neuroscience and micro-scale cognitive psychology that have profound implications for education. Later we ask why educators have not picked up on these. The question here is: why have cognitive psychologists been unsuccessful in bringing them to the attention of educators? This was Bruer's suggestion [9] for bridging the scale gap; why didn't it work?

3.3.1 Specific questions

Two specific findings about learning are used to probe this issue:

- Durable learning requires practice over an extended period of time, and well beyond the point of accurate recall or performance.
- It is very difficult to correct errors in durable learning.

These are stated in behavioral terms because they are not hard to see in micro-scale behavior, and in fact were more-or-less well known in cognitive psychology well before the neuroscience era.

Neuroscience revealed mechanisms behind these observations: learning is in part implemented through physical changes in the brain. Interconnections are strengthened, developed, or degraded, but these changes are temporary unless extensively reinforced. And if durable learning is found to be erroneous then correcting it requires dismantling a physical structure.

[[incomplete]]

3.4 Ineffective neuroscience

The need for—and current lack of—insightful and appropriate guidance is illustrated by analysis of recent neuroscience studies. Section 3.4.1 describes studies of multiplication; 3.4.2 concerns solving simple equations; ?? discusses the understanding of errors; and 3.4.4 concerns word problems.

3.4.1 Multiplication

In this section we review two fMRI studies of integer multiplication, [31] [48]. These are relatively clear and are the sort of study one could imagine educators trying to use in some way. However both omit subtle but crucial issues, particularly concerning errors, and one cannot imagine educators compensating for this.

3.4.1.1 Krueger, Landgraf, et al.

The study reported in [31] finds activity in five main areas. Lacking “theories which specify cognitive processes that are detailed enough to be examined by neuroimaging” (quote from [17], above), they simply note which areas are active in each of a sequence of time blocks, and from the coincidences infer correlated activity.

The first concern is that the study uses three tasks identified as having increasing difficulty: multiplication of two 1-digit integers; one 1-digit and one 2-digit integer; and two 2-digit integers. In fact these tasks have qualitatively different cognitive and mathematical structures:

- multiplication of two 1-digit integers is simple fact recall and input-output.
- multiplication of a 1-digit and a 2-digit number requires two multiplication facts, and short-term storage and addition of outcomes. The addition requires shifting one output by one place, usually a single 1-digit addition, and occasionally dealing with an overflow.
- multiplication of two 2-digit numbers engages the standard algorithm. Here it is used to organize and combine two 1×2 -digit products, each generated as above.

Difficulty increases because successive tasks have qualitatively different sub-tasks, not because they are more of the same. Moreover the algorithm used

for the 2×2 -digit multiplication is more representative of later mathematics than is single-digit product memorization. An important question potentially addressed by this study is: how does 2×2 -digit multiplication *differ* from the simpler cases?

More generally, there is an urgent need to understand how components of mathematical algorithms interact cognitively. It seems likely that some could be redesigned to reduce cognitive difficulty, and this could have profound educational consequences. Substantial exploration of the multiplication algorithm would require products with three or more digits in each factor, but this would be pointless without sophisticated mathematical input (see §2.1).

Another concern is that [31] was envisioned as a test of mental arithmetic, so participants were unable to do the scratch work usually employed in multiplication. In particular in the 2×2 digit task, the output from the first 2×1 multiplication had to be held in internal working memory rather than written. This may have introduced an artifact: instead of being written it may have been held in internal memory *associated with writing*, and this may partially account for the observed recruitment of the left precentral gyrus (subjects were right-handed). Their interpretation of this as being connected with use of fingers in counting is doubtful, and could be misused if wrong.

A general context for this concern is that most mathematical procedures use written intermediate results as external working memory. External memory has a tradeoff: it is more accurate and durable, but requires input-output processing and shifts in attention focus. A key part of algorithm design is to optimize written components for this use (and other things; see §??). Neuroscience studies could certainly contribute to this, but the main point is that excluding scratch work renders studies of all but the very simplest tasks useless.

A third concern is more neurological than mathematical. There seem to be semi-autonomous facilities, for instance in the anterior cingulate cortex, that check for conflicts and inconsistencies (see §??). This means that generating an outcome, and recognizing whether or not a proposed outcome is correct, can be substantially different neural activities. A lack of coordination between these activities seems to underlie some learning difficulties [20]. However many mathematical neuroscience experiments (including this one) tacitly assume that correct alternatives are identified by generating an outcome and comparing. For that matter, this assumption is used to justify wide use of multiple-choice tests in mathematics education. This assumption is problematic and urgently needs to be tested.

3.4.1.2 Rosenberg-Lee, Lovett, and Anderson

[48] describes an fMRI study of 3×1 - and 5×1 -digit multiplication, comparing two different strategies to predictions of ACT-R computer models (see [2], [5], and the Wikipedia entry). Differences from the study above:

- The models give predictions that enable detailed imaging (Granger analysis not needed), and the imaging essentially supports the model.

- Subjects gave answers rather than identified them among choices.
- The use of a single-digit factor avoids the cognitive complexity of the full multiplication algorithm, so this should have been described as a study of a *component* of multiplication, not the full activity. This is a concern about possible misinterpretation, not the science.
- The single-digit factor restriction, and comparison of two strategies, provides a clearer picture of this component of the algorithm.

The main shortcoming of [48] is that only correct responses were analyzed. There is no information on how errors occur or how to avoid them, and the design prevented error-checking, see §??.

3.4.2 Solving equations

Most studies of equation-solving have been neurological explorations without significant educational goals. Lack of a coherent overall context for this activity¹ makes large-scale goal formation difficult, but small-scale goals were available.

[5] and [52] seem primarily designed to show that ACT-R programs [2] can effectively model the activities. This suggests that ACT-R is ready to be challenged by questions with important consequences, but that not much will happen without such challenges. In detail, [52] compares solving equations with symbolic and numerical coefficients, and finds relatively little difference. This supports the idea that a lot of arithmetic is more symbolic than numerical, but the task was too simple to test this effectively. [5] compares solving numerical equations to a symbol-manipulation task that is less relevant than one might have liked. In both of these studies errors would probably have been more revealing than correct work, but the studies were limited to correct solutions because, so far, the models are.

A study of elementary calculus [32] showed activation patterns similar to algebra. The routines used are essentially algebraic so mathematically these are just more-complicated algebra problems, and apparently the brain sees them the same way. They are significantly more complicated than the tasks used in other trials, but more subject sophistication seems to be needed to draw useful conclusions.

3.4.3 Errors

[[significance]] [45] studies errors in solving simple numerical equations. However the conclusions are weak in a number of ways:

- Coefficient arithmetic has a significant error rate. How much of the observed error was due to arithmetic, how much to the solving component, and how much (if any) to interference between these tasks?

¹Equation-solving should probably be seen as symbolic pattern recognition and manipulation.

- The task was mental (i.e. did not allow scratch work). See the next section for an explanation why this is problematic.
- It would have been useful to know if there were patterns in the errors that got noticed and triggered correction. However the protocol inhibited corrections.

There have been many studies of conflict alerts and the patterns so far are summarized in [13]. Most of these studies concern perceptual tasks irrelevant to mathematics. They are useful guides

Internal conflict alerts originate in the region where the questionable cognition takes place. For instance the fMRI study [45] shows a correlation between errors in solving simple numerical equations, and reduced preliminary activity in a region in the prefrontal cortex associated with procedural fact retrieval. Presumably insufficient activity increases the likelihood of some sort of loading error. However their data also shows significantly *increased* activity during and after an erroneous outcome, and ‘Error-Related Negativity’ (ERN) is seen at this time in analogous EEG studies. Confusion or conflict due to the loading error seems to cause formulation of a problem report and the ERN results from dispatch of this report to the Dorsal Anterior Cingulate Cortex (DACC). The DACC apparently determines what sort of conflict might result from the problem, identifies the relevant control region in the Dorsolateral Prefrontal Cortex (DLPFC), and forwards an amplified report to that region for consideration. It seems to be the DLPFC that actually issues the alert, determines the level of awareness, and perhaps provides preliminary plans for response.

[22] focuses on internal conflict alerts. Consistently effective alerts, referred to as “introspective awareness”, was found to correlate with gray matter volume in a region in the prefrontal cortex. Presumably this is the control region relevant to the task in question. In any case it seems to be the bottleneck and (as observed in [22]) effective error correction may require enough training that neural plasticity leads to an increase in volume. A similar finding for mathematical error correction would have profound educational significance because error correction is largely absent from current curricula.

3.4.4 Word problems

Contemporary mathematical practice reflects this by splitting real-world applications into two parts: modeling (essentially a translation into symbolic form suitable for mathematical analysis); and mathematical analysis of the model.

Conventional wisdom in elementary education rejects the separation of modeling and analysis. As suggested in (4), word problems are thought of as a different *format* rather than a different activity. Neuroscience investigations that accept this, e.g. [19], [51], are ineffective. They misinterpret neural evidence that these really are different activities, and find behavioral equivalence because they followed the educational practice of restricting to problems with trivial mathematical core. A critical comparison with more complex problems should give a very different picture; see §2.2.2 for a proposal.

[35] provides an extreme example. This study compares two strategies for young children working extremely simple word problems that require a size comparison. One uses a symbolic translation to connect to innate number sense, the other uses pictures to enable visual comparison. The strategies do not apply to other problem types, and the use of innate abilities is a dead end.

3.4.5 Summary

Neuroscience experiments genuinely effective for education will require appropriate educational and psychological expertise and, crucially, cognitively-informed mathematical sophistication. Lack of these has prevented cognitive neuroscience from having much impact.

3.5 Dangerous neuroscience

Incautious interpretation of neuroscience findings can be counterproductive, not just ineffective. This is illustrated with material from a 2008 profile of Stanislas Dehaene, by Jim Holt in the *New Yorker* magazine [29].

First, a general principle:

“The fundamental problem with learning mathematics is that while the number sense may be genetic, exact calculation requires cultural tools—symbols and algorithms—that [...] must be absorbed by areas of the brain that evolved for other purposes. The process is made easier when what we are learning harmonizes with built-in circuitry. If we can’t change the architecture of our brains, we can at least adapt our teaching methods to the constraints it imposes.”

This is certainly true as stated, but most educators would interpret it as “adapt our teaching *goals* to the constraints imposed by brain architecture”. They would want to build on innate number senses and connection to spacial sense, as mapped out by Dehaene and others, and avoid some of the obviously unnatural material now taught. These inclinations are shared by some “mathematics educational neuroscientists” [12].

A content-aware interpretation would be: understand skills needed in the long term; find out how they are implemented in brains of successful users; and design teaching methods to develop this implementation as quickly and painlessly as possible.

The educational and content-aware interpretations differ: there are unnatural things that really are needed for long-term success. Returning to [29]:

Our inbuilt ineptness when it comes to more complex mathematical processes has led Dehaene to question why we insist on drilling procedures like long division into our children at all. There is, after all, an alternative: the electronic calculator. ‘Give a calculator to

a five-year-old, and you will teach him how to make friends with numbers instead of despising them,' he has written. By removing the need to spend hundreds of hours memorizing boring procedures, he says, calculators can free children to concentrate on the meaning of these procedures . . .

Boring memorization is indeed a problem, and this solution is already widely used in the US.

In the long term, however, this is the wrong solution. The reasons are described next, and some are expanded in the experimental section.

First, calculators make numbers friendly in a very superficial way: they all seem the same and have no valuable structure. A basic feature of the place-value notation, for instance, is that multiplication or division by ten can be accomplished by moving the decimal point. Hand-arithmetic students know this because it is a big time-saver when it can be used, and it is an integral part of multi-digit multiplication and division algorithms. Calculator students do not, because it is useless: on a calculator moving the decimal point would be accomplished by multiplying or dividing by ten. Similarly when a calculation calls for both multiplying and dividing by the same number, hand-arithmetic students will cancel them to avoid both operations. Calculator students almost never do this, partly because they do not have written intermediates that they could scan for opportunities, and partly because it has so little payoff that such scanning is not worthwhile. The result is that calculator students frequently have much weaker number sense.

The second problem is more subtle. For long-term purposes, internalizing the algebraic structure of numbers in a way that extends to symbols is more important than fast or perfect numerical multiplication. In the past, much of this internalization seems to have been a subliminal consequence of the fact that much of the manipulation in the boring and unnatural algorithms is essentially symbolic, see §2.3. Symbols may even seem friendlier than numbers because one doesn't have to recall numerical multiplication facts. Calculator students never see these quasi-symbolic manipulations, and so do not get this subliminal exposure. Further, encoding operations as keystrokes seems to make them inaccessible to abstraction, and because symbols cannot be manipulated the same way, symbols seem completely different from numbers. The result is that calculator students frequently have weaker symbolic skills.

The educational reasoning is:

1. Mathematics is an abstraction of systematic structure in the physical world;
2. We have a great deal of intuitive understanding of the physical world, either innate or learned;
3. Therefore, our intuitive understanding can and should serve as a basis for developing mathematical knowledge and skills;

However this seems to be wrong in many ways. We return to the philosophical assertion (1) later. (2) is true: it is now well-known to cognitive psychologists that we do have an innate version of physics, but it is non-Newtonian and must be *overcome* to learn the real thing [20], [7]. In other words (3) is wrong for physics. It should not be a big surprise that it is wrong for mathematics too.

We do have some innate number sense but, as discussed above, it is quite insufficient for mathematics and again must be avoided or overcome. We have some sense of space and geometry, and in antiquity it was found that with a little prodding this could be used to do Euclidean plane geometry. This is still about all we can do with it two thousand years later, and it is not a good foundation for the real thing. In short, conclusion (3) is wrong. Once scientific or mathematical understandings are established, then intuitive ideas can be retrained and recruited to enrich this understanding, but trying to do this too soon inhibits rather than advances reliable learning. Belief (3) was abandoned in professional practice about a century ago, see §3.7.2 and [5].

Whether mathematics is “really” an abstraction of the physical world, as in (1) above, is a pointless philosophical debate. Cognitively this actually seems to be *backwards*: It seems to be more effective to think of the physical world as, roughly speaking, a murky implementation of a fragment of mathematics.

3.5.0.1 Summary

Dehaene offers “broad brush messages” of the sort Goswami [25] observes are needed by educators. They lack the nuance and precision of the original science, but surely Dehaene should be qualified to formulate broad-brush messages that are true to the science. They are also quite compatible with dominant conventional wisdom in education, so they are instances of “successful” collaboration between neuroscience and education. They are nonetheless counterproductive in the long term because they are insensitive to the needs of mathematics.

It is puzzling that Dehaene supported a version of mathematics on the basis that it is easy to learn, without considering whether it is effective. He surely knows that our cognitive structures are more attuned to Aristotelian physics than Newtonian, but it is doubtful that he would promote the former on that basis. For that matter, the old philosophical approach to cognition is much easier than the science he does, but he would not accept this as a justification. There seems to be something about mathematics that invites confusion.

It is important, however, to recognize that the difficulties Dehaene identifies are real. The easy, neuroscience-endorsed calculator alternative may be counterproductive, but memorization is still boring and the antique approach still leaves a lot to be desired. Is there a way to use calculators that does not undercut subliminal learning of algebraic structure? Is there a way to internalize multiplication facts without explicit memorization?

3.6 Ineffective education

[disconnected, due to incomplete editorial reorganization]

3.6.1 Lack of scientific skills

The core problem is that education is not a science. This is not for lack of trying: systematic efforts to develop ‘education science’ go back more than a century, and progress of a sort has been made. Research went from a top-down view dominated by social theory and forceful personalities, thorough behavioral models of children essentially as small animals, and later tried to come to grips with the complexity of actual classroom practice and human learning. See [33] for a detailed recounting of the general story and [30] for mathematics in the US. Unfortunately the obstacles are immense and the complexity overwhelming:

- Children are surely the most complicated and difficult experimental subjects possible;
- There are strong constraints on how the subjects can be controlled or manipulated, and on how much control investigators have over actual practice;
- There are many important variables that cannot be measured, let alone controlled;
- The situation is so complex and the theories so vague that scientific precision in terminology is impossible; and
- theories and ‘facts’ are not precise or strong enough to support logical analysis or deductive reasoning.

Recent education research in the US has attempted to be more scientific by use of large-scale trials, statistical analysis, etc. more-or-less modeled on medical studies. The resemblance is rather superficial, however, and many of the practices would be considered unethical in medicine. For instance:

- The double-blind protocol found to be necessary for reliable medical conclusions is impossible in education. But rather than accept the medical conclusion that findings will be biased and unreliable, educators conclude that it somehow doesn’t matter.
- Small studies often have sample sizes two orders of magnitude too small to justify the statistical analysis.
- Large-scale trials often have little more than the name in common with the pilot studies, but outcomes are still interpreted as validations of the pilot design.
- Significant variables such as per-student cost and teacher expertise are known but not reported, and others such as student disruption are sometimes controlled without mention.

- Complexity is controlled by severely restricting—in advance—possible interpretations of the data.

One can sympathize with the difficulty and admire the effort, but the consequence is still that educational researchers do not develop scientific sensibilities or discipline.

Educational research outside the US tends to be more qualitative and thoughtful, but not more scientific. Influences include dubious and conflicting psychological theories, simplified abstract models of “the student”, and deeply-held convictions derived from classroom practice, particularly at elementary levels. A complicating factor is that humans have powerful instinctive responses to children. Some educational theory seems as much informed by the emotional response of teachers as by dispassionate facts about children. In any case the effort is still not data-driven or logic-constrained, and educators do not develop the facilities to deal with data-driven and logic-constrained scientific material.

Some of the more explicitly philosophical approaches to education verge on the hilarious: see [42] for use of Wittgenstein to defend education against a caricature of neuroscience.

There is an extensive literature on educational abuse of cognitive neuroscience. In 2006 Goswami [25] observed the inability to make use of complex or nuanced information, but expressed hope that broad-brush or big-picture messages might lead to better outcomes. The 2010 article of Alferinka and Farmer-Dougana [3] describes more extensive misuse, with less optimism. So far neuroscience does not *have* broad-brush or big-picture formulations that avoid the need for scientific precision (more about this below). Again, this means the mainstream educational community cannot be expected to use it carefully or correctly.

3.7 Mathematics and learning

The author feels, particularly after developing the examples in later sections, that a cognitively-oriented understanding of learning difficulties of real students is the primary qualification needed for a genuinely productive neuroscience collaboration. It would also be useful to understand modern cognitive strategies that enable humans to do mathematics. These are explained in this section.

3.7.1 Teaching vs. diagnosis

The mainstream educational community, and teachers at all levels, are more focused on teaching than learning, or in other words more on information delivery than diagnosis of problems with receiving the information. There is a good reason for this: teaching is essentially concerned with learning in *groups*, and resource constraints forbid much individualized attention².

²See [44](c) for further discussion.

By *diagnosis* we mean one-on-one sessions, usually initiated by the student, and intended to isolate and fix a specific difficulty. The goal is to provide a brief, targeted intervention that will enable the student to resume working on his own. To accomplish this the helper should listen more than talk, and not jump to conclusions about the difficulty. Expert teachers find both of these difficult. Diagnostic work provides a complex, fine-grained and individual view of learning. This view is rather different from the teacher's perspective and from standard educational theory, and seems considerably more relevant to cognitive issues and neuroscience.

Diagnosis in this sense is (in a nutshell) the procedure used by the help staff in the Math Emporium at Virginia Tech, a facility providing computer-based and computer-tested mathematics courses to around 10,000 students per semester, and now in its thirteenth year; see <http://www.emporium.vt.edu>. The author has spent over 1,000 hours in diagnostic work with students in the Math Emporium, and found it far more revealing than thirty years of classroom teaching. This is the primary experience drawn upon in developing the proposals later in the article.

3.7.2 Modern mathematics

Professional practice changed profoundly in the early twentieth century. It became better adapted to mathematics and consequently more powerful, but other aspects of the change are more important here.

Modern mathematical methods are more systematic, deliberate, and precise, and less dependent on intuition and heuristics. A curious consequence is that strategies for human use have developed: systematic methods admit strategies; intuition either works or it doesn't.

It is significant that mathematical human-use strategies evolved without conscious direction or understanding. Up through the nineteenth century mathematics was quite influenced by philosophy, but the early twentieth-century transition included a break with philosophy. Because the strategies evolved without interference they could adapt, in ways we do not understand, to human abilities and limitations that we also do not understand. Cognitively-oriented study of these strategies can therefore reveal quite a bit about human cognition [5].

At present these cognitive strategies are used by only a few tens of thousands of professionals, but since they address general cognitive issues they should, in principle, be a rich resource for new educational practices. They certainly could be a rich resource for educational neuroscience. The discussion in §3.4.1, ?? and examples later in this article should illustrate this point. However explicit understanding of these strategies is extremely rare, and in the end less vital than the diagnostic understanding of learning problems discussed above.

3.7.3 Educational hostility

Mathematics education remains modeled on obsolete practices of the nineteenth century and before; see [5] for a detailed discussion. This is no doubt one reason

it has had trouble improving on nineteenth-century outcomes. It is also part of the reason so few students make the transition from school mathematics to the twentieth century.

Attempts to introduce a bit of modern mathematics into education, for instance ‘new math’ in the US, have been failures. The big problems were not mathematical: the large-scale dissemination of ‘new math’ was so poorly designed and executed it might have failed even if the goal had been to give away candy. However the education community saw the failure as proof that modern mathematics is suitable only for freaks. Mainstream educators remain deeply hostile to modern methodologies. The hostility includes human-use strategies: one of the most powerfully effective is the concise self-contained definition, but this is universally rejected by educators.

A consequence of this hostility is that mathematicians who become involved in pre-college education are required to “check their weapons at the door”: buy into the idea that nineteenth-century methods are somehow kinder, gentler, and more appropriate. This is easier than one might think, because the features that make modern mathematics powerful are internalized, not articulated. In any case the result is that mathematicians involved in mainstream education have some of the same drawbacks as educators, regardless of their mathematical credentials. See [8], [56] for examples.

3.7.4 Theoretical incoherence

Theories of mathematics and mathematical practice are as incoherent and inconsistent as educational theories. One reason is that the philosophers and historians who might develop explicit theories are still concerned with pre-modern practice, and their incoherence reflects incoherence in actual practice that made change necessary. Ironically, modern practice is inaccessible to philosophical investigation *because* it is more effective: it enabled a rapid increase in technical difficulty that made it opaque to outsiders.

Mathematicians writing about mathematics are as incoherent as philosophers. They have internalized the methodology so well that it has become transparent, and it seems to be a general principle that people cannot figure out how they do things that they do well. Somewhat like birds and flying.

The point for the present discussion is that current descriptions of mathematics, no matter what the source, are not good resources for neuroscience³.

3.7.5 Summary

It seems that the qualification most important for neuroscience collaboration is a diagnostically-based and mathematically sophisticated understanding of cognitive learning problems of real students. This is rare. An understanding of human-use strategies of modern mathematics could also be valuable, but is even more rare.

³The author hopes [5] will be an exception.

The main point is that the relevant expertise will not be found in educators, in mathematicians involved in mainstream pre-college education, in the philosophy or history of mathematics, nor even in the writings of mathematicians themselves. A consequence seems to be that ideas must be evaluated directly on their own merits, rather than on the credentials of the proposers.

3.8 Appendix: Technical difficulty, and consequences

The section begins with a brief review of the technical difficulties of brain imaging. We see that these difficulties impose strong constraints on how the work is conducted and on the feedback needed to make progress. It also makes good use of the outcomes a challenge.

3.8.1 Technical difficulties

The first problem is that the brain is encased in the densest bone in the body. In principle, sensors could be implanted inside the skull, but this is invasive, expensive, and current sensors are unsuitable for all but the most urgent human applications. Education-oriented imaging must be done from outside. As a result all techniques must deal with signal attenuation and distortion by the skull, and the inversion problem (reconstruction of internal activity from external data) requires difficult blends of mathematics, physics, and anatomical knowledge. No method has completely satisfactory inversion: see [57] for an illustration of how better anatomical knowledge could improve interpretation of fMRI data, for example. Inversion methods for MEG are particularly primitive.

Next, brains are busy places, and signals relevant to the question at hand must be extracted from the general hustle and bustle. The faintness of individual signals is suggested by the fact that all this busy activity uses, on average, less than 20 watts of power. A dim bulb, so to speak.

A great deal remains inaccessible. None of these methods give information about neurotransmitter activity, for instance. Neurotransmitters are profoundly important and individual differences effect both cognition and imaging, but there is currently no way to anticipate or compensate for such effects and they must be treated as noise. This, no doubt, is one reason statistical aggregation is necessary for useful outcomes.

Different methods have their own specific difficulties:

- Positron emission tomography (PET) gives good images but requires ingestion of radioactive substances. Total exposure is low compared to X-ray tomography, but the radiation is higher-energy and the target is an organ that we don't want to damage. It is also expensive.
- The usual explanation of PET is that it tracks glucose uptake associated with increased neural activity. This is not literally true and exactly what

it does track, and whether it really correlates with glucose delivery, is a matter of debate. For practical purposes it may be more important to understand how it correlates with other imaging data.

- Functional magnetic resonance imaging (fMRI) requires high magnetic fields. The machines are large, noisy, and expensive to operate. The data should be similar to PET data, and nicely complementary to EEG or MEG. Unfortunately the high field strength makes it difficult or impossible to use these methods at the same time, so data from a single trial usually cannot be correlated. There has been some progress with simultaneous EEG and high-field fMRI [28]. Low-field fMRI that might address some of these problems is in development [10], but has a long way to go.
- fMRI primarily images the so-called Blood Oxygen Level Dependent (BOLD) response. The presumption is that deoxygenated hemoglobin indicates energy consumption, therefore neural activity, and therefore cognitive use of the area. However most activity is anaerobic and aerobic processes recharge local energy stores rather than directly power the activity. One consequence is that the BOLD response is rather sluggish and the link to activity is indirect. Further, direct comparison with implanted electrodes in monkey brains [49] indicates that this signal can be triggered by *anticipated* use as well as the physiological stress that follows use. This means standard assumptions about the link are not completely reliable.
- Diffusion Spectrum Imaging (DSI) is a variation on MRI that detects the Brownian movement of water molecules [26, 27]. DIS can be used to image fiber structure and connectivity in white matter because water is constrained to move within fibers. This is very powerful information but the technical requirements are high: very high fields, high gradients, long acquisition times, and complex mathematical analysis. There are less-demanding and less-informative variants (diffusion tensor imaging, etc., [26]).
- Near infrared spectroscopy (NIRS) exploits the fact that tissues containing relatively little blood—in particular the skull—are translucent in the near infrared (650–950 nm). This window includes absorption lines for oxygenated and deoxygenated hemoglobin, and these can be used to detect energy-intensive activity in the outermost few mm of the brain [39], [1]. The thick diffusive layer (skull) and multitude of artifacts (cardiovascular and respiratory activity, head movement, scalp effects, etc.) limit imaging to very coarse resolution. On the other hand the equipment is modest and unobtrusive compared to other methods, so NIRS is attractive for large-scale studies or individual assessment when high-resolution pilot studies have shown what to look for [19].
- Magnetoencephalography (MEG) uses magnetic fields resulting from current loops in the brain. These fields are weak and detection requires elab-

orate shielding and liquid-helium-cooled sensors a few millimeters from the scalp.

- Inversion techniques for MEG are still primitive, and it is a stretch to call the output “images”. Fields detectible outside the skull must be generated by relatively large currents, but the anatomical structures that support these currents are not fully understood and cannot be inferred from the field data. DSI, which does image these structures, might eventually give a ‘wiring diagram’ that would enable better inversion.
- There is a lot of individual variation not related to the questions at hand. People with tinnitus, for example, have significantly different MEG profiles [53]. Again, at present these differences are treated as noise and degrade the signal.
- Electroencephalography (EEG) makes use of electric fields. These are easier to detect than magnetic fields but the data is more complex: the fields are distorted by the skull and scalp; fluctuating reference levels; and artifacts from pulse and cardiovascular electrical activity. Eye-blinks also cause strong artifacts because there is a significant potential difference between the front and back of the eye. As with MEG, genuine images cannot yet be extracted.
- Both MEG and EEG have response fast enough to enable real-time tracking of neural activity, while the slower image rate of PET and fMRI give time-averaged results. Fast response offers opportunities but also additional challenges. Making sense of data that involves vision, for instance, requires tracking and compensating for eye movement. Eyes move a lot, and the brain pre-processes optic nerve signals to produce stable perceptions. This compensates for limitations of light receptors and gives better visual perception, but the pre-processing produces complex and mostly irrelevant signals.
- Tracking eye movement can give information on attention focus, and pupil dilation sometimes correlates with cognitive resource allocation [34].

3.8.2 Needs of neuroscience

Neuroscience studies are expensive, with costs orders of magnitude higher than traditional educational studies of comparable scope. To attract funding, and to avoid wasting it, education-oriented experiments must be designed so there is a good chance there will be a signal above the noise level, and that this signal will be meaningful. Moreover the technical difficulties mean that experimental design must be quite insightful and targeted to have a good chance of success.

As De Smedt et.al. [17] put it (p. 102):

“few attempts have been made to study more complex and higher order mathematical skills. . . . A particular challenge of this research

is that it requires educational and psychological theories, which specify cognitive processes that are detailed enough to be examined by neuroimaging.”

More precisely, insights into cognitive processes are needed from *somewhere*. Recall that these are *mathematical* processes, so it seems likely that mathematical sophistication will play a key role. It is now very doubtful that educational theories will be helpful in any way. Psychological theories are either off-base, or too tentative and unfocused to be much help.

3.9 Fragments

Contemporary educational theory, by contrast, follows the more top-down model of classical philosophy. The theory is “informed” by low-level experience, but also by social and political convictions, current psychological theories, etc., so the connection is tenuous.

In these terms Bruer’s question is: how can new micro-scale information be incorporated in a top-down system? His suggestion was to rely on an external agency (cognitive psychology) that does have a bottom-up structure. It appears that this won’t work. External recommendations are easily ignored, particularly if the conclusions are unwelcome (§3.3), and lack of contact with intermediate levels in education means important issues will be neglected. In particular, anything specific to the subject matter will be overlooked.

3.9.1 Unmet needs

As explained in §3.8, productive neuroscience research requires targeted and insightful guidance about what to look for, and what it means. Bruer hoped that cognitive psychologists could extract specific questions from the educational maelstrom, but psychologists did not rise to this challenge. There have been efforts to translate neuroscience findings into education, and to debunk misuses, but essentially no feedback to the neuroscience community.

Bibliography

- [1] Abdelnour, A. Farras; and Huppert, Theodore, *Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model*, Neuroimage **46** (2009) pp. 133143.
- [2] ACT-R: Theory and Architecture of Cognition, web site <http://act-r.psy.cmu.edu/>.
- [3] Alferinka, Larry A.; Farmer-Dougana, Valeri *Brain-(not) Based Education: Dangers of Misunderstanding and Misapplication of Neuroscience Research*, Exceptionality, 18: 1, 42–52; <http://dx.doi.org/10.1080/09362830903462573>.
- [4] Anderson, J. R.; Reder, L.M.; Simon, H.A.; *Applications and Misapplications of Cognitive Psychology to Mathematics Education*, Texas Educational Review (Summer 2000), <http://act-r.psy.cmu.edu/papers/misapplied.html>.
- [5] Anderson, J.; Qin, Yulin; Sohn, Myeong-Ho; Stenger, V. A.; Carter, C. S. *An information-processing model of the BOLD response in symbol manipulation tasks* Psychonomic Bulletin and Review, **10** (2003) 241–261.
- [6] Ansari, Daniel and Coch, Donna *Bridges over troubled waters: education and cognitive neuroscience*, TRENDS in Cognitive Sciences **10** (2006) pp. 146–151.
- [7] Barton, Kevin; Fugelsang, Jonathan; Smilek, Daniel; *Inhibiting beliefs demands attention*, Thinking and Reasoning, **15** (2009) pp. 250–267.
- [8] Bass, H. *Mathematics, Mathematicians, and Mathematics Education*, Bull. Amer. Math. Soc. **42** (2005) pp. 417–430.
- [9] Bruer, J. T. *Education and the brain: A bridge too far*, Educational Researcher **26** (1997) 4–16.
- [10] Burghoff, Martin, et. al. *On the feasibility of neurocurrent imaging by low-field nuclear magnetic resonance*, Applied Physics Letters **96**, 233701 (2010).

- [11] Campbell, Stephen R. *Defining mathematics educational neuroscience* preprint 2010?
- [12] _____, *Embodied Minds and Dancing Brains: New Opportunities for Research in Mathematics Education* in B. Sriraman, L. English (eds.), *Theories of Mathematics Education*, Advances in Mathematics Education, Springer (2010) pp. 309–331.
- [13] Carter, Cameron; van Veen, Vincent; *Anterior cingulate cortex and conflict detection: An update of theory and data*, Cognitive, Affective, and Behavioral Neuroscience **7** (2007) pp. 367–379.
- [14] Dehaene, S. *The Number Sense: How the mind creates mathematics* Oxford University Press 1997.
- [15] Dehaene, S. et al. *Sources of mathematical thinking: behavioral and brain-imaging evidence*, Science **284** (1999) 970–974.
- [16] DeLoache, J. et al. *Do Babies Learn From Baby Media?*, Psychological Science, in press, 2010.
- [17] De Smedt, Bert, et. al., *Cognitive neuroscience meets mathematics education* Educational Research Review **5** (2010) 97–105.
- [18] _____, Program for 2009 Bruges conference ‘Cognitive neuroscience meets mathematics education’, http://www.earli.org/resources/ASC/Program_EARLIASC_NeuroscienceMathEducation_Bruges2009
- [19] Dresler, T.; Obersteiner, A.; Schecklmann, M.; Vogel, A. C. M.; Ehrlis, A. C.; Richter, M. M.; Plichta, M. M.; Reiss, K.; Pekrun, R.; Fallgatter, A. J. *Arithmetic tasks in different formats and their influence on behavior and brain oxygenation as assessed with near-infrared spectroscopy (NIRS): A study involving primary and secondary school children*. Journal of Neural Transmission **116** (2009) pp. 1689–1700.
- [20] Dunbar, K., Fugelsang, J., and Stein, C. *Do nave theories ever go away?* In M. Lovett, and P. Shah (Eds.), *Thinking with Data: 33rd Carnegie Symposium on Cognition*, Erlbaum pub. (2007) pp. 193–205.
- [21] Fischer, Kurt; Goswami, Usha; Geake, John; *The Future of Educational Neuroscience*, Mind, Brain, and Education **4**, (2010) pp. 68–80.
- [22] Fleming, S.; Weil, R.; Nagy, Z.; Dolan, R.; Rees, G.; *Relating introspective accuracy to individual differences in brain structure*, Science **329** (2010) pp. 1541–1543.
- [23] Gabrieli, John D. E., et al., *Dyslexia: A New Synergy Between Education and Cognitive Neuroscience*, Science **325** (2009), 280–283.
- [24] Geake, J, *Educational neuroscience and neuroscientific education: in search of a mutual middle-way* Research Intelligence 2005, pp. 10–13.

- [25] Goswami, Usha *Neuroscience and education: from research to practice?* Nature Reviews Neuroscience 7 (2006), 406–411.
- [26] Hagmann P; Jonasson L; Maeder P; Thiran JP; Wedeen VJ; Meuli R; *Understanding diffusion MR imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond*, Radiographics 26 Suppl 1 (2006) pp. S205-223.
- [27] Hagmann P; Cammoun L; Gigandet X; Meuli R; Honey CJ; Wedeen, VJ; Sporns, O; *Mapping the structural core of human cerebral cortex* PLoS Biol 6(7) (2008) pp. 1479-1493.
- [28] Herrmann, Christoph S.; Debener, Stefan, *Integration of EEG and fMRI Simultaneous recording of EEG and BOLD responses: A historical perspective*, International Journal of Psychophysiology 67 (2008) pp. 161–168.
- [29] Holt, Jim, *Numbers Guy: Are our brains wired for math?* New Yorker Magazine, March 3, 2008; retrieve from http://www.newyorker.com/reporting/2008/03/03/080303fa_fact_holt.
- [30] Klein, David *A Brief History of American K-12 Mathematics Education in the 20th Century*, Chapter 7 of ‘Mathematical Cognition: A Volume in Current Perspectives on Cognition, Learning, and Instruction’, p. 175-225. Edited by James Royer, Information Age Publishing, 2003. Preprint version retrieved from <http://www.csun.edu/vcmth00m/AHistory.html>.
- [31] Krueger F., Landgraf S., van der Meer E., Deshpande G., Hu X, *Effective connectivity of the multiplication network: A functional MRI and multivariate Granger causality mapping study*, Human Brain Mapping (Wiley), 2010.
- [32] Krueger, F.A.; Spampinato, M. V.; Pardini, M.; Pajevic, S.; Wood, J. N.; Weiss, G. H.; Landgraf, S.; Grafman, J.; *Integral calculus problem solving: An fMRI investigation*, Neuroreport 19 (2008) 1095–1099.
- [33] Lagemann, Ellen Condliffe, *An Elusive Science: the troubling history of education research*, U. Chicago Press 2000.
- [34] Landgraf S, van der Meer E, Krueger F, *Cognitive resource allocation for neuronal activity underlying mathematical cognition: A Multi-Method Study* Int J Math Ed., 2010.
- [35] Lee, K.; Lim, Z. Y.; Yeong, S. H. M.; Ng, S. F.; Venkatraman, V.; Chee, M. W. L. *Strategic differences in algebraic problem solving: Neuroanatomical correlates* Brain Research 1155 (2007) pp. 163–171.
- [36] Liefoghe, B.; Barrouillet, P.; Vandierendonck, A.; Camos, V. *Working Memory Costs of Task Switching* Journal of Experimental Psychology: Learning, Memory, and Cognition 34 (2008) pp. 478–494.

- [37] McCandliss, Bruce D. *Educational neuroscience: The early years* Proc. Nat. Acad. Sci. **107** (2010) pp. 8049–8050.
- [38] Monsell, Stephen *Task switching*, TRENDS in Cognitive Sciences Vol.7 No.3 March 2003
- [39] Obrig, Hellmuth; Villringer, Arno *Beyond the Visible Imaging the Human Brain With Light*, Journal of Cerebral Blood Flow and Metabolism **23** (2003) pp.1–18.
- [40] Periañez, J. A.; Maestú, F.; Barceó, F.; Fernández, A.; Amo, Carlos; Alonso, T. O. *Spatiotemporal brain dynamics during preparatory set shifting: MEG evidence* NeuroImage **21** (2004) pp. 687–695.
- [41] Pashler, Harold, *The Psychology of Attention*, MIT Press 1998.
- [42] Purdy, Noel; Morrison, Hugh, *Cognitive neuroscience and education: unravelling the confusion*, Oxford Review of Education **35** (2009) pp. 99–109.
- [43] Quinn, Frank *The Nature of Contemporary Mathematics*, current draft available at http://www.math.vt.edu/people/quinn/history_nature
- [44] _____ *Essays on Mathematics Education*, preprint at <http://www.math.vt.edu/people/quinn>
- a *Contemporary proofs for mathematics education*, to appear in the proceedings of ICMI Study 19.
- b *Student computing in mathematics: interface design*
- c *Teaching vs. learning in mathematics education*
- d *K-12 Calculator woes*, Notices of the Amer. Math. Soc. May 2009, p. 559.
- [45] Ravizza, S. M.; Anderson, J. R.; Carter, C. S. *Errors of mathematical processing: The relationship of accuracy to neural regions associated with retrieval or representation of the problem state* Brain Research, **1238** (2008) pp. 118–126.
- [46] Richert, R.; Robb, M.; Fender, J.; Wartella, E. *Word Learning From Baby Videos*, Archives of Pediatric and Adolescent Medicine **164** (2010) p. 432.
- [47] Roseberry S; Hirsh-Pasek K; Parish-Morris J; Golinkoff RM; *Live action: can young children learn verbs from video?* Child Dev. **80** (2009) pp. 1360-1375.
- [48] Rosenberg-Lee, M., Lovett, M., Anderson, J. R. *Neural correlates of arithmetic calculation strategies* Cognitive, Affective, and Behavioral Neuroscience, **9** (2009) pp. 270–285.

- [49] Sirotin, Yevgeniy; Das, Aniruddha *Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity*, *Nature* **457** (22 January 2009) 475-479.
- [50] Sohn, Myeong-Ho; Ursu, Stefan; Anderson, John R.; V. Stenger, Andrew; Carter, Cameron S. *The role of prefrontal cortex and posterior parietal cortex in task switching*, *Proc. Nat. Acad. Sci.* **97** (2000) pp. 13448–13453.
- [51] Sohn, M. H.; Goode, A.; Koedinger, K. R.; Stenger, V. A.; Fissell, K.; Carter, C. S.; Anderson, J. R; *Behavioral equivalence, but not neural equivalence—Neural evidence of alternative strategies in mathematical thinking*, *Nature Neuroscience* **7** (2004) 1193–1194.
- [52] Stocco, A.; Anderson, J. R. *Endogenous Control and Task Representation: An fMRI Study in Algebraic Problem Solving* *Journal of Cognitive Neuroscience* **20** (2008) pp. 1300–1314.
- [53] Weisz, Nathan; Moratti, Stephan; Meinzer, Marcus; Dohrmann, Katalin; Elbert, Thomas; *Tinnitus Perception and Distress Is Related to Abnormal Spontaneous Brain Activity as Measured by Magnetoencephalography* *PLoS Medicine* (www.plosmedicine.org) **2** (June 2005) 546–553.
- [54] Willingham, Daniel, Collected articles at <http://www.danielwillingham.com/articles.html>
- a *Ask the Cognitive Scientist: Is it true that some people just can't do math?* *American Educator*, Winter 2009-2010, pp. 14–39.
- b *Ask the Cognitive Scientist: “Brain-Based” Learning: More Fiction Than Fact*, *American Educator*, Fall 2006.
- c *Ask the Cognitive Scientist: Have Technology and Multitasking Rewired How Children Learn?*, *American Educator*, Summer 2010.
- [55] Willingham, Daniel; Lloyd, John; *How Educational Theories Can Use Neuroscientific Data* *Brain Localization and Education*, **1** (2007) pp. 140–149. Retrieve from [54] above.
- [56] Wu, Hung-Hsi, *What's Sophisticated about Elementary Mathematics?* *American Educator*, Fall 2009, pp. 4–14.
- [57] Wu, S. S.; Chang, T. T.; Majid, A.; Caspers, S.; Eickhoff, S. B.; and Menon, V. *Functional Heterogeneity of Inferior Parietal Cortex during Mathematical Cognition Assessed with Cytoarchitectonic Probability Maps* *Cerebral Cortex* **19** (2009) pp. 2930–2945.

Chapter 4

Contemporary Proofs for Mathematics Education

January 2010

Introduction

It is widely known that mathematics education is out of step with contemporary professional practice: Professional practice changed profoundly between about 1890 and 1930, while mathematics education remains modeled on the methodologies of the nineteenth century and before. See [5] for a detailed account.

Professional effectiveness of the new methodology is demonstrated by dramatic growth, in both depth and scope, of mathematical knowledge in the last century. Mathematics education has seen no such improvement. Is this related to continued use of obsolete methodology? Might education see improvements analogous to those in the profession, by appropriate use of contemporary methods?

The problematic word in the last question is “appropriate”: Adapting contemporary methods for educational use requires understanding them in a way that relates sensibly to education, and until recently such understanding has been lacking. The thesis here is that the description of contemporary proof in [5] could be useful at any educational level. Use of contemporary definitions is similarly illustrated in *Contemporary Definitions for Mathematics Education*, in [8].

According to [5], contemporary proofs are first and foremost an enabling technology. Mathematical analysis can, in principle, give the right answer *every time*, but in practice people make errors. The proof process provides a way to minimize errors and locate and fix remaining ones, and thereby come closer to achieving the abstractly-possible reliability.

This view of proof is much more inclusive than traditional ones. “Show work”, for instance, is essentially the same as “give a proof”, while the anno-

tations often associated with proofs appear here in “formal proofs” (Section 4.1.2), as aids rather than essential parts of the structure. To emphasize the underlying commonalities, the word “proof” is used systematically in this essay, but synonyms such as “show work” are appropriate for use with students.

The first section carefully describes proof and its components, but the essence is: “A transcript of work with enough detail that it can be checked for errors.” The second section gives examples of notations and templates designed to let students easily generate effective work transcripts. Good template design depends, however, on deep understanding of student errors. The third section illustrates how carefully designed methods can remain effective for “long problems” well outside the scope of usual classroom work. The final section describes the conflict between contemporary methodology and the way real-world (word) problems are commonly used. Changes and alternatives are suggested.

4.1 Proofs, Potential Proofs, and Formal Proofs

Too much emphasis on the correctness of proofs tends to obscure the features that help *achieve* correctness. Consequently, I suggest that the key idea is actually “potential proof”, which does not require correctness. Variations are described in Sections 4.1.1–4.1.2, and the role of correctness is described in Section 4.1.3. Some educational consequences are discussed in Section 4.1.4, *The Role of Diagnosis*; others occur later in the essay.

4.1.1 Potential Proof

A *potential* proof is a record of reasoning that uses reliable mathematical methods and is presented in enough detail to be checked for errors.

Potential proofs are defined in terms of what they *do* rather than what they *are*, and consequently are context-dependent. At lower educational levels, for instance, more detail is needed. Further, the objective is to enable individual users to get better results, so even in a single class different students may need different versions. Commonalities and functionality are illustrated here, but individual needs must be borne in mind.

4.1.1.1 Example, Integer Multiplication I

Multiply 24 and 47 using single-digit products.

Solution:

$$\begin{array}{r}
 2\ 4 \\
 4\ 7 \\
 \hline
 1\ 1 \\
 1\ 4 \\
 1\ 6 \\
 8 \\
 \hline
 1\ 1\ 1\ 1
 \end{array}$$

This is essentially the traditional format, and is designed to efficiently support the algorithm rather than display mathematical structure; see Section 4.1.2.1 for an alternative. It is also not annotated, so it is not a *formal* proof in the sense of Section 4.1.2. Nonetheless, it provides a clear record of the student's work that can be checked for errors, so it is a potential proof that the product is 1111.

4.1.1.2 About the Example

The example is not a proof because it contains an error. However:

- The error is localized and easily found. Ideally, the student would find and fix it during routine checking
- The error is not random, and a possible problem can be diagnosed: 11 in the third line is the *sum* of 4 and 7, not the product.
- The diagnosis can be used for targeted intervention. If the error is rare the student can be alerted to watch for it in the future. If it resulted from a conceptual confusion then teachers can work with the student to correct it.

4.1.1.3 About the Idea

In the last decade I have spent hundreds of hours helping students with computer-based practice tests. In the great majority of cases they more-or-less understand how to approach the problem and have a record of the work they did, but something went wrong and they can't find the error. The goal is to diagnose the error, correct it, and perhaps look for changes in work habits that would avoid similar errors in the future.

Sometimes the student's work is easy to diagnose: Intermediate steps are clearly and accurately recorded; the reasoning used in going from one to the next can be inferred without too much trouble; the methods used are known to be reliable; etc. In other words it is what is described here as a potential proof. In these cases the mistakes are often minor, and the student often catches them when rechecking. Sometimes I can suggest a change in procedure that will reduce the likelihood of similar mistakes in the future (see Section 4.2 on Proof Templates). The occasional conceptual confusions are well-localized and can usually be quickly set right.

In most cases my students' work does not constitute a potential proof. Problems include:

- Intermediate expressions are incomplete or unclear. For instance when simplifying a fragment of a long expression it is not necessary to copy the parts that do not change, but without some indication of what is going on it is hard to follow such steps and there are frequently errors in reassembly.
- Steps are out of order or the order is not indicated, for instance by numbering.
- Too many steps are skipped.
- The student is working “intuitively” by analogy with an example that does not apply.
- Notations used to formulate a problem (especially word problems) are not clear.

All these problems increase the error rate and make finding errors difficult for either the student or a helper. If not corrected they limit what the student can accomplish.

The point here is that “potential proof” is to some extent an abstraction of the work habits of successful students. The same factors apply to the work of professional mathematicians, though their role is obscured by technical difficulty and the fact that checking typically proceeds rapidly and almost automatically once a genuine potential proof is in hand.

4.1.2 Formal Potential Proof

A *formal* potential proof includes explicit explanation or justification of some of the steps.

The use of justifications is sometimes taken as part of the definition of proof. Here it appears as useful aid rather than a qualitatively different thing: The objective is still to make it possible to find errors, and formality helps with complicated problems and sneaky errors.

The best opportunities for formal proofs in school mathematics are in introducing and solidifying methods that in standard use will not need formality. This process should improve elementary work as well as make the formal-proof method familiar and easily useable when it is really needed. The next example illustrates this.

4.1.2.1 Example, Integer Multiplication II

Multiply 24 and 47 using single-digit products.

Solution:

Explanation	Result
write as polynomials in powers of 10	$(2 \times 10^1 + 4 \times 10^0)(4 \times 10^1 + 7 \times 10^0)$
set up blank form for output	$10^2(\quad) + 10^1(\quad) + 10^0(\quad)$
enter products in the form, without processing	$10^2(2 \times 4) + 10^1(2 \times 7 + 4 \times 4) + 10^0(4 \times 7)$
compute coefficients	$10^2(8) + 10^1(30) + 10^0(28)$
recombine as a single integer	$800 + 300 + 28 = 1128$

4.1.2.2 Comments

This example uses a “structured” format for proof, see [4], [5]. I have not had enough experience to judge the benefits of a standardized structure.

The procedure follows the “template” for multiplication of polynomials described in Section 4.2.1. (See Section 4.3.1 for a version used to multiply large numbers.)

Writing in expanded form with explanations clarifies the procedure. Once the procedure is mastered a short-form version can be used:

$$10^2(\underbrace{2 \times 4}_8) + 10^1(\underbrace{2 \times 7 + 4 \times 4}_{\substack{14 \quad 16 \\ 30}}) + 10^0(\underbrace{4 \times 7}_{28})$$

$$800 + 300 + 28 = 1128$$

In this form:

- The numbers are not rewritten explicitly as polynomials because the coefficients can be read directly from the decimal form. Some students may have to number the digits to do this reliably.
- The extra space in the outer parentheses after the powers of ten indicates that the blank template was set up first.
- The products for the coefficients were entered without on-the-fly arithmetic (explained in Section 4.2.1).
- Individual steps in the arithmetic are indicated, as is the final assembly.

Thus, when the method is familiar, a compressed notation provides an effective potential proof that the outcome is correct.

4.1.2.3 Example, Solutions of Linear Systems

For which values of a is the solution of the system *not* unique?

$$\begin{aligned}x + ay + 2z &= -1 \\3y + az &= 2 - a \\4x + y &= 13\end{aligned}$$

Solution:

The solution to a square linear system is not unique exactly when the determinant of the coefficient matrix is zero. The coefficient matrix here is

$$\begin{pmatrix} 1 & a & 2 \\ 0 & 3 & a \\ 4 & 1 & 0 \end{pmatrix}$$

Row operations $R_3 = R_3 - 4R_1$ and $R_3 = R_3 - \frac{1-4a}{3}R_2$ do not change the determinant and reduce this to a triangular matrix with $R_3 = (0, 0, -8 - a\frac{1-4a}{3})$. The determinant of a triangular matrix is the product of the diagonal entries, so the determinant is

$$(1)(3)\left(-8 - a\frac{1-4a}{3}\right) = -24 - a(1-4a) = 4a^2 - a - 24$$

This is zero for $a = (-1 \pm \sqrt{385})/8$.

4.1.2.4 Comments

This example is a bit less detailed than the previous one in that some calculations (effects of the row operations and application of the quadratic formula) are not recorded. Presumably they are on a separate paper, but because the operations themselves are recorded the calculations can be completely reconstructed. At the level of this example, students should be able to reliably handle such hidden steps and explicit display should not be necessary.

An alternative evaluation of the determinant might be: “Cramer’s rule applied to the second row gives $(+1)(3)(-4 \times 2) + (-1)(a)(1 - 4a) \dots$ ”.

Cramer’s rule involves adding up: a sign times the entry times the determinant of the matrix obtained by omitting the row and column containing the entry. The expression reflects this structure, with the 2×2 determinants evaluated. Giving relatively unprocessed expressions like this both reduces errors (by separating organization from calculation) and allows quick pin-pointing of them when they occur. For example, it would be possible to distinguish a sign error in the second term due to a misunderstanding of Cramer’s rule, from a sign error in the evaluation of the sub-determinant.

Students will not give this sort of explanation without examples to copy and quite a bit of guidance. This guidance might include:

- When using a theorem (e.g. nonzero determinant if and only if unique solutions), say enough about it to inspire confidence that you know a precise statement and are using it correctly. Confused statements indicate that conceptual errors are likely in the future, even if this wasn't the problem in this case.
- In particular, mention of the theorem is an essential part of the work and must be included even in short-form versions. (For additional discussion of style in short-form proofs, see *Proof Projects for Teachers* in [8].)
- In lengthy calculations, rather than showing all details, describe the steps and carry out details on a separate sheet. The descriptions should be explicit enough to enable reconstruction of the details. Organizing work this way both reduces errors and makes it easier to check.

It can be helpful to have students check each others' work and give explicit feedback on how well the layout supports checking. The eventual goal is for them to diagnose their own work; trying to make sense of others' work can give insight into the process.

4.1.2.5 Further Examples

For further discussion, and examples of elementary formal proofs concerning fractions and area, see *Proof Projects for Teachers* [8].

4.1.3 Proof and Correctness

A *proof* is a potential proof that has been checked for errors and found to be error-free.

Work that does not qualify as a potential proof cannot be a proof even if the conclusion is known to be correct. In education, the goal is not a correct answer but to develop the ability to routinely get correct answers; facility with potential proofs is the most effective way to do this. Too much focus on correctness may undercut development of this facility.

This is usually not an issue with weak students because potential proofs are an enabling technology without which they cannot succeed. Weak students tend to have the opposite problem: the routines are so comforting and the success so rewarding that it can be hard to get them to compress notation (e.g. avoid recopying) or omit minor details even when they have reached the point where it is safe to do so. Similarly, some persist in writing out formal justifications even after they have thoroughly internalized the ideas.

Strong students are more problematic, because the connection between good work habits and correct answers is less direct. I have had many students who were very successful in high-school advanced placement courses, but they got by with sloppy work because the focus was on correctness rather than methodology. Many of these students have trouble with engineering calculus in college:

- The better students figure it out, especially with diagnostic support and good templates (Section 4.2). Most probably never fully catch up to where they might have been, but they are successful.
- Unfortunately a significant number were good enough to wing it in high school and good enough to have succeeded in college with good methodological preparation, but are not good enough to recover from poor preparation.

All students stand to benefit from a potential-proof-oriented curriculum rather than a correctness-oriented one, but for different reasons. Gains by weak and mid-range students are likely to be clearest.

4.1.4 The Role of Diagnosis

The thesis of this article is that the reliability possible with mathematics can be realized by making mathematical arguments that can be checked for errors, *checking them*, and correcting any errors found. Other sections describe how checkable arguments could become a routine part of mathematics education. However they won't produce benefits unless *checking* also becomes a routine part. To be explicit: Diagnosis and error correction should be key focuses in mathematics education.

- Answers are important mainly as proxies for the work done. Incorrect answers indicate a need for diagnosis and correction. Ideally, *every* problem with a wrong answer should be diagnosed and corrected.
- Mathematics uniquely enables quality, so the emphasis should be on quality not quantity. In other words, doing fewer problems to enable spending more time on getting them right is a good tradeoff.
- An important objective is to teach students to routinely diagnose their own work. The fact that diagnosis is possible and effective is the essence of mathematics, so teaching self-diagnosis is mathematics education in the purest sense.

Ideally, teachers would regularly go through students' work with them so students can see the checking process in action. Students should be required to redo problems when the work is hard to check, not just when the answer is wrong. As explained in the previous section, the goal is to establish work habits that will benefit students; however students respond to feedback from teachers, not to long-term goals.

4.1.5 Other Views of Proof

There are many other—and quite different—views of the role of proof (c.f. [1], [11], [12]). These generally emphasize proofs as sources of understanding and insight, or as repositories of knowledge.

The basic difference is that I have emphasized proofs as an enabling technology for users. Most other views focus on “spectator proofs”: arguments from which readers should benefit, but that are not intended as templates for emulation. Both views are valid in their own way, and this should be kept in mind when considering specific situations.

What counts as user-oriented or spectator-oriented, and the mix in practice, varies enormously with level. In school mathematics—as illustrated here—almost everything is designed for emulation. Spectator proofs play little or no role. Issues that might be addressed with spectator proofs (e.g., how do we know the multiplication algorithm really works?) are simply not addressed at all.

At intermediate levels, college math majors for instance, spectator proofs play a large role. They provide ways for students to learn and develop skills long before they can be emulated. At the research frontier the primary focus is again on user-oriented work. It is a nice bonus if an argument functions as a spectator proof (i.e., is “accessible”), but if the argument cannot be fleshed out to give a fully-precise user-oriented proof it is unsatisfactory.

Misunderstanding these different roles of proof has led to conflict and confusion. For example, Thurston [12] justified his failure to provide a proof of a major claim by observing that the technology needed for a good spectator proof was not yet available. This point resonated with educators since they have a mainly spectator-oriented conception of proof. However Thurston was responding to criticism [2] that he had failed to provide a user-oriented proof for use in the research community. An inability to provide a spectator proof was not accepted as justifying the failure to provide any proof at all. The problem was later declared unsolved, and complete proofs were eventually provided by others (see [5]).

4.2 Proof Templates

Students learn mainly by abstraction from examples and by imitating procedures. It is important, therefore, to carefully design examples and procedures to guide effective learning.

A “proof template” is a procedure for working a class of problems. Design considerations are:

- Procedures should clearly reflect the mathematical structures they exploit. This makes them more reliable and flexible, and often provides subliminal preparation for more complex work.
- Procedures should minimize problems with limitations of human cognitive abilities. For example, conceptually distinct tasks such as translating word problems, organizing a computation, or doing arithmetic, should be separated.
- Efficient short-form versions should be provided.

Examples in this section explain and illustrate these points.

4.2.1 Polynomial Multiplication

1

4.2.1.1 Problem

Write $(3z^2 - z + 5a)(z^3 + (2 - a)z^2 - a)$ as a polynomial in z . Show steps.

4.2.1.2 Step 1: Organization

There are three terms in each factor, so there will be nine terms in the product. Some organizational care is needed to be sure to get them all. Further, we would like to have them sorted according to exponent on z rather than producing them at random and then sorting as a separate step. To accomplish this, we set up a blank form in which to enter the terms. A quick check of exponents shows that all exponents from 0 to 5 will occur, so the appropriate blank form is:

$$\begin{array}{r} z^5[\quad \quad \quad] + z^4[\quad \quad \quad] + z^3[\quad \quad \quad] + \\ z^2[\quad \quad \quad] + z^1[\quad \quad \quad] + z^0[\quad \quad \quad] \end{array}$$

Next, scan through all possible combinations of terms, one from each factor. (Use a finger to mark your place in one term while scanning the other.) For each combination, write the product of coefficients in the row with the right total exponent. The result is:

$$\begin{array}{r} z^5[(3)(1) \quad \quad \quad] + z^4[(3)(2 - a) + (-1)(1)] + z^3[(-1)(2 - a) + (5a)(1)] + \\ z^2[(3)(-a) + (5a)(2 - a)] + z^1[(-1)(-a) \quad \quad \quad] + z^0[(5a)(-a) \quad \quad \quad] \end{array}$$

Note the products were recorded with *absolutely no* arithmetic, not even writing $(3)(1)$ as 3. Reasons are:

- Organization and arithmetic are cognitively different activities. Switching back and fourth increases the error rate in both, with sign errors being particularly common.
- This form can be diagnosed. We can count the terms to see that there are nine of them and the source of each term can be identified. The order of scanning can even be inferred, though it makes no difference.

Note also that every term is enclosed in parentheses. This is partly to avoid confusion, because juxtaposition is being used to indicate multiplication. The main reason, however, is to avoid thinking about whether or not parentheses are necessary in each case. Again, such thinking is cognitively different from the organizational task and may interfere with it.

¹This material is adapted from the polynomial problem list in [9].

4.2.1.3 Step 2: Calculation

Simplify the coefficient expressions to get the answer:

$$3z^5 + (5 - 3a)z^4 + (6a - 2)z^3 + (7a - 5a^2)z^2 + az + (-5a^2)$$

In this presentation the only written work is the organizational step and the answer. More complicated coefficient expressions, or less experienced students, would require recording some detail about the simplification process. A notation for this is shown in the arithmetic example in Section 4.1.2.1.

4.2.1.4 Comments

- The separation of organization and computation makes the procedure reliable and relatively easy to use.
- The close connection to mathematical structure makes the procedure flexible. It is easily modified to handle problems like “Find the coefficient on z^3 ” or “Write a product involving both x and y as a two-variable polynomial”.
- Variations provide methods for by-hand multiplication of integers (Section 4.1.2.1) and multiplication of large integers using a calculator (Section 4.3.1).
- If the baby version in Section 4.1.2.1 is used to multiply integers, then students will find the polynomial version familiar and easy to master.
- Similarly, students who work with polynomials this way will find some later procedures (e.g., products of sums that may not be polynomials, or iterated products like the binomial theorem) essentially familiar and easier to master.

This procedure should be contrasted with the common practice of restricting to multiplication of binomials, using the “FOIL” mnemonic². That method is poorly organized even for binomials, inflexible, and doesn’t connect well even with larger products. In particular, students trained with FOIL are often unsuccessful with products like the one in the example.

4.2.2 Solving Equations

This is illustrated with a very simple problem, so the structuring strategies will be clear.

²First, Outer, Inner, Last.

4.2.2.1 Problem

Solve $5x - 2a = 3x - 7$ for x .

Annotated Solution:

Explanation	Result
Collect terms: move to other side by adding negatives	$5x - 3x = -7 + 2a$
calculate	$\underbrace{(5 - 3)}_2 x = -7 + 2a$
move coefficient to other side by multiplying by inverse	$x = \frac{1}{2}(-7 + 2a)$

4.2.2.2 Comments

The primary goals in this format are efficiency and separation of different cognitive activities (organization and calculation).

The first step is organizational: we decide that we want all x terms on one side and all others on the other. Collecting x terms can be accomplished by adding $-3x$ to each side. However it is inefficient to do this as a separate calculation step because we know ahead of time what will happen on the right side: we have chosen the operation exactly to cancel the $3x$ term. Instead we think of it as a purely organizational step: “move $3x$ to the other side...”. To keep it organizational we refrain from doing arithmetic (combining coefficients) and include “by adding negatives” to the mental description.

The second step is pure calculation.

The final step is again organizational, and the description is designed to emphasize the similarity to the first step.

Finally, the steps are guided by pattern-matching: The given expression is manipulated to become more like the pattern $x = ?$. (See the next section for another example.)

4.2.3 Standardizing Quadratics

This is essentially “completing the square” with a clear goal.

4.2.3.1 Problem

Find a linear change of variables $y = ax + b$ that transforms the quadratic $5x^2 - 6x + 21$ into a standard form $r(y^2 + s)$ with s one of $1, 0 - 1$, and give the standard form.

This is done in two steps, each of which brings the expression closer to the desired form. A short-form version is given after the explanation.

4.2.3.2 First Step

Eliminate the first-order term with a change of the form $y_0 = x + t$.

Square the general form and multiply by 5 to get $5y_0^2 = 5x^2 + 10tx + 5t^2$, which has the same second-order term as that of the given quadratic. To match the first-order term as well we need $10t = -6$, so $t = -3/5$ and $y_0 = x - 3/5$. Moving the constant term to the other side gives $5y_0^2 - 5t^2 = 5x^2 - 6x$. Use this to replace the first- and second-order terms in the original to transform it to

$$5(y_0)^2 \underbrace{-5(-3/5)^2 + 21}_{-\frac{9}{5} + \frac{105}{5} = \frac{96}{5}} \quad (4.1)$$

4.2.3.3 Second Step

Factor out a *positive* number to make the constant term standard.

$$5y_0^2 + \frac{96}{5} = \frac{96}{5} \left(\underbrace{\frac{5^2}{96} y_0^2}_{(\frac{5}{\sqrt{96}} y_0)^2} + 1 \right) \quad (4.2)$$

The number factored out must be positive because we had to take the square root of it.

Comparing with the goal shows the standard form is $\frac{96}{5}(y^2 + 1)$ with $y = \frac{5}{\sqrt{96}}y_0 = \frac{5}{\sqrt{96}}(x - \frac{3}{5})$.

4.2.3.4 Short Form

$$5 \underbrace{(x+t)^2}_{y_0} = 5x^2 + \underbrace{10t}_{-6} x + 5t^2$$

So $t = -3/5$.

$$5y_0^2 \underbrace{-5(3/5)^2 + 21}_{\frac{96}{5}} = \frac{96}{5} \left(\underbrace{\frac{5}{96} 5y_0^2}_{(\frac{5}{\sqrt{96}} y_0)^2} + 1 \right)$$

So $y = \frac{5}{\sqrt{96}}y_0 = \frac{5}{\sqrt{96}}(x - \frac{3}{5})$ and the form is $\frac{5}{96}(y^2 + 1)$.

Methods must be introduced with explanations, but compression is necessary for routine use. It is important for teachers to provide a carefully-designed short format because the compressions which student invent on their own are rarely effective.

For example, it is often necessary to simplify a fragment of an expression. The underbrace notation here indicates precisely which fragment is involved and connects it to the outcome. I have never seen a student do this. Usually, the student either writes fragments without reference or rewrites the whole expression.

Experience often reveals errors that need to be headed off by the notation. In the work above, the notation

$$5y_0^2 \underbrace{-5(3/5)^2 + 21}_{\frac{96}{5}}$$

clearly indicates that the sign on $-5(3/5)^2$ is part of the fragment being simplified. Many students seem to think of this sign as the connector between the expression fragments, and hence do not include it in the sub-expression. It then gets lost. This is a common source of errors, and may well have resulted in the student making an error in this case. Providing a clear notation and being consistent in examples will avoid such errors.

4.2.3.5 Pattern Matching

Routine success requires that at any point the student can figure out “What should I do next?” In the problem above there is a direct approach: Plug $y = ax + b$ into the given quadratic, set it equal to $r(y^2 + s)$, and solve for a, b, r, s . This can be simplified by doing it in two steps, as above, but even so it requires roughly twice as much calculation as the method given above. This is a heavy price to pay for not having to think.

By contrast, the suggested procedure uses pattern matching to guide the work. It can be summarized as “What do we have to do to the given quadratic to get it to match the standard pattern?” In the first step we note that the given one has a first-order term and the pattern does not. We get closer to the pattern by eliminating this term, getting something of the form $Ay_0^2 + B$. If B is not 1, 0, or -1 we can get closer to the pattern by factoring something out to get $C(Dy_0^2 + s)$ with standard s . The only thing remaining to exactly match the pattern is to rewrite Dy_0^2 as a square, and whatever result we get is the y we are seeking.

Pattern matching is a powerful technique, a highly-touted feature of computer algebra systems, and humans can be very good at it. Much of the work in a calculus course can be seen as pattern-matching. Students could use it more effectively if teachers presented the idea more explicitly.

4.2.4 Summary

Carefully-designed procedures and templates for students to emulate can greatly improve success and extend the range of problems that can be attempted. Important factors are:

- Procedures should follow the underlying mathematical structure as closely as possible. Doing so reveals connections, provides flexibility, and expands application. It also ensures upward-compatibility with later work, and frequently provides subliminal preparation for this work.
- Ideas that guide the work, pattern matching for example, should be abstracted and made as explicit as possible for the level.

- Procedures should separate different cognitive tasks. In particular, organizational work should be kept separate from computation.
- Short-form formats that show the logical structure (i.e., are checkable) and encourage good work habits should be provided.

Good test design can also encourage good work habits. For example:

- Ask for a single coefficient from a good-sized product like the example in Section 4.2.1. This rewards students who understand the organizational step well enough to pick out only the terms that are needed.
- A computer-based test might ask for an algebraic expression that *evaluates* to give the coefficient³. The students could then enter the unevaluated output from the organizational step. This approach rewards careful separation of organization and calculation, by reducing the time required and reducing the risk of errors in computation.

4.3 Long Problems

Current pre-college mathematics education is almost entirely concerned with short, routine problems. Advanced-placement courses may include short tricky problems. However, much of the power of mathematics comes from its success with long *routine* problems. Because the conclusions of each step can be made extremely reliable, many steps can be put together and the combination will still be reliable. Further, carefully-designed methods for dealing with short problems will apply to long problems equally well.

Long problems have an important place in elementary mathematics education. They give a glimpse into the larger world and illustrate the power of the methods being learned. They also reveal the need for care and accuracy with short problems. It is not clear how long problems might be incorporated into a curriculum, but group projects are a possibility. The examples here are presented as group problems about multiplication and addition of large integers (with calculators) and logic puzzles.

4.3.1 Big Multiplications

The goal is to exactly multiply two large (say 14- or 15-digit) integers using ordinary calculators. This cannot be done directly so the plan is to break the calculation into smaller pieces (e.g., 4-digit multiplications) that can each be done on a calculator, and then assemble the answer from these pieces. The method is the same as the by-hand method for getting multi-digit products from single-digit ones, and uses a notation (like that of Section 4.1.2.1) modeled on polynomial multiplication.

³Tests with this kind of functionality are a goal of the EduTeX project [10].

The number of digits in each piece depends on the capability of the calculators used. The product of two 4-digit numbers will generally have 8 digits. We will be adding a list of these, but no more than nine, so the outcome will have 9 or fewer digits. Four-digit blocks will therefore work on calculators that can handle nine digits. Eight-digit calculators would require the use of three-digit blocks.

4.3.1.1 Problem

Multiply 638521988502216 and 483725147602252, using calculators that handle 9 or more digits, by breaking them into 4-digit blocks.

4.3.1.2 Step 1: Organize the Data

Write the numbers as polynomials:

$$\begin{aligned} 638521988502216 &= 2216 + 8850x + 5219x^2 + 638x^3 \\ 483725147602252 &= 2252 + 4760x + 7251x^2 + 483x^3 \end{aligned}$$

where $x = 10^4$.

The power-of-ten notation should be used even with pre-algebra students, because it is a powerful organizational aid. The exponent records the number of blocks of four zeros that follow these digits.

4.3.1.3 Step 2: Organize the Product

The product of two sums is gotten from all possible products, using one piece from each term. Individual terms follow the rule $(ax^n)(bx^k) = (ab)x^{n+k}$, which we use to organize the work. The product will have terms x^r for $r = 0, \dots, 6$ and seven individuals or teams could work separately on these.

For instance, the x^2 team would collect the pairs of terms whose exponents add to 2: x^0 (x not written) from the first number and x^2 from the second, then x^1 from the first and x^1 from the second, etc. They would record:

$$x^2(2216 \times 7251 + 8850 \times 4760 + 5219 \times 2252)$$

This is an organizational step; no arithmetic should be done. The students can infer how the pieces were obtained, and can double-check each other to see that nothing is out of place and no pieces were left out.

4.3.1.4 Step 3: Compute the Coefficient

Carry out the arithmetic indicated in the second step, using calculators. If the students can use a memory register to accumulate the sum of the successive products then the output is the answer, $x^2(69947404)$. If the multiplications

and addition have to be done separately then the notation of Section 4.1.2.1 can be used:

$$x^2 \left(\underbrace{2216 \times 7251}_{16068216} + \underbrace{8850 \times 4760}_{42126000} + \underbrace{5219 \times 2252}_{11753188} \right)$$

$$\underbrace{\hspace{15em}}_{69947404}$$

Again, different students or teams should double-check the outcomes.

4.3.1.5 Step 4: Assemble the Answer

At this point the group has found the product of polynomials,

$$4990432 + 30478360x + 69947404x^2 + 91520894x^3 + 45154399x^4 + 7146915x^5 + 308154x^6$$

and the next step is to evaluate at $x = 10^4$, or in elementary terms translate the powers of x back to blocks of zeros, and add the results. The next section gives a way to carry out the addition.

4.3.2 Big Additions

The goal is to add a list of large integers using ordinary calculators. This cannot be done directly, so the plan is to break the operations into smaller pieces (e.g., 6-digit blocks) that can be done on a calculator and then assemble the answer from these pieces. The procedure is illustrated with the output from the previous section.

4.3.2.1 Problem

Use calculators to add $4990432 + 30478360 \times 10^4 + 69947404 \times 10^8 + 91520894 \times 10^{12} + 45154399 \times 10^{16} + 7146915 \times 10^{20} + 308154 \times 10^{24}$ using 6-digit blocks.

4.3.2.2 Step 1: Setup

			4	990432
			304783	600000
		6994	740400	
	91	520894		
	451543	990000		
714	691500			
308154				

Here we have written the seven numbers to be added in a column with aligned digits. Vertical lines are drawn to separate the 6-digit blocks, and we omit blocks that consist entirely of zeros. We have not, however, omitted zeros at the end of blocks because doing this would mix organizational and arithmetic thinking.

4.3.2.3 Step 2: Add 6–digit Columns

			4	990432
			304783	600000
		6994	740400	
	91	520894		
	451543	990000		
714	691500			
308154				
308868	1	517888	1	590432
1	143134	1	045187	

Each column is added separately, for instance by five different students; again, the outcomes should be double-checked.

Most of the sums overflow into the next column. We have written the sums of the even-numbered columns one level lower to avoid overlaps. Since there are fewer than nine entries in each column, the sum can overflow only into the first digit of the next column to the left.

4.3.2.4 Step 3: Final Assembly

Add the sums of the individual columns:

308868	1	517888	1	590432
1	143134	1	045187	
308869	143135	517889	045188	590432

In this example the final addition is easy, because the overflow from one column only changes one digit in the next. This happens in most cases; if examples are chosen at random, it is very unlikely that students will see more than two digits change due to overflow.

Students should realize, however, that digits in sums are unstable in the sense that, very rarely, an overflow will change *everything* to the left. Teachers should ensure that students encounter such an example, or perhaps challenge them to contrive an example that makes the simple-minded pattern crash. This phenomenon illustrates the difference between extremely unlikely events and mathematically impossible ones, and the “low-probability catastrophic failures⁴” that can occur when the difference is ignored.

4.3.3 Digits in Big Products

The goal here is to find a specific digit in a product of big numbers, and be sure it is correct. An attractive feature of the formulation is that careful reasoning and understanding of structure are rewarded by a reduction in computational work.

⁴A term from the computational software community, where this is a serious problem.

The least-thought/most-work approach is to compute the entire number and then throw away all but one of the digits. I give three variations with increasing sophistication and decreasing rote computation. In practice, students (or groups) could be allowed to choose the approach that suits their comfort level. More-capable students will enjoy exploiting structure to achieve efficiency. Less-capable ones will be aware of the benefits of elaborate reasoning, but may see additional rote computation as a safer and more straightforward.

4.3.3.1 Problem

Find the eighteenth digit (from the right, i.e. in the 10^{17} place) in the product $52498019913177259058 \times 33208731911634712456$.

4.3.3.2 Plan A

We approach this as before, by breaking the numbers into 4-digit blocks and writing them as coefficients in a polynomial in powers of $x = 10^4$. These are 20-digit numbers so there are five 4-digit blocks and this gives polynomials of degree 4 (powers of x up to x^4). The product has terms up to degree 8.

The eighteenth digit is the second digit in the fifth 4-digit block ($18 = 4 \times 4 + 2$). When working with polynomials in $x = 10^4$ this means it will be determined by the terms of degree x^4 and lower (the coefficient on x^5 gets 20 zeros put after it, so cannot effect the 18th digit).

Plan A is to compute the polynomial coefficients up to x^4 , combine as before to get a big number, and see what the 18th digit is. This gives a significant savings over computing the whole number because we don't find the $x^5 \dots x^8$ coefficients.

4.3.3.3 Plan B

This refinement of Plan A reduces the work done on the x^4 coefficient.

We only need the 18th digit, so only need the second (from the right) digit in the coefficient on x^4 . To get this we only need the product of the lowest two digits in each term. To make this explicit, the terms in the coefficient on x^4 are:

$$x^4(9058 \times 3320 + 7725 \times 8731 + 9131 \times 9116 + 8019 \times 3471 + 5249 \times 2456)$$

But we only need the next-to-last digit of this. If we write the first term as $(9000 + 58) \times (3300 + 20)$, then the big pieces don't effect the digit we want. It is sufficient just to compute 58×20 .

This modification replaces the x^4 coefficient by

$$x^4(58 \times 20 + 25 \times 31 + 31 \times 16 + 19 \times 71 + 49 \times 56)$$

Lower coefficients are computed and the results are combined to give a single number as before. This number will have the same lower 18 digits as the full product, and in particular will have the correct 18th digit.

4.3.3.4 Plan C, Idea

Plans A and B reduce work by not computing unneeded higher digits. Here, we want to reduce work by not computing unneeded *lower* digits. The overflow problem makes this tricky, and some careful estimation is needed to determine how bad lower-digit overflows can be. This is a nice opportunity for good students to exploit their talents.

1. The coefficients in the product polynomial have at most nine digits (products of 4-digit numbers have at most 8 digits, and we are adding fewer than nine of these in each coefficient). The x^2 term therefore has at most $9 + 2 \times 4 = 17$ digits. This can affect the 18th digit only through addition overflow.
2. The plan, therefore, is to compute the coefficients on x^4 and x^3 , combine these to get a number, and see how large a 17-digit number can be added before overflow changes the 18th. We will then have to estimate the x^2 and lower terms and compare this to the overflow threshold.
 - If the lower-order terms cannot cause overflow into the 18th digit, then the 18th digit is correct.
 - If lower terms might cause overflow, then we will have to compute the x^2 coefficient exactly, combine with the part already calculated, and see what happens. In this case, we will also have to check to see if degree 0 and 1 terms cause overflow that reaches all the way up to the 18th digit. This is extremely unlikely: These terms have at most $9 + 1 \times 4 = 13$ digits, so overflow to the 18th can only happen if the 14th through 17th digits are all 9.
 - In this unlikely worst-case scenario we will have to compute the lower-order terms too.

4.3.3.5 Plan C, Setup and Compute

The x^3 coefficient and Plan B version of the x^4 coefficient are:

$$\begin{array}{l} x^4 \quad (58 \times 20 + 25 \times 31 + 31 \times 16 + 19 \times 71 + 49 \times 56) \\ x^3 \quad (9058 \times 8731 + 7725 \times 9116 + 9131 \times 3471 + 8019 \times 2456) \end{array}$$

Computing gives $200894863x^3 + 6524x^4$. Substituting $x = 10^4$ gives

$$(200894863 + 65240000) \times 10^{12} = 266134863 \times 10^{12}.$$

The 18th digit (from the right) is 1. It is not yet certain, however, that this is the same as the digit in the full product.

4.3.3.6 Plan C, Check for Overflow

The 17th digit in 266134863×10^{12} is 3. If the top (i.e. 9th) digit in the x^2 coefficient is 5 or less then adding will not overflow to the 18th digit. ($3+5=8$, and overflow from the x^1 and x^0 terms can increase this by at most one).

The next step is to estimate the top digit in this coefficient.

1. The x^2 coefficient has three terms (from x^0x^2 , x^1x^1 , and x^1x^0).
2. Each term is a product of two 4-digit numbers, so each has at most 8 digits. In other words the contribution of each term is smaller than 10^9 . Adding three such terms gives a total coefficient smaller than 3×10^9 .
3. When we substitute $x = 10^4$ we get a number less than 3×10^{17} . The top digit is therefore at most 2.
4. Since the top digit of the lower-order term is smaller than the threshold for overflow ($2 \leq 5$), we conclude that the 18th digit found above is correct.

We were fortunate: If the 17th digit coming from the higher-order terms had been 7, 8, or 9 then we could not rule out overflow with this estimate. For borderline cases I describe a refined estimate that gives a narrower overflow window.

The actual coefficient on x^2 is 131811939. Knowing this, we see that a 17th digit 7 would not have caused an overflow, while a 9 would have increased the 18th digit by 1, and 8 is uncertain. This conclusion can be sharpened by using more digits: If digits 15-17 are 867 or less, then there is no overflow; if they are 869 or more then there is an overflow of 1; and the small interval between these numbers remains uncertain. As noted above, in rare cases lower-order terms have to be computed completely to determine whether or not overflow occurs.

4.3.3.7 Grand Challenge

Use this method to find the 25th digit of the product of two fifty-digit numbers.

4.3.4 Puzzles

We will not explore them here but logic puzzles deserve mention as opportunities for mathematical thinking (see Wanko [13], and Lin [3]). These should incorporate an analog of proof: a record of moves that enables reconstruction of the reasoning and location of errors. The notation for recording chess moves (see [Wikipedia](#)) may be a useful model.

A minor problem is that the rules of many puzzles are contrived to avoid the need for proof-like activity and should be de-contrived.

For example, the usual goal in Sudoku is to fill entries to satisfy certain conditions. The final state can be checked for correctness and—unless there is an error—would seem to render the record of moves irrelevant. A better goal is to find *all* solutions. If the record shows that every move is forced, then the

solution is unique. However, if at some point no forced moves can be found and a guess is made, all branches must be followed. If a branch leads to an error, that branch can be discarded (proof by contradiction). If a branch leads to a solution, then other branches still have to be explored to determine whether they also lead to solutions. This would be made more interesting by a source of Sudoku puzzles with multiple solutions.

Notations and proof also enable collaborative activity. All members of a group would be given a copy of the puzzle, and one appointed “editor”. On finding a move, a member would send the notation to the editor as a text message. The editor would check for correctness and then forward the move to the rest of the group. Maintaining group engagement might require a rule like: Whoever submits a move must wait for someone else to send one before submitting another.

4.4 Word Problems and Applications

This essay concerns the use of contemporary mathematical methodology in education. Up to this point the ideas have been unconventional and possibly uncomfortable but more-or-less compatible with current educational philosophy.

There are, however, genuine conflicts where both contemporary methodology and direct experience suggest that educational practices are counterproductive, not just inefficient. Some of the methodological conflicts are discussed in this section. A more systematic comparison is given in *Mathematics Education is Poorly Adapted* in [8], and conflicts in concept formation are discussed in *Contemporary Definitions for Mathematics Education* in [8]. Historical analysis in [5] indicates that many educational practices are modeled on old professional practices that were subsequently found ineffective and were discarded.

4.4.1 Word Problems and Physical-World Applications

The old view was that mathematics is an abstraction of patterns in the physical world and there is no sharp division between the two. The contemporary view is that there is a profound difference and the articulation between the two worlds is a key issue. The general situation is described in [5]; here I focus on education.

4.4.1.1 Mathematical Models

In the contemporary approach, physical-world phenomena are approached indirectly: a *mathematical model* of a phenomenon is developed and then analyzed mathematically. The relationship between the phenomenon and the model is not mathematical, and is not accessible to mathematical analysis.

4.4.1.2 Example

A beaker holds 100 cc. of water. If 1 cc of X is added, what is the volume of the result?

Expected solution: $100 + 1 = 101$ cc.

4.4.1.3 Discussion

The standard expected solution suppresses the modeling step. Including it gives:

Model: volumes add.

Analysis: $100 + 1 = 101$ so the model predicts volume 101 cc.

The analysis of the model is certainly correct, so it correctly predicts the outcome when the model applies: for example, if X is water. If X is sand, salt, or alcohol, then the volume will be more than 100 cc. but significantly less than the predicted value of 101 cc. If X is metallic sodium a violent reaction takes place. When the smoke clears, the beaker will contain considerably less than 100 cc., and may be in pieces.

In the latter instances, the prediction fails because the model is not appropriate. This is not a mathematical difficulty. In particular no amount of checking the written work can reveal an error that accounts for the failure. One might try to avoid the problem in this case by specifying that X should be water, but discrepancies could result from differences in temperature. Even elaborately legalistic descriptions of the physical circumstances cannot completely rule out reality/model disconnects.

The point is that the reality/model part of real-world applications is essentially non-mathematical. Applications have an important place in mathematics courses, but the reality/model aspect should not be represented as mathematics.

Equally important, modeling and analysis of the model are different cognitive activities. Failing to separate them increases error rates, just as happens with organization and calculation (see Section 4.2.1). Students may be successful without separating modeling and analysis when the mathematical component is trivial, as is often the case in pre-college courses these days. However difficulties grow rapidly as complexity grows, so in practice failing to separate these activities makes significant applications impossible.

4.4.2 Applications

Mathematics is brought to life through applications. In this context the word “application” is usually understood to mean “physical-world application”. However, such applications alone do a poor job of bringing elementary mathematics to life. After explaining why, I suggest that there are better opportunities using applications from within mathematics.

4.4.2.1 Difficulties with the Real World

The main difficulty with physical-world applications is a complexity mismatch. In one direction, there are impressive applications of elementary mathematics, but they require significant preparation in other subjects. On the other hand there are easily-modeled real-world problems but these tend to be either mathematically trivial or quite sophisticated.

Examples of applications of elementary mathematics:

- One can do interesting chemistry with a little linear algebra, but the model-building step requires a solid grasp of atomic numbers, bonding patterns, etc. The preparation required is probably beyond most high-school chemistry courses and certainly beyond what one could do in a mathematics course.
- There are nice applications of trig functions to oscillation and resonance in mechanical systems, electric circuits, and acoustics. Again, subject knowledge requirements makes these a stretch even in college differential equations courses.
- Multiplication of big integers, as in Section 4.3.1, plays an important role in cryptography, but it is not feasible to develop this subject enough to support cryptographic “word problems”.

Problems with easily-modeled situations include:

- It is difficult to find problems that are not best seen as questions in calculus or differential equations (or worse).
- Special cases may have non-calculus solutions, but these solutions tend to be tricky and rarely give insight into the problem.
- Even as calculus problems, most “simple” models lead to mathematical questions too hard for use in college calculus.
- Our world is at least three-dimensional. Many real problems require vectors in all but the most contrived and physically-boring cases.

In other words, real-world problems should be part of a serious development of a scientific subject in order to be genuinely useful. The next section describes difficulties that result when this constraint is ignored.

4.4.2.2 Bad Problems

The practical outcome of the complexity mismatch described above is that most word problems—in the US anyway—have trivial or very constrained mathematical components and the main task is formulation of the model (e.g., the example in Section 4.4.1.1).

Some elementary-education programs exploit this triviality with a “key-word” approach: “When a problem has two numbers, then the possibilities

are multiplication, division, addition or subtraction. Addition is indicated by words ‘added’, ‘increased by’ ...”. The calculator version is even more mindless, because the operations have become keystrokes rather than internalized structures that might connect to the problem: “Press the “+” key if you see ‘added’, ‘increased by’, ...”

The higher-level version of this can be thought of as “reverse engineering”: Since only a few techniques are being tested, one can use keywords or other commonalities to figure out which method is correct and where to put the numbers.

Other problem types amount to translating jargon: Replace “velocity” with “derivative”, “acceleration” with “second derivative”, ...

- In other words, there is so little serious contact with any real-world subject that translation and reverse-engineering approaches that *avoid* engagement are routinely successful, and are fast and reliable. Students who master this skill may enjoy word problems, because the trivial math core makes success easy.
- The errors I see make more sense as translation problems than conceptual problems. A common example: When one is modeling the liquid in a container, liquid flowing *out* acquires a negative sign, because it is being *lost* from the system. Translators miss the sign, students who actually envision the situation should not.
- Some of my students despise word problems, regarding them as easily-solved math problems made hard by a smokescreen of terminology and irrelevant material. These students may be weak at this cognitive skill, or they may be thinking too much and trying to engage the subject. In any case, the most effective help I can offer is to show them how to think of it as an intelligence-free translation problem.
- Finally, many problems are so obviously contrived that they cannot be taken seriously. The one that begins “If a train leaves Chicago at 2:00...” has been the butt of jokes in comic strips.

Conventional wisdom holds that word problems engage students and provide an important connection to real-world experience. This notion is abstractly attractive, but the difficulties described above keep it from being effective in practice. Further, a curriculum justified by, or oriented toward, word problems is likely to be weak, because weak development is good enough for immediately-accessible problems.

4.4.3 Mathematical Applications

A common justification for word problems is that mathematics is important primarily for its applications, and math without applications is a meaningless formal game. I might agree, with the following reservations:

- Goals should include preparation for applications that will not be accessible for years, not just those that are immediately accessible.
- “Application” should be interpreted to include applications *in mathematics* as well as real-world topics.

The application of polynomial multiplication to multiplication of big integers in Section 4.3.1, and the refinements developed in Section 4.3.3 to minimize the computation required to find individual digits, are examples:

- These two topics clearly have genuine substance, and they support extended development.
- Unlike physical-world topics, they are directly accessible, because they concern mathematical structure that has already been extensively developed.
- The multiplication algorithm (4.3.1) does have real-world applications, even if these are not accessible to students. In any case, it is a good example of the kind of mathematical development that has applications.
- The single-digit refinement (4.3.3) is a very good illustration of a major activity in computational science: carefully exploiting structure to minimize the computation required to get a result.
- The Plan C variation (4.3.3.4) provides an introduction to numerical instability and “low-probability catastrophic failure” of algorithms. This is a major issue in approximate (decimal) computation but is completely ignored in education.
- Both projects significantly deepen understanding of the underlying mathematical structure, and develop mathematical intuition.

The main objection to mathematical applications is that, because they lack contact with real-world experience, they do not engage students. I believe this underestimates the willingness of students to engage with almost anything if they can succeed with it. Further, the more obviously nontrivial the material, the more pride and excitement they get from successful engagement.

Student success is the key, and the key to success is methods and templates carefully designed to minimize errors. In other words, methods informed by contemporary approaches to proof.

Bibliography

- [1] Hanna, G.; Barbeau, E. *Proofs as bearers of mathematical knowledge*, Mathematics Education **40** (2008) pp. 345–353.
- [2] Jaffe, A.; Quinn, F. *Theoretical Mathematics: Towards a synthesis of mathematics and theoretical physics*, Bulletin of the American Mathematical Society, (1993)
- [3] Lin, Chun-Chi *How can the game of hex be used to inspire students in learning mathematical reasoning?*, Proceedings of ICMI Study 19 (2009).
- [4] Mannila, Linda; Wallin, Solveig, *Promoting students' justification skills using structured derivations*, Proceedings of ICMI Study 19 (2009).
- [5] Peltomaki, Mia; Back, Ralph-Johan, *An Empirical Evaluation of Structured derivations in high school mathematics*, Proceedings of ICMI Study 19 (2009).
- [6] Quinn, Frank *Education web page* <http://www.math.vt.edu/people/quinn/education/>
- [7] _____ *The Nature of Contemporary Mathematics*, Draft July 2010, retrieve from [6].
- [8] _____ *Essays on Mathematics Education*, Draft July 2010, retrieve from [6].
- [9] _____ *AMS Working Group on Preparation for College Mathematics* web site, <http://amstechnicalcareers.wikidot.com>
- [10] _____ *The EduT_EX project*, Wiki at <http://www.edutex.tug.org>.
- [11] Rav, Y. *Why do we prove theorems?*, Philosophia Mathematica **7**(3), (1999) pp. 5–41.
- [12] Thurston, W. *On proof and progress in mathematics* Bull. Amer. Math. Soc. **30**, (1994) pp. 161–177.
- [13] Wanko, Jeffrey J. *talking points: experiencing deductive reasoning through puzzle discussions*, Proceedings of ICMI Study 19 (2009).

Chapter 5

Proof Projects for Teachers

October 2008

Introduction

This note¹ outlines projects for college students who may become elementary or secondary teachers. This is *experimental* in various ways: It was written to test and illustrate ideas in [5] about proofs, definitions, abstractions and mathematical methods, but it has not been tested in practice. Accordingly, it is a resource or starting point, and not intended to be used in this form.

Intended Use

Each project should be done as a unit without interruption. If students are immersed in a topic the ideas will become familiar and easy to work with. If there are interruptions then it is harder to develop this familiarity and the work will be harder. Further, in the sections that are covered all problems should be worked. Skipping material also slows development of familiarity and will make later work harder.

Problems could be worked in groups. It is a good idea to go over proofs in groups to be sure that the sense is communicated correctly, see §5.1.4.1 (Style in Short-Form Proofs).

Topics

Fractions in commutative rings are developed in §5.1. The general treatment includes polynomial fractions (rational functions) and many other things for little more effort than needed for a careful development of integer fractions. The section on Grothendieck groups, §5.1.12, provides interesting perspective:

¹Written for Fou-Lai Lin, with thanks for hospitality at ICMI Study 19, May 2009.

the same construction with additive rather than multiplicative notation, and what happens when you allow division by zero.

A formula for the area of the region enclosed by a closed piecewise linear path is developed and explored in “Area”, §5.2. The development is relatively elaborate, in part because students do not have a previous definition to compare it with. Winding numbers are used to describe the general case as “area counted with multiplicity”, and the polyhedral Jordan Curve theorem comes out of the development.

The study of area continues in §5.3. In §5.3.1 custom-made rings are used to explore possible extensions of the formula. Section 5.3.2 uses a vector and matrix-product description of the area formula to explore, among other things, how to “morph” a polygon to shift the ordering of the vertices by one place, without changing area. Finally §5.3.3 describes ways of extending a closed polygonal path to a map of the 2-disk into the plane, and interpreting the area formula as giving the (signed) area of the image.

The final section outlines a way to introduce derivatives shortly after students begin working with polynomials.

Audience Levels

This essay contains material for three different levels:

- *For Students*: descriptions or examples of materials that might be given to school children.
- *For Teachers*: comments, problems, etc. addressed to prospective school teachers (students in a “proofs for teachers” course).
- *For Educators*: addressed to higher-education faculty (instructors in a “proofs for teachers” course).

Level organization is discussed at the beginning of each section. Level shifts in text are indicated with the following somewhat clumsy method:

————— *For Educators* —————

Additional comments for educators, mainly concerned with design of learning programs that include formal definitions and proofs, are collected in §5.6.

————— *End, For Educators* —————

5.1 Fractions

Material in this section is generally addressed to prospective teachers (students in a “proofs for teachers” course).

The objective is to investigate fractions in a general context that includes both integer fractions and rational functions (fractions of polynomials). This illustrates the use of abstraction, and unexpected features to be found in good definitions of even the simplest objects.

5.1.1 Commutative Rings

We suppose R is a commutative ring. This means there are addition and multiplication operations that obey all the same basic rules as the operations in the integers. There are, of course, axiomatic formulations of these rules (commutative, distributive etc.) but they are already familiar so you can work without thinking about them explicitly. Examples are given in §5.1.2.

————— For Educators —————

5.1.1.1 Internalization of Structure

The point is that internalization of *structure* in good special cases can be transferred to the general case. In this case internalization of arithmetic structure in integers, real numbers, and polynomials will transfer to commutative rings. Many teachers are likely to be uncomfortable with this. They should be pushed to do “business as usual” and prevented from thinking explicitly about the ring axioms. Worrying about axioms increases the cognitive complexity of the development ([5], §4.3) and makes it unnecessarily difficult.

Related to this, one of the goals is to subliminally show teachers that it is students’ subliminal internalization of the formal structure of arithmetic that has long-term power, not the “meaning” or numerical outcomes.

————— End, For Educators —————

5.1.1.2 Notation Comment

A *ring* (without “commutative”) has addition and multiplication operations that obey the basic rules *except* multiplication may not commute (ab is generally different from ba). After long experience mathematicians have found that this is a more basic structure so it gets the short name. If multiplication does commute then “commutative” is added.

5.1.2 Examples of Commutative Rings

1. the Integers, \mathbf{Z}
2. the Integers modulo a number n , denoted \mathbf{Z}/n
3. polynomials with real coefficients,

$$\mathbf{R}[x] = \left\{ \sum_{i=0}^n a_i x_i \mid \text{some finite } n \text{ and } a_i \text{ real} \right\}.$$

4. Laurent polynomials (negative exponents are allowed),

$$\mathbf{R}[x, x^{-1}] = \left\{ \sum_{i=-n}^n a_i x_i \mid \text{some finite } n \text{ and } a_i \text{ real} \right\}.$$

5. formal power series (“infinite polynomials”) denoted

$$\mathbf{R}[[x]] = \left\{ \sum_{i=0}^{\infty} a_i x_i \mid a_i \text{ real} \right\}.$$

The last three are examples of general constructions: if one starts with a commutative ring R then polynomials, series etc. with coefficients in R give new commutative rings denoted $R[x]$, etc. Real-coefficient polynomials are given as the example because these are important in the study of functions and calculus but much of what we do holds more generally.

Because we are working abstractly we can study fractions in all these examples at once with the same effort needed to study integer fractions.

5.1.3 Preliminary Definition of Fractions

The key property of a fraction is that $b \frac{a}{b} = a$. We make this official:

Definition, preliminary version If a and b are in R then $\frac{a}{b}$ is a name for the solution of the equation $b \times (?) = a$.

This is “preliminary” because there are serious problems with it, see §5.1.7, and the final version has a restriction on b to ensure it makes sense. However this discussion is postponed until after some practice work with inverses.

5.1.3.1 Notation Comment

The definition says $\frac{a}{b}$ is a *name* for an object. Objects can have several names. For instance the integer fraction $\frac{1}{4}$ also has the decimal name 0.25. The fraction name encodes the equation it satisfies, just as \sqrt{a} encodes the fact that it is the nonnegative solution of the equation $x^2 = a$.

If we care about the connection to the integers then $\frac{1}{4}$ is a good name. For example $\frac{1}{4} + \frac{7}{28} + \frac{197}{23}$ is a name for a number. It has a fraction name and we would want this if we care about the connection to the integers. If we do not care about this connection then it would be easier to use decimal names.

5.1.3.2 School Presentation

The “solution of an equation” description is a fast and easy approach for people familiar with equations. However integer fractions are treated in schools long before “equations” are introduced. This may mean the approach is unsuitable for use in schools, but here is a possibility:

For Students

Fractions are defined by a *property* rather than directly. The special property of $\frac{a}{b}$ is “times b gives a ”. For example the defining property of $\frac{2}{3}$ is “times 3 gives 2” or $3 \times \frac{2}{3} = 2$.

Example Problem: Express $\frac{12}{3}$ as an integer. *Solution:* To get rid of the 3 denominator we have to find a multiplication by 3. Divide 12 by 3 to find $12 = 3 \times 4$, then

$$\frac{12}{3} = \frac{3 \times 4}{3} = 3 \times \frac{4}{3} = 4$$

Example Problem: Express $\frac{17}{3}$ as an integer plus a proper fraction (numerator smaller than denominator). *Solution:* To get rid of the 3 denominator we have to find a multiplication by 3. As above we divide 17 by 3 but now we get a remainder: $17 = 3 \times 5 + 2$. This gives

$$\frac{17}{3} = \frac{3 \times 5 + 2}{3} = 3 \times \frac{5}{3} + \frac{2}{3} = 5 + \frac{2}{3}$$

End, For Students

5.1.4 Inverses

If the equation $bx = 1$ has a solution *inside* R then it is called the *inverse* and written b^{-1} .

For example, every nonzero real or rational number has an inverse in the same ring. Examples and significance of inverses is discussed after the following problem.

Sample Problem: Fractions and Inverses: Show that if b has an inverse then $\frac{a}{b} = a b^{-1}$.

To illustrate what “show” means we give the solution.

Solution, Long form:

1. The hypothesis is that b has an inverse, which unpacks to: there is b^{-1} with $b^{-1}b = 1$.
2. The conclusion unpacks to: $\phi = a b^{-1}$ is a solution for the equation $b\phi = a$.
3. We check to see if the unpacked version of the conclusion is true for $\phi = b^{-1}$: $b(ab^{-1}) = (bb^{-1})a = (1)a = a$. The information in (1) is used in the middle step. Since the unpacked version of the conclusion is true so is the packed version, and we are finished².

“Unpacking” is described below. Note the arithmetic in the last step is business as usual, even though it is going on in some unspecified commutative ring. We do *not* spell out the axioms needed to justify it. Also, the first two steps are routine unpackings that we know in advance that we will have to do. They require a bit of care but no real thought. After that the core step (3) is easy.

Solution, Short form: According to the definition of $\frac{a}{b}$ we need to show $b(\frac{a}{b}) = a$. But this is immediate from the definition of the inverse and standard rules of arithmetic.

²There is actually something still missing, see Problem 5.1.9.1.

5.1.4.1 Style in Short-Form Proofs

Students should carefully compare the long and short forms of the proof above. The long form is the “official” version while the short form is a compressed version that can be routinely expanded to get the long form. However short-form proofs are acceptable only if they really can be expanded. This means *style* is important in short-form proofs: minor errors in calculation or use of words may cause doubt that the person who wrote the proof really could have written a valid long-form proof. Students who have trouble with style in short-form proofs should practice writing out long forms and then compress them. After some practice they should be able to write directly in the short form.

5.1.5 Unpacking Definitions

“Unpacking” is the use of definitions to translate statements to more primitive forms that can be worked with directly. Eventually the objects become familiar enough that they can be worked with directly and unpacking is no longer necessary, but until then we unpack.

For example a fraction $\frac{a}{b}$ is defined *indirectly* as a solution to an equation. Statements about fractions are unpacked by clearing denominators to remove the indirection, see steps (1) and (2) in the long-form solution above. This unpacking will be appropriate until after the exercises in §5.1.9.3, at which point it should be too routine to need explicit mention and can be “left to the reader”.

5.1.6 More about Inverses

5.1.6.1 Significance

Inverses may or may not make fractions uninteresting.

- There is not much point to decimal fractions like $\frac{4.209}{22.8888}$ because we can compute inverses (carry out the division).
- Exact real fractions like $\frac{1}{\pi}$ are useful because they retain a connection to the meaning of the number.
- The polynomial $(1-x)$ has inverse $\sum_{i=0}^{\infty} x^i$ in the formal power series ring. (Check to see that the product really is 1.) The fraction $\frac{1}{1-x}$ is usually more useful than the inverse. It defines an easily-computed function of x for $x \neq 1$, while the series form defines a function only for $|x| < 1$ and it is not easy to compute or work with.

5.1.6.2 Problem: Inverses in Standard Rings

- Determine which numbers have inverses in the integers mod n .
- Show that a real-coefficient polynomial has an inverse in the ring of polynomials if and only if it is constant and nonzero.

- (Hard) Show that a formal power series has an inverse in the formal power series ring if and only if it has nonzero constant term. This extends the example given in the previous section.

5.1.7 Difficulties with the Preliminary Version

We return to the definition of fractions. The preliminary version given above has problems:

- We need to know there *is* a solution somewhere. It is usually not in the original context.
- We need to know there is *at most one* solution.

These are very different problems. It turns out that because the name maintains a connection to R we can almost ignore the existence problem. It does have to be addressed eventually, see §5.1.10.

If there is more than one solution then it is hard to make sense of $\frac{a}{b}$ as a single thing. This has to be addressed immediately and we do that next.

5.1.8 Zero divisors

An element b in the ring is called a *zero divisor* if there is a $r \neq 0$ in R with $rb = 0$.

5.1.8.1 Problem: Zero Divisors

- Show that if b is a zero divisor then there is an element a so $b\phi = a$ has more than one solution.
- Show conversely that if there is a with more than one solution then b is a zero divisor.
- Show that bc is not a zero divisor if and only if neither b nor c is a zero divisor. Hint: unpack using the conclusions just above.

The first two points imply that “non-zero-divisor” is exactly what we need for a fraction to make sense.

5.1.9 Final version of the Definition

If b is not a zero divisor then $\frac{a}{b}$ is a name for the solution of the equation $b\phi = a$. If b is a zero divisor then $\frac{a}{b}$ not defined.

Fractions with denominator a zero divisor, for example $\frac{3}{0}$, are undefined (mathematically illegal expressions) because they genuinely don’t make sense. Trying to use them leads quickly to errors. For instance the definition requires that $\frac{3}{0} = \frac{4}{0}$, but this is a problem because $3 \neq 4$.

One can *force* division by zero to make sense, by use of an equivalence relation. Something like this is done in a slightly different context in the section

on Grothendieck groups, §5.1.12. When applied to fractions the result is disappointing, see the third problem in §5.1.12.7. This gives another explanation of why we are stuck with the don't-divide-by-zero rule.

5.1.9.1 Problem: Fractions and Zero Divisors

We now see that the proof in §5.1.4 is incomplete: in order to be sure the fraction $\frac{a}{b}$ makes sense we must verify that if an element has an inverse then it is not a zero divisor. Rewrite both the long and short forms of the proof in §5.1.4 to include this.

5.1.9.2 Problem: Zero Divisors in Standard Rings

1. Show for examples (1), (3), (4), (5) in §5.1.2 that the only zero divisor is 0.
2. Find an explicit form of the zero-divisor condition in the second example in terms of the modulus n . Compare this with the invertibility condition in §5.1.6.2.
3. Laurent polynomials allow finitely many negative and positive exponents. Series allow infinitely many positive exponents. A natural generalization is series that are infinite in both positive and negative directions. However these rings are tricky to work with because they have a lot of zero divisors. As an example show that $(1-x)\sum_{i=-\infty}^{\infty} x^i = 0$. Generalize this: if r is a nonzero real number find a bi-infinite series whose product with $(r-x)$ is zero.

5.1.9.3 Problem: Standard Fraction Facts

Here the standard fraction facts are shown to hold for fractions in any commutative ring.

1. Find a fraction expression for the sum $\frac{a}{b} + \frac{x}{y}$ (be sure to check the zero-divisor condition for the answer. Note that there is an implicit hypothesis that the fractions make sense: b and y are not zero divisors. Use this and Problem 5.1.8.1)
2. Find a fraction expression for the product $\frac{a}{b} \frac{x}{y}$.
3. Find a fraction expression for the fraction $\frac{a/b}{x/y}$. What is the condition required for this to make sense? (i.e. when is $\frac{x}{y}$ not a zero divisor?)
4. Show that if c is not a zero divisor then $\frac{ca}{cb} = \frac{a}{b}$.

Since the rules are the same, people who can work accurately with integer fractions should also work with general fractions without explicitly referring to either rules or the definition.

5.1.10 Rings of fractions

We return to the existence problem mentioned in §5.1.7: if b is not a zero divisor in R then $\frac{1}{b}$ seems to make sense, but it is not an element of R unless b has an inverse. What, or where, is it? In fact we use fractions to define a new ring.

5.1.10.1 Definition

The *ring of fractions* of a commutative ring R , denoted here by $\text{Frax}(R)$, is the set of $\frac{a}{b}$ with b not a zero divisor, with two such being equivalent if they solve the same equation $b\phi = a$.

Equivalence means, for instance, that $\frac{a}{b}$ and $\frac{ac}{bc}$ are considered the same object even though they are different symbolic expressions.

The addition and multiplication formulas 5.1.9.3 are now no longer identities for preexisting objects but actually used to *define* addition and multiplication in the ring of fractions. Strictly speaking one should verify that these operations satisfy the rules required in a commutative ring. We will not do this because they follow easily and routinely from the rules in R .

5.1.10.2 Examples

We can now recognize some standard systems as being rings of fractions.

- The rational numbers are the ring of fractions of the integers.
- The rational functions³ are the ring of fractions of the real polynomial ring.

5.1.10.3 Problem: Fractions and Zero Divisors

- Show that $\frac{a}{b}$ is a zero divisor in the ring of fractions if and only if a is a zero divisor in the original ring.
- Show that every element in the ring of fractions that is not a zero divisor has an inverse.
- Show (conversely) that if every non-zero-divisor in R has an inverse, then the natural inclusion $R \subset \text{Frax}(R)$ is a bijection.
- Describe the ring of fractions of the integers mod n (see §5.1.9.2).

The natural inclusion in the third problem is defined by $a \mapsto \frac{a}{1}$. “Bijection” means that every element in $\text{Frax}(R)$ comes from exactly one element in R .

³“Rational functions” really should be called “polynomial fractions”. They are very useful as functions, but identifying them as elements in a ring of fractions is more fundamental. Note that the formal power series ring §5.1.2 Example (5) also has a ring of fractions but these generally cannot be interpreted as functions.

5.1.11 Ring Homomorphisms

To explore the relationship between rings and their rings of fractions we need a definition. A function $f: R \rightarrow S$ between two rings is said to be a *ring homomorphism* if it preserves the multiplication and addition operations and their units:

- $f(a + b) = f(a) + f(b)$
- $f(ab) = f(a)f(b)$, and
- $f(0) = 0, f(1) = 1$.

5.1.11.1 Examples of Ring Homomorphisms

Many of the standard relationships between the examples of rings in §5.1.2, for instance mod- n reduction going from the integers to the integers mod n , are ring homomorphisms. We add a few more:

- Fix a real number r . Show that evaluation at r defines a ring homomorphism from the polynomial ring to the real numbers.
- Show that the inclusion $R \subset \text{Frax}(R)$ of a commutative ring into its ring of fractions, defined by $a \mapsto \frac{a}{1}$, is a ring homomorphism.

5.1.11.2 Naturality?

Applying a ring homomorphism f to an equation $b\phi = a$ gives $f(b)f(\phi) = f(a)$. Interpreting these as defining equations for fractions seems to show that f preserves fractions: $f(\frac{a}{b}) = \frac{f(a)}{f(b)}$. This should mean f induces a ring homomorphism on the rings of fractions: $\hat{f}: \text{Frax}(R) \rightarrow \text{Frax}(S)$. However:

- Find the error in this proposed construction.
- Find a really obvious ring homomorphism between two of the examples in §5.1.2 that does not extend to the rings of fractions.
- Fix a real number r . Determine which polynomial fractions $\frac{p(x)}{q(x)}$ do *not* give a real fraction (and therefore not a real number) when evaluated at r .

5.1.12 Grothendieck Groups

Historically, fractions were first introduced as ratios, then used to encode and work with rational numbers. It was an unexpected bonus that they gave a way to mass-produce new rings by adjoining multiplicative inverses. The Grothendieck construction uses the same idea in a simpler context: *additive* systems without additive inverses.

The most familiar example is the natural numbers, and the construction adjoins additive inverses to produce the integers. There are many other examples but they require sophisticated preparation. This is why the additive version was described so much later than the fraction construction.

5.1.12.1 Commutative Semigroups

A commutative semigroup is a set with a binary operation, denoted $+$, that is associative and commutative. Denote the set by N , then specifically:

- $a + b$ is defined for all a, b in N ;
- $a + b = b + a$; and
- $a + (b + c) = (a + b) + c$.

There is a strong convention that an operation is entitled to be denoted “ $+$ ” *only* if it has these properties. This means we can do arithmetic as usual with $+$ operations, and don’t have to explicitly think about the rules.

5.1.12.2 Examples

- The natural numbers with the standard addition operation. This is denoted by \mathbf{N} .
- The natural numbers with operation given by minimum:

$$\min(a, b) = \begin{cases} a & \text{if } a \leq b \\ b & \text{if } b \leq a \end{cases} .$$

This is sometimes called the *tropical* semigroup structure⁴.

- If R is a commutative ring then R with the multiplication operation is a commutative semigroup. Caution: the use of multiplicative notation when thinking of it as a ring, and $+$ for the same operation when thinking of it as a semigroup, is an endless source of confusion.
- The non-zero-divisors in a ring, again with multiplication as the operation, also form a semigroup. The third property in §5.1.8.1 is needed to see that this is true.

5.1.12.3 Terminology

A *group* is a set with a binary operation *with a unit element and inverses*. “Semigroup” weakens this by dropping the requirement that inverses exist. *Commutative* means that the operation is commutative, just as with rings. Mathematical experience suggests that the most fundamental object is a (possibly non-commutative) group. This therefore gets the short name, and related objects are described by modifying the name, just as with commutative rings (see §5.1.1.2).

⁴See the “tropical geometry” entry on Wikipedia.

5.1.12.4 A Difficulty

Recall that a fraction $\phi = \frac{a}{b}$ is defined to be the solution to the equation $b\phi = a$. We would like to similarly define a *difference* $\phi = (a - b)$ as the solution to the equation $b + \phi = a$, but there is a problem with this.

Recall that there are many solutions to the defining equation of a fraction if the denominator is a zero divisor. The solution was to only define fractions with non-zero-divisor denominators. The analog for addition is cancellation: $b + a = b + c$ implies that $a = c$. The strict analog of the fraction construction therefore only defines differences $a - b$ if b satisfies the cancellation condition.

We didn't mind having a condition on denominators of fractions. We do mind having a condition on negative objects (we want to be able to subtract without conditions). This requires a modification of the construction: requiring the defining equation to hold only after addition of the same term to each side⁵. Adding such a term is called "stabilization".

5.1.12.5 The Construction

Suppose $N, +$ is a commutative semigroup. $G(N, +)$ is defined to be equivalence classes of pairs of elements a, b , with equivalence classes written $[a - b]$:

- $[a - b]$ is equivalent to $[a' - b']$ if there is c so that $a + b' + c = a' + b + c$. (Think $a - b = a' - b'$, clear negative signs by adding b, b' to each side, then stabilize by c to avoid the cancellation problem).
- The operation $+$ is defined on equivalence classes by $[a - b] + [c - d] = [(a + c) - (b + d)]$.

5.1.12.6 Problem: Identities in $G(N, +)$

- Show that $+$ is well-defined: if $[a - b] \simeq [a' - b']$ then $([a - b] + [c - d]) \simeq ([a' - b'] + [c - d])$.
- Show that $[a - a] \simeq [b - b]$ for all a, b , and that this equivalence class is a unit for the operation: $[a - b] + [c - c] \simeq [a - b]$. We follow tradition by denoting the equivalence class $[a - a]$ by 0.
- Show that $[b - a]$ is an additive inverse for $[a - b]$.

The outcome is that $G(N, +)$ is a commutative group (i.e. commutative semigroup *with inverses*). The construction adjoins inverses.

⁵The analogous modification of the fraction construction is to allow multiplication by an arbitrary element. The effect is to enable division by anything, including zero divisors if there are any. The next-to-last example in §5.1.12.7 illustrates what happens.

5.1.12.7 Problem: Examples

- Show that the function $[a - b] \mapsto a - b$ defines an isomorphism⁶ from the Grothendieck group of the natural numbers with standard addition, to the integers.
- Describe the Grothendieck group of natural numbers with the tropical (min) operation.
- Suppose R is a commutative ring. Show that the Grothendieck group of R , with multiplication as the operation, is trivial (everything equal to the identity element). This semigroup is an example in §5.1.12.2 where there is a warning about notation problems with the operation. The best way to proceed is to translate the definition of $G(R, \times)$ into multiplicative notation, and see it as a modification of the definition of fractions.
- Describe the Grothendieck group of the *nonzero* integers, with multiplication as operation.

Generalize the example just above to show:

5.1.12.8 Proposition

Suppose R is a commutative ring. Then the group of invertible elements in the ring of fractions of R is the Grothendieck group of the semigroup of non-zero-divisors of R .

After unwinding the definitions you should see this as essentially obvious. The impressive-sounding statement is the result of having two terminologies for essentially the same construction.

5.2 Area

The object is to explore a formula for areas of polygonal figures in the plane, using coordinates of the vertices.

5.2.1 Polygonal closed paths

Suppose (p_0, p_1, \dots, p_n) are points in the plane \mathbf{R}^2 . The *oriented closed path* with these as vertices is obtained by joining p_i to p_{i+1} for $0 \leq i \leq n$ where, if $i = n$, we set $p_{n+1} = p_0$.

Closed means it goes back to the starting point: this is the effect of the $p_{n+1} = p_0$ convention. *Oriented* refers to the preferred direction on the path

⁶“Isomorphism” here means one-to-one and onto, and takes $+$ in one group to $+$ in the other.

coming from the order of the vertices. Problem: Draw⁷ a few of these by choosing points at random, numbering them, and then connecting them, drawing arrows to indicate direction on the edges. Do one with only two vertices.

5.2.2 The Project

Suppose P is an oriented closed path with vertices p_0, \dots, p_n . Denote the coordinates by $p_i = (x_i, y_i)$ and define

$$A(P) = \frac{1}{2} \sum_{i=0}^n (x_i y_{i+1} - x_{i+1} y_i). \quad (5.1)$$

The project is to show that if P is a closed polygonal path then $A(P)$ is the area enclosed by P . More precisely, find conditions under which this is true.

5.2.2.1 Problem: Example

If this works then it makes areas easy to compute when coordinates of the vertices are known. Find $A(P(t))$ for quadrilaterals $P(t) = ((0, 0), (1, t^2), (1 - t, 1 - t), (t^2, 1))$ when $-1 \leq t \leq 1$. Draw a few of these to see what is happening in this family. Determine the t at which $A(P(t))$ attains its maximum.

5.2.2.2 Notes

- This may seem unlikely: how can something that only uses the path give the area? Also, the definition of “path” allows self-intersections so does “enclosed” even make sense? Or is it just for simple closed paths?
- The statement can’t be right even for simple closed curves: areas should be nonnegative, but $A(P)$ can be negative: reversing the order of the vertices in P reverses the sign of $A(P)$. This has to be sorted out.
- It is useful to observe that A is defined for curves that aren’t closed. In fact it is defined for a single edge, and

$$A(P) = \sum_{e_i \text{ edges of } P} A(e_i) \quad (5.2)$$

5.2.2.3 Problem: Test Cases

To clarify what is going on, try some special cases (with pictures!). Compute both A and the area.

- A triangle⁸ with p_0 at the origin and p_1 on the positive x axis. This has vertices $((0, 0), (x_1, 0), (x_2, y_2))$ with $x_1 > 0$. Note there are several cases depending on whether (x_2, y_2) is above, on, or below the x axis.

⁷Draw on paper, with a pen or pencil. There is something about actual drawing that significantly aids learning, and students really are expected to do this.

⁸Do this symbolically. Do *not* put numbers in for x_1 , etc.

- A trapezoid (4 vertices) with p_0, p_1 as above and the segment from p_2 to p_3 horizontal.

You should see there is a sign problem: the definition needed to get the sign right is *counterclockwise orientation of the boundary*. Essentially it means that if you imagine yourself moving along the boundary in the direction specified by the orientation then the “inside” of the region is on your left⁹.

5.2.3 A Difficulty, and a Strategy

Area is not defined in school mathematics. Students are taught formulas for areas of simple figures, but these are obtained from basic examples (especially rectangles) and justified heuristically. Without a definition, or at least a reasonably general way to compute, there is almost nothing to connect with. How can we expect to relate A to area under these circumstances?

In fact we see that this is a problem in the usual development of mathematics. The first real definition of area is given in multivariable calculus: the area of a region is the double integral of the function 1 over the region. This definition makes many calculations easy, and gives the formula $A(P)$ via Stoke’s theorem. This is connected with earlier work by showing it gives the familiar answers for triangles, circles, etc. It is *not* shown to agree with an earlier *definition* of area because, of course, there wasn’t one. In essence, earlier work becomes obsolete and is discarded.

Our objective is still to connect A to area, and to do this without going through calculus. The plan is to list properties that area—however it is defined—should have. We then verify that the function A has these properties. If the properties are strong enough to completely determine the area of a polygonal region it will follow that A must be area.

5.2.3.1 Properties of Area

Area of polygonal regions should satisfy:

Invariance under rigid motions Rotation and translation do not change area;

Additivity if a region is split into two pieces then the total area is the sum of the areas of the pieces; and

Standard triangles areas of triangles with one edge on the x axis is one-half (length of base) times (height).

A list like this is always a bit dangerous: we are working blind, and we might assume more than is actually true. In that case deductions made from the assumed properties will lead to a contradiction and the whole effort will collapse. To minimize risk we try to get the job done with the weakest possible assumptions.

Eventually, by using A , we will see that area has many additional properties that we would not dare to assume when working blind.

⁹This is imprecise but good enough for the present. See §5.2.6 for a precise version.

5.2.3.2 Rough Argument for Sufficiency

We suggest why these properties should be enough to determine the area of a region. It should be possible to divide a polygonal region into triangles. Using additivity the whole area is the sum of the areas of these triangles, so we only have to show that areas of triangles are determined. Any triangle can be translated so one vertex is at the origin, and then rotated so another vertex (and therefore an edge) is on the x axis. Since area is unchanged by rigid motions, the area of general triangles is determined by areas of these special cases. But areas in these cases is specified in the standard-triangle property.

This argument is not solid. The part about determining area of general triangles is OK (i.e. essentially a proof), but subdivision of regions into triangles needs to be done carefully. There might be a subtle difficulty with this that could require restriction to special polygonal regions, for instance convex ones. However the proper next step is to see if A has the properties. The reason is that if this fails then we can conclude that A does not give areas, or the whole approach has to be modified, and we don't have to worry about the subdivision argument. If A passes then we can return to the subdivision argument.

5.2.4 Properties of A

Terminology used in the problems is explained in the following section. After that we compare with to determine exactly what remains to be done to complete the project.

5.2.4.1 Problem: Properties of A

Suppose P is a polygonal region with vertices p_i for $i = 0, \dots, n$, and $p_{n+1} = p_0$. Prove the following:

Translation Invariance If q is any point in the plane then the region $P + q$ with vertices $p_0 + q, p_1 + q, \dots$ satisfies $A(P) = A(P + q)$.

Matrix Transformations If R is a 2×2 matrix then $A(RP) = \det(R) A(P)$. See Notes below.

Additivity Suppose Q is a polygonal path beginning and ending at vertices of P , see Notes below. Denote the two regions obtained by splitting P along Q by PQ_1 and PQ_2 . Then $A(PQ_1) + A(PQ_2) = A(P)$.

Subdivision Suppose q is a point on the edge between p_i and p_{i+1} . Denote by P_q the region defined by inserting q in the vertex list between p_i and p_{i+1} . Then $A(P) = A(P_q)$.

Cyclic Permutation The region obtained by cyclically permuting the vertices, i.e. $(p_i, \dots, p_n, p_0, p_1, \dots, p_{i-1})$, has the same A value.

5.2.4.2 Notes

- In Rotation Invariance, a 2×2 matrix operates on points in the plane by $\begin{pmatrix} a & b \\ c & d \end{pmatrix}(x, y) = (ax + by, cx + dy)$. In particular, rotations do not change A because they have determinant 1. There are matrices like the shear transformations $\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}$ that are not rotations but have determinant 1 and so also don't change A . When we know that A is area it will follow that area is not changed by these either.
- In Additivity, Q is of the form $(p_i, q_1, \dots, q_k, p_j)$. Suppose $i < j$. Then the split regions are $(p_i, \dots, p_j, q_k, \dots, q_1)$ and $(p_j, \dots, p_i, q_1, \dots, q_k)$. What happens if $j < i$ or $j = i$? Hint: use the edge-sum description in equation (5.2).
- There are two points to the Subdivision and Cyclic Permutation properties. First, A is defined using a specific ordering of specific vertices. These properties show that A depends on the underlying geometric figure (and the direction on the boundary), not specific vertices. These properties will also be needed to divide regions into triangles.

5.2.4.3 Taking Stock

The function A has all the properties we could want, and more: additivity works even if the cutting curve intersects original, or if it lies *outside* the original region. This strange behavior has something to do with it being able to take negative values. In any case the precise relation to area is still unclear.

Referring back to the discussion in §5.2.3.2, we see that all the pieces are in place except for the argument about cutting a region up into triangles. In particular A is now known to give area of positively oriented triangles. Evidently the cutting argument is where the negative-value and crossing problems get sorted out. This means a logically complete version of the argument will have to be fairly complicated.

Draw some pictures to explore what can go wrong. The curve could be a wild scribble, or have lots of sharp points, or wind back and fourth like a maze puzzle, or all of these. It is hard to imagine a cutting strategy that would do a good thing in all these cases.

Instead of trying to find a strategy for cutting a region into smaller pieces we make it bigger. It is relatively easy to prove the result for convex polygons. The general simple closed case can be done by filling in concave areas until it becomes convex.

5.2.5 Orientations and Convex Polygons

We develop ideas that lead to definitions of both convexity and orientation. These are used to prove that A gives the area of the region enclosed by a positively oriented *simple* (i.e. no self-intersections) polygon. The actual result is weaker than it sounds because we don't have a good definition for "region

enclosed”. Eventually this will be provided by winding numbers but here we use a more elementary, and much more cumbersome, approach.

The standard definition of a convex *region* is that if two points are in the region then the straight line between them is also in the region. Here we would want to apply this to the region enclosed by the curve. However we can’t do that because—again—we don’t have a criterion for which points are inside and which are outside. Again we use an approach special to the situation.

5.2.5.1 Extremal Vertices and Edges

Suppose P is a closed polygonal path. A vertex is called *extremal* if there is a bi-infinite straight line that intersects the polygon in exactly this vertex. A line is called an *extremal line* if it intersects the polygon in at least two points and there are no vertices on one side of the line.

For example suppose y_i is the smallest y coordinate of any vertex. The horizontal line (t, y_i) for $t \in \mathbf{R}$ passes through the vertex p_i and there are no vertices below this line. If p_i is the only vertex on the line then p_i is an extremal vertex. If there are other vertices on the line then this is an extremal line.

The outermost vertices on an extremal line are extremal vertices. For instance take the lowest vertex. Rotate the line about this vertex toward the side with no vertices. This moves the line away from the later vertices, but does not immediately introduce new intersections. The slightly rotated line therefore identifies the vertex as extremal.

5.2.5.2 Problem: Example

Draw a complicated, jagged, closed polygon. Locate the extremal vertices and draw the extremal lines with a ruler.

5.2.5.3 The Convex Hull

A new polygon can be constructed from the extremal vertices by ordering them so adjacent ones lie on an extremal line. This is called the *convex hull* of the polygon, and is the smallest convex set containing the polygon. There are two orderings that do this (up to cyclic permutation) and we want to specify one as being “positive”.

Obtain particular extremal lines by translating and rotating the x axis so that the polygon lies in the translated upper half plane (nonnegative y coordinates). Each such extremal line contains two extremal vertices: the ones corresponding to minimal and maximal x coordinate. Order the vertices by placing the one with maximal x coordinate just after the minimal one. Then this is the positive (or counterclockwise) orientation of the convex hull.

Referring back to the test cases in §5.2.2.3, we see that all the vertices on the triangles and quadrilaterals there are extremal. The positively oriented examples are the ones with nonnegative y coordinates, while the others are negatively oriented (opposite of positive). Note that this definition of orientation

exactly determines the sign of A for these examples. In particular if P is a positively oriented triangle then $A(P)$ is the area of P .

5.2.5.4 Problem: Areas of Convex Polygons

Define a polygon to be *convex* if all vertices are extremal. Show that if P is a positively oriented convex closed polygon then $A(P)$ is the area of the region enclosed by P .

To do this choose a vertex (say p_0) and consider the triangles (p_0, p_i, p_{i+1}) . Show that each of these is positively oriented and their union is P . Then use additivity.

5.2.6 Simple Closed Polygons

A closed polygon is called *simple* if no edge intersects another edge or vertex except at its endpoints.

We can define “positive orientation” for simple polygons: suppose e_i and e_{i+1} are adjacent extremal vertices, ordered using the positive orientation of the convex hull. Denote the line through these points by E . P splits into two polygonal paths from e_i to e_{i+1} , both contained in the half-plane on one side of E . Since there are no self-intersections in P one of these is contained inside the region bounded by the other and E . We define P to be positively oriented if e_i is before e_{i+1} in the ordering on the *inner* path.

We postpone the argument that there really is an “inner” and “outer” path. It is also true that the inner path cannot contain any other extremal vertices (a line through a vertex has to intersect the outer curve). If P is positively oriented then this implies the ordering on the vertices of P gives the natural positive order on the extremal vertices, and this in turn implies that the definition of positive does not depend on which pair of extremal vertices are used. All these things can be proved with the techniques at hand, but somewhat awkwardly. A short clear proof will be possible after winding numbers are developed.

5.2.6.1 Proposition: Areas of Simple Polygons

Show that if P is a simple closed polygon then the area of P is $(-1)^k A(P)$, where $k = 0$ if P is positively oriented, $k = 1$ otherwise.

It is sufficient to show that $A(P)$ is the area if P is positively oriented, because reversing orientation changes the sign of A .

Proceed by induction on the number of vertices of P , starting with triangles ($n = 3$). The induction step is: suppose the proposition is true for all simple polygons with fewer than n vertices, and suppose P has exactly n . Then show that the proposition is also true for P . Note that if P is not convex then there are adjacent extremal vertices e_i and e_{i+1} so that the edge between them is not a union of edges of P . There is a path in P between these vertices that does not contain any other extremal edges. Define Q by replacing this path in P with

the direct edge between e_i and e_{i+1} . Observe that Q has fewer vertices than P , and argue that P is obtained from Q by cutting along the replaced path.

5.2.7 Winding Numbers

Winding numbers count how many times a curve in the plane “goes around” various points. Winding numbers are the starting point for remarkable developments in analysis, topology and other areas, and provide a complete understanding of how A relates to area.

The rough idea is that a general polygonal curve (with intersections) does, in a sense, enclose a region. But this region can overlap itself, and—due to folding—can have pieces with negative orientation and therefore negative “area”. The function A gives the area when counted with multiplicity of overlaps, and winding numbers give the multiplicity. The formal statement is given in §5.2.7.5, but properties of winding numbers must be developed first.

5.2.7.1 Definition

Suppose P is a polygonal path, a is a point not on P , and R is a ray beginning at a that does not pass through any of the vertices of P . The *winding number* of P around a is the number of intersection points of R with edges of P , *counted with signs*.

Signs are assigned to intersection points as follows. Represent the ray as obtained from the positive x axis by translation and rotation. The translation and rotation act on the whole plane, and we can define the *positive side* of the ray to be the image of the upper half plane. Next recall that edges have a preferred direction coming from the order on the vertices. Define an intersection point to have sign $+1$ if the edge direction goes from the negative to the positive side of the ray. Sign -1 corresponds to a positive-to-negative crossing.

Note there is an intersection associated with each crossing edge. It might be that multiple edges cross at the same point on R , in which case this point is counted multiple times.

5.2.7.2 Problem: Draw an Example

Let R be the positive x axis, thought of as a ray beginning at 0. Draw a complicated closed curve that winds around 0 but with occasional changes in direction so some crossings of R are positive and others are negative. Then trace along the curve in the preferred direction and record the direction at each crossing point with an arrow. Add to determine the winding number.

5.2.7.3 Problem: Winding Numbers are Well-Defined

Show:

1. Any ray starting at a and missing the vertices of P gives the same winding number.

2. If a and b are joined by a polygonal path disjoint from P then P has the same winding numbers around a and b .

These are consequences of:

Lemma: suppose $R(t)$ is a 1-parameter family of rays with origins disjoint from P . Then $R(0)$ and $R(1)$ give the same winding number.

In part (1) of the problem, a 1-parameter family can be obtained by rotating about the point a . In part (2), start with a ray at a and use translations to move the whole ray so that the endpoint moves along the path. Caution: the statements in the problem are a bit imprecise. Be very precise about what these arguments actually prove, and show how they fit together to give the conclusion.

To prove the lemma, *assume* that there are only finitely many values of t for which $R(t)$ contains vertices of P . For a random 1-parameter family this may not be true. It is true for the families actually used (rotations and translations) and can be arranged without much difficulty in general, but there is not much benefit to going into detail here. Then:

- Argue that on an interval of t values that does not contain a vertex intersection, the number doesn't change. This should be easy.
- Argue that the number also doesn't change when the parameter passes through a value with vertex intersections. Do this with pictures of the various ways edges can enter and exit a vertex, and how a ray could sweep through the picture. Don't get too formal, but be sure you have all possibilities represented. Caution: the ray could contain an edge of P !

5.2.7.4 The Polygonal Jordan Curve Theorem

The polygonal path property in part (2) of the problem above inspires the following definition: A region is said to be *connected* if any two points can be joined by a polygonal path. When applied to complementary regions this means a path disjoint from the original curve.

Problem. Show that a *simple* closed polygonal curve divides the plane into exactly two connected regions:

- a (unbounded) region in which the winding number is 0; and
- a region in which the winding number is 1 if P is positively oriented, -1 otherwise.

To prove this construct polygonal paths by going along a ray to the first intersection point with P , then following along beside P to an intersection point with another ray. Evaluate the winding number by starting near an extremal vertex.

The general Jordan Curve Theorem asserts that a *continuous* simple closed curve divides the plane into two connected regions. The proof for continuous curves is much more difficult than the polygonal case.

5.2.7.5 Theorem

A closed polygonal curve P divides the plane into a finite number of connected subregions. $A(P)$ is the sum over these, of the area of the subregion times the winding number of P around a point in the subregion.

According to part (2) of Problem 5.2.7.3, the winding number is constant on a connected region, so it doesn't matter which point in the subregion is used.

To organize the proof, introduce the notation $W(P)$ for the weighted area sum.

- Both $A(P)$ and $W(P)$ are defined for *collections* of closed curves, not just single closed curves. We use this in a cutting argument that splits curves into pieces. (Another approach is given in §??).
- Both are additive with respect to cutting and unions of multiple curves.
- By quoting previous results we can conclude that they are the same for simple closed P .

There are several ways to complete the proof from this point. One approach is by induction on the number of complementary regions, for a single closed curve. The induction starts with two complementary regions (the simple closed case) because we know A and W agree for these.

For the induction step suppose the statement is known for n or fewer regions, and suppose P has $n+1$. Consider a segment of P between two self-intersection points. Cut along this segment to convert P into two closed curves. Together these have the same complementary regions as P (because the segment is in P), and winding numbers for the union are the same as for P . Therefore neither A nor W is changed by this. However we now have two pieces which must both be "smaller" than P , so the induction hypothesis applies to both pieces.

This needs refinement. If P traces over itself several times then the cutting procedure can give pieces with the same complementary regions. To fix this we need something more subtle than a simple count of regions. Try using a weighted count: the sum of the absolute values of the winding numbers. If that doesn't work try inducting on the number of self-intersections. However beware that it may trace over itself somewhere, so have segments of self-intersection and not just isolated points.

5.3 More About Area

This section gives explorations and elaborations.

5.3.1 Area and Rings

In this section and the next we try to get more elaborate versions of area by using the same formula in different rings. One works, one doesn't.

5.3.1.1 Complex Area?

Think of the plane \mathbf{R}^2 as complex numbers, with $p = (x, y)$ thought of as $p = x + iy$.

Now suppose $p_k = x_k + iy_k$ are vertices in a polyhedral path. The cross term $x_k y_{k+1} - x_{k+1} y_k$ in the definition of $A(P)$ is very nearly the imaginary part of the complex product $p_k p_{k+1}$. To get the sign on the second term right, recall that complex conjugation is defined by $\overline{x + iy} = x - iy$. Then $x_k y_{k+1} - x_{k+1} y_k = \text{Im}(\overline{p_k} p_{k+1})$. Define

$$A_{cx}(P) = \sum_{k=0}^n \overline{p_k} p_{k+1},$$

then the previous definition is the imaginary part, $A(P) = \text{Im}(A_{cx}(P))$.

Problem Show that the real part of $A_{cx}(P)$ is not invariant under some area-preserving transformations (i.e find one). Consequently A_{cx} does not qualify as a generalized area.

5.3.1.2 Polygons in Motion?

Suppose we have a one-parameter family of polygonal curves $P(t) = (p_0(t), \dots, p_n(t))$ defined for t in an interval (a, b) . The objective is to apply the formula defining A to both coordinates and derivatives of coordinates, to see if this encodes something useful. The first step is to construct a ring where the invariant will be defined.

The ring $\mathbf{R}[\delta]/(\delta^2 = 0)$ Extend the real numbers by adjoining δ with $\delta^2 = 0$ rather than $i^2 = -1$ as in the complex numbers¹⁰. More precisely, the multiplication in this ring is given by

$$(a + \delta b)(x + \delta y) = ax + \delta(ay + bx).$$

Denote this ring by $\mathbf{R}[\delta]/(\delta^2 = 0)$.

Problem: Zero Divisors Find the zero divisors (see §5.1.8) in $\mathbf{R}[\delta]/(\delta^2 = 0)$. We know there is one because we put it there ($\delta^2 = 0$).

Encoding Derivatives Suppose $f(t)$ is a real-valued function defined and differentiable on an interval (a, b) . Define $f^\delta: (a, b) \rightarrow \mathbf{R}[\delta]/(\delta^2 = 0)$ by $f^\delta(t) = f(t) + \delta Df(t)$, where Df denotes the derivative.

Problem: Products. Show that if f, g are both defined and differentiable on (a, b) then $(f \times g)^\delta = f^\delta \times g^\delta$.

Clarification of notation: $(f \times g)^\delta$ is the δ construction applied to the ordinary product of real-valued functions. $f^\delta \times g^\delta$ is the product in the ring $\mathbf{R}[\delta]/(\delta^2 = 0)$. Multiplication is written explicitly (i.e $f \times g$ rather than fg) to avoid confusion with composition of functions. Unpack carefully.

¹⁰The complex number i is sometimes written $\sqrt{-1}$. Could the “number” δ be written $\sqrt{0}$?

Definition of A^δ . Now suppose $P(t) = (p_0(t), \dots, p_n(t))$ is a family of polygons defined for t in (a, b) , and suppose that all the coordinate functions are differentiable. Define $A^\delta(P)$ by using the formula (5.1) used to define A , but in the ring $\mathbf{R}[\delta]/(\delta^2 = 0)$ using the extended coordinate functions $x_i^\delta(t), y_i^\delta(t)$.

Problem: Example. Find $A^\delta(P)$ (as a function of t) for the example in §5.2.2.1. As an intermediate step write out the extended polygon $P^\delta(t)$.

Problem: Describe A^δ . Show that if $P(t)$ is a differentiable one-parameter family of polygons then

$$A^\delta(P)(t) = A(P(t)) + \delta D(A(P))(t).$$

In words, this means that when we work in the ring $\mathbf{R}[\delta]/(\delta^2 = 0)$ the formula for A gives the derivative of the area as well as the area itself.

5.3.2 Area and Vectors

Here we shift point of view and express a polygonal path as a pair of vectors. See §?? for an account of where this came from.

Instead of thinking of a sequence of pairs of numbers $(x_1, y_1), \dots$ we could think of it first as a $n \times 2$ matrix

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix}$$

and then as two vectors $X = (x_1, x_2, \dots)$ and $Y = (y_1, y_2, \dots)$ of the same length. These should be thought of as column vectors ($n \times 1$ matrices) but to save space they are usually written as rows.

5.3.2.1 A as a Product

Separate the two terms in the definition of A , 5.1, shift the index of one, recombine and factor out the x term:

$$\begin{aligned} 1/2 \sum_{i=0}^n (x_i y_{i+1} - x_{i+1} y_i) &= 1/2 (\sum_{i=0}^n x_i y_{i+1} - \sum_{i=0}^n x_{i+1} y_i) \\ &= 1/2 (\sum_{i=0}^n x_i y_{i+1} - \sum_{i=0}^n x_i y_{i-1}) \\ &= 1/2 \sum_{i=0}^n x_i (y_{i+1} - y_{i-1}) \end{aligned}$$

The last expression is a dot product of the vector X and something obtained by shifting the coordinates of Y . To express this last we need a notation.

Define the *Right Rotation* of a vector by $R(y_1, y_2, \dots, y_n) = (y_n, y_1, y_2, \dots, y_{n-1})$. This is represented by multiplication by the $n \times n$ matrix with 1s just below the diagonal, a 1 in the upper right corner, and all other entries 0. We therefore usually write the function as a matrix product (indicated by a dot), $R \cdot Y$. Again note Y is considered as an $n \times 1$ matrix.

Problem Show that the left rotation R^{-1} is the transpose R^t .

We can therefore write the vector with i^{th} coordinate $y_{i+1} - y_{i-1}$ as $(R^{-1} - R) \cdot Y$. With all this in hand we get

$$A(X, Y) = X^t \cdot (R^{-1} - R) \cdot Y$$

where $A(X, Y)$ denotes (no surprise) the function A applied to the polygonal path with the indicated vertices.

5.3.2.2 Caution about the Transposition of X

The X is transposed to make it a $1 \times n$ matrix so the product is defined. This is quite important. For example if we rotate both X and Y one place to the right then we get the same path but starting at a different vertex. The area is unchanged and this should be visible from the formula. Begin with routine expansion of the definitions:

$$\begin{aligned} A(R \cdot X, R \cdot Y) &= (R \cdot X)^t \cdot (R^{-1} - R) \cdot R \cdot Y \\ &= X^t \cdot R^t \cdot (R^{-1} - R) \cdot R \cdot Y \end{aligned}$$

But $R^t = R^{-1}$ so the middle terms reduce to $(R^{-1} - R)$ and we get the expression for $A(X, Y)$.

There is a caution about this argument. We have both R and R^{-1} in the expression and the idea is to cancel them. *However* matrix multiplication usually doesn't commute so we can't move the R^{-1} past the $(R^{-1} - R)$ just by general principles. You can see that the formula works anyway by multiplying it all out.

There is a shortcut: it is easily seen that powers of a single matrix (including negative powers if they exist) *do* commute with each other. In this case only powers of R are involved so in fact the product $R^{-1} \cdot (R^{-1} - R)$ *does* commute. Therefore in this case we can just commute the R^{-1} and R to be adjacent and cancel them. It is simpler to see the identity this way, and this sort of thing will be useful when we have to deal with bigger products.

Warning: if you use this fact you *must* say "because powers of a matrix commute...". Otherwise it will look like a common error (forgetting matrix multiplication almost never commutes) and *should* be counted wrong: this error is so dangerous it must be caught every time.

The point demonstrated here is that we get cancellation because the transpose on X^t changes R to R^{-1} . It also reverses the order of the product (this is a property of transposition and is *not* commuting). If we had been sloppy and omitted the transpose then the expression would be wrong. The biggest danger with something like this is that it is often not clear what the problem is. If we can't find anything interesting, is it because there is nothing interesting to find, or because an incorrect expression can't find it? This is less trouble for students because they can check the answer, but it is a constant hazard for mathematicians. See §??.

5.3.2.3 Area-Killing Vectors

The question here is: are there “area-killing vectors” Y so that $A(X, Y) = 0$ for *all* X ?

The standard inner product $X^t \cdot Y$ is “nondegenerate” in the sense that $X^t \cdot Z = 0$ for *all* X only if $Z = 0$. Therefore the killer vectors are ones with $(R^{-1} - R) \cdot Y = 0$, or in other words with $R^{-1} \cdot Y = R \cdot Y$.

- Interpret this as a relationship between entries of Y .
- Show that if Y has *odd* length then all entries of Y are the same. If all y coordinates of points in P are the same then they lie on a horizontal line. It is certainly reasonable that the area should be 0 no matter what the x coordinates.
- Argue, using invariance under rotation, that any P whose vertices lie on a line must have zero area.
- Figure out what $R^{-1} \cdot Y = R \cdot Y$ implies if Y has *even* length.
- Interpret the result in terms of horizontal lines, then get a more general statement by rotation.
- I didn’t expect this outcome, and had to draw some pictures to feel comfortable with it. Do this, starting with Y of length 4.

5.3.2.4 Bilinearity, and Midpoint Polygons

An immediate consequence of the product formulation is that $A(X, Y)$ is linear in both X and Y . Explicitly this means

$$A(r_0 X_0 + r_1 X_1, Y) = r_0 A(X_0, Y) + r_1 A(X_1, Y)$$

(real coefficients r_i) and similarly for Y . If we take combinations of X_0, X_1 and of Y_0, Y_1 we get a 4-parameter family of polygons, and a formula for their areas in terms of the four basic ones ((X_0, Y_0) , (X_0, Y_1) etc.). This is a little too free-form to be really interesting, so we explore a classical formula where the various polygons are closely related.

Suppose $P = (X, Y)$ is a polygonal path. The *midpoint polygon* of P has vertices the midpoints of the edges. Explicitly, the i^{th} vertex is $\frac{1}{2}((x_i + x_{i-1}, y_i + y_{i-1}))$. In vector notation this becomes $\frac{1}{2}(X + R \cdot X, Y + R \cdot Y)$. We can also write $X + R \cdot X$ as $(I + R) \cdot X$, where I is the identity matrix.

1. Show that A of the midpoint polygon is given (up to a constant) by

$$X^t \cdot (I + R^{-1}) \cdot (R^{-1} - R) \cdot (I + R) \cdot Y.$$

In particular find the constant, and justify the form of the first R term.

2. Show that the product of the R terms in this expression is equal to

$$2(R^{-1} - R) + (R^{-2} - R^2).$$

Rather than multiply it all out, commute the first and second terms (*don't* forget to write the magic words for this) and just multiply out the first and third.

3. Interpret $X^t \cdot (R^{-2} - R^2) \cdot Y$ as A of a polygon with the same vertices as the original, but used in a different order. (This is called the “skip” polygon. If there are an even number of vertices it is actually two polygons.)
4. Put these together to get a formula for the area of the midpoint polygon in terms of the areas of the original and the skip polygon.

Is the use of the midpoint (i.e. coefficients $\frac{1}{2}$) really essential in this formula? Try other combinations of the polygon and its rotation:

1. Expand A of $r(X, Y) + s(R \cdot X, R \cdot Y)$, and simplify as above.
2. For which values of r, s can this be expressed in terms of the areas of the original polygon and the skip polygon, and what is the expression?

5.3.2.5 Zero–Area Skip Polygons

In §5.3.2.3 we found criteria for polygons to have zero area. We use the coordinate-free versions, e.g. “all vertices lie on a line”.

1. Use these to find conditions on a polygon that ensure the associated skip polygon has zero area.
2. The most interesting case is when the number of vertices is divisible by 4, though the case $n = 4$ is pretty trivial. Use the criterion to draw some polygons with 8 vertices with zero–area skip polygons.

5.3.2.6 Morphing

We want to “morph” P to $R(P)$ without changing the area. Then the formula at the end of §5.3.2.4 for linear combinations is

$$A(rP + sR(P)) = (r^2 + s^2)A(P) + rs A(\text{skip}P) \quad (5.3)$$

If we want this equal to $A(P)$ it becomes

$$(r^2 + s^2)A + rs A_{\text{skip}} = A$$

and the question is whether we can go from $(r, s) = (1, 0)$ to $(0, 1)$ in the set of (r, s) that satisfy this equation.

This breaks into two cases: $A = 0$ and $A \neq 0$.

If $A = 0$ we need $r = 0$ or $s = 0$, so the solution set is the union of the coordinate axes. We can go from $(1, 0)$ to $(0, 1)$ in this set, but note we have to go through $(0, 0)$ to do it. In other words the morph shrinks P to a point and then expands it back out to $R(P)$.

Now suppose $A \neq 0$, and define $k = A_{\text{skip}}/A$. The equation becomes:

$$(r^2 + s^2) + krs = 1$$

Problem. Describe this solution set for various values of k , and in particular find values where qualitative behavior changes:

1. For which k are there two different paths from $(1, 0)$ to $(0, 1)$?
2. Is there a value for which there is no morphing path?
3. Are there any restrictions on the values of k coming from actual polygons?

The last question requires constructing examples. It is not worth spending a whole lot of time on, so skip it if you don't see what to do reasonably quickly.

If you have access to software that displays polygons, do the following:

1. Explicitly parameterize part of the solution set to get a morphing path (Use polar coordinates and parameterize by the angle from the x axis).
2. Watch various examples morph and see if you can qualitatively describe any of the behavior, e.g. in terms of k .
3. You will see that some examples move around, and that the movement depends on location (translate to see this). The reason is that the polygons that can be obtained by linear combination depend on position, so these morphs go through different spaces. It would take us too far afield to explore this but two directions might be:
 - Is there a special location with a nice morph (something like “centroid at the origin”)?
 - If we enlarge the space available for morphing to something like $rP + sR(P) + (u, v)$, with four free parameters, then adjusting parameters could compensate for different positions. However the level set for constant area is then three-dimensional. The nice thing about a one-dimensional level set is that there is a more-or-less unique morphing path. How could a “nice” path through the three-dimensional set be chosen?

5.4 Derivatives

This is a first draft of a proposal to introduce derivatives early in the school curriculum, about the same time polynomials are introduced. There are two

5.5 Comments for Students

5.5.1 Route to a Formula

The Area and Vectors material in §5.3.2 is not deep but I stumbled on it accidentally and inefficiently. The story may be reassuring to students.

I heard about the midpoint polygon area formula in a lecture by Thomas Banchoff on use of dynamic geometry software. He illustrated how students might test and (laboriously) prove the result geometrically for convex polygons. It seemed a good exercise for the approach here, though I have since learned that in his course he also gave an analytic treatment similar to this one.

First I wrote out the expression (5.1) explicitly using the average description of the midpoints. The algebra was a mess but it did work out. I also tried it for other points on the edges: things of the form $tx_i + (1-t)x_{i-1}$, not just the midpoint $t = \frac{1}{2}$. This worked too, though it took a while to get the algebra right.

I started looking for a better way to organize it. It requires combining x and y coordinates in different ways, which seemed unnatural geometrically. One could (using rotations) think of the coordinates as two orthogonal projections rather than the standard coordinates. The complex numbers have specific orthogonal projections built in, via complex conjugation. I thought this might be a clue that the complex formulation in §5.3.1.1 might be a good setting. I spent a fair amount of time trying to see something good in this, and failed. I might have quit too soon, though, see [2].

I went back to looking for patterns in the sum expression. I had previously noticed the reorganization and factoring used in §5.3.2.1 but hadn't thought anything of it because I assumed it would be a bad idea to separate the coordinates (e.g. into vectors X and Y). This time I knew separation had to be part of the story. This freed me to recognize the expression as a dot product. It did not take long to get the tidy description using matrix products and the rotation matrices.

The matrix formulation showed that A defines an anti-symmetric bilinear function (anti-symmetric means $A(X, Y) = -A(Y, X)$, which happens here essentially because the orientation gets reversed). A standard question about such things is degeneracy: how many “area-killing” Y there are, in the sense of §5.3.2.3. Describing solutions of $(R^{-1} - R) \cdot Y = 0$ is a linear algebra problem.

The intelligence-free approach to linear algebra is to write out the matrix for $(R^{-1} - R)$ and use row operations to reduce it to row-echelon form. I did this. It took a while but I did it. When I got the solutions it was immediately obvious that they could have been found much faster using the rotation argument sketched in §5.3.2.3. I was a little annoyed with myself for not having seen the slick argument sooner (I don't really enjoy row operations), but this kind of thing happens all the time in research.

Next I wondered about how area changes as P is moved to $R(P)$ through linear combinations. The first observation was that if the skip polygon has zero area then going along an arc of the circle $r^2 + s^2 = 1$ (in the notation

of §5.3.2.6) doesn't change area. This is the case $k = 0$ in §5.3.2.6. I tried to explore other cases by using an explicit parameterization with polar coordinates: $(r, s) = (\rho(\theta) \cos(\theta), \rho(\theta) \sin(\theta))$. This quickly got complicated, and the parameterizations obscured what was going on. After a while I gave this up and developed the non-parametric version that appears in the text.

The point is that forty years experience as a research mathematician did not prevent me from making mistakes and straying into blind alleys, even in elementary material. The experience did, however, enable me to spot mistakes and blind alleys fairly quickly so I could correct them, or try something different. I believe this illustrates the best goals in learning mathematics:

- *Not* to never make mistakes, but to learn to watch for and recognize them, and then to fix them.
- *Not* to always see what to do, especially if it is clever, but to learn to try things and watch for clues that they could be improved or are unproductive and should be abandoned.

5.6 Comments for Educators

These are comments for developers and instructors of courses for prospective school teachers. They mainly concern the use of formal definitions, and associated proof strategies.

5.6.1 Formal Definitions and Unpacking

The use of formal definitions, and the unpacking–packing routine described in §5.1.5 has been standard practice in professional mathematics for over a century. This practice is a compromise between the way people think and the requirements of mathematics:

- As a practical matter, people have to work more-or-less intuitively with conceptual units.
- Mathematical success requires complete reliability.

The challenge, therefore, is to find ways to develop completely reliable intuition. Explicitly and consciously unpacking formal definitions while developing derived properties seems to work pretty well. In fact the effectiveness of this process was probably a key factor in the explosive growth in scope and complexity of mathematics in the last century

The usual routine is: when a definition is introduced, work explicitly with it for a while, typically by deriving secondary properties. After a certain point you should wean yourself (or your students) from the unpacking routine and rely more on intuition and secondary properties. If the intuition is not ready, unpack a little longer.

There are further general comments following an analysis of the projects.

5.6.2 The Fraction Project

The structure of the first project is designed to make the definition–unpacking routine as easy as possible when it is needed, and avoid it when this can be done safely.

5.6.2.1 Commutative Rings

The development takes place in commutative rings. There is a formal definition of these rings and in principle students could go through the formal definition–unpacking routine to develop familiarity with them. However the rules of arithmetic are essentially the same as for integers and real numbers, and students already have fully reliable intuition for these rules. The unpacking routine is not needed, and going through it would increase complexity without any real benefit.

This is why the chapter opens with “There are, of course, axiomatic formulations of these rules (commutative, distributive etc.) but they are already familiar so you can work without thinking about them explicitly.” This is also why this setting was chosen for the development.

5.6.2.2 Fractions

The definition of fractions is introduced in preliminary form in §5.1.3 and with a subtle problem addressed in §5.1.9. The preliminary–final division is used to call attention to the role of zero divisors, and provide an opportunity to develop these to the point that they can be worked with intuitively before tackling the subtle version.

The solution-of-equation definition is how fractions are defined in most texts for teachers, see McCrory [3, 4] for a discussion. It is effective in considerably more generality than originally intended, and as a bonus provides a way to manufacture new rings from old (rings of fractions, §5.1.10). Benefits like this are usually taken to mean that the definition is mathematically “right”.

This definition should be contrasted with the ones proposed for student use, even by mathematicians. The description in Wu [7] is so diffuse it is hard to know where it starts and ends, and even if students get an “understanding” they certainly cannot work with it with any precision. In mathematical terms it is a roadblock rather than a gateway.

Students should continue to unpack the definition of fraction up through the demonstration of the standard facts in §5.1.9.3. Afterwards they should have internalized the idea accurately enough to skip most of the unpacking in the problems on rings of fractions, §5.1.10.1.

5.6.2.3 Inverses and Zero Divisors

These concepts are introduced for use in the development of fractions. They are not particularly tricky, and should not be hard to internalize.

Inverses are introduced in §5.1.4. This is not a totally new concept, but student’s preexisting intuitions may not be sufficiently reliable so we go through the development process anyway. We also want to illustrate the development process itself, and previous familiarity with inverses helps with this. However after identifying inverses in standard rings in §5.1.6.2, intuition should be developed enough for general use.

Zero divisors are introduced in §5.1.8. They should be unpacked in Problem 5.1.9.1. Unpacking may or may not be necessary in the description of zero divisors in standard rings, §5.1.9.2, and should be unnecessary after that.

5.6.2.4 Rings of Fractions

Fraction rings are introduced to encourage students to expand their understanding of the fraction concept. The original definition is for a *single* fraction. Here the focus is on the *set of all fractions* and the definition becomes a procedure for producing this set. This is a change of perspective more than a new definition.

Fraction rings explain where fractions are defined. A solid answer, in other words, to “what is a fraction?”. Unfortunately the answer is “a fraction is an element of the set of all fractions”. Answers of this form sound—and usually are—stupid, but here it turns out to be profound.

Sometimes changes of perspective require as much practice and unpacking as genuinely new concepts. This will vary from student to student and can be a difficulty when students are working in groups. Students who “get it” are sometimes impatient with those who don’t.

5.6.2.5 Ring Homomorphisms

These are also largely a change of perspective. A full definition is given and students should unpack it when working through the examples. After that, however, they should realize that they have already worked with many examples and should be able to relax about details.

One objective is to connect with something they know. Polynomial fractions, when regarded as functions of a real variable, may not be defined at some values of the variable. This is now seen as an instance of a general phenomenon: a ring homomorphism (evaluation at a number) that does not take a fraction to a fraction because the denominator becomes a zero divisor.

5.6.2.6 Grothendieck Groups

This is again a concept–broadening change of perspective. There are two novelties: working with additive rather than multiplicative notation, and using brute force to fix the zero–divisor problem.

Changing notation (here from \times to $+$) causes serious conceptual dislocation even though it is mathematically inessential. To accommodate this we essentially repeat the development, starting with the definition of commutative semigroup. This should be unpacked while working with the examples in

§5.1.12.2, but this should be enough to establish good contact with previous experience.

The zero-divisor problem for fractions described in §5.1.7 is revisited in additive notation in §5.1.12.4, and the alternative fix is described.

Students may vary widely in their need to unpack the definition of Grothendieck group when verifying the identities and working through the examples.

The connection with fractions is nailed down firmly in the last two examples in Problem 5.1.12.7, and Proposition 5.1.12.8, by applying the construction to the multiplicative operation in a ring. To avoid notational confusion students should translate the whole Grothendieck definition to multiplicative notation. This should broaden the students' perspective on notation and the nature of mathematical operations as well as the fraction construction.

5.6.3 The Area Project

The main objective of the area project is to show how working without a definition makes a subject more difficult and limits what can be done. Conversely, a definition—or even a good formula—can be quite powerful and can open up rich and unexpected possibilities.

It may be that a definition or effective formula for area is impossible at the school level. In that case it seems particularly important that *teachers* realize that something important is missing.

- If a student has trouble understanding area it may be because the treatment is defective. Teachers should be able to draw on a deeper understanding for explanations, not just repeat something that did not work.
- Small changes in presentation or teacher attitude may help students make the transition to better treatments in later courses.

5.6.3.1 Generalized Areas

The discussion of complex area in §??, and “dynamic” area of parameterized families in §?? have several objectives:

- To suggest that definitions and formulas are not fixed and static, but can be a starting point for exploration.
- To emphasize that there is nothing disgraceful about an exploration that is unsuccessful. It should at least shed light on what makes the successful versions work.
- To re-enforce the idea implicit in the fraction project that flexibility about number systems (rings) can be very useful. Here we see that rings with special properties can be designed to test or enable extensions of a formula.

5.6.3.2 Infinitesimals

As an aside we mention that the ring $\mathbf{R}[\delta]/(\delta^2 = 0)$ used in §?? is related to old attempts to define derivatives using infinitesimals.

Think of δ as a very small number rather than a formal symbol. The derivative $f'(t)$ is then approximately $\frac{f(t+\delta)-f(t)}{\delta}$. The formula for $f^\delta(t)$ in §?? then becomes

$$f^\delta(t) = f(t) + \delta f'(t) \simeq f(t) + \delta \frac{f(t+\delta) - f(t)}{\delta} = f(t + \delta).$$

This is an attractive heuristic formulation but turns out to be unsatisfactory as a definition. In particular there are severe difficulties getting a precise meaning for “approximately”, and as a result the infinitesimal approach has some of the same drawbacks as a heuristic definition of area.

Historically, limits ($\delta \rightarrow 0$) were discovered to provide a powerful and flexible definition for the derivative. These replaced infinitesimals in professional mathematics well over a century ago. However taking δ to 0 makes the formula $f(t) + \delta f'(t)$ nonsense.

A way of making precise sense of infinitesimals was finally discovered by Abraham Robinson and others in the 1960s, almost three hundred years after Newton and Leibnitz introduced them as heuristic tools. Robinson’s work uses a very sophisticated version of the ring $\mathbf{R}[\delta]/(\delta^2 = 0)$. In principle this means infinitesimal formulas can be used again. However the fine print needed to make it work is so subtle and tricky that this has turned out to be impractical.

The conclusion—again—is that heuristic or intuitive formulations are unsatisfactory for ambitious development and calculation.

5.6.4 Cautions about Definitions and Internalization

Formal definitions are essentially never used in K–12 mathematics. Here we address some justifications given for this, and related potential misunderstandings with these projects.

1. Naive or innate intuitions are never sufficiently precise. We are able to work with pre-existing intuition about arithmetic in commutative rings in §5.1.1 because this intuition is not naive: it has been acquired through a great deal of disciplined practice with ordinary numbers and algebra. Problems due to K–12 use of a naive idea of area, see §5.2.3, are more typical. By comparison with what can be done with a definition, it is hard to compute, hard to describe in alternate ways, and many properties, e.g. invariance under skew transformations, see §5.2.4.2, remain hidden¹¹.
2. Internalizing properties of an object does *not* render the definition unnecessary, and being able to forget the definition is *not* an objective of the internalization process. In fact a good test of successful internalization is

¹¹A precise definition of area may be impossible in K–12, but realizing there is a problem may open the way for a better partial treatment, see §??.

that the student should be able to reproduce a completely precise statement of the definition. This means it is always available for explicit use if there is any doubt, and even experts have to do this from time to time.

3. Another confusion comes from the perception that experts work with derived properties and intuition, not definitions. This suggests that definitions, unpacking, etc. could be dispensed with, and intuition developed directly from derived properties. Unfortunately this is not the case: reliable intuition is developed by *deriving* the properties, not from the properties themselves. Trying to skip the development process usually leads to dysfunctional understanding that sooner or later will cause trouble, see the comments in §5.2.3. There is no logical reason for this: it seems to be a feature of human learning. It also seems to be particularly true for less-capable students (i.e. they benefit most from explicit development).
4. There is a temptation to break concepts into small pieces to make them easier to absorb. However because the pieces have to be assembled after absorption, this actually increases complexity and makes the subject more difficult. Difficult concepts can be approached in stages, c.f. the development of fractions, if the overall conceptual unity is kept clear.
5. Unpacking definitions is essentially routine and algorithmic. Many students actually enjoy it once they get used to it. However full unpacking is supposed to be a temporary expedient used during the development of reliable intuition, and some students have to be discouraged from continuing when it is no longer appropriate.

5.6.5 Summary

- Formal definitions provide a repository, training ground, and anchor for intuition.
- Reliable intuitions *incorporate* definitions rather than rendering them unnecessary.
- Good definitions and accompanying development are designed to promote development of reliable intuition.
- Good definitions frequently have unexpected benefits, see the comments at the beginning of §5.1.12 and the end of §5.2.3.1.

Bibliography

- [1] (References to be expanded)
- [2] Banchoff, Thomas; *Midpoint Polygons*, 1999 course materials at <http://techhouse.brown.edu/mdp/midpoint/>.
- [3] McCrory, Raven; *Mathematicians and Mathematics Textbooks for Prospective Elementary Teachers*; Notices of the Amer. Math. Soc. Vol. 53 (2006) pp. 20–29; link <http://www.ams.org/notices/200601/fea-mccrory.pdf>.
- [4] Response to letter about the above; Notices of the Amer. Math. Soc. Vol. 53 (2006) p. 526; link <http://www.ams.org/notices/200605/commentary.pdf>.
- [5] Quinn, Frank; *The Nature of Contemporary Mathematics*, retrieved from <http://www.math.vt.edu/people/quinn/education/index.html>.
- [6] Wilkerson-Jerde, Michelle H.; Wilensky, Uri *Understanding Proof: Tracking Experts Developing Understanding of an Unfamiliar Proof* Proceedings of ICMI Study 19 (2009).
- [7] Wu, H. *Fractions, Decimals, and Rational Numbers*, retrieved from <http://math.berkeley.edu/wu/NMPfractions4.pdf>.

Chapter 6

K–12 Calculator Woes

January 2009, revised February 2009

Introduction

In the third grade my daughter complained that she wasn't learning to read. She switched schools, was classified as Learning Disabled, and with special instruction quickly caught up. The problem was that her first teacher used a visual word-recognition approach to reading, but my daughter has a strong verbal orientation. The method did not connect with her strongest learning channel and her visual channel could not compensate. The LD teacher recognized this and changed to a phonics approach.

My daughter was not alone. So many children had trouble that verbal methods are now widely used and companies make money offering phonics instruction to students in visual programs.

The concern here is with serious learning deficits associated with calculator use in K–12 math. Calculators may not be making contact with important learning channels. Are they the latest analog of visual reading?

For brevity connections are presented as “deductions” (*this* about calculators causes *that* in learning). However the deficits described are direct observations from many hundreds of hours of one-on-one work with students in elementary university courses¹. The connections are after-the-fact speculations. If the speculations are off-base the problems remain and need some other explanation.

6.1 Disconnect from Mathematical Structure

Calculators lead students to think in terms of algorithms rather than expressions. Adding a bunch of numbers is “enter 12, press +, enter 24, press +, . . .” and they do not see this—either figuratively or literally—as a single expression

¹At the Math Emporium at Virginia Tech, www.emporium.vt.edu.

$12 + 24 + \dots$. Algorithms are much less flexible than expressions: harder to manipulate, generalize or abstract, and structural commonalities are hidden by implementation differences². The algorithm mindset has to be overcome before students can progress much beyond primitive numerical calculation.

6.2 Disconnect from Visual and Symbolic Thinking

Calculatorkeystroke sequences are strongly kinetic. But this sort of kinetic learning is disconnected from other channels: touch typists, for instance, often have trouble locating keys. Many students can do impressive multi-step numerical calculations but are unable to either write or verbally describe the expressions they are evaluating. Their calculator expertise is not transferred to domains where it can be generalized.

Even among high-achievers calculators leave an imprint in things like parenthesis errors. The expression for an average such as $(a + b + c)/3$ requires parentheses. The keystroke sequence does not: the sum is encapsulated by being evaluated before the division is done. Traditional programs also encourage parenthesis problems³ but they seem more common among calculator-oriented students.

6.3 Lack of Kinetic Reinforcement

It is ironic that calculators might be too kinetic in one way and not enough in another but this seems to be the case with graphing. In some K-12 curricula graphing is now almost entirely visual: students push keys to see a picture on their graphing calculators, and are tested by selecting from several pictures. They never pick up a pencil and draw a curve. Many students seem unable to internalize qualitative geometric information from purely visual input. Even some of our advance-placement students are now unable to draw or verbally describe the qualitative shape of an exponential or quadratic function.

That purely-visual instruction might have this effect should not be a surprise. Many people know they can improve comprehension of written material by copying it by hand. Kinetic reinforcement may be even more important for qualitative geometry than for text.

6.4 Lack of Verbal Reinforcement

People with strong verbal orientations often have to be able to read an expression out loud before they can understand it.

²For further analysis of this problem see *Beneficial high-stakes math tests: an example* at www.math.vt.edu/people/quinn/education.

³See the Teaching Note on Parentheses at <http://amstechnicalcareers.wikidot.org>.

My daughter went back to school in her early thirties and had to take statistics. At the beginning this was a disaster, but then I taught her how to parse and read the expressions out loud. The material became easy and she finished near the top of the class.

Now that I know what to look for, and how to look for it, it seems to me that many students would be helped by verbal reinforcement. Unfortunately this is rare in any approach to math: teachers talk more than listen and rarely make students read out loud, especially when they don't want to expose their ignorance. I cannot tell if calculators contribute to this problem but they certainly aren't part of the solution.

6.5 Summary

We have clever new technology but the same old brains. It turns out that some of the dreary things involved in by-hand math actually connected with ways our brains learn, and the ways calculators are used to bypass drudgery has weakened these connections and undercut learning.

If the explanations offered are correct then there are several further conclusions. First, some learning benefits of traditional courses are largely accidental and a more conscious approach should significantly improve learning in these as well. Second, calculators are not actually evil, but we must be much more sophisticated in how such things are designed and used⁴. But most of all, learning must now be the focus in education. Not technology, not teaching, not learning in traditional classrooms, but unfamiliar interactions between odd and variable features of human brains and a complex new environment.

⁴See *Student computing in math: interface design* at the site in footnote (2) for an attempt at such a design.

Chapter 7

Student Computing in Mathematics: Interface Design

January 2009

Introduction

I taught my first computer math course in 1975 and was convinced that it was the wave of the future. However it was atypical—a very small class at Yale—and later attempts were unsatisfactory: either too labor-intensive, or weak outcomes, or (usually) both. I have helped develop computer-*based* and computer-*tested* courses, but, ironically, the students still use by-hand computation.

Others have had similar experiences. Even after much tinkering, few college courses use more than calculators.

Calculators are widely used in K-12 math but many college teachers now associate calculator training with deficiencies in symbolic skills, number sense, and geometric understanding. Reasons calculators might undermine higher learning emerge from the analysis in this essay. Current calculators may be nearly the worst possible learning environment. In any case this cannot be considered successful.

Why has student computation been so problematic? The first problem seems to be a lack of understanding of how people learn. The second is a lack of new learning goals that computation should make accessible.

7.0.1 Learning and the Interface

We understand *teaching* very well, but teaching evolved to produce learning in a traditional classroom and, in math, using traditional techniques. It has turned out that many features of traditional good teaching are artifacts of these environments, *not* features of learning. In particular simple substitution of technology for hand work in traditional lesson plans has turned out to be a poor strategy.

§7.1 (Guiding Principles) gives a list of odd features of human learning and the contortions needed to fit mathematics into it. The math material is relatively general (support symbolic and abstract thinking, etc.) intended to guide design of a learning-friendly interface. More content-specific material is discussed in the sequel, [Student Computing in Mathematics: Functionality](#).

The basis for this analysis is experience with computer-tested and computer-based courses. These are a useful context for the study of learning because students are the primary actors in formulating and implementing their learning strategies. They can be guided but not as rigidly channelled as in traditional classes. I have spent a lot of time watching students learn in this context and their approaches are very different from what I tried to make them do in classrooms for thirty years.

§7.2 (Interface Design) applies these principles to find input methods, formats, and interactivity designed to maximally support learning. Such an interface should, it seems, be quite different from ones now in common use. Among the unexpected conclusions is that ordinary copy-and-paste is counterproductive: it does not leave a record that can be diagnosed for errors; and as a purely kinetic activity it undercuts the use of symbols to represent other expressions. Learning-friendly alternatives are proposed in [§7.2.1.3](#) and [§7.2.1.4](#).

7.1 Guiding Principles

This section gives some painfully-acquired insights about technology-enhanced mathematical learning. The focus here is on human learning and generalities about mathematics needed for interface design in the next section. Principles used to guide functionality of the computational environment are discussed in the sequel.

The point of view is as important as specific insights: there is much more to be learned and as we gain experience we must be alert to new insights that further shape design.

7.1.1 Learning, not Technology

Our objective is to help human beings learn. Humans are the bottleneck: technology design should be completely driven by the needs of human learning, *not* by availability, limitations, or capability of technology. Examples:

- Calculators are cheap and effective; shouldn't we exploit their availability? No. Calculator design is constrained by low cost and small size, and they are intended for calculation, not learning. They can be effective in meeting modest short-range goals (e.g. in K-12) but are counterproductive in the long run.
- Full-featured computer algebra systems have amazing capabilities; shouldn't we use them to "empower" students? No. They are designed for high-end professional use, not learning. Students can easily get lost in full-featured interfaces, and learning to mechanically use powerful black-box functions usually fails to develop understanding that transcends the particular interface or enables flexible general use.

To get good learning we must first understand learning, then we must design technology specifically to support learning. Anything off-the-shelf, or easily adapted from something off-the-shelf, is almost certainly unsatisfactory.

7.1.2 Symbolization

Elementary mathematics divides roughly into conceptual and mechanical (computational) steps. The goal of providing a computational environment is to enable focus on the conceptual activity. We discuss the division, then how it should be organized.

7.1.2.1 Conceptual Activity

Conceptual activities include organizing information and representing it ways suitable for computation. We refer to this as "symbolization" for several reasons:

- Material must be organized as symbolic expressions to be mechanically manipulated. Numbers are considered symbols here since their special properties only come into play in computational steps.
- Representing things as symbols is part of abstraction and a vital part of mathematical thinking.

We expand on this last point. The human mind is quite limited in the number of things that can be tracked or manipulated at one time. Fortunately the individual things can be very complicated. Thinking about complicated material therefore proceeds in two stages. The first stage is to identify good intermediate abstractions to serve as conceptual units; pack as much structure as possible into these units; and then develop enough automatic familiarity so we can actually see and use them as units. The second stage is to think about interactions between a relatively small number of these abstract units.

Representing things as symbols is only a small instance of the construction of conceptual units but it plays a vital role in early learning of the methodology. In particular if it is designed correctly it can serve as a model for later, more elaborate instances.

7.1.2.2 Computation

What counts as mechanical computation—to be farmed out to a machine—depends strongly on level. Why, and what to do about it, is the subject of the sequel, [Student Computing in Mathematics: Functionality](#). However for the most part computation involves manipulation or evaluation of symbolic expressions. In other words, computation acts on the output of symbolization.

7.1.2.3 Organization

As noted above our objective in providing computational support is to enable focus on symbolization. This can best be done by separating the two activities as much as possible.

Traditional problem-solving usually alternates the activities: the general practice is to compute as soon as a computable chunk of the symbolization is complete. These interruptions fragment symbolization, distract from organization, and conceal mathematical structure. A computational environment should be exploited to consolidate activities: symbolize first, then compute all at once, at least for elementary work. See the essay [Beneficial High-Stakes Math Tests: An Example](#) for an illustration of this in a specific simple example. In particular it is not a good strategy to simply use machines for the computational steps in traditional approaches.

Calculator work largely bypasses symbolization and mixes organization and computation so thoroughly that higher-order learning is inhibited. See the reference just above for an illustration. Another can be found in §3.1 of [Task-oriented Math Education](#).

7.1.3 Modes of Learning

The first general point is that human learning is strange and complicated, and while there are commonalities there are also great differences between individuals. A design goal is to support the commonalities while leaving flexibility for individual preferences.

The second point is that over time the complexities of learning become transparent. Character recognition and parsing of mathematical expressions become automatic. The use of a keyboard to obtain characters on a screen becomes second nature. This means the way experienced people do or learn things is not a guide to how neophytes learn, and effective tools for experienced users can be barriers for beginners.

We give examples in four cases: visual/kinetic reenforcement, visual/verbal reenforcement, the role of imitation, and subliminal learning. In each case we see that technically inconvenient things may have to be done to connect with human learning.

7.1.3.1 Kinetic Reinforcement

There are interactions between visual input (reading, seeing) and kinetic input (writing, drawing, copying) that seem to be important for memory and internalization of certain types of structure.

Examples:

- Mathematical notation is precise and delicate: omitting or wrongly locating a symbol, or misreading a statement, can completely change or destroy meaning. Copying a problem by hand seems to improve comprehension and reduce errors, and many traditional students are taught to do this as a matter of good practice. Copying often does not make sense in an electronic environment. We must be alert to problems caused by lack of kinetic feedback, and may have to find a substitute.
- Very young children learn to generate and manipulate symbols kinetically (by drawing them). This gets linked to visual recognition and alternate entry modes (keystrokes) to eventually form a seamless unit not dependent on drawing. Drawing does not translate easily to electronic environments, but trying to reduce or eliminate it from early learning will probably cause many children a lot of difficulty.
- Kinetic feedback seems to be vital for some students in internalization of geometric structure. Specifically students taught to graph functions by hand usually internalize the qualitative features of graphs of quadratic functions, exponentials, simple rational functions etc. This internalization is used strongly in later work involving plane and three-dimensional shapes, multiple integrals, and the like. Students taught with graphing calculators and tested by identifying a graph among several alternatives have been trained completely visually. They have never, or rarely, picked up a pencil and drawn the curve. And many of them cannot: apparently they have not internalized qualitative features from the purely visual approach. Omitting kinetic reinforcement puts these students at a disadvantage.

There are interesting examples in other subjects. Coloring books are a common adjunct to anatomy texts. Apparently kinetic feedback from coloring in muscles, bones etc. is helpful in fixing these complicated structures in memory.

Kinetic reinforcement might be incorporated in technology by, for instance, requiring students to trace over a computer-generated graph before allowing them to use it or submit it for a grade.

We caution that kinetic involvement alone is not a goal. Calculator arithmetic is intensely kinetic but counterproductive because it replaces rather than reinforces work with structure and symbols.

7.1.3.2 Verbal Reinforcement

Some people have strong verbal orientations. People who move their lips when reading, for example, are translating from visual input to kinetic (movement of

lips and tongue). Their best comprehension is through the kinetic channel, and comprehension falls if they have to rely on visual input alone (trying to read while eating, or trying to read something—like mathematics—that they can't pronounce).

Errors that involve confusing or substituting symbols that sound the same indicate verbal involvement. Confusing m and n , or M and m is probably a verbally-based error, while confusing p and q is a (dyslectic) visual error.

Verbal is *not* the same as auditory: there is a profound difference between speaking and hearing. Some students can transcribe lectures but can't read their notes out loud. Children in rural areas can listen to standard English on television for many hours each day but only be able to speak an incomprehensible local dialect. Hearing and speaking tend to be more tightly linked in later life but this is an example of the transparency that can conceal basic features of learning.

A corollary of this point is that audio or audio-visual materials are not a substitute for student verbalization. I do not have any clever ideas on how to incorporate verbalization into technology. For the present this has to remain a job for teachers.

Implications in math education:

- Children should probably be taught to read (out loud) what they have written. They should be encouraged to listen to what they say, and make sure it is what they meant to say. In other words, run stuff through the verbal channel to check it for accuracy.
- Students should be taught *how* to read material out loud. For instance reading expressions involving parentheses can be tricky and this is probably related to the frequency of parenthesis-related errors. This difficulty is *not* a justification for avoiding parentheses since this leads to serious problems later on.
- Reading is closely related to parsing because reading requires linear organization. I have seen verbally-oriented students completely stumped by notations such as $\sum_{i=1}^n \frac{i}{2^i}$ that make use of positional information. They usually find it tractable when taught how to parse it so they can read it out loud.

7.1.3.3 Imitation

People learn a lot by watching other people do things *provided* they can see how it is done. A teacher working a problem at a blackboard provides a model that can be imitated. Exactly the same information presented using prepared overhead slides, projected computer output, or PowerPoint cannot be imitated and therefore deprives students of this primitive and innate learning mode.

There is a particular role for imitation in mathematics. Since we do things one-at-a-time the construction of a mathematical expression is a linear dynamic process.

- Mathematics itself is usually not linear so this process involves a linearization, ideally one that follows mathematical structure.
- Parsing an expression, for instance to read it out loud (see the previous section), also involves a linearization.

These two linearizations are often different. If students are not given examples to imitate then they almost always try to use the parsing linearization, and this is less efficient and more prone to errors than structural linearizations.

The following examples are fairly complex so opportunities for confusion will be clear. Students will have similar problems with much simpler expressions:

- The structural linearization used to expand $(a + b + c)10e^{x-5}$ has first step $a(\dots) + b(\dots) + c(\dots)$. We then enter the complicated expression inside each pair of parentheses. The verbal linearization requires switching back and forth between a, b, c and the complicated expression, and offers many more opportunities for error.
- The structural linearization used to construct $\sum_{i=1}^n \frac{i}{2^i}$ begins with $\Sigma(\dots)$. Filling in the parentheses is usually the next step and the limits come last. In the verbal linearization the summation variable and limits come first, not last. This invites errors like $\sum_{i=1}^n \frac{n}{2^n}$.

It is easy to see implications for machine-based examples and presentations: they should be dynamic and emphasize the thinking behind each step. It is less clear how this should influence design of a computational environment.

7.1.3.4 Subliminal Learning

Students can potentially learn from anything, and it is often unclear exactly what they are learning. A consequence is that everything should be designed so that if students do learn from it then they will learn correctly.

For instance by-hand arithmetic involves a lot of symbol manipulation and shows mathematical structure in action. It may be that students internalize it and this prepares them for algebra. Calculator arithmetic avoids symbols and hides structure and so does not provide the same opportunities for subliminal learning.

As another example we describe how very young children can be offered an opportunity to absorb a sophisticated mathematical viewpoint. Mathematicians think of “addition” in functional terms: any binary operation that is associative, commutative, has a neutral element and inverses is entitled to be called “addition” and represented by “+”. One point of the abstraction is that work habits appropriate for integers and other elementary examples are equally valid for any other “additive” system.

Now imagine showing a child that pretty much any two expressions that can be entered into the computational environment can be combined using +. In general this is just a property of these things, but you can see what it is good for in special cases: if you combine numbers representing lengths of two sticks

using $+$ then you get the total length of the two sticks joined together. In other words, finding lengths of joined sticks is an *application* of $+$ when it is applied to numbers. It is *not* the definition, and $+$ is *not* limited to numbers that can be interpreted as lengths. This is not a point that should be made explicit to children, but the approach presents it in a way that it can be absorbed subliminally.

7.1.4 Process, not Answers

The particular virtue of mathematics is that correct use of correct methods gives correct answers. The focus in learning mathematics should therefore be on methods and their use. The implication for the current context is that a learning environment for mathematics should “show work” in the sense of providing a record of the methods and reasoning used, and this record must be usable for locating and targeted correction of reasoning errors. A wrong answer only indicates that an error was made, not the nature of the error, and without a diagnosable transcript the only recourse is to repeat the work and hope for a better result the next time.

“Correction” here means fixing errors of reasoning or mathematically incorrect methods, not conformity with a standard template. An alternative but mathematically correct approach does not need correction.

7.1.4.1 Learning from Process

In any other subject there is enough imprecision in terms or context, or possibility of unanticipated factors, that careful logical reasoning can fail to give a correct conclusion. It is still worthwhile because it greatly improves the *chances* of getting a correct conclusion, and I believe that experience with careful reasoning in complex logical systems is the greatest general benefit of studying mathematics. In this sense the medium is the message.

7.1.4.2 The Role of Answers

Mathematics also has the virtue that incorrect reasoning usually gives an identifiably incorrect conclusion. This means correct answers can be a useful proxy for correct reasoning. However this is only true if students are using correct methodology. Independent checks on methodology should be possible, and the reasoning itself should be available for diagnosis and correction when the answer is wrong.

7.1.4.3 Diagnosis and Error Correction

Ideally *every* error should be diagnosed and corrected. This would catch misunderstandings immediately, before they can be reenforced by repetition. It would also encourage students to work carefully to avoid triggering the diagnosis process.

Our best goal is for students to learn to detect, diagnose, and correct their own mistakes. Teacher diagnosis and correction should offer models students can imitate. Self-diagnosis can also be promoted by providing diagnostic aids for worked-out problems, see [Task-oriented Math Education](#), and the Teaching Notes on the [AMS Technical Careers](#) web site. This is an aspect of problem design rather than the learning environment, but it can only be effective if there is a record that students could try to diagnose.

7.2 Interface Design

This section is concerned with structuring the interactions between student and machine. Objectives established in the previous section include focus on organization and construction of mathematical expressions (§7.1.2 Symbolization); supporting learning modes such as kinetic reinforcement (§7.1.3.1); and producing a diagnosable record (§7.1.4).

Much of the student-interface interaction looks like elementary programming. This is implicit (or subliminal) rather than explicit, and is enforced by the structure of the interface. This is not an accident: programming requires extensive symbolization and explicit use of structure and so is a good model for mathematical learning. In fact it is completely compatible with the primary learning objectives to take development of fluent use of high-level programming languages as an additional long-term objective.

The section is divided into Input Modes, concerned with immediate interactions between student and machine; Windows and Sessions, describing structure of interactions, and Input Formats.

7.2.1 Input Modes

The primary input mode should be writing or drawing directly on the screen with a stylus. Reasons include:

- For young children this avoids indirect input and provides kinetic reinforcement for number and symbol formation;
- for all students it provides kinetic reinforcement of graphic work (§7.1.3.1); and
- it enables easy and intuitive addition of commentary and reference tags.

7.2.1.1 Character Recognition

The interface has to provide character recognition, but it should probably should not learn the user's style. Reasons include:

- It is appropriate to expect reasonably clear character formation, and for mathematical material it may be better for the interface to be a bit picky about characters than to have to override inappropriate guesses. “Guess” could be provided as a button.

- An adaptive system leads to non-portable input habits: they won't work on other machines, and in particular not on secure systems used for testing. Some individual calibration will be necessary, for instance for left- and right-handed differences, but this should be kept to a minimum to avoid becoming a stumbling block.

It is not so important that text entry be portable, and careful math mode would be available as a backup, so these considerations do not apply to text.

Anyone concerned about asking students to adapt to an interface should reflect on how well they have adapted to a really terrible text interface: entering text on a numeric keypad with their thumbs!

7.2.1.2 Keyboard

A standard keyboard should be provided for fast entry of text. However there should be *no function keys*:

- Functions may be disabled.
- Functions should be considered parts of mathematical expressions, internal to the symbolization process and recorded in the transcript, not as external objects living on a keypad. Logarithms should be obtained by writing “Log” and evaluating, not by pushing a button.
- In order to separate conceptual and computational steps we want students to construct expressions, possibly including functions, and then evaluate them. Immediate evaluation (via a function key) defeats this.

For similar reasons the interface should generally *not* depend on palettes for insertion of symbols, patterns, or functions. It might provide lists or browsers from which material can be copied in appropriate ways.

7.2.1.3 Copy-and-Paste via Symbols

Standard copy-and-paste has the same invisibility and symbolization-defeating problems as function keys and so should be strictly limited. We suggest substitutes for some of the functionality.

The first replacement for copy-and-paste is symbolic assignment. The student can select an expression in a static window and assign it a name, for instance by writing “A=” in the selection area. The expression can be used in an input window by entering the name, “A”. When the name is referenced the selection and name assignment should be frozen to preserve a record.

Example: A company has 47 employees with an average salary of \$37,867. What is the total payroll of the company?

The student can select 47 and write “emp=” in front of it, then select 37,867 and write “sal=” in front of it. He then can enter “emp*sal” in the input window and evaluate to get the answer. Alternately he could enter “payroll = emp*sal” to have the output accessible for later use. Note this scheme subliminally supports symbolic thinking, see §7.1.3.4, and provides a record of the work.

7.2.1.4 Copy-and-Paste via Tracing

The second replacement for copy-and-paste is provided by tracing “templates”. The student selects something in a static window and drags it to a work window. A dimmed copy appears. This cannot be evaluated or further selected, but the student can trace over it to get a functional copy.

- For young students this provides kinetic reinforcement of character and symbol formation, and construction of expressions.
- This provides kinetic reinforcement of graphic material, see §7.1.3.1.
- Expressions or graphic material can be modified rather than traced exactly. Symbols could be changed to change the input into the expression, for instance.

Again the selection area should be frozen and linked to the template to provide a record.

7.2.1.5 Copy-and-Paste in an Input cell

Input cells in a Work window (§7.2.2.3) should be an exception to these restrictions. These cells serve as “scratchpads” and there currently seems to be no reason to disable full copy-and-paste *within* such a cell.

7.2.2 Windows and Sessions

There should be three standard window types: Data, Work, and Preview.

7.2.2.1 Data Windows

These are static in the sense that they cannot be edited, but annotations and selections can be made in an overlay.

Data windows can have form boxes in which material can be entered, for example answers when the data window displays a test. Form boxes should be assigned names so material can be entered either directly (by stylus or keyboard) or by assignment. For instance in the payroll example in §7.2.1.3 the answer box might be assigned the name “**answer5**” and the answer could be recorded by entering “**answer5 = emp*sal**” in the Work window.

Note this design makes symbol use natural and helpful so it encourages symbolization.

7.2.2.2 Preview Window

The preview window nicely formats expressions but does not manipulate them. Expressions to be previewed are selected (in the usual way) and a Preview button is activated.

- Error messages are issued, for instance when parentheses are unbalanced.

- The formatting displays how expressions will be interpreted. For instance the expression $2^5 x$ will be previewed as 2^5x . If 2^{5x} was intended then the mistake will be evident and appropriate parentheses added.
- Complex expressions should routinely be proofread using Preview. For instance the TeX expression $\Sigma_{i=1}^n \frac{i}{5^i}$ is previewed as $\Sigma_{i=1}^n \frac{i}{5^i}$. If this is not what is intended then the input expression can be edited.
- Some incomplete expressions should be previewed as expressions with boxes for missing material rather than giving an error message. For instance the incomplete Mathematica expression `Integrate[, {y, }]` should be previewed as $\int_{\square}^{\square} \square dy$.

Preview material cannot be edited directly, nor can it be copied. If the source window is static then the original selection can be assigned a name or can be used to form a template. If the source is in an active input window then it can be edited.

7.2.2.3 Work Windows

Work windows are divided into alternating Input and Output cells.

The bottommost Input cell is active, and can be edited, previewed, and evaluated. Results of evaluation appear in the Output cell immediately below. The Output cell cannot be edited. The active Input and Output cells are dynamic so cannot be annotated and the material in them cannot be selected for copying.

Input and Output cells other than the bottommost are static (have been frozen) and cannot be edited. They can be annotated in the overlay, parts selected and copied, etc.

The active Input cell can be repeatedly edited, previewed and evaluated. It becomes inactive (is frozen) when a new Input cell is opened at the bottom of the window or an End Session button is activated.

Input and Output cells can be deleted. Links and symbolic-copy material from a deleted cell disappear with it. If the bottommost remaining cell is an Input cell then it becomes active.

7.2.2.4 Sessions

A session consists of working in an Input cell, freezing it by opening a new Input cell, and repeating as needed until the session is closed.

When a session is closed the Data and Working windows are linked and saved together as a Data window. This preserves the work record because it can no longer be edited. This record can be diagnosed for errors and annotations can be added to record the diagnosis. It can also be used as a source to rework problems using a new Work window. Correct fragments from the previous session can be spliced in to reduce repetition and focus on corrections. All this provides support for the diagnosis and targeted learning described in §7.1.4.

Closing the session may activate additional features. For instance if the Data window contains a test then closing the session should activate scoring functions to grade the test, and a section containing answers and diagnostic hints should become accessible.

7.2.3 Input Formats

We have discussed how symbols and other material should be entered in the interface. The topic here is how these objects should be arranged to be accepted for processing. We will mostly be concerned with symbolic material. There seems to be a role for an input format for graphics but it is far from clear what should be involved. There should also be formats for data entry, for instance tables of numbers, but these will be special-purpose and infrequently used. Entering a lot of numbers is not a useful or real-life activity and data will usually be accessible in electronic format.

The design is driven by concern for learning and based on watching and working with students on computer projects in calculus and vector geometry. A particular conclusion is that *writing* and *reading* need to be separated. A symbolic input format must be easy to write and edit; it need not be easy to read. The beautiful two-dimensional formats of typeset mathematics are easy to read but not easy to write (for machine use) and not easy to edit. There is not going to be a satisfactory single format. Instead we optimize formats for specific uses and use Preview and other tools to negotiate between them.

7.2.3.1 Linear, Primitive, Explicit, Verbose

These are characteristics needed to make the format easy for inexperienced users to write and edit. Sophisticated or experienced users with other needs should use a different system.

“Linear” means in particular that the format should not incorporate positional structure ($a \wedge 5$, not a^5). Positional data entry invites mistakes and frustrating misinterpretations. Positional representations are harder to edit. Finally, copying part of a positional representation can fragment formatting instructions and lead to bizarre results and obscure crashes.

“Primitive” essentially means limited to text. A good rule of thumb is that it should be possible to send an input expression by email as text. We might accept Unicode rather than ASCII so some symbols could be considered text. However most mathematical functions should be spelled out in some way.

“Explicit” means everything in the expression must be visible. Invisible material, for instance formatting instructions, is dangerous and not worth the trouble. “Unambiguous” should be part of this. For instance multiplication is better denoted by $*$ than \times or a dot since the latter two are easily misinterpreted.

“Verbose” means that instructions and function names should be spelled out in ways that are easy to guess and remember. For instance to get an integral one should write out “Integral”, and then provide arguments. Abbreviated function names can be entered faster by experienced users but add a coding/decoding

layer that is difficult and distracting for students.

To repeat, the objective is a format that students can easily learn to write and edit. The features listed above seem to help with this but do not guarantee success. In particular, students learn most easily and naturally from examples and hints, not explicit instruction. If students have trouble learning basic use of the format from examples then the format needs improvement.

7.2.3.2 Mathematically Correct

It may seem odd that this needs explicit mention but traditional notations and ways of thinking are sometimes imprecise, rely on context, or are heuristic rather than really correct. In such cases correctness requires a break with tradition.

For instance “=” is traditionally used in several different ways. The expression

$$y = ax^2 + bx + c$$

may indicate an *assignment*: the symbol y is given the value of the expression on the right side. Or it may indicate a *test*: a relation that is either true or false, as in “Intersection points of the two curves are points (x, y) that satisfy ...”. Further, an assignment can be *immediate*, in which case y is given the current value and not effected by later changes in a, b, \dots ; or *delayed*, in which case y is a shorthand for the right-side expression and it’s value at any particular time reflects current values of a, b, \dots . It could even be an *implicit* assignment intended to specify a or x , etc.

This notational ambiguity means a traditional expression containing “=” is incomplete and must be accompanied by text indicating the meaning. Confusion results when, as is too often the case, the text is omitted¹. This is bad enough in common practice and unacceptable in a machine environment.

A mathematically correct format must have different notations for these different meanings, or support them in other ways (e.g. implicit assignment might be done with a “solve” function rather than a variation on “=”). This will conflict with ambiguity in traditional notation and language, but we should see this as a feature (who needs self-inflicted notational confusion anyway?) rather than a flaw.

7.2.3.3 Graphics Input

An important objective in studying functions is to develop a feel for qualitative features of their graphs. What does re^{st} look like as a function of t , independent of the values of r, s ? How about $r + (t - s)^n$? Beautiful computer-generated graphs in specific cases are not a substitute for qualitative understanding.

It would be nice to have a full syntax and computational context for qualitative graphic information, but for starters it would be useful just to have an

¹The confusion can even include a failure to recognize this as a notation problem. W. Byers in *How mathematicians think: using ambiguity, contradiction, and paradox to create mathematics*, (Princeton U. Press 2007) argues that this reflects a basic ambiguity in mathematics itself!

input format. Suppose we ask a student to draw a graph with the qualitative features of an exponential function. How can we extract (mechanically) the qualitative features of graphic input to determine if the drawing is reasonably correct?

Note that we really do want students to draw the graph by hand. Kinetic reinforcement seems to be essential for some students (§7.1.3.1), and is probably important for all, so this is another case where technical difficulty or convenience cannot be allowed to override educational concerns.

7.3 Summary

The long-term goal is to better prepare K–14 students for advanced learning in mathematics, science, engineering, and other technical disciplines. It seems obvious that this should include systematic use of machine computation, but most attempts have been counterproductive and none have been fully satisfactory.

The core problem seems to be that current computational environments do not support the complex oddities of human learning. This essay describes a rough draft for an interface design driven by this complexity and the structure of mathematics. The final form will no doubt be different from this draft but it should also be clear that it will be profoundly different from any current interface. It is also clear that development of such an interface is a delicate and sophisticated undertaking.

The sequel, *Student Computing in Mathematics: Functionality* concerns the need to carefully limit functionality of the computational environment.

Part II

**The Course/Curriculum
Level**

Chapter 8

Task–Oriented Math Education

October 2008 Draft

Introduction

“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.

8.1 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.

8.1.1 Descriptions

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

8.1.1.1 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.

8.1.1.2 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.

8.1.2 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.

8.1.3 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.

8.2 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.

8.2.1 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 §8.6.3 Tracked Courses. However for simplicity we focus on non-terminal courses.

8.2.2 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 §8.4.1.

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.

8.2.3 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 §8.3 More About Tasks.

8.2.4 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.

8.2.4.1 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 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.

8.2.4.2 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 §8.5.2 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.

8.2.4.3 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.

8.2.5 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.

8.2.6 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.

8.3 More about Tasks

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

8.3.1 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".

8.3.2 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.

8.3.3 Task Constraints

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

8.3.3.1 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.

8.3.3.2 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.

8.3.4 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.

8.3.5 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 §8.5.4 Development.

8.3.6 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 §8.5.4 Development.

8.3.7 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.

8.4 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.

8.4.1 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.

8.4.2 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.

8.4.2.1 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 §8.5.2.2 Classroom Teachers and §8.6.3 Tracked Courses.

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

8.4.2.2 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.

8.4.2.3 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 unpre-

dictable 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.

8.4.2.4 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.

8.4.3 Student behavior

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

8.4.3.1 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.

8.4.3.2 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.

8.4.4 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 §8.6 Educational Opportunities.

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!

8.4.4.1 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.

8.4.4.2 Group activities

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

- 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.

8.4.5 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.

8.5 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 §8.6 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.

8.5.1 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.

8.5.2 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.

8.5.2.1 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 §8.2.4.2 Help Opportunities.

8.5.2.2 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 §8.4.2.1 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.

8.5.2.3 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.

8.5.2.4 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 §8.6 Educational Opportunities.

8.5.3 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.

8.5.3.1 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.

8.5.3.2 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.

8.5.3.3 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 §8.5.4.1 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.

8.5.4 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 §8.3.5 Software Generation, but one is that full development need only be done once. Maintenance and refinement should continue indefinitely but are relatively inexpensive.

8.5.4.1 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.

8.5.4.2 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 §8.6.3 Tracked Courses.

8.5.4.3 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.

8.5.4.4 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 given them adequate control before they could give up control over software design.

8.5.5 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.

8.6 Educational Opportunities

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

8.6.1 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.

8.6.2 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.

8.6.3 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 (§8.4.2.1 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.

8.6.4 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.

8.7 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.

Chapter 9

Downstream Evaluation of a Task–Oriented Calculus Course

Data for the study finally received October 28, 2009. We hope to complete most of the analysis by the end of December.

Chapter 10

Beneficial high–stakes math tests: an example

November 2008

Introduction

High–stakes tests influence teaching and learning. When learning is poor they provide discipline and motivation for improvement. When learning is good the influence tends to be bad because the focus shifts from learning to test performance. Recent widespread introduction of high–stakes tests is, in effect, a judgement call: the general level of learning is so poor that the discipline enforced by tests will outweigh bad effects in the few previously–good cases. Roughly speaking we accept a cap on the top to get a floor under the bottom¹.

The thesis here is that high stakes tests—when very carefully done—can influence teaching and learning in positive ways. Counterproductive influences are the result of poor tests, not of high–stakes tests per se.

Beneficial high–stakes tests require a completely different approach to testing. Every aspect, from how the tests are given, to problem design, down to the level of computer code, must be driven by sophisticated educational and mathematical wisdom. But this wisdom must be correct in an almost mathematical sense, and in particular not determined by conventional wisdom or ideological convictions.

10.0.1 Outline

To illustrate this thesis we give an example worked out in detail.

- First we identify reasons that students taught with calculators have symbolic-reasoning deficits: for instance calculator routines conceal mathematical

¹See [The K–12 math test conundrum](#) for a brief discussion

structure. But then we discover that by-hand arithmetic actually has the same problems to a lesser degree.

- We suggest a change in the way problems are worked—with or without calculators—that would address this.
- We describe a modification to test design that would make the new approach directly effective for test-taking. This provides motivation for teachers and students, though the motivation provided by a yearly test would be a bit remote.
- The envisioned test-generating system could also provide course tests and plentiful practice materials, adjusted to be appropriate for different levels and locales. This would give constant reinforcement and feedback and quickly spread improvements.
- We then discuss high-level issues in implementing such a test system. One is that it would spread mistakes just as quickly as improvements.
- For this and other reasons content for such a system must be developed in an open and non-commercial way that can respond quickly to feedback and can draw on the wisdom of the entire community. Large-scale test *administration* might still be a commercial activity.
- This particular suggestion requires a change in the functionality of tests as well as in content. We describe how to use modern electronic formats and programming tools to achieve this.
- We hope to have sample tests with these features available at the Joint Mathematics Meeting in Washington DC January 4–8, 2009.
- Implementing this change would ideally lead to significant changes in the way K–12 math is taught, including systematic use of parentheses from the very beginning of arithmetic.
- The discussion of instructional changes also illustrates use of web reference materials for teacher support.

10.0.2 Other Issues

The discussion here is based on analysis of a single issue and many others will have to be similarly understood to get a complete picture. For instance here we see how to organize elementary material to *set the stage* for abstract and symbolic reasoning but have not tried to work out how such reasoning should be taught and tested. A few more examples:

- The analysis here focuses on single-step problems. How should multi-step problems be handled?

- Students taught with graphing calculators often have geometric–reasoning deficits. What is behind this and how can it be avoided?
- Calculators have inadvertently caused serious problems and we are only now beginning to recognize them and sort them out. Can we figure out how to graduate to modern computer–algebra systems without having to suffer through turbocharged versions of the same problems?

Fortunately these do not have to be tackled all at once. If our analyses are based on thorough and accurate understanding of mathematical structure then we can expect the solutions will fit together and reinforce.

10.0.3 Web Resources

Web resources can help explain and support instructional change. This is illustrated with links to the web site of the American Mathematical Society Working Group on Preparation for Technical Careers, abbreviated [AMSTC](#).

10.1 Analysis of the Problem

A specific word problem is used to illustrate how calculators, and to a lesser extent traditional approaches, fail to support development of higher–level reasoning. The natural mathematical view suggests a remedy but also makes clear some of the difficulties that will be encountered.

10.1.1 A Sample Test Question

The example is:

Problem Three children collect acorns for an art project. Dick finds 7 acorns; Jane finds 13; and Warren amasses 40 acorns. The teacher puts all the acorns in a bowl and then divides them evenly among the three children. How many acorns does each child have for the project?

10.1.2 Mathematical approach

A mathematician would write $(7 + 13 + 40)/3$.

The structure of the situation is clearly reflected in the structure of the expression and it is a small step to write $(\# + \# + \#)/3$, where $\#$ is used as a placeholder for the numbers of acorns collected by a child. This abstraction is easily accessible and will have a subliminal influence even if it is not made explicit.

The generalization to an arbitrary number of children is conceptually easily but the placeholder notation is unsatisfactory because it does not display the linkage between the place and the child. To do this we use $\frac{1}{n}(A_1 + A_2 + \cdots + A_n)$. The underlying structure is clearly the same as for three children so the only

difficulty is with the notation, and this should seem reasonable because it solves a problem (dangerous imprecision of the $\#$ formulation).

A more subtle feature of the notation $\frac{1}{n}(A_1 + A_2 + \cdots + A_n)$ is that the sum $(A_1 + A_2 + \cdots + A_n)$ is seen as a unit that can be manipulated even though the operation has not been carried out. The best version, $\frac{1}{n}\sum_{i=1}^n A_i$, builds on this and can be made accessible if students are explicitly taught how to parse it and read it out loud.

Finally the expression makes sense for any type A_* that can be added and then divided by an integer. Students who think of polynomials as a fancy sort of numbers (as do mathematicians) will use exactly the same expression to find the average of a collection of polynomials.

10.1.3 Calculator approach

The student presses keys 7, +, 13, +, to get 20, then 40, +, to get 60, then \div , 3, =, to get 20.

This is an *algorithm* rather than an expression. Students easily see how to generalize it to handle more cases but it does not explicitly display the mathematical structure and cannot be generalized or manipulated as an expression. Further there is no notation for these algorithms so they must be remembered rather than recorded. This makes it difficult to point out structural similarities in work done at different times, or, better, have students recognize similarities because the expressions have the same structure.

Calculator-trained students have to see polynomials as new and different things. The structural similarity between integers and polynomials has been hidden: numbers are algorithmically manipulated by calculator while symbols require rules that seem strange and tedious because they have not already been internalized, for instance through by-hand arithmetic. These students will have difficulty seeing any similarity or connection at all between the solution of the acorn problem and the average of a set of polynomials.

10.1.4 Traditional approach

Traditional students will write $7 + 13 + 40$, but then encapsulate this as a unit by carrying out the operation rather than with parentheses. They then divide the sum, 60, by 3.

We now see the traditional approach as half-way between the mathematical and calculator versions.

On the plus side the sum is seen as a unit as in the mathematical version. It can be generalized to $A_1 + A_2 + \cdots + A_n$ and then to $\sum_{i=1}^n A_i$. By-hand arithmetic provides a lot of hands-on experience with mathematical structure (of addition and multiplication) so it has been internalized and this makes the transition to symbols and polynomials relatively easy. After this the sum makes sense for polynomials and other symbolic expressions.

On the negative side the full solution is still an algorithm rather than an expression: division by 3 takes place after encapsulating the sum by evalua-

tion rather than with parentheses. The unevaluated sum is not presented as an object that can be manipulated. Students learn to approach problems by alternating organization (setting up the sum or the division) and operations (adding, dividing). This works in school math because problems are designed to be worked this way but is counterproductive in the long run because it disrupts mathematical structure and invites errors.

10.2 Diagnosis and a Remedy

We saw in the previous section that calculator-oriented math education thoroughly mixes the organizational and computational components of problem-solving, and that this undercuts learning in a number of ways:

- It does not provide unevaluated arithmetic expressions that display mathematical structure, provide templates for generalization and abstraction. See [AMSTC/Products of Sums](#) for another example.
- These unevaluated expressions also aid in diagnosis of mistakes, see [AMSTC/Diagnostic Aids](#).
- It hides the functional similarity of numbers and symbolic expressions such as polynomials.
- Mixing cognitively different tasks degrades both and increases error rates, see [AMSTC/Separation of Tasks](#).

We then saw that the traditional approach with by-hand arithmetic actually has some of the same deficiencies and therefore also undercuts learning albeit to a lesser degree. In other words calculators did not cause the current abstract and symbolic reasoning deficits, but their use enabled expansion of bad practices that worsened deficits that had been invisible.

One conclusion is that “go back to the old ways” is not a satisfactory solution: we need a new approach that exploits this new understanding. For instance complete comfort with parentheses seems to be vital but traditional elementary math education has a parenthesis phobia, see [AMSTC/Parentheses](#).

10.2.1 Remedy

The proposal is for teachers and tests to encourage students to separate the organizational and computational components of problems by explicitly using unprocessed intermediate expressions.

To clarify this we give an example.

10.2.2 Better Sample Question

We revise the word problem of §10.1.1 to illustrate how the remedy might be implemented in a test, and give a variation for class use.

10.2.2.1 Test Version

Three children collect acorns for an art project. Dick finds 7 acorns; Jane finds 13; and Warren amasses 40 acorns. The teacher puts all the acorns in a bowl and then divides them evenly among the three children. Give *an arithmetic expression* that evaluates to give the number of acorns each child has for the project.

Problems like this must be explained and used with care:

- “Raw-output” expressions are not well defined, and there will be a great many correct expressions, including the numerical outcome (20), that are logically correct.
- For (machine) scoring purposes the expression is considered correct if it evaluates to give the correct outcome. In other words the student only has to set it up and the test will take care of the arithmetic. In §10.3.2 we suggest taking this literally: when an expression is entered the result of evaluation is automatically displayed.
- The point is that the student’s best strategy is to do only the organizational component of the problem and enter the resulting expression without doing any processing. This minimizes time and exposure to errors.
- The student’s best strategy therefore implements the proposed remedy: organization and processing are separated and attention is focused on the intermediate expression where structure etc. is displayed.

The test formulation is not immediately suitable for use outside a computer test environment because the unevaluated expression is not well-defined and cannot easily be checked for correctness. A class version could be:

10.2.2.2 Classroom Version

Three children collect acorns . . .

1. Set up an arithmetic expression that gives the number of acorns each child has for the project, but don’t do any arithmetic.
2. Evaluate this expression.

An answer would be considered correct if the response to (1) is not obviously bogus and has the right structure (in this case something like $(\# + \# + \#)/3$), and the number in (2) is correct. Otherwise it is considered incorrect:

- If the number is incorrect then the expression can be considered more carefully. A correct expression indicates that there was a mistake in evaluation and more evaluation drill may be called for. An incorrect expression suggests an error in setup.

- Determining that the form of an expression is wrong relies on human pattern-recognition skills. The instructions to students is that this part of the answer is to be used for diagnosis when something goes wrong; see [AMSTC/Diagnostic Aids](#). If the expression is unsuitable for diagnosis then the answer is unsatisfactory even if the number is right.

This last point goes against the idea that a correct number justifies everything, but requiring an organizational and diagnostic step really is important. In particular I believe that if students are consistently required to get the right raw form then using calculators to do the evaluation should not cause problems.

10.3 Test Design and Implementation

§10.1 gives an analysis of a single issue of a whole constellation. Rather than consider the issue in isolation we build on the analysis in [Task-oriented Math Education](#).

Test features already identified as useful are summarized in §10.3.1 and we add to this those of §10.2.2.1, Better Sample Question. Specific format and programming proposals are made in §10.3.3.

10.3.1 Learning Tasks

The analysis in [Task-oriented Math Education](#) suggests that test features should include:

- Computer-based (presented and worked in an electronic format);
- Software-generated (not assembled from a database of problems);
- Multiple-try (many equivalent instances rather than static);
- Instances can be used for practice, and provide diagnostic aids, reference links, etc. after scoring (designed as a learning environment);

We refer to tests with these features as “Learning Tasks” to emphasize that learning, not assessment, is the primary objective.

10.3.2 Functionality

The example in §10.2.2.1 requires the following:

- a form box in which the student enters an arithmetic expression;
- an “Evaluate” button so that when it is activated:
- the result of evaluation appears in a *different* box, and
- when the test is scored the contents of these boxes is frozen (and correct answers, diagnostic aids etc. appear).

The evaluation appears in a different box so the expression can be preserved. Further, if there is an error—missing parenthesis for example—an error message should appear in the evaluation box. The expression can then be diagnosed, edited, and re-evaluated.

10.3.3 Formats and Programming

We suggest formats that provide the functionality described above, and more.

10.3.3.1 Test Format

A test (or learning task) is an Adobe PDF document. PDF supports web links and forms, and embedded javascript can be used to process or evaluate material entered in form boxes. Javascript gives access to a wide range of mathematical functions so the functionality described in the previous section is easily obtained.

Other benefits are:

- Tests can be self-scoring and fully functional via embedded javascript without depending on a server or test system.
- For-credit tests generally will be linked to a test system for security and recording grades, but other than this will have exactly the same functionality as a free-standing practice test.
- Answers and diagnostic aids, revealed when a test is scored, can also have javascript functionality and can include web links to reference material.
- The content generating system is independent of for-credit administration systems. Content should be provided through a single open-source or public domain system, while security and database software for administration might be commercial products offered by a number of companies.

10.3.3.2 Source Code

LATeX is an effective source code for functional PDF documents. It can be compiled directly to PDF by the PDFTeX program², or compiled to PostScript and then Distilled to PDF.

- The hyperref package included in standard LATeX installations provides intra-document and web linking, and basic support for HTML forms.
- The [AcroTeX education bundle](#) developed by D. P. Story provides further support for HTML forms, and facilities for embedding document-level javascript in a PDF document.

Story's system actually provides packages for producing self-scoring PDF tests. These are not flexible enough (off-the-shelf) to provide the functionality described here, but almost all the work is done.

²This document was produced in this way.

10.3.3.3 Producing Source Code

The problem generators used in the Math Emporium at Virginia Tech are written in Mathematica. Current versions produce problem source code designed to be processed by Mathematica and distributed as web pages from a server. Most of them could be easily modified to produce LaTeX source code.

This means the technology and methodology for producing the source code is, in a sense, already established. However this work is done in-house and there is little publicly available material on it. A more problematic point is that developing problem generators is a very high-level activity and requires mathematical and educational sophistication.

10.3.4 Advanced Functionality

Eventually more functionality will be needed than can reasonably be provided by javascript embedded in a PDF document. For instance multi-step problems will require some sort of iterative computational support. This should be done with a separate computational environment. The test would interact with the environment to specify the appropriate level of functionality and receive output, but would not itself provide the functionality. For a draft description of such a computational environment see [Student Computing in Mathematics: Interface Design](#) and [Student Computing in Mathematics: Functionality](#)

10.4 Conclusions

High-stakes tests may improve minimum competency but current tests influence instruction in ways that depress achievement at only modestly higher levels.

In this article we worked through an example to explore what might be involved in designing tests that would actually improve teaching and learning. Our conclusion is that it is possible in principle, but getting it to work will be very challenging.

This single example required:

- recognizing a subtle problem unnoticed or denied by large parts of the education community;
- mathematically sophisticated analysis of causes that revealed an unexpected flaw in traditional elementary instruction;
- figuring out how a test might provide context and motivation to fix the flaw;
- being willing to have every single aspect of testing, from administration strategies to computer-based formats, driven by instructional needs;
- having sufficient experience and technical expertise to see how to implement the design.

The objective is not to get the test design exactly right—this is unrealistic—but to get close enough that equally careful and sophisticated field-testing would lead to an effective version. A less-sophisticated analysis or a priori constraints on test design would lead to a system that no amount of field-testing could fix.

Chapter 11

Economics of Computer–Based Mathematics Education

Introduction

I taught a computer calculus course in 1975. Since then I have followed with great interest other attempts to use computers, and participated in a number of them, but nearly all were unsuccessful in the sense that they had faded away within a few years.

For the last decade I have been privileged to work with one of the few really successful programs and this perspective leads me to believe that the main program–killer is economics, not difficulty with either computers or education. Specifically:

- To be successful, a new instructional method must be *less expensive* (particularly in faculty time) than standard methods. Consequently, educational outcomes assessments are relevant *only* if economic constraints are satisfied.

Many educators—and education funding agencies—strongly oppose this point of view on philosophical grounds. Unfortunately—as people with business experience know well—it is a statement of fact, not a philosophical issue. Denying it does not keep it from killing programs, it only keeps people from understanding why they died. The objective here is not to argue that economic constraints are a good thing but that we will not be successful until we accept them and learn to work within them.

The basic point is that computer use, or any other innovation, has costs, and these have to be balanced by savings elsewhere in the program.

The Math Emporium

The Math Emporium at Virginia Tech¹ was one of the first large math computer learning facilities, and to the best of my knowledge is still the largest by a factor of two. It serves over 6,000 students per semester with 550 computers and a yearly help staff budget over half a million dollars. At the time of writing (2009) it is in its thirteenth year of operation.

11.1 Educational Models

This section describes models for computer use in mathematics courses. These are given roughly in order of efficiency, with a brief discussion of economic factors. Numerical data is given in the next section.

11.1.1 On-line

The cheapest model is the on-line course. We do not use this because our experience is that dropout and failure rates are unacceptably high for a residential state university. To put it another way, success with our student population, as defined by institutional goals, depends heavily on easily-available, in-person help. We have experimented with online help but found it to be more expensive and significantly less effective. Presumably these problems will eventually be overcome, but at present on-line and on-site computer courses are significantly different in problems and goals and it is important to understand this.

11.1.2 Gigantic Lectures

We do not use gigantic sections with a single professor and flocks of cheap (undergraduate) graders so there is no data for this model. A shortage of large lecture halls is one reason, but we have tried it in the past and found the outcomes to be unsatisfactory. Computer-based courses with flocks of undergraduates working as computer-side helpers is cheaper and has better outcomes.

11.1.3 Computer-based

The first model currently in use is the computer-based course. The savings here is the time or salary of the classroom teacher. Some of this is redirected into help, and in fact students in our computer-based courses have access to more—and more timely—one-on-one human help than students in our traditional classes. We have precalculus, elementary linear algebra, and calculus for the life sciences in this format.

¹See <http://www.emporium.vt.edu>

11.1.4 Computer–tested

The second model is a traditional lecture course but with all assessment done by machine. The savings is the time the teacher or assistants would spend preparing and grading tests, homework, etc. In a large–enrollment, multi–section course these savings can be substantial. We have sections of a calculus course in this format and another in development.

Our computer–tested course has quite a bit of on–line reference material to support the learning–environment design of the tests, so it could be thought of as an evolutionary step toward a computer–based course. In fact in some sections as much as 30% of class used it as a computer–based course (skipped the lectures and did fine).

11.1.5 Computer–Enriched

A common model we do not use is the computer–enhanced traditional classroom. This is always over budget because it is an add–on with no compensating savings. At one time the department had the goal of computer–enhancing every class, and many classes were run in this mode. However we were unable to sustain the uncompensated extra load, and this is now voluntary and rare.

11.1.6 Computer Labs

Another model we have largely abandoned is the out-of-class computer lab, worksheet, or group project. As an add–on this is also unsustainably over budget: support costs of the computer component may be as high as for a completely computer–based course. I remain enthusiastic about the educational benefits of these activities and hope eventually to incorporate some form of them, but in the short term they are unusable for economic reasons.

11.1.7 Small Traditional Classes

The traditional ideal is a classroom with an experienced professor and 15 or so students. This was still possible when I was a student and it remains my personal favorite. I hope we will always be able to offer some upper–level courses in this mode, but it has been over–budget for nearly half a century and cannot be offered to the vast majority of our students. From this perspective the whole computer initiative is economically motivated. The fundamental goal is to do better than huge sections taught by adjuncts, but within the budget that forces such measures in the first place.

11.2 Program Costs

The following table gives relative costs per student credit hour in the Mathematics Department at Virginia Tech in 2003–04. These costs are the ones under

the control of the department (mostly salary and wages). Computers, for instance, come from a different budget and if we quit using computers the money could not be transferred to salaries, so computers are not included. Costs were computed by adding up actual salaries or wages, dividing by enrollment to get dollars per student credit hour and then normalizing so the traditional situation has cost 100. Traditional classes assume 40 students per section.

Traditional, tenure-track	100 (200) ^{note1}
Traditional, graduate student	60 ^{note2}
Traditional, 30% prof/70% instructor	57
Traditional, all instructor	39 ^{note3}
Computer-tested, 30% prof/70% instructor	32 ^{note4}
Computer-based (3,500 students)	18 ^{note5}

11.2.1 Notes

1. 100 corresponds to half salary, assigning the other half to research. Using the whole salary (as Deans sometimes do) gives a normalized cost of 200.
2. Graduate student figures are higher than instructors because they have lower loads and the cost includes tuition.
3. “Instructors” at Virginia Tech generally do not have PhDs, and have twice the teaching load of professorial-rank faculty. They are not “adjuncts” in the usual sense because they are full-time members of the faculty with open-ended appointments. In 2004 approximately 1/3 of the undergraduate student credit hours in our department were taught by instructors.
4. Teachers have 100 or more students per section or more sections, but no assessment responsibilities. This is generally considered a good tradeoff.
5. All materials on computers; tutoring and at-machine personal help is available but there is no classroom teacher.

The cost shown for computer-based or computer-tested courses does not include cost of developing materials, just as traditional course costs do not include textbooks.

Our materials are very stable and require little maintenance after development. I personally feel that more-frequent assessment and upgrades would be appropriate, and that this should be included in course costs. This would lead to modest increases in the numbers given above.

11.2.2 Scale and Efficiency

There are costs associated with the Emporium and not any particular course. Examples are:

- server and student-computer maintenance personnel;

- a director of operations;
- floor helpers to assist students working at the machines; and
- floor supervisors.

To explain this last, the Emporium is over an acre in size. It turns out that a full-time supervisor is needed to organize and efficiently deploy helpers to areas where they are needed. No doubt this would be obvious to the manager of a department store, but it was something we had to learn.

Courseware maintenance costs are associated to particular courses. Our costs after initial development and refinement are minimal, but whether the materials are commercial or developed in-house, they are software and require some maintenance.

The important point here is the way these expenses scale. Facility-related expenses (helpers, etc.) have a significant lower bound, but increase only slowly with the total number of students. Course-related expenses are nearly independent of the number of students in the course. In contrast traditional costs scale essentially linearly with the number of students. If a traditional section has 40 students then 200 students require 5 sections, 1000 students require 25 sections, etc.

Our experience shows that at large scales great efficiencies are possible with computer-based or tested courses. At smaller scales fixed costs will tend to dominate. For example it may be difficult to break even with fewer than 500 students in computer-based mode. This probably means sustainable computer-based education can only be achieved by deliberate institutional initiative: individual efforts cannot reach break-even levels.

11.3 Outcomes

Since costs of the Emporium program is substantially below those of traditional alternatives, it is appropriate to ask about educational outcomes. Are outcomes at least as good as with traditional alternatives, or have outcomes been sacrificed to get lower cost? This is an unexpectedly subtle question: it turns out that the greatest benefits have been outside the program.

11.3.1 Outcomes in the Program

It is almost impossible to make intellectually honest comparisons of courses before and after conversion to computer formats. Grades in some early courses did go up, but a significant amount of material had been omitted and when the computer versions were redone with equivalent content the grades went back down.

There is one exception: a course that was run with both traditional and computer-tested sections for five years, with common final exams. An extensive

comparison of performance in subsequent courses is in progress². Preliminary analysis suggests essentially equivalent outcomes. Detailed analysis seems more likely to indicate useful changes (e.g. “move the test schedule up one week”) than a significant difference.

I suspect that the “no-statistical-difference” phenomenon is a feature of students rather than of educational approaches. Students tend to do what is necessary to get their desired outcomes. More-effective approaches are likely to lead to increased student free time (which we do not measure) rather than better outcomes.

In any case I believe that Emporium courses would probably rate “acceptable” if evaluated purely on educational outcomes. I also believe they could be significantly improved. Unfortunately, upgrades are considered extra rather than part of ongoing course costs, so are discouraged by economic considerations.

11.3.2 Outcomes outside the Program

The efficiency data show that the Math Emporium has been a cash cow for the department and the university, by handling large numbers of students significantly below budget.

Shortly after the Emporium went into full production Virginia suffered a financial reverse. Much of this was taken from higher education, and the Virginia Tech math department lost 20% of its faculty positions even while enrollments increased. Ordinarily this would have led to increases in teaching loads, larger classes, and curtailment of upper-level and graduate offerings. However Emporium efficiency enabled the department to maintain the rest of the program and actually *reduce* teaching loads for research faculty!

The consequence is that the Emporium has had very positive effects on educational outcomes in the department as a whole, but primarily through indirect support for non-Emporium parts of the program. Ignoring economic aspects and evaluating outcomes only in the Emporium would overlook these very significant benefits.

11.4 More about Economics

“Economics” here is used as a shorthand for the relationship between costs and output. “Costs” are in dollars, sometimes translated into faculty time, and “output” is in student credit hours, with no reference to quality. It is not my idea to use these measures: they reflect realities of funding, at least in state universities. Both state legislatures and tuition-paying parents feel strongly that they are already paying enough for education. Resources are much more likely to be reduced than increased, and the job of the faculty should be to maximize quality within this constraint.

²*Downstream Evaluation of a Task-Oriented Calculus Course*, see web page for current version.

Consequences of this picture are explored in this section.

11.4.1 Goals and Evaluations

The points at the beginning of the article can be reformulated as:

- Economic measures should appear explicitly among the evaluation criteria for educational research and program development.
- Evaluations by other criteria should be conditioned on having appropriate economic outcomes.

At present, for example, grant applications usually require extensive assessment of educational outcomes and no cost accounting. Trials are usually resource-rich (grant supported), so good outcomes demonstrate only that one can do well with plenty of resources. This is neither surprising nor useful. In fact the great majority of approaches explored this way depend on these additional resources, and consequently will have no impact in the real world. They are even counterproductive when they are taken as justifying educational goals that are unrealistic with available resources.

At the other end of the spectrum it may turn out, particularly during startup, that a program has economic savings but educational outcomes mixed or weaker than richly-supported programs. Educational outcomes alone may suggest the program is unsuccessful and should be terminated. A better conclusion is that it has passed the hardest barrier and the next step is refinement to improve outcomes.

11.4.2 Internal and External Resources

Budget structures frequently complicate economic planning and evaluation. For purposes of discussion I distinguish between “internal” or flexible, and “external” or inflexible budgets.

For example, in discussion of economics of the Math Emporium the question frequently arises, “don’t the computers cost a lot?” I believe the program would be economically viable even if equipment costs were counted, but in fact this is irrelevant. The equipment budget is a separate part of the funding provided to the University. We can use it—or not—for equipment, but it cannot be used for salaries. This is a rigid, or external resource.

It can be dangerous to overlook this distinction. For example we use slightly more expensive machines with a stable and maintenance-friendly (non-Windows) operating system. Cheaper machines would have reduced external costs but increased internal costs since more department-supported personnel would have been required to maintain them. In our case the external budget is significantly better able to handle the difference.

11.4.3 An Example

To illustrate these points we analyze the NSF “VIGRE” program circa 2004 when the first version of this article was written. Large grants were made to departments that enriched their educational programs. Undergraduate research projects were the most expensive, at least an order of magnitude over budget, but others were also expensive and the net effect was to raise the cost (in faculty time) of “production” of student credit hours.

However the grants supported things like postdocs, not the enrichments. In our terms the grants were inflexible or external resources provided as bribes to encourage additional expenditure of internal resources. Internal-resource overruns had to come out of something else. Possibilities were:

- faculty research time budgets were reduced (i.e. the research effort pays);
- time spent on un-enriched educational duties was reduced (other students pay);
- some of the salary budget was shifted to adjuncts or other lower-cost ways to cover the extra hours; or faculty voluntarily work overtime (faculty and their families pay).

These programs may have been “enriched” but violating the constant-budget constraint stressed the systems in ways that one would think would be unattractive to the NSF. Naturally, the enrichments disappeared as soon as funding ran out, or as faculty become unwilling to continue making sacrifices. This approach to inducing long-term changes in instructional practice failed for simple economic reasons that should have been obvious³.

Some of the worst features of the VIGRE program have been modified, but the Education Directorate still seems to be committed to promotion of expensive methodologies and denial of economic constraints. Consequently their programs are unlikely to have long-term effect and their resources largely wasted.

11.5 Summary

The Math Emporium at Virginia Tech demonstrates that a computer-based undergraduate mathematics education program can be successful both economically and—eventually—educationally. Careful analysis of this success reveals some common causes of failure:

- Being over-budget particularly in demands on faculty time, for instance by using expensive course designs, will lead to failure.
- Inadequate attention to economic factors in planning and evaluation can lead to failure.

³This was obvious to some outside the Education directorate. A mathematics program officer described the approach as “homeopathic funding”.

- Undertaking program development without adequate support, and programs too small to take advantage of economies of scale will be problematic.

More generally, undergraduate education is highly budget-constrained, and failure to recognize this is probably the main reason there have been so few really successful innovations.

Chapter 12

Levels in a mathematics course

October 2008

Introduction

Our students are not well served by traditional course design. They come to us with diverse backgrounds, interests, degrees of engagement, and ability, but courses are one-size-fits-all: uniform assignments, tests, and—in principle—uniform grading scales. Few students receive optimal instruction and a significant number are seriously out of place. Student/course mismatches shortchange students and reduce course effectiveness.

The traditional way to address this problems is to offer several courses on essentially the same material but at different levels, say Standard and Advanced. However each course still has the problems on a lesser scale, and there are misplaced students: the best students in Standard should be in Advanced, and those who make F or D in Advanced should have been in Standard. We explain in §12.2 that this is an inherent problem with multi-level courses, and in particular cannot be addressed with better placement tests. There are also D and F students in Standard who should be in a lower level if one were available, and the best students in Advanced should be in a higher level. Offering multiple levels helps but doesn't solve the problem.

Our suggestion, in a nutshell, is to offer tracks at different levels *in a single class* rather than in separate classes. This would improve mobility between levels and avoid the misplacement problem. And if it can be made to work at all it should be possible to offer tracks at more levels than practical with separate courses.

The essay begins with discussions of the problems to be addressed. Performance diversity in a single class is discussed in §12.1 Outcome Diversity, and

problems with resolving this through placement classes at different levels are described in §12.2 Placement Tests Are Not The Answer.

The main idea is described in §12.3 Tracked Courses, using perspective from §§ 12.1–12.2 on the problems to be avoided. As usual with a clever idea the real question is whether or not it is practical. Some of the many difficulties are described in §12.4 Implementation. It is doubtful that this could be implemented in a traditional class with current student/teacher ratios. There are, however, long-shot scenarios, and the benefits would be so great that these are worth considering.

12.1 Performance Diversity

The bottom line in a math course is end-of-course performance. For example, students who make grades of F or D were in some way not well matched to the course and probably should have been in a different level. Outcomes can't be used to identify these students at the beginning of the course, and the ramifications of this are discussed in the next section. Here we discuss problems resulting from having students who *will eventually fail* whether we can identify them ahead of time or not. We also describe problems of a very different nature at the other end of the spectrum.

12.1.1 Under-performing students

Roughly speaking, under-performing students are those who end up with grades of F or D in a course. Actual grades are not quite the right measure because grades are often adjusted to avoid having a lot of Fs and Ds. A better description is: students who would have gotten F or D in the absence of such adjustment.

The big problem associated with under-performing students is that changes made to accommodate them undercut learning of other students. We begin with under-performing students in Advanced courses.

Skill-oriented math classes are important because skills—the ability to work problems—are vital for success in later coursework, and eventually for use of mathematics in technical professions. Grades in a skills course are supposed to reflect acquisition of skills. If there are a lot of students who do not acquire skills then in principle there should be a lot of failing grades. Before the 1970s (roughly) this was standard practice and these courses often did have high failure rates.

High failure rates are now considered unacceptable. Many factors contribute to this but one is the realization that skills courses are neither appropriate nor necessary for many students. Under-performing students may be misplaced, for instance to “offer them an opportunity”, not dumb. There are limits to how much it is reasonable to punish them for simply being in the wrong place or unable to take advantage of an opportunity. In any case a number of adjustments have been made to lower failure rates:

- expectations have been generally lowered;

- imprecise “understanding” may be accepted when skills are poor; and
- practices like extra credit, grade curves, dropping the lowest test, and soft homework scores are used to disconnect performance and grades.

The result is that course goals and grades have become ambiguous. Grades no longer indicate acquired skills, and it is unclear even to the best students that skills should be the key objectives. In effect the course goal has been changed to include “useful exposure”, and while this may provide general life benefits it does not prepare students for advanced work.

Standard-level courses also have students who get F or D grades, and course goals are distorted by adjustments made to accommodate them. These distortions cause less long-term damage than in Advanced courses because skills and preparation are not the main objectives. Further, these students are usually not “misplaced” in the sense of being in the wrong course because there usually is no lower-level course. We do *not* want to interpret “misplaced” to mean “don’t belong in any math course”. This is a qualitatively different issue than being in the wrong course and it is not appropriate for us to address it here.

12.1.2 Over-performing students

Over-performing students are ones who would have been successful in a higher-level course. Since this concerns hypothetical outcomes in a course they didn’t take, we can’t identify specific students as over-performers. In particular a top grade in the course they did take does not reliably identify over-performers.

Problems associated with over-performing students differ depending on level, and differ from under-performing problems in that they do not effect course goals.

12.1.2.1 Over-performing students in Standard courses

Traditionally the main reservations about multi-level course offerings concern insufficient upward mobility: students who at some point got put in the Standard level and can’t get out even though they would have done fine in Advanced. These students have, in a sense, been shortchanged by being deprived of opportunities available at the higher level.

This is an individual-benefit problem for specific students and does not have a negative impact on learning of other students.

12.1.2.2 Over-performing students in Advanced courses

These students are identified in even more hypothetical terms: success in some sort of “Gifted” course that generally isn’t even offered.

The big problem in this area is societal. The huge role technology now plays in our lives means we need a significant number of extremely capable people trained to the full extent of their ability. Our educational system is not meeting this need. The few specialized high schools that do offer “Gifted”

courses produce far more than their share of first-class scientists and engineers, even allowing for selective admissions. This suggests that the lack of very high-level courses in K-12 is a major part of the problem.

These students do not have a negative impact on other students. Further they have plenty of other opportunities so being shut out of the top echelons of science and technology is not a serious individual-benefit problem. It is only the societal problem that is severe.

12.1.3 Summary

Under-performing students impair the effectiveness of our educational system. Over-performing students are being denied opportunities, and in some cases being shut out of urgently needed technical leadership roles. Addressing these problems would seem to require offering instruction at more levels, and enabling easier and more appropriate mechanisms for mobility between levels.

12.2 Placement tests are not the answer

Most educators feel that if misplaced students are a problem then better placement tests are the solution. A key point is that placement has to be done at the beginning of the course, even if problems are most directly related to end-of-course performance. Justifications are:

- Placement decisions are limited by what we can measure. We can seek the best predictors of performance within this constraint, but there is no point in complaining about the limits imposed by this constraint.
- Since performance predictions are necessarily imprecise we should give students the benefit of the doubt. In other words deny admission to an Advanced course only if we are pretty sure the student will fail, and placement instruments need only be good enough for this.
- For whatever reason, educators put a lot of faith in placement tests.

All of these arguments are flawed. Placement is limited by beginning-of-course measures only if it has to take place at the beginning of the course. Mid-course placement is one of the advantages of tracked courses; we expand on this in §12.3. Giving students the benefit of the doubt maximizes individual opportunity but also leads to distortion of course goals and reduction in overall learning, as explained in the previous section. The last point concerns belief rather than an argument, and the objective in this section is to show this belief is unfounded.

12.2.1 False positives and negatives

Placement decisions can fail in two ways: false positives are students who get Advanced placement but turn out to be under-performers; false negatives are

students denied Advanced placement but who would have been successful at that level.

If the placement system has more than 10% false positives then, as explained in §12.1.1, Advanced teachers have little choice but to weaken the link between credit and performance. Therefore a high false positive rate undercuts the skills orientation of the course. Since the course then no longer meets stated goals, credit for it gives misleading input for later placement decisions and raises false positive rates in later courses.

The usual way to keep the false positive rate low is to have higher requirements. But this inflates the false negatives and in practice makes skills courses unduly inaccessible. The extreme is a placement test that can only be passed by those who already know the material. The false positive rate is near zero and the course would go very well, but the false negative rate is near 100% and the course serves no educational function.

Real-life placement methods are too imprecise for there to be any satisfactory balance between false positives and negatives.

12.2.2 Tests are Untested

There is little solid data on effectiveness of placement tests because the self-fulfilling way they are used makes them almost impossible to evaluate. False positive rates tend to be masked by instructors' changing grading criteria to keep failure rates acceptably low. False negatives are practically impossible to assess and usually ignored.

There are a great many factors that effect performance but are not measured by tests: procrastination, short attention span, poor work habits, not to mention alcohol, drugs, and the emotional turmoil of youth. It should seem silly to even hope for a test with false positives under 10% and an acceptable false negative rate. Nonetheless many educators seem to take it as an unexamined article of faith.

12.2.3 An Example, and Gateway tests

We describe a real-life example. Our second-semester engineering calculus course has 25–30 sections each semester. Some years ago a series of brief computer-based “skills” tests were introduced to assess learning consistency across sections. The first of these measured entry skills and was essentially a placement test, though it was not used that way. Data from several thousand students showed an impressive correlation between scores and course outcomes. This probably could have been used to justify using the test for placement, but the statistics hid an asymmetry. Essentially all students who failed the course had failed the skills test but the converse did not hold; most who failed the skills test did fine in the course. The test had a low false positive rate but a very high false negative rate.

The story takes an interesting twist. These tests are multiple-try. Each individual test is different, students can get unlimited practice copies, and the

proctored version can be taken multiple times with the best score counting. The data showed that students who initially failed but kept trying until they passed did almost as well in the course as those who got a perfect score the first try. This was taken to mean that entry skills are not immutable things that can only be measured and sorted, but somewhat malleable.

The test is now used as a skill-boosting “gateway”. Students who sign up for the course *must* pass the skills test in the first week to stay in the course. Most pass on the first try, but:

- A few percent of enrollees drop out without attempting the test for credit. Presumably they have decided—after looking at practice tests—that they will not be successful, so the test is helping with *self*-placement!
- A tiny number, less than 1%, attempt the test but are unable to eventually pass.
- The remainder—the false negatives of the original test—have to work to get their skills up to speed but do manage it.

Instead of a filter the test has become an instructional tool.

The tidy outcome in this example may depend on pre-filtering by the university admission process. Even so it does not reduce false negatives and positives enough to solve the basic problems of placement.

12.3 Tracked Courses

In a multi-level course students are sorted and placed in different classes for which they receive different credits. Tracked courses reverse this: students enter a single class, sort themselves into tracks as the course progresses, and only at the end of the course is a decision made about the credit received.

12.3.1 Basic Plan

For simplicity we describe a course with two tracks: Upper and Lower.

- Students entering the course are not assigned to a track, and beginning classes are not specialized to either track.
- Tentative track assignments are based on performance on the first major test. Lower-track students who want to be in the Upper track can retake an equivalent test to try to get the necessary score. Students who qualify for the Upper track can, if they insist, be reclassified as Lower-track.
- Subsequent tests are track-specific. Upper-track students with unsatisfactory scores are reassigned to the Lower track, but again with an opportunity to improve the score.

- If there are several sections of the course then sections can specialize after the first test. Students might have to change sections to be in a class appropriate for their current track.

At the end of the course each student will be in one of the levels, and will have test and other course scores. Grading and credit is handled as follows:

- Upper-track students receive Upper-track credit that qualifies them for more-advanced later courses, and grades A, B, or C depending on scores. Students who would have gotten grades indicating unsatisfactory Upper-track performance have dropped to the Lower track where more appropriate standards can justify better grades.
- Lower-track students receive lower-track credit and the usual A–F grades, unless there is a yet lower level or track for the under-performers.
- There is an element of choice for Upper-track students: any Upper track grade can be converted to an A in the Lower track.

The choice offered in the last point provides a safety net. Students interested in law or medicine or seeking admission to elite college or graduate programs often avoid serious math courses to avoid damage to their grade point average. This is unfortunate because these students are often quite capable of Upper-track work. They might even be lured into a technical profession: many people in mathematics and science ended up there because they took a tough course and liked it.

An important feature of this design is that marginal students make their own decisions. Students struggling to stay in the Upper track may decide they aren't *that* interested in technical careers anyway, and change their goal to getting an A in the Lower track. This is certainly better than going limp and dragging down the whole class. If they are determined to stay in the Upper track they are motivated to work harder and rise to the right challenge. Finally this decision is made in small steps—one test at a time—so they can see exactly what is required and make an informed decision.

12.3.2 Introduction to Proofs course

The previous section is implicitly aimed at K–12 and the first few years of college. This section suggests that the approach could also be useful at higher levels.

In traditional college math sequences there is a shift of emphasis after calculus from problem-solving to more abstract and conceptual reasoning, “proofs” for short. Most students find the transition to proofs uncomfortable and by and large only math majors attempt it. This is unfortunate since the generalized reasoning skills acquired this way are germane to cutting-edge work in any science or engineering field. Some top software companies, for instance, recruit PhD mathematicians on the principle that it is easier to teach someone with

high-level logical skills to use computers than it is to teach a computer expert to think on a higher level.

Until relatively recently the custom was to introduce students to proofs in a sink-or-swim way in courses on real or complex analysis or modern algebra. This was tough on students but satisfactory numbers made it through. Expansion of graduate programs in the 1970s and later weakening of lower-level education made this approach unworkable and many programs introduced “Introduction to Proofs” courses to help with the transition.

There is a new difficulty with the problem-proof transition. Further softening in lower-level courses has meant that there are fewer students with the preparation and discipline to make the transition, even among math majors, and even with help from a Proofs course. Faced with the need to keep numbers up and programs viable, some departments have softened their proofs courses. In effect they offer exposure credit to boost low skills scores. Naturally this degrades the end product.

Some university undergraduate math programs are now almost incapable of producing students that would qualify for their own graduate programs. Elite graduate programs sustain quality by recruiting foreign students. Many less-elite graduate programs are being softened to be accessible to Americans because the alternative is to close down. In other words the upper end of our mathematics educational system is starting to erode.

Using tracks, for instance Professional and General, in a proof course would help with this. The General track would be quite satisfactory for prospective K-12 teachers and the less math-intensive sciences. The Professional track would require the discipline needed for further math and math-intensive science and engineering, without harming the General-track students.

12.3.3 Both Tracks and Levels

We have used “multi-level” for separate courses with placement at the beginning of the term, and “tracked” for a combined course with placement at the end of the term. Up to this point the two have been compared directly in order to make the differences clear. In practice, however, the two approaches are complementary and often would be used together.

- A class with a serious skill component will spend a lot of time doing things non-skills students generally dislike. A lower track would reduce the impact of skills materials but would not make it more relevant. A non-skills course can focus more on interest and enrichment. Separate courses for the two levels are appropriate.
- There could be tracks in each level, with significance partly defined in terms of subsequent level placement. Upper-track credit in the upper level would be required to qualify for the upper level in the next course; lower-track students would move to the lower level.

- This does re-introduce the mobility problem that is one of the virtues of tracks. It should be less problematic because students have had a lot of input into their placement, but some sort of upgrade process should be provided.

12.4 Implementation Problems

We list a few obstacles to implementation of tracked courses. Familiar problems such as developing texts and syllabi are not discussed.

12.4.1 Resource Constraints

Informal use of tracks was common in the one-room-schoolhouse days because there were too few students to justify separate classes. This is rarely possible now because it requires unrealistically low student/teacher ratios, willing and well-behaved students in K-12, and may require lower content density in college courses.

Formal introduction of tracks will not solve the student/teacher ratio problem. Teachers with typical-size classes are rarely able to focus on a subgroup for an extended time. Further, if test preparation, grading, etc. are done by hand then tracks could double the time required for this. Trying to introduce tracks in such cases will predictably lead to failure and should not be attempted.

Possible exceptions are:

- “Gifted” tracks in upper-level courses. The student/teacher ratios are generally low, students are cooperative, and very few students would be involved in the upper track.
- Lower tracks in low-level courses. These could be accomplished simply by changing the grading scale at the lower end, without changing materials or presentations.
- Computer-tested courses.

12.4.2 Institutional Barriers

Most institutions will be uncomfortable with the idea of students signing up for a course without knowing which course it is. They may also be uncomfortable with leaving course-credit decisions (i.e. end-of-course placement) in the hands of the faculty. Resistance by credit score-keepers (Registrars et al.) will make trials of the approach difficult.

12.5 Conclusion

Traditional course structure evolved to support a single goal, and the traditional single goal was good outcomes for a relatively small number of students.

Education now has another, conflicting goal: modest outcomes for essentially everyone. The traditional structure has been unable to do justice to both goals at once. Tracks may provide a way to resolve this by offering several grading criteria and letting students play a significant role in deciding which goal and associated grading criterion is best for them.

Tracks should be relatively easy to implement in computer-tested courses. The extra burdens of course administration and multiple assessments make the approach infeasible in most traditionally-tested courses.

Chapter 13

Teaching vs Learning in Mathematics Education

Feb. 2009

Introduction

Most educators see teaching and learning as two sides of the same coin: we teach so they will learn, end of story. It was hard to compare while everyone was doing more-or-less the same thing. Technology has changed this however, and I'll give examples that suggest we are far too focused on what happens on our side of the desk. It looks as though teaching and learning were never as closely linked as we wanted to think, and the gap will widen unless we really focus on students and learning, particularly long-term learning, and *not* through the lens of teaching¹.

13.1 Goals vs. Responsibilities

The way we organize it, math begins with arithmetic and the rest of the subject is built on this. Arithmetic instruction should, therefore, provide a foundation for learning in the rest of mathematics. We need some careful terminology to describe how this should work.

13.1.1 Generalities

Teaching or learning *goals* are usually understood as short-term, specifying deliverables, and determined by the teacher or course designer. Teaching goals specify that a teacher should do certain things, while learning goals specify that

¹This is not a new point, see [Association for Educational Communications and Technology](#). We, as a community, might have avoided a lot of grief if we had paid more attention to it.

students should end up with certain things. Traditional goals in arithmetic mostly concern working problems.

The objective “provide a foundation for further learning” does not qualify as a goal in this sense so we refer to it as a *responsibility*. More explicitly, responsibilities are long-term or downstream, defined operationally rather than explicitly, and not a matter of choice. In principle, goals should be chosen so that responsibilities are fulfilled. It is certainly not clear how “work problem” goals end up meeting “provide foundation” responsibilities, but in traditional courses it seems to work.

A final, and absolutely vital, general point is that “students” are individuals. We have goals for and responsibilities to individual students, and different individuals might need substantially different goals and responsibilities. Discussions that don’t stay grounded in this reality encourage one-size-fits-all thinking that is a real disservice to students.

13.1.2 Calculator arithmetic

Returning to arithmetic, there have been two recent developments. First, calculators enable more students to achieve “work problems” goals more easily and with greater accuracy; and second, “understand what they are doing” has replaced rote computation as a goal². The good news is that in these programs goals are being met better than ever. The bad news is that long-term responsibilities are not being met. Number-sense and symbolic-skills deficits in students from these programs were a major issue in the K–12 “math wars” and are a serious concern at the college level.

Apparently a disconnect developed between goals and responsibilities. What happened and what can we learn from it?

The first lesson from the disconnect is that “work problems” by itself is evidently not enough to “provide a foundation”. Apparently there was something about the *way* traditional students work problems that was important³. But rejecting calculators is not a satisfactory response. We urgently need to understand *how* by-hand arithmetic supports later learning. Perhaps we can fix the calculator approach by adding the missing factor to teaching goals. This might improve the traditional approach as well⁴.

The second lesson from the disconnect is that there are several ways to address responsibility problems. Responsibilities of one level can be thought of as preparing students to accomplish goals at the next level. If goal changes at the lower level no longer meet this responsibility then one response is to adjust the lower-level goals. However it is also possible to change the definition of “responsibility” by changing the goals *of the higher level*. This was the strategy in K–12 calculator-oriented curricula. They adjusted goals at all levels to “take advantage of calculator skills” and de-emphasize traditional goals not supported

²Reference to NCTM standards?

³For a guess see “K–12 calculator woes”.

⁴If the guess in the previous footnote is right then traditional approaches are indeed far from optimal.

by calculators. The result was a system with internally consistent goals and responsibilities.

The goal-changing approach to responsibility eventually fails. College courses have responsibility for preparation for study in science, engineering and advanced mathematics. These responsibilities are determined by the demands of the subjects and can't be negotiated. Meeting these responsibilities strongly constrains choices of short-term goals in college courses. Working down the chain, college course goals should establish end-of-curriculum responsibilities for K-12. There we have a train wreck: the calculator-oriented K-12 community (at least) seems to have no understanding of, nor interest in, these external responsibilities.

So far this lesson seems to concern responsibility, but there is a teaching/learning core. *Responsibilities* concern learning because the teacher is not in the picture when responsibilities fall due. However the K-12 education community is intensely teacher-oriented. The "responsibility" idea is not part of the world view and even hard to formulate sensibly.

A comforting corollary of this last point is that the school/college mismatch comes from a lack of understanding rather than conscious irresponsibility. This is further illustrated by a common K-12 response to college-level complaints: we should follow their lead and adjust our teaching goals to "take advantage of new skills" rather than bemoan the decline of old ones. Our unwillingness to do so looks like a reactionary attachment to the past; it doesn't occur to them that it might result from constraints of downstream responsibilities.

The third lesson from arithmetic concerns why taking understanding as a teaching goal did not improve outcomes, and in particular why it did not replace the mysterious benefit of hand arithmetic. The reason is not deep. Over the millennia mathematicians have found that in order to support learning "understand" must be given a rather strong meaning, including "make the solving of problems straightforward". K-12 educators use the word in a much weak sense that does not imply skills. They use a meaning already known (by mathematicians) to be dysfunctional for mathematical learning!

To connect this to the teaching/learning theme note that the teaching point of view suggests a lot of flexibility in choosing goals. One can choose the meaning for "understand" and there is no obvious reason why one should not take a weak one that is easy to achieve. Furthermore one can formulate teaching goals to address any definition.

The learning point of view is much more constrained. First, the meaning used for a word must accomplish longer-term responsibilities. Second, in order to incorporate something into learning goals it must be visible in outcomes, i.e. be testable in some way. Even *trying* to use the math-ed definition for "understand" would have revealed it's inadequacy. This doesn't solve the problem though. The mathematical meaning satisfies responsibilities and is testable but is far too demanding to be useful at this level. We are taken back to the point in the first lesson: we need a functional explanation of *how* by-hand arithmetic supports later learning before we can identify good learning goals.

13.1.3 Computer calculus?

I used calculator arithmetic in the example above because the response of the K–12 community amplified rather than fixed the problem, and consequently it gives a relatively clear signal at the college level. Similar things can happen at any level. For instance, do calculus students learn more from by-hand techniques of integration than just how to evaluate integrals? Use of computer integration packages will let us meet teaching goals more easily and more often, but will they undermine long-term learning?

I have learned that I am not smart enough to anticipate something like the calculator–arithmetic problem. Without this example I might have used computers to screw up my integral calculus course and not known there was a responsibility failure unless someone at the graduate level or in a client discipline pointed it out. But I should be smart enough to learn from the example. If I screw up integral calculus now it will be *my* fault.

13.2 Improved teaching vs. Improved learning

College classes over 100 are now common. Traditional teacher–student interactions are impossible, but a number of technologies have been developed to provide substitutes. One is classroom polling (clickers). A recent article in the journal *Science*⁵ describes how a biology professor used clickers to show that students can learn by talking to each other. Another approach assumes students have laptop computers, as is now common in our engineering courses. Software enables the professor to send material to all students or specific students; receive questions or comments from students; import material from student computers to assess later or to display and discuss, etc. Some professors are quite enthusiastic about this.

These technologies improve the teaching experience. Do they improve learning? Is having active discussions or getting students to talk to each other really an effective use of class time⁶?

The answer depends on why 100 students were crammed in the room in the first place. We know that students learn less in big classes so this practice is either restricted to courses with modest goals or is forced by economics: resources are so scarce that we cannot afford to split the class into two sections of 50, or three sections of 34. In classes with modest learning goals student engagement may be as important as actual learning and these practices may be appropriate. In classes with more ambitious goals (often the case in math) these practices may be too inefficient to be appropriate.

We expand on the efficiency concern. When a teacher interacts with one student he is to some degree neglecting the others. Interacting with one in a class of 100 is to neglect 99. Further, interactions with any one student will be

⁵“Why peer discussion improves student performance on in-class concept questions”, *Science*, vol. 323 pp. 122–124.

⁶The benefits of student interactions are not in question and did not need rediscovery. But would it be better to promote this outside the classroom?

very rare so this does not address individual needs for help. It may not be quite this bad: if ten students benefit from the interaction, and only 50 were really following the lecture anyway, then student benefit decreases only by a factor of 5. And note that “benefit” is not the same as “engage”. Students can be engaged (or entertained) without getting any particular benefit, and again this can be positive if it works⁷ and there are no efficiency concerns. However for learning purposes, teacher–student interactions in large classes are inefficient at best, do not effectively address student needs, and are usually a massive waste of student time.

The extreme learning–oriented view is this: think of the teacher’s time, or maybe the teacher’s salary, as a resource. Is a traditional class the best way to use this resource to get learning? In some cases there are already computer-based systems that would do better. The message I think we should be getting, in math anyway, is that there is a point beyond which teaching in the traditional sense is no longer a satisfactory path to learning, even if it can be made “engaging”.

13.3 Computer teaching vs. Computer–based learning

Most courseware is developed by experienced educators, which is to say people with a lot of classroom expertise. It shows: most computer courses are modeled on traditional courses and the computer is seen as an “electronic teacher”.

Ten years spent watching students trying to deal with courseware has convinced me that this point of view is wrong. Students have to take an active role in computer-based learning. They seem to have “learning instincts” in the sense that there are consistent behaviors when they are ready to go to the next stage, get stuck, etc. Sometimes there are several different patterns. The point is that none of these patterns match classroom practice.

We have to think of the learner as the center of the process. Not think “what should we have her do next” but “how might she want to approach the next task?” Watch and find out rather than extrapolate from classroom experience. And then make sure the way is clear and tools designed to work the way she wants to use them are at hand.

13.4 Information delivery vs. Diagnosis

What is a teacher’s core mission? Most would give some version of “information delivery” and most classroom practices fit this description.

Students now have many sources of information. I have seen students look something up on Wikipedia rather than try to find it in the course text. Web materials and computer courseware can do a good job of providing information

⁷I’m not sure my students were particularly entertained.

in a variety of media and at convenient times. Are teachers irrelevant, or is there a better description of the mission?

I believe our principal mission should be “help with *problems* of information delivery”. Students learn relatively easily but the learning is usually flawed. What we can do that machines cannot is diagnose and fix learning errors. The key, again, is a shift of emphasis from teacher to learner.

The computer-side help system in the Math Emporium⁸ illustrates this point. In a nutshell the help goal is “fix and run”. The helper listens carefully to diagnose the student’s specific problem, says the minimum needed to get them past it, and leaves.

Experienced teachers have a hard time doing fix-and-run. They want to say “let me explain this to you” and give a mini-lecture. The answer to the student’s problem is in there somewhere but neither the teacher nor the student know where. The teacher didn’t diagnose the specific problem, and the student has probably already heard a lecture that didn’t work. Or the teacher will say “I’ll show you how to work this problem”. The student’s work, good as well as bad, is discarded. The new solution may help but the student is often left with a flaw that will surface again later. It is very hard for experienced teachers to listen instead of talk, but this is the key to learner-oriented education.

I myself have thirty years of classroom teaching whispering in my ear “give your insightful lecture”. As with advice to my children about their boyfriends and girlfriends, I’ve had to learn that an insightful lecture is often not the best path to learning.

13.5 Summary

Technology has enabled us to make some pretty bad mistakes. In the long run this is all right if we recognize and correct these mistakes. But one of the lessons seems to go to the very core of the way we see ourselves: teaching is not the same as learning, and changes that we think improve teaching may actually degrade learning. Can we make the transition from “teachers” to “learning facilitators”?

⁸Virginia Tech Math Emporium, <http://www.emporium.vt.edu>

Part III

The Subject Level

Chapter 14

Professional Practice as a Resource for Mathematics Education

Chapter 15

Updating Klein’s ‘Elementary Mathematics from an Advanced Viewpoint’: content only, or the viewpoint as well?

April 2010

15.1 The Question

The basic point of Felix Klein’s famous 1908 book was that the rigors of professional practice had required mathematicians to develop much clearer and deeper views of mathematics than those of the ancients, and these clearer views could be a powerful resource for elementary education. He gave rich and convincing demonstrations of this idea in his book and other work, and this has been one of the main influences in the area.

Mathematics has moved on since 1908, and the ICMI Klein Project¹ was formed to develop a modern version of Klein’s book. The question here is: should the objective be more mathematics from *Klein’s* “advanced viewpoint”, or should the viewpoint be updated as well? The point is that the rigors of professional practice in the intervening century have pushed contemporary mathematics as far beyond Klein’s viewpoint as his was beyond that of, say, the sixteenth or seventeenth century². Could *contemporary* viewpoints be a corre-

¹See <http://www.kleinproject.org/>.

²In fact much of twentieth century mathematics would have been literally incomprehensible to Klein because he was committed to an ineffective approach to concept formation, see §15.3.

spondingly better resource for elementary education?

At present we do not know how to formulate such a viewpoint, and there are substantial barriers to even exploring the idea. The real question for the Klein Project seems more modest but would still be a challenge: should the viewpoint be made *upward-compatible*, to reduce barriers to eventual modernization?

General perspective on the issue is given in the next section. Section §15.3 illustrates, in the context of concept formation, what might be involved in a contemporary approach to elementary education³.

15.2 Perspective

This section concerns generalities. The following sections, on concept formation and methods of working, clarify what might be involved in a contemporary viewpoint on elementary mathematics, and the final section discusses what upward compatibility would involve.

15.2.1 Historical Baggage

Modernizing the “advanced viewpoint” on elementary mathematics is an emotionally charged topic, and this is partly due to Klein.

Much of the changeover to contemporary methodology took place early in the twentieth century. The transition was traumatic for the community and Klein was one of many vocal and determined opponents of these changes⁴. At the time it seemed to be a matter of philosophy and taste, but in hindsight we see that the changes were *forced* by increasing difficulty of the mathematics and ambition of the profession. Klein lost the struggle because his viewpoint was inadequate for twentieth century mathematics. In any case his viewpoint is that of the late nineteenth century; much of it was professionally obsolete by the time he wrote his book on education; and Klein was well aware of this and not happy about it.

It seems that Klein was determined not to lose the same struggle at the elementary level. He built his well-honed arguments against the new methodology into the philosophical foundations of his viewpoint, and in effect demonized the new methods. He was so successful that most educators still view them with fear and loathing. Moreover his viewpoint—by design—cannot be modernized without actually *reputiating* a good deal of it, and the very idea is met with hostility.

One result of this hostility is that it is virtually impossible for educators to get funding for, or publish, research on the use of contemporary mathematical methods. This is a barrier that needs to be lowered.

³A great deal more detail, explanation, justification, and suggestions are given other essays, see the Education, and History and Nature of Mathematics pages at <http://www.math.vt.edu/people/quinn/>.

⁴See for example §4.2.1.1 (pp. 197199) and §4.8.3 (p. 277) in Jeremy Gray's *Plato's Ghost*, Princeton U. Press 2008.

15.2.2 Ineffective Response

Articulate arguments against the new methodology were developed early in the twentieth century and have been refined since. In contrast there are almost no coherent arguments for, or even accounts of, this methodology. The ones I have seen are not even convincing to me.

There are reasons for this incoherence:

- If they think about it at all, mathematicians feel that the obvious power and success of contemporary mathematics makes external justification or defense unnecessary.
- Explicit *internal* justification or defense is also unnecessary. These practices became standard because they are much more effective, and natural selection defends them in the sense that people who try to do advanced mathematics using the old methods usually fail quickly.
- Previous methodology was usually the result of conscious philosophical investigation so was well-articulated at an early stage. Contemporary practices are the end result of a natural selection process that not only was not consciously driven but was *opposed* by philosophers.
- Finally, the methodology is most effective when it is internalized and invisible to the user. Invisibility makes effective explicit description difficult, and attempts so far have been incoherent, off-base, or both.

In any case the articulate arguments are still the ones against contemporary methodology. The people making these arguments—philosophers, educators, hold-overs from the nineteenth century, cognitive psychologists, some physicists and applied mathematicians, etc.—are ridiculed and dismissed as irrelevant by the professional community. But they have not been effectively answered, and their views are still accepted outside the mathematical community.

15.2.3 Is Change Needed?

Attempts to use modern methodology in education have been unsuccessful and this is generally interpreted as showing that Klein was right to oppose it. However there are reasons to be cautious about this conclusion.

The first point is that traditional mathematics education has not been satisfactory either. There are bright spots but overall results fall far short of needs. Twenty years ago in the US the NCTM developed “Standards” that strongly emphasized Kleinian ideals such as building on intuition, preferring understanding to rote work, and connecting to applications and the physical world. This has been enormously influential, but outcomes have declined. Excuses for the decline include poorly prepared or indifferent teachers, disengaged parents, lazy and distracted students, bad materials, in fact everything *except* the basic methodology.

Mathematics itself has been incredibly successful in the last century, but had to abandon this methodology to achieve success. Perhaps the continued use of nineteenth-century methodology is one reason education has been unable to improve much on nineteenth-century outcomes.

Another concern is the “pipeline” problem⁵. All school children study mathematics but very few go far enough to meet needs in subjects that depend on serious use of mathematics. The US has depended on mathematical high-achievers from other countries to make up its own deficit for almost half a century, and this dependence is growing because the pipeline losses are growing.

My concern is that the current pipeline *by design* has a nearly complete blockage at the transition from elementary to advanced mathematics. Basing elementary education on nineteenth-century principles means that *every student* who wants to go much beyond calculus must go through the wrenching methodological change that Klein himself could not manage. If we want better throughput then we need a more modern pipe, or at least one designed to connect better with the modern pipe later in the system. In other words, elementary education with a viewpoint either based on, or upward compatible with, contemporary methods.

The pipeline problem may be connected to another issue. It is often lamented that teachers of school mathematics rarely have any exposure to advanced topics. But their education courses teach them that the methods needed for success in such courses are “wrong”. What they learn in education may make them fail in mathematics!

The final point concerns failures of “new math” attempts to use modern methodology in education. These were designed by mathematicians, and we have also seen poor results—most recently in California—when mathematicians try to design traditional programs. It seems to me that educational incompetence and underestimation of the difficulty of innovation *by mathematicians* are more than enough to explain these failures. The methodology should not be blamed. This is not to say it is the answer, just that it has not had a good test and should not be ruled out.

15.2.4 Conclusions

The first conclusion is that it is not yet possible to develop a contemporary viewpoint on elementary mathematics.

- Lack of success to date means there are currently no good models, and suggests that any such development will be long and difficult.
- Traditional hostility to the idea means that the methods themselves are foreign to nearly all educators.

The second conclusion is that there may be some urgency to development of a viewpoint that is upward-compatible, at least in the sense that it is not

⁵See the Pipeline Project of ICMI at <http://www.mathunion.org/icmi/other-activities/pipeline-project/>.

hostile to contemporary methodology.

- The current entrenched hostility makes it impossible for educators to even try contemporary ideas. Actual innovation from within the community is unlikely for at least a generation after hostility has abated.
- This hostility also means innovation cannot come from outside the community, even if by some miracle a mathematician were to get it right.

15.3 Example: Concept Formation

The old and contemporary methodologies differ substantially in their approach to development of mathematical concepts. The approaches, and arguments for and against them, are sketched in this section.

15.3.1 Definitions

19th century: Intuitive ideas, either innate or abstracted from experience with the physical world, can be refined to give concepts on which elementary mathematics can be based. Experience with elementary mathematics develops new intuitions that can be refined to give a basis for more advanced work.

Contemporary: Concise formal definitions are like seeds that contain the DNA of a concept. Working with the definition is like planting the seed and tending the sprout. Physical or intuitive context may clarify the purpose of the mature plant and guide development, but if there is a conflict between preconceptions and DNA, DNA wins.

15.3.2 Arguments for the old approach

The old approaches have been obsolete in professional practice for a century. As a result arguments for educational use of old approaches tend to be just as much arguments *against* the new. Further there are two versions: the strong form asserts (following Klein) that the change in professional practice was a mistake. The weak form is that regardless of the benefits to professional practice, the nineteenth century remains a better model for elementary education.

The main arguments are:

1. Connections to intuition and the physical world makes concepts easily accessible, particularly to young children.
2. Axiomatic definitions, in contrast, are artificial and not easily accessible because they are unnatural in a cognitive sense.
3. Connections to the ‘real world’ makes the importance and meaning of a concept clear, and this should be reinforced with word problems and applications.

4. Axioms are matters of convention and more likely to stifle intuition and understanding than support them. Their significance is unclear and indeed to the extent that they are abstract and disconnected from reality they may not have any real meaning.
5. Many mathematicians, including some of the most powerful, depend on intuition and understanding rather than axioms. The best educational goals should be this sort of intuition and understanding, not axioms.

15.3.3 Arguments for the contemporary approach

These arguments are my own because, as mentioned above, I find most of what was previously available incoherent and unconvincing. I would appreciate feedback if they still seem incoherent. They are organized roughly as responses to the arguments above.

1. Formal definitions will not always be appropriate, but when they are they give more precise and more usable concepts, much faster. The reasons have to do with primitive features of human learning⁶:
 - It is easier and faster to learn something new than to find and repair errors in pre-existing ideas. Beginning anew with a definition, and taking care that the concept develops accurately, takes advantage of this. Beginning with a vague intuition commits one to a lengthy and difficult repair process.
 - Repeated finding and repair of conceptual errors often leaves conceptual scars (associations with or confusions about the process) that inhibit full internalization and fluent use.
2. Mathematical concepts based on definitions often have powerful features that were not part of anyone's intuition⁷. As a result the mathematics often illuminates naive intuitions far more than naive intuition illuminates mathematics.
3. Axioms certainly can be arbitrary, but the standard ones appropriate to elementary mathematics are not. They have been carefully crafted over long periods to optimize speed and precision of the concept-development process. They are also designed to maximize the effectiveness of the concept, sometimes in ways that are not evident for a long time.
4. Mathematical concepts should be thought of more as tools to use to *achieve* understanding, than as things to be understood. In other words the appropriate goal is to learn to use them rather than "understand" them.

⁶See *Cognitive Neuroscience and Mathematics Education* for discussion and a plan for further exploration.

⁷For an example see the discussion of fractions in *Proof Projects for Teachers*.

- When concepts are well-internalized people do feel that they “understand” them and they have “meaning”. But these are psychological expressions of mastery. They may indicate that a goal has been reached but are not themselves goals.
 - Abstraction or elaboration may make a concept a more effective tool, and thereby enable more understanding. It may also make the concept harder to internalize. In practice this leads to tradeoffs between power and accessibility. The common view that abstraction makes a concept “harder to understand” obscures the utility aspect and prevents finding a good balance.
5. Finally it is true that many powerful mathematicians work intuitively. It is false that these intuitions are the ones they started with. These intuitions are the result of extensive and intense work with axiomatic definitions that are so completely internalized that they have become invisible. Vast experience results in complicated manipulations being handled by the subconscious, and this appears to the conscious mind almost as direct perception or magically effective “understanding”. Unfortunately there are no shortcuts to this blessed state. I have seen many attempts to get “high-level understanding” without extensive low-level work, at levels from elementary algebra to graduate study to professional mathematics, and they have all lead to dysfunctional concepts and dead ends.

Part IV

The Educational Theory Level

Chapter 16

Dysfunctional Standards Documents in Mathematics Education

December 2004, revised November 2008

Introduction

This essay grew out of a meeting held in Park City in July 2004 to review K–12 Mathematics Standards Documents from the 50 US states. The objective was to look for commonalities and in particular determine the extent to which the 1989 NCTM publication *Curriculum and Evaluation Standards for School Mathematics* had acted as a template for the development of such commonalities. In fact this publication has been very influential, but it has been interpreted in so many ways—particularly in the upper grades—that the commonality seems to be more of language than content. Further there seemed to be little reason to think this influence has made these documents more effective.

Here we back up a bit and ask: what is the point of a standards document? Who is supposed to read it, and what do they need from it? In §16.1 we list some jobs to be done. These jobs fall to the standards document by default—nothing else can do them—so standards writers cannot decide whether or not to take on these jobs, they can only decide how well to do them.

§16.2 concerns standard high-stakes tests. These seem to be necessary as the “enforcement arm” of the standards process¹, but current tests are often counterproductive and linkage to standards is poor.

Even this brief discussion makes clear that genuinely effective documents will be very difficult to develop and are beyond the scope of current procedures. We may have to seek other ways to improve our educational system.

¹See [The K–12 Math Test Conundrum](#) for a brief discussion.

16.1 Roles for Standards Documents

Standards documents provide organizational frameworks and common reference points for teachers, administrators, curriculum developers, textbook writers, test developers, and indeed anyone interested in the system. To be effective they must be interpreted unambiguously and consistently by all parties.

16.1.1 Tests

The most problematic relationship is between standards and tests. In principle the high-stakes system-wide tests coming into wide use should reflect standards. In practice standards are vague and often unrealistic statements of goals that give little hint how they should translate to a test.

For example some state standards include enough probability and statistics to justify the use of a college final exam as the state test. This is clearly unrealistic, and in fact very little of the material appears on actual tests. Teachers obviously find old tests much better guides to the outcomes expected. Subsequent tests are expected to be consistent with the old ones, and to accomplish this the test designers refer not to the standards, but to the old tests. As a result old tests become *de facto* standards. To the extent the official standard differs from tests it becomes irrelevant.

There are many well-known disadvantages to teaching to a test, and some are discussed in §16.2. But with high-stakes testing this will happen unless a teacher can use the official standard to anticipate tests in detail: what will *not* be covered as well as what might. At the very least this would require a large number of sample problems and careful attention to what can be realistically accomplished in a typical classroom.

16.1.2 Textbooks

One of the main influences of standards documents is guiding the selection of textbooks. But conformity to standards is difficult to determine in the best of circumstances and nearly impossible when standards are unclear. To deal with this some publishers and state departments of education have developed a bizarre convention: the publisher prints at the top of each page the standard ostensibly addressed by the material on the page. The department of education then checks to be sure each standard appears at the top of at least one page. This can hardly be thought of as quality control.

Linking texts and standards can also enforce a disconnect between texts and teaching: teachers take old tests as *de facto* standards, tests differ from official standards, and texts conform to official standards. In any case the standard does not serve as a useful common reference point.

Matching texts to standards would be possible if standards are detailed and stable, and worthwhile if they also correspond to classroom practice.

16.1.3 Coordination

A common observation in comparisons of educational systems is that US programs have far more repetition. Some of this is by design but some may be a consequence of the standards system.

Frequently material is described in multi-year “bands” or “threads” rather than for years or smaller increments. This means teachers cannot count on material having been mastered until the end of the band. Or put another way, it is consistent with the standard for teachers early in the band to pass students who have not assimilated the material, essentially guaranteeing that it will have to be repeated before it can be used later.

To avoid repetition a standard must not only specify the learning goals in a class, but also enforce discipline in getting it done by specifying that the material *not* be repeated in non-remedial classes later in the curriculum. In the absence of standard curricula, standards documents are the only way this sort of coordination can be accomplished.

Standards can coordinate content as well as timing. Material learned in early grades is needed in later grades. Material learned in later grades is needed in college or the workplace. Unfortunately it is common to find that teaching methods or simplified problem sets focused on a particular level do not effectively support the needs of later levels. Standards documents are the ideal place to address this.

For instance when specifying that multiplication of multi-digit numbers should take place at one level, the standards might also recommend that this be done in a way that will support multiplication of polynomials at a later level. Or when students first learn to factor polynomials they usually see many with integer roots because these are easy to do. Some students get the unfortunate impression that quadratics usually have integer roots. A standard could have a warning about this and require a significant number of problems with irrational roots.

For standards to be successful in coordinating a program they must be detailed, explicit, and stable.

16.1.4 Process and Outcomes

The discussion above suggests that to be effective a standard should describe testable outcomes in considerable detail. In some cases testable outcomes and non-testable supporting activities occur in different courses or grade levels and organizing this may be part of the job of a standards document. However for the most part prescribing non-testable activities is likely to be counterproductive:

- Most teachers feel that figuring out how to meet testable goals via non-testable activities is a pedagogical or curricular issue properly the domain of teachers.
- On a practical level anything not explicitly labeled “not tested” is a liability for teachers: there is always a risk that some test writer will figure

out how to test it in an unexpected way.

- Non-testable goals may act as loopholes through which students receive credit and advance without acquiring skills. Lower skill expectations are certainly appropriate for some students but this should be explicit and managed rather than hidden in loopholes.

As an illustration of the last point, it sounds right to say “understanding is more important than rote mechanical skills.” However there are several ways to interpret this. College teachers would take “understanding” to include effective skills, so an inability to work problems implies a lack of understanding. In contrast standards documents almost universally use “understand” to mean “exposed to but not expected to work problems with”. “Know” is frequently used the same way though occasionally it means “able to reproduce” (as in “the student will know the formula for the area of a rectangle”) or “able to identify among three alternatives” (on a multiple-choice test). In this interpretation “know” and “understand” are not linked to testable skills².

Students, teachers and parents may reasonably infer that “understanding” is a separate—and possibly superior—pathway to success, distinct from mere skill acquisition. They feel cheated that high-stakes tests and college teachers do not reward such understanding. However when the rubber hits the road in later courses or real life, the skills needed are the ones that can be tested. Students promoted on the basis of nonfunctional understanding are at a disadvantage.

These problems can be avoided if standards documents focus on testable outcomes. Sometimes more careful use of language may help. However this not the whole solution since legalistic precision often leads to legalistic obscurity, and is more useful for fixing blame than for preventing problems. A better approach would be to illustrate every testable expectation with a representative sample problem, and explicitly link untestable activities to testable ones later in the curriculum.

16.1.5 Mathematical Structure

The points above are not subject-specific and may apply to other problem-oriented subjects. Mathematics does have some subject-specific features: first it is *cumulative* in that *essentially all* knowledge and skills learned at one level will be needed at later levels. Second it has a lot of *abstract logical structure*. Teaching mathematics, and therefore any document that structures the teaching of mathematics, should be consistent with these features.

We expand on the role of structure. In practice most math problems are routine applications of mechanical skills. These skills really are needed in later work, and few K–12 students are able to effectively learn abstract structure, so skills are an appropriate focus. However it is abstract structure that makes mechanical routines work, and the better they reflect the structure the better

²The lack of agreement on meaning of terms is further explored in [Communication between the mathematical and math-education communities](#)

they work. Further most students internalize abstract structure if it is clearly displayed in routine work. This internalization makes it easier to progress to deeper work based on similar ideas, and eventually to the ideas themselves.

As an example, the arabic digit representation of numbers replaced roman numerals not because Arabs conquered Romans, but because it works better. And it works better because it is more closely aligned with deeper mathematical structure: the same structures used to manipulate arabic-style digits are used to manipulate polynomials. This is why students taught to work with numbers using algorithms that cleanly reflect this structure find the transition to polynomials relatively painless. If number work obscures the structure (e.g. with certain addition tricks, or mechanical aids such as an abacus, slide rule or calculator) then students tend to see polynomials as a new and difficult subject. They have learned to deal with trees, but without absorbing the viewpoint needed to see the forest.

K-12 students are not tested on abstract structure so it is up to teachers to make sure structure is clearly reflected in the materials. Standards documents could help by making explicit the key abstract structures involved in a particular topic, pointing out where else these ideas appear in the curriculum, and offering sample problems that display the structure.

16.1.6 Summary

Ideally a standards document will provide an effective common reference point for all concerned parties. Specifically:

- Test constructors should see what sort of problems are appropriate, and further see how problems might probe absorption of underlying mathematical structure.
- Teachers should be able to anticipate tests in detail, and see the underlying structures (general principles) the tests are supposed to support. Ideally the document should be more useful for this than a test derived from it.
- Teachers should see with some precision what the students have already done in earlier courses and so should not be repeated in non-remedial courses.
- Teachers should see how skills to be acquired in their course—and ways of thinking underlying these skills—will be needed later. More generally the document should coordinate connections between the material and the structure of mathematics.
- Textbook writers should see how to expand the material in ways useful to students and teachers.
- The focus should be on testable outcomes and content. Methods used to achieve these outcomes should be left to teachers, curriculum developers, and other education professionals. In particular the document should not be, or resemble, a set of lesson plans.

Unfortunately most standards documents are developed in politicized and often contentious processes that overlook most of these points and cannot address any of them effectively.

16.2 More About Tests

High-stakes tests provide enforcement and accountability for the implementation of standards. They are intended to powerfully influence learning so great care ought to be taken to ensure this influence is beneficial. As with so much else this is a job that by default falls to the standards document.

16.2.1 Tests as instruments of terror

System-wide tests are typically given once, though a few systems have “second chance” administrations. Stakes are high for both students and teachers so teachers (and occasionally parents) emphasize this to motivate students to prepare. Stress levels are high. Test formats, grading criteria, and even question types are different from those typically used in class and this is another source of confusion and stress.

In these circumstances strong students usually do consistently well and weak students do consistently poorly. Outcomes for average students tend to be less reproducible: repeated tries at equivalent tests give scores with significant spread that seems random. Any given score doesn’t correlate well with anything, so in particular cannot correlate with learning.

In some communities there is vocal opposition to high-stakes testing, and the drawbacks noted above are often given as reasons to end it. The reality is that testing is here to stay, but this does not mean the drawbacks are not real or not important. Problems should be honestly acknowledged and fixing them should be an urgent priority for test designers and administrators. However it is hard to imagine how this could happen without guidance from the standards document.

16.2.2 Tests as defective standards

Tests are traditionally thought of as assessment instruments and not part of the educational process itself.

When there is a lot of material assessment tests usually spot-check at random: if the student does not know what will be omitted then comprehensive learning is needed for reliable good performance. Similarly if generic problems are time-consuming then tests may use artificially simplified cases. If the student does not know how problems may be simplified then again comprehensive learning is needed for reliable good performance.

The traditional disconnect between tests and learning does not hold for system-wide high-stakes tests. Old tests are available and carefully scrutinized and new ones are expected to be consistent with them. Tests become *de facto*

standards so simplifications or omissions are incorporated in and weaken the curriculum. The converse to this is that to avoid weakening the curriculum standard tests would have to be harder and more comprehensive than they are now. Clearly our approach to testing must change dramatically if this is not to be a prescription for massive failure.

A related problem is that current multiple-choice tests tend to drive curricula away from abstract and symbolic work. Symbolic expressions have structure that may give shortcuts to identification of correct answers and it is common practice to hide such structure by numerical evaluation. πr^2 for instance is instantly recognizable as the area of a circle, while 16.6 is not obviously the (approximate) area of a circle of radius 2.3. This leads to high-stakes tests dominated by approximate numerical problems, and this in turn de facto establishes the goal of the course as success with numerical problems. Students come out knowing exactly what to do with a problem involving a circle of radius 2.3, but are stumped by the same problem when the radius is given as “ r ”. This is particularly acute in curricula emphasizing use of calculators. These students have missed the benefits of math as an introduction to abstract logical reasoning, and are at a disadvantage in college courses.

16.2.3 Tests as suppressors of quality and diversity

In the last 40 years the US K–12 system improved in some ways, going from one in which many children dropped out to one with a realistic hope that none need be left completely behind. However there have been costs including a decline in achievement levels: since the priority is now to get everyone over one bar it has to be set low. Resources are focused on weak students since they are at risk of failure and good students are not; a great shift from the Sputnik era goal of boosting the best.

Declining preparation of high-school graduates has driven a corresponding decline in post-secondary achievement and American students have nearly disappeared from top achievement levels. Our better graduate schools are populated by high-achieving students from other countries and our leading scientists, engineers, and educators are increasingly international. Significant parts of our way of life are now maintained by importing high-quality K–12 and undergraduate education.

Dependence on foreign educational systems for high-quality preparation is a threat to our national security and prosperity. Eventually we must do better with our own good students. Any real movement in this direction would have to be supported—if not started—by standards documents and system-wide tests.

For example a state might have two levels of tests, say “general” and “college prep”. A bad score on the college prep test could be converted to a good score on the general test, so no one would “fail” college prep. College prep tests and standards would organize development of more-demanding courses and therefore increase diversity in the system.

It should be emphasized that the need is for better preparation in high-school subjects such as algebra, geometry and trigonometry, *not* in topics such

as calculus and statistics. Very few schools have the resources to do a college-quality job with college-level subjects. Mediocre or mechanical courses (driven for instance by the AP calculus test) give little advantage to college students, and certainly do not make up for weak preparation in algebra and geometry.

16.2.4 Summary

Consistency is the overwhelming concern in traditional high-stakes test design. Tests must be similar in content and scores should be as consistent as possible from one administration to the next. This is difficult and expensive but test developers do impressively well at it. In contrast current tests show little or no evidence of concern for the effect they have on the instructional program. It may be that design criteria and pressures during the development process make this impossible, but the end result is a consistently negative influence and no reason for hope that traditional approaches to test construction will produce anything else.

16.3 Conclusion

The 2004 version of this article went on to suggest ways to make standards documents more functional. However at the time of the revision in late 2008 the NCTM and many states have revised their documents and other organizations including the [College Board](#) have issued Standards, all perpetuating the defects discussed above. The federal No Child Left Behind regulations has further polarized and obscured many issues. The [National Mathematics Advisory Panel](#) identified a few of the problems but was unable to come to a firm conclusion on most of them. It no longer seems reasonable to hope for significant change in the way Standards Documents are constructed.

It seems remotely possible that good tests could be developed outside the Standards system, see [Beneficial High-Stakes Math Tests: An Example](#).

Chapter 17

Math / Math-Education Terminology Problems

February 2009

17.1 A Search for Meaning

A few years ago a draft K–12 Standards Document arrived at the AMS for review. This happens from time to time and while as far as I can tell AMS feedback has no effect, it is flattering to be asked. However this Document was accompanied by a guide for reviewers that included the question:

“Do the standards specify a range of cognitive skills to be expected, including some range of the following?”

- Remembering: recognizing, recalling
- Understanding: selecting, interpreting, illustrating, classifying, summarizing, inferring, comparing, explaining
- Applying: using, executing, implementing, computing, translating
- Analyzing: differentiating, organizing, attributing, synthesizing
- Evaluating: checking, critiquing, justifying
- Creating: generating, hypothesizing, planning, designing, constructing”

Say what?? Are these ranges of cognitive skills or ranges of synonyms?

Math educators generally reject use of careful definitions so one cannot just look these up. However there is an extensive literature from which we could try to infer meanings, and we can see how these things actually play out in students. Two conclusions emerge: first, as expected, these are for the most part synonyms

and reflect a richness of language rather than of content. A more troubling conclusion is that when these terms do have specific meanings they are quite different from the meanings used in the mathematical community¹.

17.2 Misunderstanding Understanding

Every discipline develops terminology adapted to the discipline. Specialized meanings for common terms lead to “talking past each other” communication failures. We illustrate this with the term “understand”.

The mathematical community has evolved a rather strong meaning for “understand”: roughly “complete mastery” including full facility with working problems. Weaker meanings have been found to be dysfunctional in the sense that they do not provide a foundation for further mathematical learning.

The educational community has a much weaker meaning for this term. My guess is that it reflects something about human learning: people learn some things (e.g. inferring patterns from examples) quickly and easily. Fixing errors in this natural learning is a different process and much harder, so it makes sense to have terms for the first step. “Understand” may be one of these. At any rate the math-ed meaning for “understand” is closer to “show evidence of exposure”. Teachers can say “you can’t work the problems but I see that you basically understand, so I can give you partial credit”. And when students get to the college level they say “I really do understand it, but just can’t work problems. Can’t you give me partial credit?”

There are similar mismatches with most other terms. Does “recall the quadratic formula” mean “know and be able to use the quadratic formula” or “recall having seen the quadratic formula”? Does “know multiplication facts” mean “know there is a multiplication table” or “be able to multiply numbers with facility”? Terms such as “synthesizing”, “justifying”, “creating”, “discovering”, etc. refer to highly-structured activities that have little in common with the mathematical meanings.

17.3 Right, Wrong or Different?

To a degree these terminology issues can be seen as cultural: they have their meanings, we have ours, and it is neither necessary nor appropriate to declare one or the other “wrong”. We just have to be mindful of the differences and very careful when trying to communicate.

There are, however, cases where one meaning really is wrong. The slogan “we should put less emphasis on rote learning and mechanical calculation, and more emphasis on understanding” has strongly influenced math education in the last few decades. It is certainly very attractive. But remember that there is a job to be done: students should emerge with a good foundation for further

¹For a more detailed discussion with a different objective see [Communication between the mathematical and math-education communities](#).

mathematical learning. We are not at liberty to use any convenient meaning for “understand” but must use one that actually gets this job done. The math-ed meaning is dysfunctional in this regard and so—in the context of the slogan—is actually wrong.

I do not believe that use of a dysfunctional meaning for “understand” is an evil plot designed to cripple higher education in mathematics, even if it is working out that way. The K-12 system is rather self-contained and the curriculum adapts to whatever students can do. The failure to “provide a foundation” only becomes unavoidable and acute at the college level. K-12 educators are pretty unresponsive to complaints from the college level, but in their defense it must be said that these complaints are often incoherent.

17.4 Plea

It is important to realize that the real problems will not have terminology solutions. Mathematical understanding is too demanding to be appropriate at the school level. The mathematically-adapted meaning for “understand” might make the “understanding, not rote calculation” slogan correct but would also make it unrealistic. The proper goal may be a mostly-subconscious *template* for mathematical understanding. The math-ed meaning will likely play an important role in its development. Making effective sense of something like this would take deep insights into both human learning in general and the needs of long-term learning in mathematics. It is likely to require cooperative effort by both the math and math-ed communities.

Unfortunately, we won’t be able to formulate or agree on the real problems, much less solve them, until we sort out the terminology issues.

Chapter 18

Communication between the Math and Math–Education Communities

June 2006, edited October 2008

Introduction

The mathematical community is primarily concerned with developing new mathematics and training in the professional use of mathematics. The US math-education community is concerned with teaching mathematics at least three hundred years old to the general population. There is not much overlap between these primary concerns and so—until recently—little need for systematic communication.

In recent decades the US math-ed community, largely under the leadership of the National Council of Teachers of Mathematics (NCTM), has developed an innovative, coherent and forcefully articulated approach to K–12 math education. At the same time students graduating from this system have shown a significant decline in preparation for technical careers. For decades the US has imported technical training to make up the shortfall but demand has now outpaced imports and technical jobs are being exported. This is an impending national disaster. Being coherent and innovative is not the same as being right, and the current K–12 program seems to be seriously flawed in this regard.

Part of the problem is that the K–12 focus has been on the weakest students (No Child Left Behind) at the expense of high-achievers. This focus is so complete that most math educators either deny there is a high-end problem (skills are “different”, not worse) or believe that more of the same will fix it.

The academic mathematical community has to deal with the output of the K–12 system so is much more sensitive to the problem. Many now believe that some sort of action is needed. The most vocal proposals have been simplistic and unconstructive: “game over, return to the 19th century” or “adopt the proven program used in (country name)”. But attempts at thoughtful communication have also been unsuccessful.

This essay is concerned with recognizing barriers to communication between the two communities, and seeking ways to avoid them.

Some of the barriers are linguistic: in the second section we describe the very different interpretations given to words such as “understand”, “know”, “apply” and “recall”. But in the third section we suggest that these differences are adapted to the jobs at hand rather than arbitrary, so they cannot be settled by linguistic negotiation. Our conclusion is that successful communication on an abstract or conceptual level will be so difficult that alternatives must be found.

This requires a sharpening of the problem: exactly what is it that needs to be communicated, and to what end? The first need is to communicate about preparation of students for success in college and technical work. This is investigated in the fourth section where we suggest that it might be done on a very primitive level—essentially annotated lists of sample problems.

18.1 Language differences

Professional communities develop specialized language for precision and clarity. Naturally, different communities will give terms different specialized meanings, and while this is an obvious and well-known source of confusion it almost always takes participants by surprise. For instance a core part of the new math-ed vision is a shift from drill and rote mechanical work to conceptual skills such as “understanding” and “knowing” and conceptual activities such as “creating” and “recognizing”. These goals sound good to everyone but the words have rather different meanings in the math and math-ed communities. In this section we describe some of these differences. The next section suggests reasons for the differences.

18.1.1 Understanding

In the math-ed community “understand” does not imply “able to work problems with”, while mathematical use includes this and more. We expand on the differences, then consider the problems they invite.

The use of “understand” as something untestable and distinct from “able to use” is clear and consistent in recent math-ed literature and many (probably over 100) standards documents. The Executive Summary of the 2000 NCTM publication *Principles and Standards for School Mathematics* provides a useful and authoritative example. There are 24 occurrences of the word “understand” in the text, eleven referring to student learning objectives. Ten of these are accompanied by a phrase “and apply”, “as well as use” etc. The intent is to clarify

that students should be able to *do* as well as understand, but this also clarifies that “understand” is not taken to include doing. The eleventh appearance is in the sentence “An understanding of numbers allows computational procedures to be learned and recalled with ease.” This is one of the few explicit suggestions of a utility for understanding, but again it is clear that a failure to “learn and recall” is not taken as evidence of a lack of understanding.

In mathematics “understand” means mastery and certainly includes ability to work problems. It is correct usage to say “evidently you *don’t* understand since you can’t use it effectively.” Non-functional exposure is considered useless and not given a name.

The difference in usage means any statement containing the word “understand” will be misunderstood. Standards documents often include a phrase like “understand the quadratic formula”. College teachers think this means “know cold, be able to recite the formula instantly, apply quickly and accurately, and translate quickly and easily between the various formulations of the outcome.” Math educators are more likely to interpret it as “recognize the name and realize it has something to do with solving equations with a squared term” and expect that if anything more is intended it would be spelled out.

Awareness of language differences would help but not be a complete solution. Consider for instance “I would rather my calculus students have a mathematical-level understanding of algebra alone than a math-ed understanding of algebra, calculus, statistics and probability” (really!). Educators might understand this as an assertion about the importance of precision and “fluency” but would not know what it involves in detail.

18.1.2 Remembering, recalling, knowing

A central tenant of the new vision of math education that memorization is a superficial approach to learning and should be avoided. A corollary is that teachers should not require a level of precision that might require memorization. “Remember” and “recall” implement this idea: they indicate a “bringing to mind” that demonstrates familiarity but is tolerant of error. This is close to the meaning in common language where words such as “recite” are available to indicate “recall with precision”.

The presumption is that students’ “remembering” will become more precise as their “understanding” grows, and it is better to wait for understanding than to require memorization. Thus when a Standards document states that students should “know properties of numbers”, “recall multiplication facts”, or “remember facts about solutions of quadratic equations” the math-ed interpretation is that they should *not* be required to do so *correctly*. These are not endpoints but processes that in the fullness of time are supposed to lead to accuracy.

Deliberate fault tolerance does not make sense to mathematicians. Mathematics is empowering only if it is used exactly right, and errors in “recalling” almost always render it useless. Most of the work in teaching college mathematics is getting students to find and fix errors in their “recollections”. This is not to say mathematicians are fans of memorization. The author often teaches

a calculus class that requires use of trig addition formulas. He can't remember these formulas, but can derive them in a few seconds from properties of the complex exponential function. He would be overjoyed to have a student who could do this too, but the realistic and honest approach is to say "you need these; memorize them". Indeed mathematical experience is that understanding in the strong mathematical sense *follows from* accurate recall and extensive practice, and cannot precede them.

Some educators acknowledge the need for better "recollection" and the term "automatic recall" is starting to appear. This does not have an agreed standard meaning so use of it does not indicate a commitment to anything in particular. The fact that "memorize" and "recite" are still being avoided suggests that the outcome is uncertain.

18.1.3 Applying, evaluating

"Applying" and "evaluating" appear in Standards documents in phrases like "apply math concepts" and "evaluate mathematical statements". These phrases are also common in mathematics but indicate activities developmentally inappropriate before the second or third year of college. Indeed the shift from problem-solving to concepts and logical evaluation is the main reason so many students who start out as math majors change their minds. In this case it is obvious, at least to mathematicians, that there are profound differences in meaning in the two communities. The nature of the differences may be less clear.

Mathematicians use "apply" as a synonym for "work problems with" while the math-ed meaning is closer to "illustrate". For instance if the concept is that multiplication is related to stacking blocks then "applying" might mean stacking blocks to model a multiplication problem. If the concept is commutativity of multiplication then "applying" might mean rearranging stacks of blocks to illustrate this. This clarifies that "applying" is an understanding-enhancing activity, not a testable skill. Generally the practice in Standards and the educational literature is that if something does not contain an unambiguous phrase such as "computational fluency" or "work problems" then a non-testable interpretation is acceptable and probably intended. The examples also illustrate a mismatch in the meaning of "concept". Mathematicians think of commutativity of multiplication of numbers as a property rather than a concept and use "concept" for higher-level abstractions.

The mathematical meaning of "evaluate mathematical statements" is "demonstrate correctness or incorrectness". This is certainly a testable skill. For example students might be asked "Evaluate the statement 'addition distributes over multiplication' ". and be expected to know that to show the statement is false one should find particular values for which the two sides of the equation give different answers; and second to be able to find such values. In the math-ed interpretation demonstrations are "understanding-building" activities and not expected of students. The most that would appear on a test would be "true or false: addition distributes over multiplication".

18.1.4 Creating, discovering

“Creating” and “discovering” may provide the most extreme examples of mismatch between the mathematical and math-ed communities. These activities are very highly valued by mathematicians: the primary requirement for the PhD degree is that the student demonstrate he is capable of creativity and discovery. Undergraduate research projects are fashionable but difficult and extremely time-consuming. How could this possibly be scaled down to K–12? Mathematicians generally find the whole idea disturbing.

Quite a few math educators suggest that students should “discover” their own versions of algorithms for multiplication or division. But the standard algorithms are finely-tuned instruments developed with the difficulty and depth of experience required by the US Bill of Rights. Would government teachers ask students to discover the Bill of Rights? Probably not unless the plan was to spend half a year explaining why the discovered versions were inadequate. Would carpentry teachers show students screws and nails and ask them to discover screwdrivers and hammers? And if they did, would a shortage of competent carpenters and an epidemic of carpentry-anxiety be a mystery?

The point is that some things are simply out of reach of student discovery. The problem goes beyond that however. Professional experience is that 90% of math discovery is either dysfunctional or outright wrong, and consequently 90% of the effort in *effective* discovery is spent finding and correcting errors. It would be truly wonderful if K–12 students could experience this. However few students are willing to be wrong (and get corrected) 90% of the time and few teachers have the time or training to guide the necessary diagnosis.

The math-ed interpretation of “discovery” is quite different: either a process intended to build “understanding” but so tightly controlled by the teacher that it can’t go wrong, or a less-directed activity that is unevaluated because it lacks the refinement process needed to be effective. The outcome (e.g. an algorithm) can be tested but the discovery process itself is not a testable skill.

18.1.5 Teaching vs learning

The final terminology problem is much more profound and concerns location of responsibility in the educational process. One view centers on students: learning requires effort and it is their responsibility to put in this effort, or at the very least not disrupt efforts of others. The other view centers on teachers: teachers are providers, students are recipients, and if engagement is required then it is the responsibility of the teacher to develop it. Are teachers “learning facilitators” or are students “teaching customers”? Are grades “given” by the teacher or “earned” by the student? On the slogan level, “you can lead a student to knowledge but you can’t make him think” vs. “if the student hasn’t learned then the teacher hasn’t taught”.

In the US K–12 system responsibility is placed primarily on teachers. It is standard practice for teachers and school system to be punished if students do poorly on state tests. At the college level there is simply no way to avoid

placing primary responsibility for learning on students. This is incorporated into the way college teachers think and interpret terminology. As a result even the words “teaching” and “learning” will cause interpretation problems in K-12-college communications¹.

18.2 Mathematics and learning

In this section we suggest that the dramatically different word usages described above are adapted to their subjects: there are actually reasons for them. The mismatch is not simply linguistic and cannot be solved by linguistic compromise.

18.2.1 Mathematics

The demanding nature of mathematics is suggested by the fact that it was an organized subject of study for three thousand years before it really got off the ground. Mathematical conclusions are like legal documents: powerful if fine print is satisfied and loopholes are avoided but you can lose your shirt if you make the smallest mistake. After three thousand years of lost shirts we figured this out and learned to read and write fine print. Mathematics did not become routinely successful as a profession until this was incorporated into community norms. “Know” and “understand” came to mean “so intimately familiar with the fine print that blunders are minimized”. Rigorous standards made math slow and difficult and were resisted by many mathematicians during the changeover, but they were enforced by mathematics itself. Sloppy people were less effective and ended up marginalizing themselves. It took a century of such reenforcement for rigorous standards to win general acceptance.

The seeming ridiculously high standards of modern mathematics are simply what it takes to be successful, not a conspiracy to shut out non-members.

18.2.2 Human learning

People see patterns and connect facts quickly and with little effort. This instinctive facility is thought to have developed because it enhances survival in dangerous situations. Inevitably many of these patterns and connections are incorrect, but people do *not* recognize and correct errors either quickly or easily: the persistence of superstition and gullible belief is well known. Apparently error correction does not enhance survival.

People also have difficulty understanding abstract *explanations* of patterns. It is frequently more effective to provide examples and hints and let them find the patterns themselves.

Effective learning requires finding and fixing errors in natural learning. Young students need help with diagnosing errors. Learning to do this oneself is the key to effective learning at higher levels. At the highest level mathematicians need

¹For an extended discussion see [Teaching vs. Learning in Mathematics Education](#).

such accurate and reliable understanding that they must learn to vigorously test—almost attack—the impressions coming from natural learning.

18.2.3 Math education

“Learning” in US math education seems to correspond to the “natural learning” described above. It makes sense to have a term for this because there is quite a change of gears between this and the more disciplined error-correcting phase. It does not make sense to have no terminology for, or even awareness of, the later phase.

Disciplined areas regard natural learning as only a starting point for understanding and knowledge. The US educational community has taken a different approach: redefine “understanding” and “knowledge” in a sufficiently fault-tolerant way that natural learning is nearly sufficient. Some error-correction is still needed but instead of doing it explicitly the US practice is to cycle through the natural-learning process multiple times. This sets some students more firmly into bad habits and is a mind-numbing waste of time for the ones who got it right the first time, but does give some improvement in the middle.

Lack of concern for errors in learning seems to pervade the profession. Elementary math textbooks are packed with distractors and intellectual content is diffused. The distractors are supposed to maintain interest and enrich the learning process. They also increase the error rate. The error-tolerance of the math-ed community is so great that either they cannot see this or they regard it as a good exchange for “enrichment”. Error tolerance makes educators’ job easier and reduces the effort required of students but it also largely cuts students off from areas requiring high-precision knowledge.

How can error tolerance coexist with something as black-and-white as a math test? Not well. US students fare poorly in international comparisons. Statewide high-stakes tests are causing dislocations, though this is softened by the political need to set standards low enough that most students pass. To improve grades teachers can use simplified problems and standard phrasing in classroom tests. Credit for routine homework with low quality control provides a buffer against low test scores. Valuing “knowledge” etc. provides a loophole: teachers can say “I can give you credit even though you can’t work problems because I see you basically understand it”. Finally calculators have been a godsend: students can be trained to get good numbers via keystrokes without a disciplined grasp of detail.

There is an historical explanation for error-tolerance that long predates the NCTM vision. The old view that mathematics is good training in disciplined logical thinking is explicitly not error-tolerant. About a century ago some US educational leaders asserted that elementary mathematics should focus on and be valued for its applications. This may have been a mistake. There has been a decline in disciplined logical thinking in US society, and elementary math is now personified to many people by bizarre and contrived word problems. In any case math education was released from the constraints of rigor. Moreover this happened so many generations ago that it is deeply ingrained in the mindset,

literature, teacher training methods and community standards. Any change will be slow and painful.

18.3 Communicating about student preparation

We have argued that terminology and mindset differences are too great for abstract statements such as “students should know how to solve quadratic equations” to be successful. Here we outline a more concrete and direct approach. This has been arrived at by a process of elimination—anything else seems likely to fail—but it has a number of other significant benefits.

18.3.1 Quadratic example

18.3.1.1 Task

Students should be able to recite the quadratic formula and use it to find exact solutions to quadratic equations with either numerical or symbolic coefficients. For instance:

Solve the equation $2x^2 + ax - 5a^2 = 0$ for a in terms of x . Use this to rewrite the left-hand side in the form $-5(a - R)(a - S)$.

(A real-life treatment would continue with many more examples.)

18.3.1.2 Annotation

many mathematical procedures require solving an equation. (Examples: finding extrema, intersections of curves, solutions of some differential equations.) Quadratics are one of the very few general classes of equations that can easily be solved so they are heavily used in examples and problems. Students who can work with quadratics easily and without much thought will be able to focus on the new material. Students who have difficulty with quadratics will constantly find this a barrier to further learning.

18.3.2 The general pattern

The core of a document intended to communicate goals would be an extensive list of sample problems. These problems would be selected to illustrate key points: the example above illustrates that coefficients might involve symbols; that roots may be irrational and also involve symbols; and that the variable being solved for may not always be called “ x ”. There should be notes explaining why such problems are important; how they may be used; abstract principles underlying them; and so on, but the notes should be clearly subordinate to the problems. Alternate interpretations of notes does not justify simplifying or discarding problem types.

The web site of the AMS Working Group on Preparation for Technical Careers, <http://amstechnicalcareers.wikidot.com> is an attempt to implement this idea.

18.3.3 Benefits

The first virtue of this approach is that it avoids the modes of failure identified in earlier sections. This alone would make it worth pursuing, but it seems to have some significant further advantages.

18.3.3.1 Neutrality

The math-ed community has learned how to teach K–12 math. The mathematical community believes this learning is flawed and needs correction. Error correction may be a routine part of mathematical culture but it is not in math education. Teachers as well as students worry about the difference between “you have made an error” and “you are stupid”; between being offered correction and being disrespected. Linguistic differences exacerbate this.

Problem lists provide a neutral meeting place for the professions. Mathematicians can formulate goals without judgmental overtones or misinterpretations. Educators can see for themselves what the core concerns are, and formulate these conclusions in their own language.

18.3.3.2 Focus on outcomes

There is always tension and confusion between process and outcome. It is faster and easier to say “do this and you will come to the right place” rather than carefully describe the “right place”. But the fact that it is faster means it is more susceptible to misunderstanding, and the fact that it is easier makes it more susceptible to error. It should bring focus and clarity to the process to undertake describing the “right place” concretely in terms of sample problems.

Focus on problems would also help the mathematical community develop a coherent position. “Mathematicians” are not a coherent group with uniform views: there is a great deal of shared culture and agreement on general principles, but agreement disintegrates quickly as one gets into specific issues. The community lacks mechanisms for developing agreement and in particular lacks the forceful and articulate leadership provided by the NCTM in the math-ed community. However most of the disagreement concerns process or terminology: the right *way* rather than the right outcome. Mathematicians will probably agree that certain sample problems are good even if they disagree on *why* they are good.

18.3.3.3 Professional autonomy

Mathematicians have no business dictating in detail how K–12 math should be taught. They can and should specify the outcomes they need as a basis for further education. They can further describe interconnections of patterns and flow of mathematical ideas that might make teaching easier. If educators find these ideas useful they will use them but they should not be ordered to use them.

For instance mathematicians might specify—through sample problems—that students be able to multiply and divide polynomials “with facility”. They might further point out that the algorithms used for long division and multiplication of numbers are also needed to divide or multiply extremely large numbers using calculators or computers with limited digit capability, and again for division or multiplication of polynomials. Students who learn the simpler versions early may find the steps up in complexity or abstraction relatively easy, while those who come to the advanced versions without preparation will find them difficult. Educators may find this a convincing reason to return the algorithms to the early curriculum, or they may prefer to experiment with ways to tackle it later. As long as they accept the final goals and get the job done it shouldn’t matter. The process should be up to educators and not dictated by mathematicians.

Those who observe that dictatorial control worked well in Soviet math education should realize this was a matter of great luck in choice of dictators. The same dictators forbade the teaching of evolution and essentially destroyed effective biology outside the biological warfare programs. Dictatorial control of process is considered a failed management model and one successful example does not change this.

The need to be persuasive—rather than imposing solutions by fiat—may also encourage mathematicians to be a bit more clear and coherent.

18.3.3.4 Testing

High-stakes statewide tests are nearly universal and national tests seem increasingly likely. Preparing students for these tests has become a matter of personal survival for many teachers. They need to know what the students will face and what is needed to prepare them. In principle this information is provided in Standards documents. In practice there is such a gap between the abstract goals in Standards and problems on tests that they are useless and old tests are de facto the authoritative guides.

The other side of this coin is the dilemma facing test designers. Ostensibly tests should reflect goals set by Standards documents. However these goals are so vague and inflated that this can’t be taken seriously and again old tests serve as the main guides. The Standards document itself is effectively removed from the process and any intellectual content is lost.

A great virtue of a goals document organized around sample problems is that it connects clearly and directly to tests. For teachers the phrasing “students should be able to work problems like the following” becomes “test problems will be like the following”. For test designers, explanations of what the sample problems are supposed to illustrate become instructions on how they can be varied while accomplishing the same goals. The results should be better tests and better student preparation for them.

18.4 Conclusion

We have argued that annotated problem lists would be an effective way for the mathematical community to communicate goals for student preparation in K-12. More than that we have argued that other—more “conceptual”—approaches face such severe obstacles as to be a waste of time. The mathematical community should undertake the development of such lists as soon as possible².

² An attempt can be found at <http://amstechnicalcareers.wikidot.com>.

Chapter 19

Evaluation of Methods in Mathematics Education

March 2006, revised October 2007

Introduction

The research community devoted to the learning of mathematics has grown significantly in size and professionalism but it has not grown in success. In this article we suggest reasons for this, and in particular argue that the way new ideas are developed and evaluated has channeled the enterprise in unproductive directions.

Many of these conclusions come from work in a large computer-learning facility (over 6,000 students, 500 computers) developing computer-based courses, computer-tested classroom courses, and computer-lab additions to traditional courses. The big issues came not from computers per se but from the environments they provide. Traditional classrooms herd students down a single path and provide many tools to keep them on it. Computer environments necessarily give students more control. Factors that are locked together and invisible in classrooms come apart and must be understood separately. The author has spent most of the last decade trying to understand these new learning environments. Experience from 30 years of traditional teaching turned out to be a minefield of preconceptions rather than a useful guide.

This article describes flaws that seem to be common in math education research and curriculum development. §19.1 concerns important factors omitted from evaluations; §19.2 lists dangers of evaluation on the basis of method rather than outcomes while §19.3 describes dangers of incautious use of statistics when outcomes are analyzed. Finally, §19.4 concerns problems caused or obscured by a focus on teaching rather than learning.

19.1 Missing Criteria

Here we describe factors that play crucial roles in success or failure of educational programs but are not considered in evaluations.

19.1.1 Resources

The main reason “good” educational experiments flop in practice has to do with resources. Education has a limited budget and the task of educators is to do the best they can with the means provided. Since educators have no influence on resources there is generally no point in thinking about them. But this is a dangerous blind spot: new methods that require more resources are nonstarters, and neglecting costs in evaluations will miss this.

For instance imagine a new way to teach fractions is shown to significantly improve outcomes. Teachers are urged to adopt the method and parents and administrators are led to expect that great things can be accomplished. But as is often the case the improvements were accomplished in intensive sessions with fewer than fifteen students. There is not much hope the method would work in a typical large attention-span-challenged class, and no particular reason to think that lavishing such resources on other methods would not work as well. The high expectations are unrealistic and attempts to widely implement the method will lead to (yet another) failure.

If resource costs were tracked and factored into the evaluation then a small-scale study like the example above would be considered at best a pilot project needing large-scale field testing. Investigators should develop indicators for successful scaling, or do medium-scale trials before considering the project complete. This would be harder than current practice and there would be fewer “successful” small projects, but long-term impacts would be greater.

19.1.2 Student effort

Student effort—and a student’s willingness to expend effort—seems to be a crucial factor. This is an unexpected conclusion and we have no suggestion as to how these might be measured, but we describe the problem.

19.1.2.1 A behavior model

We offer a simple model to explain certain patterns in student behavior. This can at best be a rough guide with myriad exceptions, and in a later section we argue that assuming students have common features can be dangerous. Nonetheless the model encapsulates a lot of painful experience and offers insight into some difficulties in educational research.

The model is this: students see a grade as their objective in a course. They come to the course with a target grade and an “effort budget” they are willing to expend. They work until they either reach the target grade or expend their

effort budget. If they hit the target grade first then they quit and get extra free time. If they hit the effort limit first then they quit and accept a lower grade.

Put another way, it is rare for a student to either put in extraordinary effort to get a good grade or, if a subject is easier than expected, keep working after reaching a target. Both do occur of course, but not often enough to invalidate the model as a predictor of bulk behavior.

This model has both explanatory and predictive value, and we explore it below.

19.1.2.2 Standards versus success

It is well-known and obvious that performance of successful students can be improved simply by raising standards for grades. The drawback is that fewer students will be successful. The model suggests that students who quit because they reached their target grade will work more if standards for that grade are increased. Students who reached the limit of their effort budget will accept lower grades. Educators try to balance standards and failure rates, although views on the proper balance has changed quite a lot over time.

Pressure to avoid dropouts and high failure rates has driven a lowering of standards. Most students hit their target grade before their effort maximum. This has consequences at the next level: not only have lowered standards left students less well-prepared, but ease of success has led them to revise their effort budgets downward.

Most students are capable of, and willing to do, better work. It follows that having several tracks with different standards could improve outcomes, particularly at the upper end where our system is weakest. However the argument above suggests track selection should be based on willingness to work rather than past performance or innate ability. It is certainly unclear how this might be measured but it is a worthwhile research topic since the potential payoff is enormous. Also tracking by ability or performance is politically unpopular, and tracking by willingness to work (if it can be measured) might be more acceptable.

19.1.2.3 Equivalent outcomes

A frequent experience with full-scale trials is that different methods have statistically equivalent educational outcomes. By this measure there is no basis for preferring one to any other. However the behavior model suggests that there is a reason outcomes are similar and that some methods may have benefits that are not being measured.

Suppose two methods are being compared. One enables students to achieve a given level of mastery with less effort, but the trial is “fair” in the sense that fixed achievements are required for a given grade. According to the model, students who are already reaching their target grade won’t do better but will benefit through reduced effort. Since most students are in this group we expect very little overall outcome improvement. Thus equivalence of outcomes should

be an expected consequence of student behavior, not a reflection on program quality. If we want a clear indicator of program value we apparently must figure out how to measure student effort.

The end objective is still better outcomes. In this view this would be accomplished in two steps. First compare methods but look for reduced effort (if it can be measured) rather than better outcomes. Then, after implementing the more efficient method, increase standards to bring effort requirements back to previous levels.

19.1.2.4 Work ethic

It is obvious that students' willingness to work is a major determinant of outcomes. The arguments above suggest that as a side effect it also tends to hide or distort other factors. Is this a fact of life, or does it just emphasize the old point that part of a teacher's job is to engage students and get them to work?

Sadly, work ethic is one of the things eroding from the American character. Students are increasingly unwilling to work and harder to engage. On the whole teachers will be unable to slow this decline let alone reverse it. There have been examples of teachers who dramatically inspired their students, but they are rare enough to become subjects of major motion pictures. Blaming teachers is a prescription for failure.

19.1.3 Procrastination

Everyone knows procrastination is a problem and traditional courses are packed with preemptive measures against it: constant checked homework; frequent quizzes; lots of major tests; and generally so structured that students who aren't working can be identified and hassled. On the other hand no one seems to measure procrastination directly and there has been little thought about how it might effect new approaches to education. The problem described here may seem obvious but recognizing it was a surprise outcome of years of data mining in a relatively unstructured course: the only measure that correlated strongly with failure turned out to be a proxy for procrastination.

Many new approaches to education lack traditional defenses against procrastination. For instance a student might be allowed to choose from a variety of tools to accomplish a task, rather than be forced to do it a particular way. But the consequence of not checking use of a particular tool is that it is hard to know if *any* tool is being used. This problem can be acute in computer- or web-based assignments where—for better or worse—students must play a more active role in the learning process. Procrastinators fare poorly in such environments.

For another example consider the common complaint that most homework is pointless busywork. An alternative when there is a well-defined task (e.g. a kind of math problem) is to provide plentiful examples, give instructions to work examples until they can be done reliably, and a deadline. The ability to quit when ready provides motivation and a payoff for fast and accurate learning, and many students respond well. Procrastinators tend to put it off until too late.

Generally we can expect that any increased reliance on student initiative or reduction of lockstep control will show mixed results: non-procrastinators may do better but procrastinators will do more poorly. This is a problem we are stuck with: chronic procrastination is either a character trait or so difficult to unlearn that it might as well be. It may be a limiting factor in how much choice or control can be given to students. Alternatively it might be possible to identify chronic procrastinators and provide them with a more supportive (i.e. constricted) environment.

Finally we contrast procrastination with the limited work budget problem discussed above. In a fixed environment the distinction is not useful: students who wait until they don't have time to do the work may as well not be willing to do it at all. The difference becomes important when the environment changes. Giving students more control should benefit non-procrastinators with low work budgets but be counterproductive for procrastinators.

19.2 Process as a criterion

Educational approaches are often evaluated on the basis of methods used rather than outcomes. This is easier and faster, and is a reasonable proxy for outcomes if similar methods have been carefully evaluated in similar areas and have had good outcomes. It is an appropriate way to design new approaches as a starting point for development and testing. Unfortunately process evaluations are misused more often than not. It is particularly dangerous to have an ideological attachment to a Good Thing and “know” it will improve any program.

19.2.1 Multimedia

Video and animations are considered a Good Thing. They have wide and satisfactory use in some subjects, and any proposal that includes them gets extra credit. However the successful uses are in subjects with low expectations for testable outcomes. The generalization that these are Good Things in any area has turned out to be false: they are much less unsuccessful in areas that require concentration and have high testable outcomes.

Our students are very accomplished spectators. The entertainment industry has shortened their attention span and trained them to suspend critical thinking while in spectator mode. The advertising industry has forced them to avoid learning while in spectator mode. Consequently spectator mode is an enemy of serious learning. Anything that triggers it, including almost anything that moves on a screen, is likely to seriously degrade math learning.

19.2.2 Technology as a goal

A common process goal is that a course should “use technology.” The process has been taken as the *outcome* goal as well, and there are no educational outcomes to be assessed. This guarantees positive evaluation of the course, but

probably also guarantees that nothing of real value will come of it and any bad consequences will be overlooked. Some examples:

- Internet use is a common process goal. The justification offered is “it is important for students learn to use the tools of the information age.” This might make sense if the goal was “learn to use the tools *well*.” The internet has a low signal to noise ratio and the ability to filter and critically assess information would indeed be valuable. Students are not learning this, and will not as long as use alone is the goal. It is true that even professionals have difficulty with critical assessment, and an attempt to teach school children these skills would almost certainly fail an honest evaluation. But the difficulty of finding realistic outcome goals does not justify dodging the issue.
- Computer (as opposed to calculator) use in math is sometimes taken as a goal with justification “computers have transformed real-life use of mathematics and we should prepare our students for this.” However computers have made math more powerful, *not* easier. In fact effective computer use requires quite a bit more discipline and sophistication than standard by-hand work. Computers can solve many standard problems in a few keystrokes. But learning keystrokes instead of tedious hand methods puts students further away rather than closer to the sophistication needed for effective computer work.
- Visualization is a particularly attractive use of computers and a prominent feature of programs with “computer use” as a goal. It seems to have few benefits beyond pretty pictures. At the high end, an NSF-funded institute was established to determine if direct visualization could be a useful tool in mathematical research. The answer was “no” and when the institute was unable to find a more productive focus it was disbanded. At lower levels the experience is that people have to know what they are looking at before they can see it. In mathematics, at least, visualization seems to have limited use as a primary learning tool, but this may not become apparent until evaluation criteria graduate from *any* use to *effective* use.

A problem common to these examples is that the error-correction part of learning has been completely omitted. See the final section for further discussion.

19.2.3 Trendy methods

“Discovery”, “Reform” and “Standards-based” methods seem to be popular now. The danger here is that exciting ideas, sometimes demonstrated in pilot projects but not tested on large scales may be prematurely adopted. In some cases there have been large-scale adaptations but such extensive changes were needed that the name is the main similarity to the pilot program. Or the materials were used in a course with different methods. All too often these

modifications are overlooked and success is taken as validating the original vision. Even though they have actually failed to scale, the methods are used as the basis for process evaluations.

19.2.4 NCTM Standards

The most remarkable example of the use of process standards is the evaluation of math curricula according to how well they conform to the 1989 National Council of Teachers of Mathematics (NCTM) standards. This has been so widespread that curricula and standards documents nationwide have been profoundly influenced.

Use of process standards as a proxy for outcomes is reasonable when the processes have been shown to have good outcomes. Thus if the NCTM standards had been a distillation of best current practice then a push for general conformity would have made sense. However large parts of the NCTM document were a bold attempt to chart the way to the future rather than a distillation of the past. It was a research agenda rather than a finished product, and pushing for implementation before large-scale evaluation was a procedural error.

These standards have been at least partially implemented on the widest possible scale over the last decade and a half. For better or worse, a large-scale evaluation is now possible. Many outcomes, for instance degree of preparation for university work, have declined significantly. By presenting it as a finished product the promoters raised the stakes and narrowed the outcome to pass/fail. By the rules they themselves have established it seems to be a failure and continued promotion has triggered a backlash. Unfortunately the process of correcting large-scale failure may also sweep away any potential value as a research agenda.

19.3 Outcome measurement

Standard protocols for experimental design and assessment have developed as educational research has grown as a discipline. The most sophisticated have taken clinical trials in medicine as a model. This seems appropriate since both medicine and education have to deal with the worst possible experimental subjects, people.

The ideal educational trial begins with a method to be tested and specific outcomes (e.g. question types) that the method is expect to influence. To avoid questions about “changing the target after the arrow is shot” there are usually advance decisions about how the data is to be analyzed and how various outcomes will be interpreted. The trial itself involves two groups of students, one using the new method and one with a standard or control method. Data is gathered. Frequently there is an analysis of student characteristics to check for bias in the division into groups. Finally there is a statistical analysis of the outcomes. The care and sophistication of the statistical analysis is often taken as a key indicator of the value of the study, so this analysis is often quite elaborate.

It is unfortunate that this tidy experimental design will almost never give useful conclusions. Even if it worked exactly as expected the outcomes would be problematic because crucial parameters have been neglected, as explained in the first section of this article. But the experimental design itself is unreliable. The medical community discovered and adjusted for this long ago. The educational community needs to do so as well.

One conclusion is that educational research is still in its infancy. Recent attempts to be more “scientific” have not actually made the field more mature and effective.

19.3.1 Student variation and statistics

Statistics only gives an accurate picture when all students will be effected in roughly the same way: “one size fits all.” This is rarely the case, and we describe instances where modest overall improvement resulted from big improvement in one subpopulation of students canceling a big decline in another. The medical analog is the strangely belated realization that males are different from females and that there are environmental and genetic differences between ethnic groups.

The point made above is that a different approach to data analysis is needed when there are subpopulations with different educational needs or responses. It is also unclear how the data would be used. Rejecting methods that favor one subpopulation over another would probably leave us with no educational program at all. Offering different methods to serve different subpopulations might be effective but would require great care in how people are assigned to, or allowed to choose, different methods.

19.3.1.1 Modes of thinking

People learn through three main channels: visual, auditory and kinetic/tactile. They tend to have dominant learning channels just as they have dominant hands (left or right). At one time this was well known and a principle topic in educational research. The awareness has all but vanished, possibly because it conflicts with the homogeneity hypotheses needed for statistical analysis. The fact has not vanished however, and educational methods that emphasize one channel will still favor one of these groups over the others.

For example not long ago reading was taught largely visually through pattern recognition of letters and words. Proponents of this approach were either unaware of the needs of auditory and kinetic/tactile students or expected them to adapt much like left-handed students were once required to write with their right hands. The program may have been effective in some average way but outcomes for primarily-auditory students were unsatisfactory. There were significant numbers of high-school students with very weak reading skills. This drove a resurgence of phonics approaches that favor these students. If phonics turn out to be unsatisfactory for primarily-visual students we may see a swing back to visual methods.

At this point it might seem obvious that the solution is to offer both approaches and let students use the one that works best for them. This also has problems, with resources and placement for instance, but the show-stopper is the conflict with dominant dogma.

How do strongly kinetic/tactile students fit into the various reading programs? Generally poorly. Some adapt and go on to become surgeons, or in the author's case, a topologist. In the past many who could not function "right-handed" gravitated to trade schools or apprenticeship programs, or dropped out to work. These alternatives are now considered "left behind" and the students are retained in standard school programs. Consequently we expect to still see significant numbers of high-school students with weak reading skills. Any group likely to have kinetic/tactile orientations, athletes for instance, should be heavily represented in this group. Maybe athletes are not dumb after all, but are just not served by current programs.

Learning-channel differences may pose serious problems for web or computer-based educational materials. Current materials are intensely visual. Does this render them inaccessible to significant numbers of students with auditory or tactile primary learning channels? Current analytical techniques can't even see the question.

19.3.1.2 Placebo effects

Small-scale trials almost always have positive outcomes even when large-scale trials with the same methods are unsuccessful. Students do better in small groups and they respond to extra attention and the instructor's expectation that something good should happen. The medical analog is the placebo effect and it regularly swamps information about the drug or procedure being tested. The medical solution is the double-blind trial, but this is rarely feasible in education.

If small trials have predictable and meaningless evaluations do they have any point? Perhaps they should be formative rather than summative. Instead of conclusions like "the method works" perhaps "with the following modifications we feel the method is ready for large-scale trials." Perhaps the problem is not the use of statistics but insufficient caution, wisdom, and humility about what the numbers tell us.

19.3.1.3 Cultural and behavioral bias

"Avoid cultural bias" is usually taken to mean "replace 'cow' and 'chicken' with 'dog' and 'pidgeon' because the population is now largely urban." Or "be sure 'guns' and 'dolls' are evenly represented in word problems." Avoiding behavioral bias takes forms like "behavioral difference are normal and teachers should deal with it." However some biases are not education-neutral and avoiding or ignoring them can compromise the validity of a study.

There are subcultures that highly value education and whose students are willing and well-prepared, and subcultures that find education irrelevant and

demeaning and whose students are unwilling if not disruptive. There are students who need medication before they can control their behavior. Does “avoid bias” mean “do not report the effects of disruptive students”? Or worse, does it mean “do not report that disruptive students were excluded from the experimental group”? In any event there is essentially no discussion of disruption in the math-ed literature even though math is probably one of the most vulnerable subjects.

Disruptive students degrade any educational environment and will completely defeat some approaches. Other cultural or individual differences may impose other limits. It is not clear what, if anything, could be done about this. The point here is that we cannot know the real potential or limits of an educational method if these factors are ignored, and current methods ignore them.

19.3.1.4 Sample size

Many trials reported in the education literature are small, with as few as a dozen students. The hope is that sense can be extracted through careful statistical analysis. This would be doubtful in a physics experiment with uniform particles and is silly when dealing with people. How big should a statistically analyzed trial be?

The author analyzed a multi-section college calculus course with enrollment between 900 and 1400 per semester, divided roughly evenly between two teaching methods and tested with a common final exam. This was done for six semesters with a total of around six thousand students. Amazing variation was seen. Three subpopulations with different characteristics were identified and there were differences due to class size and teacher effects. However most of the variation remains mysterious and considerably exceeds what would be expected from random distribution. For instance one class of over a hundred students had an unusual outcome pattern, significantly different from other classes with the same teacher or the same size. There are group effects such as an attitude—good or bad—“infecting” a class, but we are wary of this as an explanation for such large differences in such large classes. We do not have an explanation, much less a way to anticipate or correct for this.

The conclusion is that there is far more variation in educational trials than would be expected if the underlying assumptions of statistical analysis were valid. A thousand students may not be enough to ensure reproducible results.

19.3.1.5 Limited imagination

The final criticism of the standard measurement protocol concerns deciding beforehand what to measure and what it should mean. It is always hard to find something one is not looking for but this practice makes it impossible. Education is too complex and human imagination too limited for this to be acceptable. Most of what the author has learned about education—nearly everything in this article—was originally well beyond his imagination. It was not clever prediction

borne out by trials, but ignorance slowly and often reluctantly dispelled by confrontation with data.

The alternative to prior decisions is data mining after the fact. The problem with this is artifacts. Randomly comparing lots of things always produces coincidences, so lots of unexpected connections will emerge but most of them are bogus. Further trials or analysis are required to eliminate bogus conclusions and sharpen real ones. This is much more trouble than running data through a statistical program but it may be necessary for real progress.

19.3.2 Goal selection

Returning to the theme of problems with measuring outcomes, we consider how goals are selected. Courses and lessons are part of an intricate whole, not free-standing entities. Goals that make sense in a limited context may either advance or undermine work at later levels, and bad choices can give great immediate outcomes but greater damage later on. Mathematics is particularly highly interconnected and vulnerable to goal-selection errors, and we describe some common ones.

19.3.2.1 Calculators

Memorizing the multiplication table is a pain and long multiplication and division are dreary. Calculators offer relief. If “accurate arithmetic” is taken as a goal then calculators are a winner. If student joy is factored in then calculators look like the best things since sliced bread. Unfortunately the view from the college level is that calculator-trained students often have significantly weaker number sense and other deficits, described below. Calculator use urgently needs to be reconsidered in spite of the glorious short-term outcomes.

Does this mean we should go back to multiplication tables? Not necessarily. For instance calculators now use keystrokes and connect directly to the motor/tactile learning channel. Perhaps tactile thinking is bad with numbers. Many students learn the multiplication table by verbal repetition so perhaps we need a connection to the auditory channel. If so then calculators with verbal data entry might solve the problem. Or if an extended expression could be entered and visually checked and edited before execution perhaps it would connect with the visual channel.

Ironically there may be problems due to an *insufficient* connection to the motor/tactile channel. Graphing calculators give students quick and accurate access to graphs of functions. They see these graphs many more times than students once did, and become adept at picking out a particular graph from alternatives on a test. On the other hand they have never drawn these curves with a pencil. When they get to multivariable calculus and have to sketch solids, or work out regions of integration, they cannot draw graphs. They also cannot articulate qualitative features (e.g. an exponential function swoops up really fast). Does this mean we should give up graphing calculators? Not necessarily.

It may be sufficient to require students to draw a picture as part of coursework and testing.

The points are that solutions at one level may lead to problems downstream; and that seeing this and finding long-term solutions might require thinking (or following data) far outside our preconceptions. If we cannot rise to this challenge then we probably should return to multiplication tables, at least for students we want to be capable of pursuing technical careers.

19.3.2.2 Over-simplification

Over-simplification of problems used to train and test students at one level can cause problems later. We describe three examples.

- Multiplying polynomials or other compound expressions is a standard task. When this is first encountered the focus is on the simplest case: two binomials, $(a + b)(x + y)$. It is common for teachers to introduce the mnemonic “FOIL” for the algorithm in this case. This increases speed and accuracy with binomials but formulates it in a way that does not generalize to larger expressions, and these students often have trouble multiplying trinomials. Describing the process in terms of associativity and distributivity may require more practice and be slower but it would make the step up to bigger problems completely routine.
- Finding roots of a quadratic corresponds to factoring it as $(x - r)(x - s)$. For simplicity this is usually illustrated and tested with quadratics with integer roots. But most quadratics do not have integer roots, and students taught this way often have trouble dealing with these when they come up in later courses. This is a much bigger problem than might be apparent from the K-12 perspective. Many methods and applications of mathematics require solving for a variable. Quadratics are one of the very few families of equations that students can easily solve and so are heavily used in examples and problems in college courses. Consequently any student who has trouble with quadratics is at a serious disadvantage.
- At a higher level, most high-school calculus courses are oriented toward preparation for the AP calculus test. Problems on this test are simplified and routine, so the course goal is to deal quickly and accurately with routine problems. Lots of mnemonics and tricks are used and the whole thing is rather mechanical. High school teachers, and more to the point AP calculus test designers, probably do not know that quite a lot of a science/engineering college calculus course is devoted to getting students to unlearn some of this.

19.3.2.3 Symbols and numbers

In K-12 work in recent years there has been a substantial increase in decimal numerical tasks and a corresponding decrease in symbolic, integer, or rational

problems. Calculators drive some of this. Teachers may believe that numbers are the real goal; symbols are just placeholders for numbers, and now that we have calculators we can do the real thing. Students can work with circles of radius 5.687 rather than “ r ”. Or a fraction like $\frac{2}{3}$ is a frustrated division, and we can now carry it out to get a “real” answer., $\frac{2}{3} = 0.6666$. Perhaps using numbers is supposed to convince students that the problems have “real-life” significance. Maybe teachers have such an investment in developing calculator skills that they want to use these at any opportunity. In addition to any of this, teachers are also responding to pressure from high-stakes tests as we describe below. In any case calculator use has blossomed.

In college courses we now see students who have trouble dealing with problems when the answer is not a number. They can handle circles with radius 5.687 but not with radius “ r ”. They have trouble with expressions with two or more symbols, and generally have weak symbolic manipulation skills.

The connection between calculators and weak symbolic skills is this: fractions and “numbers” like π and $\sqrt{2}$ are more than half-way to being symbols in the way they are handled. People who have learned to deal with $3 + \sqrt{2}$ as a root of a quadratic can routinely deal with $3 + r$. People who work with 4.414 as the root see $3 + r$ as a completely different thing. The painful algorithms used to do long division and multiplication are the same algorithms used in multiplication and division of polynomials, so the step to symbols is a minor one for students who do arithmetic by hand. It is a whole new—and complicated—world for calculator users.

Goal selection on high-stakes K–12 tests has also contributed to the decline in abstract thinking. Most tests are multiple choice, and answers are mostly numerical. Numerical answers are partly a matter of convenience for test developers. Symbols in answers provide clues: πr^2 is identifiable as the area of a circle of radius r , while 8.5 is not obviously the area of a circle of radius 2.7. Numerical answers are therefore an easy way to keep students from identifying the correct answer without working the problem. However this has consequences. It has become the primary goal of many courses to prepare students for these tests, and they therefore emphasize numbers over symbols.

We have not argued that either calculators or high-stakes tests are inherently evil, only that they are very powerful and may unintentionally have bad consequences.

19.3.2.4 The College/K–12 divide

We have observed that goal changes may lead to short-term success and long-term failure. If both are located in K-12, or both at the college level, then there is an educational research community that should, in principle, notice and correct the problem. However the examples above are of changes in K-12 that cause problems at the college level, and there is almost no communication across that divide. We briefly describe the situation.

The “new math” debacle of the 1960s drove a wedge between the K–12 and university communities. Since then research in K–12 has become quite

professional (whether or not it is on target) and the leadership has been focused and effective. College preparation is a big part of the job but they are completely confident they can accomplish it without much input from college teachers. Also, the input offered seems to be largely unsupported personal opinion. These opinions (e.g. calculators are causing problems) often conflict with articles of faith (calculators are Good) so they are seen as Wrong as well as unsupported.

Communication requires a receiver and a transmitter, and at the K-12 end the signal sounds like noise and the receiver is turned off. The problem at the college end is that there is no transmitter. More precisely, there is no mechanism for collating and sharpening individual concerns to arrive at a “conclusion of the community” let alone any good way to present such a thing to others.

The college community certainly has keen awareness of shortcomings in school math preparation but for them the important question is how to deal with it. There are strong opinions on “how did it get this way?” but these opinions are rarely carefully thought out; there is no well-developed educational research community that might extract a useful signal from the noise; and there is no leadership that might organize some other way of getting this done. The consequence is that “input” from the college community really is rarely more than individual opinion. Mathematicians certainly know that mathematical ideas must be tested with care and most will be wrong. If they approached educational ideas the same way their individual opinions might be pretty good. Unfortunately they seem to be wrong as often as anyone else.

The lack of communication is a serious problem. Here is an analog: many college math courses prepare students for work in other subjects. The needs of these subjects provide an anchor for content. Material cannot be weakened or omitted (too much, anyway) just because it does not fit well into a new educational approach because this evokes negative feedback. In principle preparation for college work should provide the same sort of content anchor for K–12 math programs. In practice the lack of communication keeps this from happening.

The conclusion is that we have no collective mechanism for dealing with problems that straddle the K-12 – college divide, and prospects for one developing any time soon are gloomy.

19.4 Teaching, learning and errors

Some problems seem to be due to, or at least hidden by, an increasing focus on teaching rather than learning. The focus itself is understandable: growing pressure for results puts attention on things that can be directly influenced, mainly teaching. Most education researchers are located in teacher preparation programs and see development of teaching techniques as their mission. The very phrase “teacher preparation” invites a focus on teaching. However in the end learning is the objective and teaching is only effective if it supports learning.

19.4.1 Teacher-centered education

When teachers are considered the main actors, students tend to be regarded as essentially all the same, if not “blank slates”. The central problem is taken to be getting students to conform to the teacher’s direction and expectations. This vision has difficulty accommodating a variety of learning styles (c.f. the discussion of tactile, visual and auditory styles above) and has many other drawbacks.

Emphasis on teaching also locates responsibility in a problematic way. Teachers, as the main actors, are responsible if students don’t learn and students are absolved of accountability. In this view the way to better performance is more pressure on teachers. This degrades the attractiveness of the profession and chases away teachers whether it improves student performance or not.

19.4.2 Error correction

The most serious problem in teacher-oriented education concerns the way errors are handled. The problem was revealed by study of new educational environments in which teachers play smaller roles, or even without a teacher in the traditional sense. Teachers are clearly not the main actors. A small amount of human mentoring is vital, but this is *not* teaching and experienced teachers often have difficulty doing it effectively. Mentoring is focused on learning, not teaching.

We describe the role of error correction in learning, and why it causes trouble. The context is that people look for, and find, patterns in their experiences. The usual explanation is that this developed as a survival skill in dangerous situations. In any case “natural learning” is a strong and largely innate part of our intelligence.

The problem is with mistakes. Any single person’s experience will have coincidences and bogus patterns and the natural learning drawn from these are wrong. But critical thinking is apparently not a survival skill: our error-correction abilities are much more primitive than the natural-learning ability. Superstitions are born easily and are notoriously hard to root out. Effective error-correction must be learned.

Learning divides roughly into corresponding stages: first getting information and seeing patterns; and then error correction in the patterns seen. Teacher-oriented education focuses on information delivery. This is the easiest part of the task and is often mechanical enough to be done by computers. Diagnosis and correction of errors is more subtle and for the foreseeable future will depend on teachers or mentors.

In brief: the really essential role of teachers is not information delivery, but diagnosis and correction of errors. This is also the hardest part of independent learning, so the best way a teacher can help a student “learn how to learn” is to be clear and deliberate about error correction.

We illustrate how this plays out in practice with example responses to a student making mistakes in a math problem:

“This is wrong. I’ll watch while you go through your work; let me know if you spot the error and I will watch for it too.”

The student is the main actor. It is clear that a mistake has been made and must be corrected, and that the student has significant responsibility for the correction. The student learns to find errors as well as getting the specific error fixed.

“This is wrong. I see your mistake and will show you how to avoid it in the future.”

The student is more passive. However the information delivered is targeted and the student sees the diagnosis process in action.

“This is wrong. Let me show you how to do it.”

The student is passive. The mentor sees an error was made, the student’s work is discarded rather than diagnosed, and the response is repetition of information. Clues about the specific error are buried in the general picture and frequently no more accessible than the first time the information was delivered.

“This isn’t quite right. I see you have the right idea but I’ll show you again.”

This phrasing is common even when they don’t have the right idea, as a gentle and encouraging way to tell them they are wrong. However it undercuts the learning process by suggesting there is no need to locate and correct an erroneous “idea”. The tiresome technical error may not seem important enough to need correction. The student is disengaged as well as passive.

These responses range from effective to unproductive, with the latter being more common today. The conclusion is that weaker critical thinking and independent-learning skills may be due in part to teaching methods in K–12: emphasis on presentation and information delivery; neglect of error diagnosis; kinder, gentler ways to deal with errors; even suggesting that untestable “understanding” may be as valuable as testable skills.

An error-diagnosis approach to learning requires cooperative students. Not only must they be willing to participate individually, but in a class setting the rest of the class must be able to work independently while one student is getting individual attention. The approach is vulnerable to disruption by unwilling or disinterested students. Consequently for researchers to even consider this approach would require them to both shift emphasis from teaching to learning and to either exclude or account for the effects of disruptive students.

19.4.3 Repetition

International comparisons reveal that in the US material is presented in shorter segments and there is much more repetition. Outcomes are also weaker. It is often suggested that to improve outcomes we should lengthen segments and reduce repetition. However short repeated segments may be a symptom rather than the root cause, and changing this without addressing root causes may worsen the situation.

Short lesson segments may work better for classrooms with little error-correction. Once a significant number of students have become dysfunctional it makes sense to take a break, give them time to forget, and try again with a new round of information delivery. Students prone to errors may have to go through the process several times before they stumble on the correct approach. It is, of course, inefficient for students who got it right the first time.

Lack of error correction may itself be a symptom rather than a root cause. Error diagnosis and correction is a one-on-one activity so in a classroom setting it requires the cooperation of all students in the class. Focus on information delivery may work better for classrooms with discipline problems.

This chain of connections, tracing bad outcomes back through repetition and lack of error correction to discipline problems, is highly speculative. The point here is not whether or not it is correct but that current mindsets and evaluation procedures prevent the educational research community from investigating such things.

19.5 An Obsolete Model

A deeper problem than those discussed above is that current mathematics education is modeled on nineteenth-century mathematics. Nineteenth-century methodology was inadequate for modern use and was substantially revised at the beginning of the twentieth century. It seems that commitment to an obsolete model not only contributes to educational problems but explains why educational researchers are unable to identify or resolve them. This is discussed in some detail in *The Nature of Contemporary Mathematics*, <http://www.math.vt.edu/people/quinn/education/nature0>.

19.6 Summary

The 1983 “Nation at risk” report described the K–12 situation at the time in dire terms:

If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen to ourselves. [...] We have, in effect, been committing an act of unthinking, unilateral educational disarmament.”

This proclamation sparked enormous activity in education research and upheavals in curriculum design, all intended to address the problem. But in many ways the problem has gotten worse. In the 1990s ambitious graduate programs in mathematics were largely populated by immigrants: we were importing high-quality K–12 and undergraduate education, critical thinking, and work ethic. Now we can no longer meet demand through imports and the high-tech jobs that require these skills are beginning to be exported.

Not only has the enormous activity been largely unproductive, but attitudes, assumptions and methodologies employed seem more likely to accelerate the

decline than arrest it. Finally the problematic attitudes and methodologies seem to be locked in place by political pressures and incorporation into policies of the NSF and other funding sources. It seems likely that our educational disarmament will continue for the foreseeable future.