The Math Academy Way

Using the Power of Science to Supercharge Student Learning

– Working Draft –

Note: If you are reading this as a PDF,
consider switching to the Google Doc
which may contain new updates.

Authored by Justin Skycak, advised by Jason Roberts.
Major acknowledgements: Alexander Smith, Sandy Roberts, Yurii Leshchenko.

First edition (working draft), updated 28 November 2024.

mathacademy.com

Authored by Justin Skycak, advised by Jason Roberts.
Major acknowledgements: Alexander Smith, Sandy Roberts, Yurii Leshchenko.

First edition (working draft), updated 28 November 2024.

www.mathacademy.com

Contents

Preface **19**

I. PRELIMINARIES 21

Chapter 1. The Two-Sigma Solution **23**

Bloom's Two-Sigma Problem 23

Talent Development vs Traditional Schooling 24

Talent Development is Prohibitively Expensive 28

Stages of Talent Development 29

| Bloom's 3-Stage Talent Development Process 29

| Bloom's Taxonomy is Often Misinterpreted 30

Key Papers 33

Chapter 2. The Science of Learning **35**

Cognitive Learning Strategies 35

The Persistence of Tradition 37

A Common Theme Preventing Adoption 39

| Theme and Examples 39

| Desirable Difficulty vs Illusion of Comprehension 41

| The Educational System Prefers Illusion 42

Technology Changes Everything 44

| Revival via Technology 44

| Necessity of Technology 44

| Accelerating Student Learning by 4x 45

Key Papers 47

Chapter 3. Core Science: How the Brain Works **49**

Sensory, Working, and Long-Term Memory 49

Design Constraints 52

Case Study: Information Flow During a Computation 53

Neuroscience of Working Memory 59

Key Papers 61

Chapter 4. Core Technology: the Knowledge Graph **63**
Understanding the Knowledge Graph 63
| Linking Topics and Prerequisites 63
| Zooming Out 64
| Course Graph 66
Using the Knowledge Graph 68
| Scaffolded Mastery Learning 68
| Additional Linkages 69
> Key Prerequisites Enable Targeted Remediation *69*
> Encompassings Enable Turbo-Boosted Learning Speed *71*
| Diagnostic Exams 73
Chapter 5. Accountability and Incentives **75**
Accountability and Incentives are Necessary but Absent in Education 75
What Happens in the Absence of Accountability and Incentives 77
| Tragedy of the Commons 77
| Grades Can't Be Trusted 79
> Evidence for Grade Inflation *79*
> Why Grade Inflation is a Problem *82*
| Resistance to Objective Measurement 84
> Radical Constructivism Rejects the Idea of Measuring Learning Objectively *84*
> Radical Constructivism is a Present Force in Education *85*
Math Academy is Held Accountable for Student Learning 88
Key Papers 89
II. ADDRESSING CRITICAL MISCONCEPTIONS 91
Chapter 6. The Persistence of Neuromyths **93**
Neuromyths are Common Misconceptions about the Brain 93
Why Neuromyths Persist 94
Key Papers 95
Chapter 7. Myths & Realities about Individual Differences **97**
People Differ in Learning Speed, Not Learning Style 97
| Learning Style Preferences are Irrelevant 97
| Working Memory Capacity (WMC) Differences are Relevant 98

> *WMC Impacts Perceived Effort* *99*

> *WMC Impacts Abstraction Ability* *100*

> *WMC Impacts Learning Speed* *101*

| Lack of Evidence for WMC Training 103

| Different Students Need Different Amounts of Practice 105

Your Mathematical Potential Has a Limit, but it's Likely Higher Than You Think 106

| Levels of Math 106

| The Abstraction Ceiling 107

| Learning Energy vs Level of Math 108

| All Can Learn Some, Many Can Learn More, but Few Can Learn All 111

> *Nature or Nurture? Both Matter* *111*

> *Speed of Skill Acquisition Matters Because Time is Limited* *113*

| Why the "All Can Learn All" Myth Persists 114

| Struggle Does Not Imply Inability 115

> *Struggle Can Be Caused by Missing Foundations* *116*

> *Struggle Can Be Caused by Ineffective Practice* *117*

> *Struggle Can Be Caused by Insufficient Practice* *118*

> *Struggle Can Be Caused by Lack of Motivation* *119*

| Analogy to Lifespans 120

| Student Bite Size vs Curriculum Portion Size 121

> *The Misconception: If Instruction is Done Perfectly, Won't All Students Learn at the Same Rate?* *121*

> *The Resolution: Under Favorable Learning Conditions, Student Bite Size Equals Curriculum Portion Size* *121*

> *Consistency with Observations* *122*

Key Papers 124

Chapter 8. Myths & Realities about Effective Practice 129

Effective Practice Does Not Emulate the Professional Workplace 129

| Direct Instruction is Needed 130

> *Definition and Importance* *130*

> *Unguided Instruction has a History of Pseudoscience* *131*

> *Unguided Instruction is Logically and Scientifically Inconsistent* *132*

> *Unguided Instruction Leads to Major Issues in Practice* *134*

| Many Hands Make Light Work… and Light Learning 136

Effective Practice Requires Effort........ 138
| There is No Such Thing as Effortless Learning........ 138
| To Oppose Effortful Practice is to Oppose Talent Development........ 140
> *Testing and Repetition are Necessary........ 140*
> *Computation is Necessary........ 141*
> *Competition Can be Helpful and is Unavoidable in the Big Picture........ 142*
> *Why the Myth Persists in Education (But Not in Talent Development)........ 143*
> *Why Talent Development is Important in Math........ 144*
Key Papers........ 146

Chapter 9. Myths & Realities about Mathematical Acceleration........ 149
Acceleration is Often Misunderstood........ 149
Developmental Appropriateness........ 150
| Advanced Study is Appropriate Once Prerequisites Have Been Mastered........ 150
| Why the Myth of Developmental Inappropriateness Persists........ 151
Depth of Learning........ 153
| Accelerated Students Learn More Material, Just as Deeply........ 153
| How We Ensure Comprehensive, Deep Learning........ 154
Continuity of Courses........ 155
| Accelerated Students Don't Run out of Math Courses........ 155
| Advanced Students Can Place Out of College Courses Beyond Placement Tests........ 156
Relevance to Students' Futures........ 157
| Learning Math Early Reduces Risk and Opens Doors to Opportunities........ 157
| Higher-Grade Math is Typically More Productive than Grade-Level Competition Problems........ 158
Key Papers........ 160

III. COGNITIVE LEARNING STRATEGIES........ 163

Chapter 10. Active Learning........ 165
Definition and Importance........ 165
Case Studies........ 167
| Case Study 1: Why Active Learning is Obvious........ 167
| Case Study 2: Most Students Don't Even Pay Attention During Lectures........ 169
| Case Study 3: How Active Learning Saved MIT's Physics Classes........ 170
| Case Study 4: If You're Active Half the Time, That's Still Not Enough........ 171

Neuroscience of Active Learning....173

Persistence of Misconceptions....174

Key Papers....176

Chapter 11. Deliberate Practice....179

Definition and Importance....179

| Deliberate vs Non-Deliberate Practice....179

| Deliberate Practice is Effective, Non-Deliberate Practice is Not....181

Effort is Required....182

| Deliberate Practice Feels Like Exercising with a Personal Trainer....182

| Cycle of Strain and Adaptation....183

Discomfort is Required....184

Long-Term Compounding....187

| Expertise is the Product of Incremental Improvements Over Time....187

| Motivational Supplements are Not Substitutes for Deliberate Practice....188

Misinterpretations of Deliberate Practice....189

Key Papers....191

Additional Resources....193

Chapter 12. Mastery Learning....195

Mastery Learning is Underused....195

Implementing Mastery Learning....197

Knowledge Frontier as Zone of Proximal Development....200

Key Papers....203

Chapter 13. Minimizing Cognitive Load....205

The Learning Staircase....205

Micro-Scaffolding....207

The Expertise Reversal Effect....210

Key Papers....212

Chapter 14. Developing Automaticity....213

Importance of Automaticity....213

| Autonomicity Frees Up Working Memory....213

| Working Memory is Limited, but Long-Term Memory is Not....214

| Expertise Requires Automaticity....216

Case Study: Computing Exponents With vs Without Automaticity

on Multiplication and Addition Facts....217
Automaticity, Creativity, and Higher-Level Thinking....220
| Automaticity is Necessary for Creativity.... 220
| Automaticity is Necessary for Higher-Level Thinking....222
| Automaticity is a Gatekeeper to Mathematical Literacy and Academic Success.... 223
Neuroscience of Automaticity....224
Key Papers....226
Chapter 15. Layering....229
Facilitation and Structural Integrity.... 229
| Facilitation....229
| Structural Integrity....230
How We Layer....231
Key Papers....233
Chapter 16. Non-Interference....235
Associative Interference....235
Non-Interference....236
Key Papers....238
Chapter 17. Spaced Repetition (Distributed Practice)....239
Retaining Knowledge Indefinitely....239
| The Spacing Effect....239
| Spaced Repetition....240
| Building Intuition About Spaced Repetition....241
| Consensus Among Researchers....242
Spaced Repetition is Underused.... 244
Spaced Repetition Improves Generalization....246
Repetition Compression....248
Calibrating to Individual Students and Topics....250
Spaced Repetition vs Spiraling....252
Key Papers....255
Additional Resources....257
Chapter 18. Interleaving (Mixed Practice)....259
Interleaving vs Blocking....259
Benefits of Interleaving....260

| Efficiency 260

| Discrimination and Category Induction Learning 261

| Experimental Support 262

Desirable Difficulty: Why Interleaving is Underused 263

Micro- and Macro-Interleaving 266

| Macro-Interleaving 266

| Micro-Interleaving 266

Key Papers 269

Additional Resources 272

Chapter 19. The Testing Effect (Retrieval Practice) 273

Retrieval is the Most Effective Method of Review 273

Spaced Retrieval Practice 274

The Testing Effect is Underused 275

Reducing Anxiety and Promoting a Growth Mindset 276

| Appropriate vs Inappropriate Usage of Timed Tests 276

| Desirable vs Undesirable Difficulties 277

| Appropriate Timed Testing Can Reduce Math Anxiety 278

| Implementing Appropriate Timed Testing 281

Key Papers 282

Additional Resources 285

Chapter 20. Targeted Remediation 287

High-Granularity, High-Integrity Remediation 287

Corrective Remediation 288

Preventative Remediation 290

Foundational Remediation 291

Content Remediation 292

Chapter 21. Gamification 295

Importance of Gamification 295

| Increasing Learning, Engagement, and Enjoyment 295

| Increasing Learning Efficiency 296

Incentivizing Quantity and Quality of Work 297

| XP-Time Equivalence 297

| Incentivizing Quantity of Work 298

| Incentivizing Quality of Work........ 299
Closing Loopholes........ 300
Progress vs XP........ 301
Key Papers........ 304
Chapter 22. Leveraging Cognitive Learning Strategies Requires Technology........ 305
High-Level Context........ 305
| The Problem: Cognitive Learning Strategies Remain Underused........ 305
| The Blame: Teachers are Victims of Circumstance........ 306
| The Solution: Technology Changes Everything........ 306
Resistance to Additional Effort........ 307
| Active Learning........ 308
| Non-Interference, Interleaving, and Spaced Repetition........ 308
> Shuffling Instructional Material........ 308
> Opening a Can of Worms on Forgetting........ 309
| Testing Effect........ 310
| Gamification........ 311
Tutoring the King's Kid: How Would You Teach if Your Life Depended On It?........ 311
Heterogeneity of Student Knowledge Profiles........ 312
| Tutoring the King's Kid vs Teaching Many Kings' Kids........ 312
| Differences in Background Knowledge........ 313
| Student Knowledge Profiles Naturally Tend Towards Heterogeneity........ 314
| Every Student in the Class Effectively Needs a Private Tutor........ 315
IV. COACHING........ 317
[In Progress] Chapter 23. Parental Support........ 319
Why Parental Support is (Usually) Necessary........ 319
The Bare Minimum: Incentives and Accountability........ 319
The Platonic Ideal........ 321
[In Progress] Chapter 24. In-Task Coaching........ 323
The Vicious Cycle of Forgetting........ 323
V. TECHNICAL DEEP DIVES........ 325
Chapter 25. Technical Deep Dive on Spaced Repetition........ 327

Fractional Implicit Repetition (FIRe)...327
| Concrete Example...327
| Visualizing Repetition Flow...329
| Partial Encompassings...329
Setting Encompassing Weights Manually...331
| Direct and Key Prerequisites are Sufficient...331
| Non-Ancestor Encompassings and Mastery Floors...332
Student-Topic Learning Speeds...333
| Ratio of Student Ability and Topic Difficulty...333
| Measuring Student Ability at the Level of Individual Topics...333
| Measuring Topic Difficulty...334
High-Level Structure...335
Chapter 26. Technical Deep Dive on Diagnostic Exams...339
Minimizing the Number of Questions...339
Knowledge Confidence and Conditional Completion...340
| Theory...340
| Example...341
| Implementation...342
Conservative vs. Aggressive Edge of Mastery...343
Supplemental Diagnostics...344
Selecting Good Diagnostic Questions...344
Chapter 27. Technical Deep Dive on Learning Efficiency...347
What is Learning Efficiency?...347
| Theoretical Maximum Learning Efficiency...347
| Theoretical Minimum Learning Efficiency...348
Factors that Impact Learning Efficiency...349
| Performance...350
| Pace...351
> Learning Efficiency vs Pace...351
> Pace vs Time to Completion...352
Chapter 28. Technical Deep Dive on Prioritizing Core Topics...355

Core and Supplemental Topics 355

The Mathematical Foundations Sequence 356

Frequently Asked Questions 359

The Practice Experience 359

| Active Learning, Scaffolding, and Automaticity 359

> How does a lesson work? 359

> Solving problems breaks my flow of learning. Is it really necessary? 360

> Don't students need to struggle for long periods of time, without too much guidance, to train their general problem-solving ability? 360

> Is automaticity really required to move up to the next level? Doesn't it just come with time? .. 362

> If worked examples are necessary to maximize learning efficiency, then why am I able to solve problems just fine without them? 363

> Why aren't Math Academy's university courses structured like typical higher math textbooks with minimal scaffolding? 364

| Spaced Repetition and Interleaving 365

> Why can't I just learn one unit at a time? Interleaving feels disorienting. 365

> Is the spaced repetition happening? I started recently and all I have is lessons. Where are the reviews? 365

> Math Academy reviews feel challenging. Aren't reviews supposed to be easy if I learned the material properly? 366

> Does Math Academy's spaced repetition system provide enough practice? I learned some new information on Math Academy but I am not confident in my ability to retrieve it from memory unassisted. 367

> After I complete a course and move to the next course, won't I forget what I've learned in the first course unless I keep on reviewing it? 368

> Given that lessons can "knock out" reviews, should students always give preference to lessons over reviews if both activity types are available? 368

> Math Academy maximizes learning efficiency if a student is willing to engage in forms of training that are highly effortful. What about for students who don't have as much energy and motivation? 368

| Remediation 369

> If a student fails a task, why does the system ask them to re-attempt the task later? Why doesn't it just peel back their knowledge profile immediately? 369

Student Behavior 369

| Usage of Paper and Pencil 369

> Should Math Academy students take notes during lessons? 369

> When should a student solve a problem in their head vs writing it out on paper? 370

| Reliance on Reference Material......373

> During a quiz, if I can't remember a "fact" (e.g., a definition or theorem) but I remember the process for using it to solve problems, should I look it up?......373

> What should a student do if they are unsure how to solve a problem despite trying their best to refer back to the supporting instructional content?......373

XP and Practice Schedules......374

| XP System......374

> If I pass a lesson but don't get full XP, does that mean I only understood part of the material? If so, how does Math Academy fill in the rest of my understanding?......374

> I can do way more than 1 XP per minute!......374

> As I progress through Math Academy's curriculum, it is gradually taking me longer to earn XP. Why does this happen, and can anything be done about it?......375

> I don't think XP is a perfect measurement of effort......376

| Practice Schedules......377

> If I have a limited amount of time to devote to Math Academy each week, should I allocate that time into longer, less-frequent sessions or shorter, more-frequent sessions?......377

> If a student completes a Math Academy course very rapidly, working several hours per day for several weeks, will they still learn the material properly?......378

> Can Math Academy be used for very casual learning, an hour or two per month?......378

Diagnostics and Curriculum......378

| Diagnostics......378

> There were way more questions than I expected on the diagnostic!......378

> Can't you improve the diagnostic algorithm to cut down on the number of questions?......379

> There is a topic that I know how to do, but the diagnostic didn't ask me about it and I didn't get credit for it......380

> After the diagnostic, what if I am asked to complete a lesson for which I have not learned a prerequisite?......380

> After the diagnostic, why did failing a single task bring my progress down multiple percent?...381

| Curriculum......382

> With short lessons, is the curriculum really comprehensive?......382

> Will taking a Math Academy course for a standardized exam (e.g., AP Calculus BC) fully prepare a student for the actual exam?......382

> Does Math Academy explain the "why" behind procedures? Are all the concepts taught first? 384

> Do your university courses have exercises with proofs or is it just computation?......385

> It's hard to believe that 5 hours a week for a year starting from basic multiplication tables will have me completely prepared for university courses. I'd prefer an explanation for people who are not familiar with the XP system......386

> Why is "holistic mode" (in which students also fill in any missing knowledge in lower-grade topics that are not prerequisites of their enrolled course) disabled for university courses?........... 387

Miscellaneous.. 388

| Math Academy Itself.. 388

> Why use Math Academy instead of self-studying a textbook or free online resource?............... 388

> Why isn't Math Academy free?...389

> Where do the exercises and content on Math Academy come from? Are they all made in-house or pulled from other materials?...390

| Features that Do Not Exist for a Good Reason.. 391

> I don't really want to do any of the learning tasks that Math Academy presents to me. There are other topics I would rather learn. Why can't I choose my own tasks?... 391

> Why can't I edit my knowledge profile?... 392

> Why isn't there an "I don't know" button on questions during tasks other than the diagnostic?..... 392

> Why doesn't Math Academy use large language models (LLMs) to engage students in conversational dialogue?...393

> If I get stuck, is there somewhere that I can ask for help or a further explanation?...................395

Glossary.. 397

References.. 409

Notes for Future Additions.. 425

Parental Support / Coaching...425

Direct Instruction..426

Knowledge Spaces... 427

Spaced Repetition Visualization.. 428

Chunking.. 430

Biological Basis of Neuroplasticity...430

Automaticity and Intuition.. 431

More Case Studies..431

Spaced Repetition in Math vs Language...431

Disambiguating Interleaving, Non-Interference, and Spaced Repetition......................... 432

Visualization of Macro-Interleaving... 432

Elaborative Interrogation..432

Optimal Manual Teaching... 432

Elaborative Interrogation..433

Coaching Science.. 433
Miscellaneous Resources to Check Out..433

Preface

This book answers the following questions:

1. What techniques exist to maximize student learning and talent development, particularly in the context of math?

2. Why are these techniques so impactful, and if they are indeed so impactful, then why are they so often absent from traditional classrooms?

3. How does Math Academy leverage these techniques?

The book is admittedly verbose with numerous lengthy quotes pulled from the literature, but we believe these receipts are vital for building trust with the reader. Too often, writeups in the field of education cite references that don't even support their claims. Consequently, when faced with the decision to (a) build credibility by quoting the literature extensively, versus (b) streamline our communication, we have chosen to lean towards credibility. It is our experience that if credibility is anything but high, no communication will occur.

The book is written with many different audiences in mind: students, teachers, parents, researchers, technologists, math hobbyists, academic coaches, just to name a handful. We have done our best to include much information of interest to those who wish to dive deep, while simultaneously structuring the information in a way that is skimmable for the surface-level reader.

The book is a working draft, about halfway done. It has been written with care but has not undergone formal proofreading.

I. PRELIMINARIES

Chapter 1. The Two-Sigma Solution

Summary: *Educational psychologist Benjamin Bloom is widely known for demonstrating that one-on-one tutoring produces vastly better learning outcomes than traditional classroom teaching, and documenting how talent development differs from traditional schooling. Math Academy is addressing these issues by creating an adaptive, fully-automated online mathematical talent development platform that emulates the decisions of an expert tutor to provide the most effective way to learn math.*

Bloom's Two-Sigma Problem

In 1984, educational psychologist Benjamin Bloom published a landmark study comparing the effectiveness of one-on-one tutoring and traditional classroom teaching. The difference was monumental: the average tutored student performed better than 98% of the students in a traditional class.

This finding led to a challenge widely known as **Bloom's two-sigma problem:** can we develop methods of group instruction that are as effective as one-on-one tutoring? (The terminology "two-sigma" comes from statistics, where the effects of interventions are often measured in standard deviations or *sigmas*. An effect size of 98% is slightly more than two sigmas.)

To quote Bloom directly (Bloom, 1984):

> *"...[T]the most striking of the findings is that under the best learning conditions we can devise (tutoring), the average student is 2 sigma above the average control student taught under conventional group methods of instruction.*
>
> *The tutoring process demonstrates that most of the students do have the potential to reach this high level of learning. I believe that an important task of research and instruction is to seek ways of accomplishing this under more practical and realistic conditions than the one-to-one tutoring, which is too costly for most societies to bear on a large scale.*
>
> *This is the '2 sigma' problem. Can researchers and teachers devise teaching-learning conditions that will enable the majority of students under group instruction to attain levels of achievement that at present can be reached only under good tutoring conditions?*
> *...*
> *If the research on the 2 sigma problem yields practical methods ... it would be an educational contribution of the greatest magnitude. It would change popular notions about human potential*

> *and would have significant effects on what the schools can and should do with the educational years each society requires of its young people."*

Bloom speculated that an equivalent two-sigma effect might be achieved by combining various evidence-based learning strategies, especially those involving different objects of change (the learner, the instructional material, the home environment or peer group, and the teacher and teaching process) and those that occur at different times in the teaching-learning process.

This is Math Academy's challenge and purpose. Math Academy is solving Bloom's two-sigma problem by bringing together many evidence-based cognitive learning strategies into a single online learning platform. Our adaptive, fully-automated platform emulates the decisions of an expert tutor to provide the most effective way to learn math.

Talent Development vs Traditional Schooling

The core philosophy of Math Academy is centered around **talent development** as opposed to traditional schooling. At surface level, the two ideas may seem similar: after all, isn't the purpose of schooling to develop students' talents? Bloom, who researched this question extensively, discovered that the answer is a resounding "no" – the differences between talent development and traditional schooling are so numerous, so striking, and so critical that traditional schooling typically cannot even be characterized as supporting talent development.

Around the same time that Bloom coined the two-sigma problem, he was also immersed in a massive study of talent development. As summarized by other researchers (Luo & Kiewa, 2020), Bloom (1985) discovered striking commonalities in the upbringing of extremely successful individuals across a wide variety of fields, leading to a general characterization of the process of talent development:

> *"Research interest on talent development was sown by psychologist Benjamin Bloom's (1985) seminal book, Developing Talent in Young People. Bloom studied 120 highly talented individuals across six talent domains and discovered common factors that led to exceptional achievements across domains. In an interview, Bloom remarked:*
>
> *'We at one time thought that the development of a tennis player would be very different from the development of a concert pianist or a sculptor or a mathematician or a neurologist. What we've found is that even though the content and the procedures may be enormously different in each field, there is a common set of characteristics in the home, the instruction, and the like. There is a very general process that seems to be central to the development of talent no matter what the field. (Brandt, 1985: 34)'"*

Bloom believed (Brandt, 1985) that this talent development process was being leveraged much more effectively in athletic than in academic contexts, and that there was an opportunity to massively elevate students' degree of learning and academic achievement by reproducing favorable conditions for talent development:

> *"I [Bloom] firmly believe that if we could reproduce the favorable learning and support conditions that led to the development of these [extremely successful] people, we could produce great learning almost everywhere.*
>
> *...*
>
> *[T]hey [educators] do a very good job in sports. There's nothing we can tell coaches in high schools and colleges. But when we get beyond sports, things are sporadic, accidental. Students may have a good teacher one year and a very poor one the next. And even in the academic subjects, all kinds of chance circumstances are at work. ... Schools do not seem to have a great tolerance for students who are out of phase with other students in their learning process."*

One of the main differences between traditional schooling and talent development, according to Bloom & Sosniak (1981), is that students are grouped primarily by age, rather than ability, and each group progresses through the curriculum in lockstep. Each member of the group engages in the same tasks, and it is expected that different students will learn skills to different levels.

> *"The school schedule and standards are largely determined by the age of the child. The curriculum and learning experiences are presumably appropriate to most students at that age or grade. While there may be some adjustments for different rates of progress and some adjustment of standards for individuals within a grade or classroom, each individual is instructed as a member of a group with some notion that all are to get as nearly equal treatment as the teacher and the instructional material can supply.*
>
> *...*
>
> *[T]he group is central in the school learning process and only minimal adjustments are made for individual children. If the group as a whole has difficulty, the teacher will reteach the task or skill until some portion of the group has learned it. But generally, all the children are not expected to learn a task or skill to the same level and little is done with the use of feedback-corrective procedures to bring all children to the same standard of accomplishment.*
>
> *Since it is not expected that each child will learn to the same standard or level, relative standards are emphasized, but the tasks are the same. Certain children are expected to learn a task to a high level while others are expected to learn it only to a much lower level."*

In talent development, however, instruction is completely individualized. Learning tasks are chosen based on the specific needs of individual students, each student must learn each skill to a sufficient level of mastery before moving on to more advanced skills. Students progress through skills at different rates, but learn skills to the same threshold of performance. Their progress is measured not by their level of learning in courses that they have taken, but rather by how advanced the skills are that they can execute to a sufficient threshold of performance.

> *"Part or all of the instruction the talented individual received was on a one-to-one basis. The pianists had weekly or twice weekly private lessons. ... The swimmers worked with many other*

> *swimmers in the pool, but the instruction was individualized and personalized. The mathematicians had much less systematic instruction in the early years, but they almost always learned alone or with one adult or peer.*
>
> *Some of the instruction each week was provided by a teacher (tutor) who diagnosed what was needed, set learning objectives, and provided instruction with frequent feedback and correctives. The teacher also suggested appropriate practice, emphasizing specific points or problems to be solved, and set a time by which the individual was expected to attain the objectives to a particular standard. At the end of the set time, the child performed and the teacher noted the gains and what had still to be accomplished, gave corrective instruction, and then gave further instruction for new material and procedures. The teacher praised and encouraged the child for his or her accomplishments, and when the standard was attained, set a new task and further objectives and standards. The cycle of learning tasks, objectives, standards, and motivation was repeated over and over as the child progressed.*
>
> *In talent development, each child was seen as unique and the teacher (tutor) set appropriate learning tasks for the child, gave rewards which the child valued or responded to, and set the pace of learning believed to be appropriate for the individual child. The child's learning rate was central and there was continual adjustment to the child learning the talent. The objectives and standards set by the teacher were always in terms of specific tasks to be accomplished in particular ways by the individual child. While the child was frequently judged in comparison with other children, emphasis was on the accomplishment or mastery of the particular learning tasks set for the individual."*

To recap, Bloom & Sosniak (1981) summarized these differences as follows:

> *"In general, school learning emphasizes group learning and the subject or skills to be learned. Talent development typically emphasizes the individual and his or her progress in a particular activity. In school group learning, little is done to help each individual solve his or her special learning problems, while in talent learning the instruction is regarded as good, at least by the parents, only if it helps the individual make clear progress, overcome learning difficulties, and move to higher and higher standards of attainment."*

They also noted that these differences are closely related to the scope of a teacher's responsibility: in traditional schooling, teachers focus on a "cross section" of many students covering a small subset of curriculum over a short period of time, whereas in talent development, teachers have "longitudinal" accountability for fewer students each learning long progressions of skills over a long periods of time.

> *"...[In talent development, the teacher] emphasizes the child's progress from lesson to lesson with the child's stage at one time as the benchmark for noting progress or gains. ... The teacher is concerned with the child's growth and progress toward what is possible at the highest level. This stems from the likelihood that the teacher will remain with the child over a number of years and also from the teacher's long term view of what is possible for the particular child.*
>
> *In contrast, the schools are arranged by courses. Although the curriculum in a particular subject may extend over a period of ten or more years, each teacher has the child only for a term, year, or course. And the teacher is responsible only for what happens during that period of time. The teacher judges each child in terms of how well he or she is doing in comparison with other*

> *children at that grade level or in that class. Each teacher at a particular grade level is primarily concerned with the teaching and learning appropriate to that grade. Little attention is paid to what the child has already learned, or to what each child will need to effectively enter the next grade or course."*

Bloom & Sosniak (1981) also observed that these differences are so critical that traditional schooling typically cannot even be characterized as supporting talent development. As Bloom describes, talent development is not only different from schooling, but in many cases *completely orthogonal* to schooling:

> *"For one portion of our sample, talent development and schooling were almost two separate spheres of their life. ... Usually the student made the adjustments, resolving the conflict by doing all that was a part of schooling and then finding the additional time, energy, and resources for talent development. ... Mathematicians found and worked through special books and engaged in special projects and programs outside of school.*
>
> *Sometimes the schools or particular teachers made minor adjustments to dissipate the conflict. Mathematicians were sometimes excused from a class they were too advanced for and allowed to work on their own in the library. Sometimes they were accelerated one grade as a concession to their outside learning.*
>
> *...*
>
> *Whether the individual or the school made these adjustments, it was clear that these adjustments minimized conflict but did little to assist in talent development. The individual was able to work at both schooling and talent development, although with minimum interaction. ... Talent development and schooling were isolated from one another. Schooling did not assist in talent development, but in these instances it did not interfere with talent development."*

And while other participants that Bloom studied had more overlap between schooling and talent development, the overlap was not always positive. Rather, it yielded a mixed bag of experiences:

> *"For a second portion of our sample, school experiences were a negative influence on their talent development. For these individuals the conflicting requirements of talent development and schooling could rarely be resolved. Schooling was truly something to be suffered through. These individuals found that their efforts in the talent field were not well received by teachers, principals, or peers.*
>
> *...*
>
> *For the third portion of our sample, we find the most encouraging role of the schools in talent development. School experiences became a major source of support, encouragement, motivation, and reward for the development of talent. ... Some individuals found private support for their development of talent from teachers or principals. These teachers or principals noticed the child's special development and recognized the quality of his or her work. ... They recognized the student's seriousness and shared with the student an eagerness for working toward very high standards and a commitment to excellence."*

The general orthogonality of schooling and talent development, and the mixed bag of positive and negative experiences resulting from any overlap between them, echo one of Bloom's quotes (Brandt, 1985) at the beginning of this chapter:

> *"...[W]hen we get beyond sports, things are sporadic, accidental. Students may have a good teacher one year and a very poor one the next. And even in the academic subjects, all kinds of chance circumstances are at work."*

Talent Development is Prohibitively Expensive

Unfortunately, in most fields and particularly in mathematics, there is no widely available solution to the lack of talent development by traditional schools other than 1-on-1 private coaching, which is prohibitively expensive for most families and schools.

To understand just how expensive this is, let's work through the cost computation. The question that we seek to answer is as follows:

> *How much would it cost to develop a student's mathematical talent to the maximum degree with a 1-on-1 private coach, assuming a reasonable amount of daily work time that is in line with the amount of time that students would be working anyway during the school year?*

In an academic subject like mathematics, 1-on-1 private coaching would be obtained from a tutor. Note, however, that we are *not* concerned with the question "how much does typical usage of supplemental tutoring cost." We are *not* supposing that the tutor functions as a supplemental assistant who helps a student through their class homework. Instead, we are supposing that the tutor functions as a main instructor, specifically, a private coach who engages the student in 1-on-1 talent development using a personalized training program that is tailored and constantly adapting to their individual needs.

We are supposing that the tutor is hired to completely replace the student's mathematical training from school, which, as a conservative estimate, is approximately 1 hour per day, 5 days per week. (This estimate is conservative because students typically have 50 minutes of class each day plus 30-60 minutes of homework.) A tutor typically charges at least \$50/hour, and \$50 × 5 days/week × 52 weeks/year = \$13,000.

This ballpark lower bound is in line with Guryan et al. (2023), who describe a successful low-cost tutoring intervention (40 minutes per school day, 1 tutor per 2 students) that cost about \$4,000 per student per year, with tutors being paid a yearly stipend of only \$16,000 (plus benefits) while working through the entire school day (6 class periods). Under these conditions, a full hour of fully individualized tutoring (1 tutor per student) each school day would cost \$12,000 per student per year (= \$4,000 × 2 × 60/40).

It's important to note that while these tutors described by Guryan et al. (2023) possessed strong math skills, they were not long-term expert coaches in the sense of the preceding discussion on talent development. Rather, tutors were *"willing to devote one year to public service – for example, recent college graduates, retirees or career-switchers – but do not necessarily have extensive prior training or experience as teachers."* Needless to say, long-term expert coaches would be far more costly and harder to find.

Additionally, while there do exist mathematical "talent search" competitions in which top competitors are selected for free talent development, only a tiny proportion of highly talented students take the exam and make the cut, and the duration of talent development that they receive is brief. To quote mathematician George Berzsenyi (2019):

> *"Participation in each of these [exams] is based on performance in the previous competition(s), and hence to a great extent the entire process is aimed at finding about 60 students for the three-week Mathematical Olympiad Summer Program (MOSP), where six students are selected to represent the United States at the International Mathematical Olympiad (IMO).*
>
> *It always bothered me to have several hundred thousand students take the AMC, learn that tens of thousands of them are talented, and then select 60 for a brief talent development program and ignore the rest, expecting them to develop their own capabilities and, if not discouraged, come back the following year to prove themselves again."*

The goal of Math Academy is to make mathematical talent development widely available to serious students who are motivated to undertake it – and bringing the $13,000 figure down to $499/year (26x cheaper) via Math Academy makes mathematical talent development accessible to many, many more people.

Stages of Talent Development

| Bloom's 3-Stage Talent Development Process

As summarized by researcher Gordon Bloom (2002), Benjamin Bloom (1985) observed that the journey to developing a talent could be divided into three phases in which the student's activity in the talent area transitioned from fun and exciting playtime, to intense and strenuous skill refinement, to developing their individual style and pushing the boundaries of the field.

> *"Bloom's (1985) research identified three phases of talent development of expert performers ... labeled the early years, the middle developmental years, and the final years of perfecting the skills.*
> *...*

> *[The early years began] when individuals were introduced to activities in their sport. ... The coach/teacher provided the performer with considerable amounts of positive feedback and approval and allowed the children to play and explore all aspects of the sport. Rewards were garnered for effort rather than for achievement, and rarely was the coach critical of the children.*
> *...*
> *In the second phase or middle years ... individuals became fully committed to their performance goals. For the tennis players, the sport became more than a "game," it became "real business." ... In the early years of development, the coaches had been good at getting the athletes interested in and excited about their sport. In the middle years, however, the athletes and their parents felt they needed someone to teach them precision and technique as well as strategy. They also needed to tailor their skills to emphasize their own personal strengths and to compensate for any weaknesses they might have. ... The cultivation of talent now became a top priority for the performer. Coaches demanded more hard work, commitment, and discipline from their athletes. The athletes' training regimens became more intense and advanced as coaches introduced them to more strenuous and strategic areas of their sport.*
> *...*
> *Later Years. Athletes who achieved high levels of success auditioned for the opportunity to work with ... an individual widely recognized as a master teacher or expert in the domain. ... [The athletes] were totally committed to their chosen activity and would do whatever was necessary to excel. ... [including making] a number of sacrifices ... such as greater expenses and often moving to a new city. ... The relationship between athlete and expert coach evolved into one of mutual respect and collegiality with both parties focusing less on instructional methods and more on tactical refinement and the development of the individual's style. ... These coaches challenged their proteges to excel beyond their perceived human capabilities. 'This was especially true of the Olympic swimmers, who were expected to exceed records beyond that ever previously accomplished by any human being. So, too, was it true of the mathematicians, who were expected to solve problems that had never been solved before' (Bloom, 1985, p. 525)."*

Math Academy carries students through the second stage of talent development, which centers around intense and strenuous skill development. In this stage, it is assumed that students are motivated, be it intrinsically or extrinsically, to engage in particularly effortful forms of practice that maximize their learning.

Note that Math Academy may not be appropriate for students who remain in Bloom's first stage and desire a form of educational "playtime," or students who have progressed to the third stage and are developing original research in mathematics.

| Bloom's Taxonomy is Often Misinterpreted

It's worth pointing out that while Bloom is widely known for Bloom's Taxonomy of Learning, this taxonomy is often misinterpreted in a way that is not aligned with Bloom's 3-Stage Talent Development Process discussed above.

The misinterpretation arises from assuming that the makeup of every year in a student's education should be balanced the same way across Bloom's Taxonomy – whereas Bloom's Talent Development Process suggests that the time allocation should change drastically as a student progresses through their education (i.e., heavily focused on the lower parts of the taxonomy in the middle years, and heavily focused on the higher parts of the taxonomy in the later years). Put simply, Bloom's Talent Development Process argues for front-loading foundational skill development and then shifting to creative production afterwards.

Why does order matter? Why not just split the time 50-50 between foundational skill development and creative production throughout the whole talent development process? The answer, to be elaborated further in chapter 8, is this:

- there's a mountain of empirical evidence that one can increase the number of examples & problem-solving experiences in a student's knowledge base,

- but a lack of evidence that one can increase the student's ability to generalize from those examples by engaging in other pedagogical techniques (that is, techniques other than equipping the student with progressively more advanced examples & problem-solving experiences).

In other words, research indicates that the best way to improve a student's problem-solving ability in any domain is simply by having them acquire more foundational skills in that domain. Below are some entry points into the literature:

> *Teaching General Problem-Solving Skills Is Not a Substitute for, or a Viable Addition to, Teaching Mathematics* (Sweller, Clark, & Kirschner, 2010)
>
> *Putting Students on the Path to Learning: The Case for Fully Guided Instruction* (Clark, Kirschner, & Sweller, 2012)
>
> *Should There Be a Three-Strikes Rule Against Pure Discovery Learning?* (Mayer, 2004)
>
> *Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching* (Mayer, 2004)
>
> *Radical Constructivism and Cognitive Psychology* (Anderson, Reder, & Simon, 1998)

Armed with this information, if your goal is to maximize your depth within a talent domain, then the optimal rational strategy is a greedy approach:

1. Augment your knowledge base with all the examples and problem-solving experiences in the intended direction, as quickly as possible.

2. Upon reaching the edge of human knowledge in that direction, and only then, switch over to creative production.

Creative production is a substantially less efficient means of acquiring skills within a talent domain, so you want to save it for the end when it's the only way to continue moving forward.

(To briefly fend off an expected critique: the act of perpetually avoiding the leap into creative production, opting instead to indefinitely "expand sideways," acquiring skills that are not foundational for the talent domain, does not constitute the above strategy.)

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6), 4-16.

Importance: *Comparing the effectiveness of one-on-one tutoring and traditional classroom teaching, the average tutored student performed better than 98% of the students in the traditional class, an effect size of two sigmas (standard deviations).*

- Bloom, B. S., & Sosniak, L. A. (1981). Talent development vs. schooling. *Educational Leadership, 39*(2), 86-94.

  Brandt, R. S. (1985). On Talent Development: A Conversation with Benjamin Bloom. *Educational Leadership, 43*(1), 33-35.

  Luo, L., & Kiewra, K. A. (2021). Parents' roles in talent development. *Gifted Education International, 37*(1), 30-40.

Importance: *There are striking commonalities in the upbringing of extremely successful individuals across a wide variety of fields, leading to a general characterization of the process of talent development. However, the differences between talent development and traditional schooling are so numerous, so striking, and so critical that traditional schooling typically cannot even be characterized as supporting talent development.*

One of the main differences between traditional schooling and talent development is that students are grouped primarily by age, rather than ability, and each group progresses through the curriculum in lockstep. Each member of the group engages in the same tasks, and it is expected that different students will learn skills to different levels.

In talent development, however, instruction is completely individualized. Learning tasks are chosen based on the specific needs of individual students, and each student must learn each skill to a sufficient level of mastery before moving on to more advanced skills. Students progress through skills at different rates, but learn skills to the same threshold of performance. Their progress is measured not by their level of learning in courses that they have taken, but rather by how advanced the skills are that they can execute to a sufficient threshold of *performance.*

- Bloom, B. S., ed. (1985). *Developing Talent in Young People.* New York: Ballantine Books.

  Bloom, G. (2002). Role of the elite coach in the development of talent. *Psychological foundations of sport*, 466-483.

Importance: *The journey to developing a talent can be divided into three phases in which the student's activity in the talent area transitioned from fun and exciting playtime, to intense and strenuous skill refinement, to developing their individual style and pushing the boundaries of the field.*

Chapter 2. The Science of Learning

Summary: *Math Academy leverages evidence-based cognitive learning strategies including active learning, deliberate practice, mastery learning, minimizing cognitive load, developing automaticity, layering, non-interference, spaced repetition (distributed practice), interleaving, the testing effect (retrieval practice), and gamification. These methods are backed by decades of research, but they clash with traditional educational practices, which are held in place by convenient misconceptions about learning. By systematically applying these strategies, Math Academy accelerates student learning by 4x, meaning that serious students learn 4x the amount of material in the same time as compared to traditional classrooms.*

Cognitive Learning Strategies

The science of learning has advanced significantly over the past century. Numerous effective cognitive learning strategies have been identified and researched extensively since the early to mid-1900s, with key findings being successfully reproduced over and over again.

At a glance, here are some of the highlights:

- **Active Learning** – students learn more when they are actively performing learning exercises as opposed to passively consuming educational content.

- **Deliberate Practice** – effective learning feels like a workout with a personal trainer and should center around individualized training activities that are chosen to improve specific aspects of one's performance through repetition and successive refinement.

- **Mastery Learning** – each individual student needs to demonstrate proficiency on prerequisite topics before moving on to more advanced topics.

- **Minimizing Cognitive Load** – because our brains can only process small amounts of new information at once, it's critical to break down skills and concepts into tiny steps.

- **Developing Automaticity** – to free up mental processing power, it's also critical to practice low-level skills enough that they can be carried out without requiring conscious effort.

- **Layering** – learning is about making connections. The more connections there are to a piece of knowledge, the more ingrained, organized, and deeply understood it is, and the easier it is to recall. The most efficient way to increase the number of connections to existing knowledge is to continue layering on top of it – that is, continually acquiring new knowledge that exercises prerequisite or component knowledge.

- **Non-Interference** – conceptually related pieces of knowledge should be spaced out over time so that they are less likely to interfere with each other's recall. New concepts should be taught alongside dissimilar material.

- **Spaced Repetition** (**Distributed Practice**) – reviews should be spaced out or *distributed* over multiple sessions (as opposed to being crammed or *massed* into a single session) so that memory is not only restored, but also further consolidated into long-term storage, which slows its decay.

- **Interleaving** (**Mixed Practice**) – the effectiveness of practice is diminished when a single skill is practiced many times consecutively beyond a minimum effective dose. Review problems should be spread out or *interleaved* over multiple review assignments that each cover a broad mix of previously-learned topics. In addition to being more efficient, this also helps students match problems with the appropriate solution techniques.

- **The Testing Effect** (**Retrieval Practice**) – to maximize the amount by which your memory is extended when solving review problems, it's necessary to avoid looking back at reference material unless you are totally stuck and cannot remember how to proceed. For this reason, it's necessary to test frequently as a part of the learning process itself.

- **Gamification** – when game-like elements (such as points and leaderboards) are properly integrated into student learning environments, students typically not only learn more and engage more with the content, but also enjoy it more. However, these gamified elements must be aligned with the goals of the course, the motivations of the students, and the context of the educational setting. Further, they need to be resistant to "hacking" behaviors that attempt to bypass learning by exploiting loopholes in the rules of the game.

The Persistence of Tradition

One might expect to find these strategies being leveraged in today's classrooms to drastically improve the depth, pace, and overall success of student learning. However, the disappointing reality is that the practice of education has barely changed, and in many ways remains in *direct opposition* to the strategies outlined above.

- Classes still march through linear sequences of topics according to a predetermined schedule. Students are tethered to the pace of the class, which means that students who get lost are continually asked to learn new topics despite not having mastered the prerequisites, and students who learn quickly are prevented from learning more advanced concepts that come later in the class schedule or in a higher grade level (even if they have already mastered the prerequisites).

- Units of related material are taught in subsequent lessons, which promotes confusion, impedes recall, and places a severe bottleneck on how many topics can be successfully taught simultaneously, thereby creating lots of friction and massively slowing down the learning process.

- After learning a topic during class and practicing it on the homework, students forget about it until it's time to study for a test – and there are only a handful of tests given throughout the entire duration of a course. After the test, students are rarely required to practice the topic again, unless it just happens that some new topic requires them to remember the old one. The end result is that students end up forgetting most of what they learn.

- All students are given the same homework and assessments. This creates opportunities for coordinated cheating, a wide-open loophole in the grading system. Many students habitually exploit this loophole to bypass learning and obtain grades that do not reflect their (lack of) knowledge.

As lamented by Weinstein, Madan, & Sumeracki (2018):

> *"The science of learning has made a considerable contribution to our understanding of effective teaching and learning strategies. However, few instructors outside of the field are privy to this research.*
>
> *In particular, a review published 10 years ago identified a limited number of study techniques that have received solid evidence from multiple replications testing their effectiveness in and out of the classroom (Pashler et al., 2007).*

> *A recent textbook analysis (Pomerance, Greenberg, & Walsh, 2016) took the six key learning strategies from this report by Pashler and colleagues, and found that very few teacher-training textbooks cover any of these six principles – and none cover them all, suggesting that these strategies are not systematically making their way into the classroom.*
>
> *This is the case in spite of multiple recent academic (e.g., Dunlosky et al., 2013) and general audience (e.g., Dunlosky, 2013) publications about these strategies."*

Kirschner & Hendrick sum it up as follows (2024, pp.275):

> *"...[M]ost students, and also many or even most teachers, don't have an accurate picture of the effectiveness of their study approach.*
>
> *After more than a hundred years of research into learning and memory, there are a few things that we know about good and less good approaches. Since the turn of this century, people have been trying to figure out how to remember as much as possible, how to ensure that we forget as little as possible, and how to do this in as little time as possible.*
>
> *The reason we have our doubts with respect to teachers is because the findings that have emerged from this research aren't yet included in textbooks for teachers (both in research in the US, as well as in the Netherlands and Flanders; Pomerance, Greenberg, & Walsh, 2016; Surma, Vanhoyweghen, Camp, & Kirschner, 2018)."*

As Halpern & Hakel (2003) emphasize more sharply:

> *"Those outside academia further assume that because we are college faculty, we actually have a reasonable understanding of how people learn and that we apply this knowledge in our teaching. ... It would be reasonable for anyone reading these fine words to assume that the faculty who prepare students to meet these lofty goals must have had considerable academic preparation to equip them for this task. But this seemingly plausible assumption is, for the most part, just plain wrong.*
>
> *The preparation of virtually every college teacher consists of in-depth study in an academic discipline: chemistry professors study advanced chemistry, historians study historical methods and periods, and so on. Very little, if any, of our formal training addresses topics like adult learning, memory, or transfer of learning.*
>
> *And these observations are just as applicable to the cognitive, organizational, and educational psychologists who teach topics like principles of learning and performing, or evidence-based decision-making. We have found precious little evidence that content experts in the learning sciences actually apply the principles they teach in their own classrooms. Like virtually all college faculty, they teach the way they were taught.*
>
> *But, ironically (and embarrassingly), it would be difficult to design an educational model that is more at odds with the findings of current research about human cognition than the one being used today at most colleges and universities.*
> *...*
> *There is a large amount of well-intentioned, feel-good psychobabble about teaching out there that falls apart upon investigation of the validity of its supporting evidence."*

These sentiments are also echoed by Rohrer & Hartwig (2020):

> *"We fear, however, that continued advocacy might fall on deaf ears. ... [E]mpirical evidence is not highly valued by many of the educators who recommend learning methods and train teachers (e.g., Robinson, Levin, Thomas, Pituch, & Vaughn, 2007; Sylvester Dacy, Nihalani, Cestone, & Robinson, 2011). Against this backdrop, it might be difficult to inspire the kind of support for evidence-based interventions like those that sparked the dramatic improvements in Western medicine over the last century. Doing so, we believe, is the most pressing challenge facing learning scientists."*

A Common Theme Preventing Adoption

| Theme and Examples

So, what happened? Why have these cognitive learning strategies been rejected by the education system? The common theme throughout the literature is that *effective cognitive learning strategies often deviate from traditional conventions, which are held in place by convenient misconceptions about learning.*

The most obvious example of this theme is active learning.

- Traditionally, classes are taught using passive learning: the instructor lectures, and students listen, maybe answering a question here and there. Unsurprisingly, this is not nearly as effective as an active learning class where students spend most of their time actively performing learning exercises.

- However, it has been shown (Deslauriers et al., 2019) that even though students in active learning classes learn more, they *mistakenly perceive* that they learn less. Active learning produces more learning by increasing cognitive activation, but students often *mistakenly* interpret extra cognitive effort (such as productive struggle and occasional confusion) as an indication that they are not learning as well, when in fact the opposite is true.

- Of course, this misconception is a convenient belief for students who want to minimize the amount of effort that they expend during class while still "feeling" as though they are learning (even if it is not really happening). It is also a convenient belief for teachers who enjoy the spotlight and art of lecturing and the "feeling" that their students are learning, do not want to nag students to stay focused during class, and do not suffer repercussions for the reality that is their students' lack of learning.

Another example of this theme is interleaving (mixed practice).

- Traditionally, homework assignments focus on a single topic (or group of closely related topics) that are practiced many times consecutively beyond a minimum effective dose. This is not nearly as effective as spreading out or *interleaving* those problems over multiple review assignments that each cover a broad mix of previously-learned topics, which is more efficient and also helps students learn to match problems with the appropriate solution techniques.

- However, it has been shown (see Rohrer, 2009 for a review) that even though interleaving promotes vastly superior retention and generalization, students again *mistakenly believe* that they are learning less due to the increased cognitive effort. Teachers can be fooled, too, because although interleaving increases performance on cumulative tests, it actually lowers performance on homework (which is otherwise artificially high if students settle into a robotic rhythm of mindlessly applying one type of solution to one type of problem).

- Again, this misconception is a convenient belief for students who want to get through homework as quickly and effortlessly as possible while "feeling" as though they are mastering new skills (even if they are unable to consistently reproduce those skills in true assessment situations). It is also a convenient belief for teachers who want to assign good homework grades and "feel" as though these grades represent their students' learning, but don't want to spend extra effort organizing a properly spaced mixed review schedule and fielding a greater number and variety of homework questions from students.

A similar example can be constructed for every cognitive learning strategy that was mentioned earlier in this chapter. In some way or another, each strategy increases the intensity of effort required from students and/or instructors, and the extra effort is then converted into an outsized gain in learning. However, the extra effort also exposes the reality that students didn't actually learn as much as they (and their teachers) "felt" they did under less effortful conditions. This reality is inconvenient to students and teachers alike; therefore, it is common to simply believe the illusion of learning and avoid activities that might present evidence to the contrary.

More generally, while "innocent until proven guilty" is a good model for a legal system, "competent until proven incompetent" is a poor model for an educational system. If students are not made to demonstrate measurable learning at each step of the way, until they are able to consistently reproduce learned skills in true assessment situations, then the most likely outcome is that very little learning will happen. Whereas the casualties of the legal system are those who are jailed without just cause, the casualties of the education system are those students who are hopelessly pushed to learn advanced skills despite not having actually mastered the prerequisites. Empowering students requires ensuring their learning, and ensuring learning requires interrogating their knowledge.

| Desirable Difficulty vs Illusion of Comprehension

This theme is so well-documented in the literature that it even has a catchy name: a practice condition that makes the task harder, slowing down the learning process yet improving recall and transfer, is known as a **desirable difficulty**. As summarized by Rohrer (2009):

> *"A feature that decreases practice performance while increasing test performance has been described by Bjork and his colleagues as a **desirable difficulty**, and spacing and mixing are two of the most robust ones. As these researchers have noted, students and teachers sometimes avoid desirable difficulties such as spacing and mixing because they falsely believe that features yielding inferior practice performance must also yield inferior learning."*

Many types of cognitive learning strategies introduce desirable difficulties – for instance, Bjork & Bjork (2011) list a few more:

> *"Such desirable difficulties (Bjork, 1994; 2013) include varying the conditions of learning, rather than keeping them constant and predictable; interleaving instruction on separate topics, rather than grouping instruction by topic (called blocking); spacing, rather than massing, study sessions on a given topic; and using tests, rather than presentations, as study events."*

As Bjork & Bjork (2023, pp.21-22) elaborate, desirable difficulties make practice more representative of true assessment conditions. Consequently, it is easy for students (and their teachers) to vastly overestimate their knowledge if they do not leverage desirable difficulties during practice, a phenomenon known as the **illusion of comprehension**:

> *"A general characteristic of desirable difficulties (such as the spacing or interleaving of study or practice trials) is that they present challenges (i.e., difficulties) for the learner, and hence can even appear to be slowing the rate at which learning is occurring. In contrast, their opposites (such as massing or blocking of study or practice trials) often make performance improve rapidly and can appear to be enhancing learning.*
>
> *Thus, as either learners or teachers, we are vulnerable to being misled as to whether we or our students are actually learning effectively, and, indeed, we can easily be misled into thinking that these latter types of conditions, such as massing or blocking, are actually better for learning. Such dynamics probably play a major role in why students often report that their most preferred and frequently used types of study activity include activities such as rereading chapters (e.g., Bjork et*

> *al., 2013), typically right away after an initial reading. Such activities can provide a sense of familiarity or perceptual fluency that we can interpret as reflecting understanding or comprehension and, thus, produce in us what we have sometimes called an* ***'illusion of comprehension'*** *(Bjork, 1999; Jacoby et al., 1994).*
>
> *Similarly, when information comes readily to mind, which frequently is the case in blocked practice, or with no contextual variation in a repeated study or practice setting, we can be led to believe that such immediate access reflects real learning when, in fact, such access is likely to be the product of cues that continue to be present in the unchanging study situation, but that are unlikely to be present at a later time, such as on an exam. As both learners and teachers, we need to be suspicious of conditions of learning, such as massing and blocking, that frequently make performance improve rapidly, but then typically fail to support long-term retention and transfer. To the extent that we interpret current performance as a valid measure of learning, we become susceptible both to mis-judging whether learning has or has not occurred and to preferring poorer conditions of learning over better conditions of learning."*

| The Educational System Prefers Illusion

As Bjork (1994) explains, the typical teacher is incentivized to maximize the immediate performance and/or happiness of their students, which biases them against introducing desirable difficulties and incentivizes them to promote illusions of comprehension:

> *"Recent surveys of the relevant research literatures (see, e.g., Christina & Bjork, 1991; Farr, 1987; Reder & Klatzky, 1993; Schmidt & Bjork, 1992) leave no doubt that many of the most effective manipulations of training – in terms of post-training retention and transfer – share the property that they introduce difficulties for the learner.*
> *...*
> *If the research picture is so clear, why then are ... nonproductive manipulations such common features of real-world training programs? ... [T]he typical trainer is overexposed, so to speak, to the day-to-day performance and evaluative reactions of his or her trainees. A trainer, in effect, is vulnerable to a type of operant conditioning, where the reinforcing events are improvements in the [immediate] performance and/or happiness of trainees.*
>
> *Such a conditioning process, over time, can act to shift the trainer toward manipulations that increase the rate of correct responding – that make the trainee's life easier, so to speak. Doing that, of course, will move the trainer away from introducing the types of* ***desirable difficulties*** *summarized in the preceding section."*

What's more, most educational organizations operate in a way that exacerbates this issue:

> *"The tendency for instructors to be pushed toward training programs that maximize the performance or evaluative reaction of their trainees during is exacerbated by certain institutional characteristics that are common in real-world organizations.*
>
> *First, those responsible for training are often themselves evaluated in terms of the performance and satisfaction of their trainees during training, or at the end of training.*

> *Second, individuals with the day-to-day responsibility for training often do not get a chance to observe the post-training performance of the people they have trained; a trainee's later successes and failures tend to occur in settings that are far removed from the original training environment, and from the trainer himself or herself.*
>
> *It is also rarely the case that systematic measurements of post-training on-the-job performance are even collected, let alone provided to a trainer as a guide to what manipulations do and do not achieve the post-training goals of training.*
>
> *And, finally, where refresher or retraining programs exist, they are typically the concern of individuals other than those responsible for the original training."*

As a result, these cognitive learning strategies often ruffle the feathers of educational traditionalists, whose immediate response is to lash out against it. Take it directly from John Gilmour Sherman (1992), a professor who implemented evidence-based learning strategies in his own classroom, only to be shut down for no reason other than his superior's unsupported opinions about how learning works:

> *"Avoiding a frontal attack, the chairman of the Psychology Department at Georgetown declared by fiat that something on the order of 50% of class time must be devoted to lecturing. By reducing the possibility of self-pacing to zero, this effectively eliminated PSI [Personalized System of Instruction] courses.*
>
> *He issued this order on the grounds that in the context of lecturing 'it is the dash of intellects in the classroom that informs the student.' No data were presented on this point! The spectacle of purporting to defend scholarship while deciding the merits of instructional methods by assertion is silly.*
>
> *The troubling aspect of all these cases was that data played no part in the decisions. It is disturbing when one has to wonder whether research on the education process makes any difference."*

Ultimately, Sherman's experiences led him to conclude that

> *"...[T]he investment in keeping things as they are may be impossible to overcome. ... Improving instruction is the goal, but only in the context of not changing anything that is important to any vested interest. ... [When the role of the teacher] does not conform to what most people think of as teaching; this is a problem and an obstacle to implementation."*

This sentiment continues into recent years. As Bjork & Bjork (2023, pp.19) reminisce:

> *"Having been asked to convey in 'our own words' what we most want students and teachers to know regarding how to apply findings from the science of learning has led us to think back on our efforts to spread the desirable difficulties gospel, so to speak. It verges on laughable that we thought 25 years or so ago that we would simply tell people about certain key findings, and they would then immediately change how they managed their own learning."*

Or, as Rohrer & Hartwig (2020) put it bluntly:

> *"...[T]he success of an intervention depends partly on whether students and teachers are willing to use it. Too often, the classroom is where promising interventions go to die."*

Technology Changes Everything

| Revival via Technology

It is unfortunate that Sherman and countless other researchers, practitioners, and proponents of evidence-based education are no longer alive to see their life's work positively transform the practice of education – and especially so for those like Sherman (1992) who eventually despaired "whether research on the education process makes any difference."

However, some did maintain hope that one day their contributions might be revived in the future when computers advanced far enough to make individualized digital learning environments technologically possible and commercially viable.

Indeed, these cognitive learning strategies are now some of the main guiding principles behind Math Academy. By leveraging these strategies to their fullest effect and capitalizing on their compounding nature, Math Academy is proud to offer a learning environment where students can learn many times more than they would otherwise in a traditional classroom.

| Necessity of Technology

In building this environment, we discovered something interesting: technology not only lets us circumvent the opposing inertia in the education system, but also helps us leverage these cognitive learning strategies to a degree that would not be feasible for even the most agreeable and hard-working human teacher. While it's true that a human teacher can reap some benefits of these strategies while maintaining a reasonable workload (and there really is no good excuse for not doing so), technology enables us to leverage these strategies to their full extent and produce even better learning outcomes than a human teacher who uses loose approximations of these strategies as much as humanly possible.

For instance, consider spaced repetition. While some curricula now adopt a spiral approach where material is naturally revisited and further built upon in later textbook chapters and/or grades, this is nowhere near the level of granularity, precision, and individualization that is required to capture the maximum benefit of true spaced repetition. Taken to its fullest extent, spaced repetition requires the instructor to keep track of a repetition schedule for every student for every topic and continually update that schedule based on the student's performance – and each time a student learns (or reviews) an advanced topic, they're implicitly reviewing many simpler topics, all of whose repetition schedules need to be adjusted as a result.

Of course, this is an inhuman amount of work. In fact, before building our online system, we actually tried performing a loose approximation of spaced repetition manually while teaching in a human-to-human classroom. It turned out that, teaching just two classes with only a handful of students in each class, it took more time and effort than a full-time job to implement a very loose approximation of spaced repetition *for the class as a whole* – not even personalized to individual students. And that's just one of many strategies that are necessary for effective teaching!

But just because fully leveraging these cognitive learning strategies requires an inhuman amount of work, doesn't mean that there's little to gain from it (especially when a century of research has shown that these strategies lead to immense improvements in learning). All it means is that the human teacher is a bottleneck to effective teaching. And what's always the solution when manual human effort is a bottleneck? Technology.

| Accelerating Student Learning by 4x

By building a system that fully leverages these cognitive strategies, we have accelerated student learning by 4x: on Math Academy, serious students learn 4x the amount of material in the same time (or the same amount of material in a quarter of the time) as compared to traditional classrooms. And that's being conservative, since our courses tend to be even more comprehensive than what you'd find in a traditional classroom. (Our courses aim to cover the superset of all content that one could reasonably expect to find in any major textbook or standard class syllabus.)

The 4x factor is a hard measurement, backed by concrete numbers:

- We measure our course length in terms of XP. One XP is approximately one minute of focused effort, give or take, depending on the individual student. We model our average student on a serious (but imperfect) student who works an average of 40 XP per weekday.

- Using the AP Calculus BC Course as a comparison, a typical student in a school will have a 50-minute class five days per week plus about an hour of homework per night. Throughout a typical 32-week school year, that's a total of (50 minutes class + 60 minutes homework average per day) (five days) (32 weeks) = 17,600 minutes. Add in a couple extra hours for each test and quiz throughout each semester and then at least 30-40 hours for practice exams and studying for the AP exam, if you want to get a 5. That will put you in the ballpark of our calculation of 24,000 minutes.

- The Math Academy AP Calculus BC Course is approximately 6,000 XP (equivalent to about 6,000 minutes) and already includes quizzes, reviews, and highly specific test prep.

The rest of the book describes how our technology accomplishes this.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Weinstein, Y., Madan, C. R., & Sumeracki, M. A. (2018). Teaching the science of learning. *Cognitive research: principles and implications, 3*(1), 1-17.

  Halpern, D. F., & Hakel, M. D. (2003). Applying the Science of Learning. *Change*, 37.

  Importance: *The science of learning has made a considerable contribution to our understanding of effective teaching and learning strategies. However, few instructors outside of the field are privy to this research – and even those who are, often do not apply it in their classrooms.*

- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society, 2*(59-68).

  Bjork, E. L., & Bjork, R. A. (2023). Introducing Desirable Difficulties Into Practice and Instruction: Obstacles and Opportunities. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 111-21). Society for the Teaching of Psychology.

  Importance: *A practice condition that makes the task harder, slowing down the learning process yet improving recall and transfer, is known as a desirable difficulty. Desirable difficulties include varying the conditions of learning, rather than keeping them constant and predictable; interleaving instruction on separate topics, rather than grouping instruction by topic (called blocking); spacing, rather than massing, study sessions on a given topic; and using tests, rather than presentations, as study events.*

- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205).

Importance: *Many of the most effective manipulations of training – in terms of post-training retention and transfer – share the property that they introduce difficulties for the learner. The typical trainer is incentivized to maximize the immediate performance and/or happiness of trainees, which biases them against introducing these types of desirable difficulties. What's more, most training organizations are set up to exacerbate this issue.*

Chapter 3. Core Science: How the Brain Works

Summary: *Cognition involves the flow of information through sensory, working, and long-term memory banks in the brain. Sensory memory temporarily holds raw data, working memory manipulates and organizes information, and long-term memory stores it indefinitely by creating strategic electrical wiring between neurons. Learning amounts to increasing the quantity, depth, retrievability, and generalizability of concepts and skills in a student's long-term memory. Limited working memory capacity creates a bottleneck in the transfer of information into long-term memory, but cognitive learning strategies can be used to mitigate the effects of this bottleneck.*

Sensory, Working, and Long-Term Memory

In order to develop a good intuitive sense of how learning can be optimized, it's crucial to understand – at a concrete, physical level in the brain – what learning actually *is*. At the most fundamental level, **learning** is the creation of strategic electrical wiring between **neurons** ("brain cells") that improves the brain's ability to perform a task.

When the brain thinks about objects, concepts, associations, etc, it represents these things by activating different patterns of neurons with electrical impulses. Whenever a neuron is activated with electrical impulses, the impulses naturally travel through its outward connections to reach other neurons, potentially causing those other neurons to activate as well. By creating strategic connections between neurons, the brain can more easily, quickly, accurately, and reliably activate more intricate patterns of neurons.

As one might expect, it is extraordinarily complicated to understand what these specific brain patterns are, how they interact, and how the brain identifies strategic ways to improve its connectivity. However, to some extent, these are just nature's way of implementing cognition – and the overarching cognitive processes of the brain are much better understood.

At a high level, human cognition is characterized by the flow of information across three memory banks:

1. **Sensory memory** temporarily holds a large amount of raw data observed through the senses (sight, hearing, taste, smell, and touch), only for several seconds at most, while

relevant data is transferred to short-term memory for more sophisticated processing.

2. **Short-term memory**, and more generally, **working memory**, has a much lower capacity than sensory memory, but it can store the information about ten times longer. Working memory consists of short-term memory along with capabilities for organizing, manipulating, and generally "working" with the information stored in short-term memory. The brain's working memory capacity represents the amount of effort that it can devote to activating neural patterns and persistently maintaining their simultaneous activation, a process known as rehearsal.

3. **Long-term memory** effortlessly holds indefinitely many facts, experiences, concepts, and procedures, for indefinitely long, in the form of strategic electrical wiring between neurons. Wiring induces a "domino effect" by which entire patterns of neurons are automatically activated as a result of initially activating a much smaller number of neurons in the pattern. The process of storing new information in long-term memory is known as consolidation. At a cognitive level, **learning** can be described as a positive change in long-term memory.

These memory banks work together to form the following pipeline for processing information:

1. **Sensory memory** receives a stimulus from the environment and passes on important details to working memory.

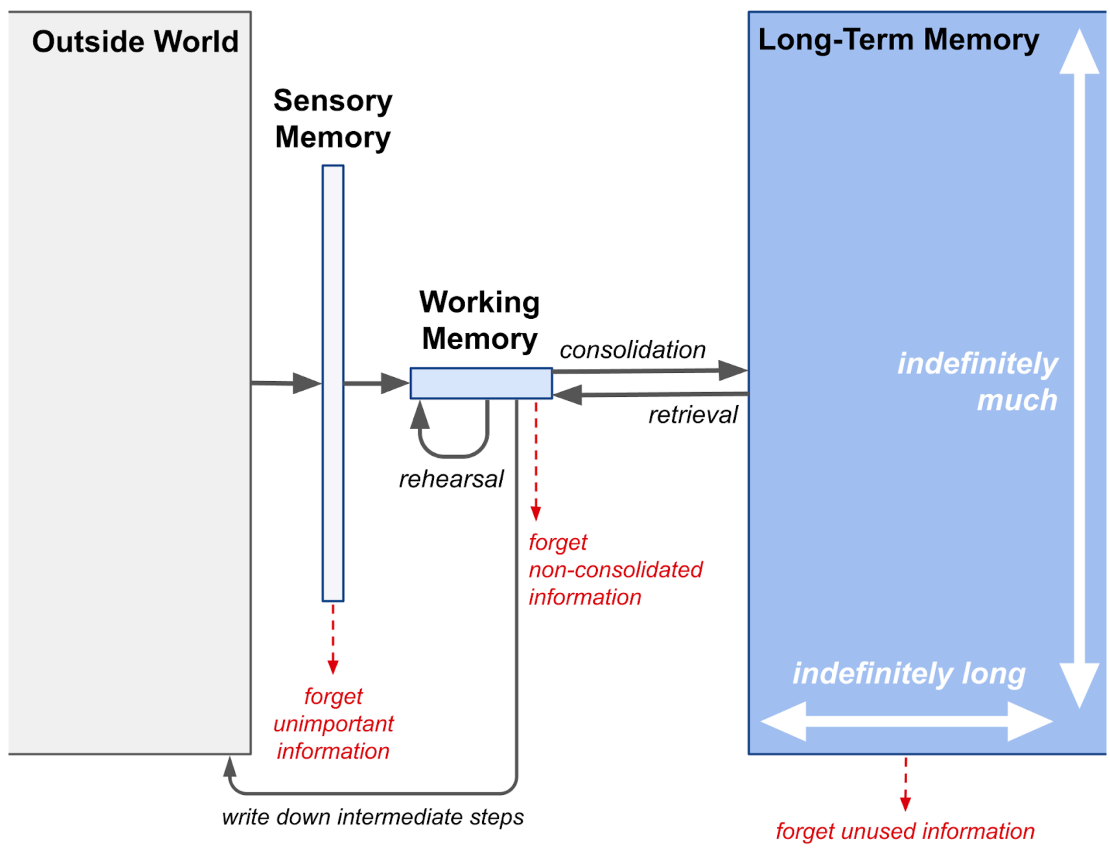
2. **Working memory** holds and manipulates those details, often augmenting or substituting them with related information that was previously stored in long-term memory.

3. **Long-term memory** curates important information as though it were writing a "reference book" for the working memory.

Note, however, that there is a crucial conceptual difference between long-term memory and a reference textbook: long-term memory can be forgotten. The text in a reference book remains there forever, accessible as always, regardless of whether you read it – but the representations in long-term memory gradually, over time, become harder to retrieve if they are not used, resulting in forgetting. The phenomenon of forgetting in long-term memory has been widely researched and can be characterized as follows (Hardt, Nader, & Nadel, 2013):

> *"...[F]orgetting refers to the absence of expression of previously properly acquired memory in a situation that normally would cause such expression. This can reflect actual memory loss or a failure to retrieve existing memory."*

However, the lower-level mechanisms underlying forgetting in long-term memory are not yet well understood.



There are two complementary perspectives by which we can think about this pipeline.

- **encoding** perspective – the pipeline converts or "encodes" information from the outside world into a representation that can be stored in long-term memory and later recalled.

- **executive function** (or **cognitive control**) perspective – the pipeline is centered around working memory, which pulls relevant information from sensory and long-term memory into an area where it can be combined, transformed, and used to guide behavior to achieve goals.

In the context of mathematical talent development, once a student is beyond the stage of learning how to read and count, we are less concerned with their sensory memory and more concerned with their long-term memory. The goal of instruction is to increase the quantity, depth, retrievability, and generalizability of mathematical concepts and skills in the student's long-term memory.

The student's working memory capacity is a bottleneck in the transfer of information into their long-term memory. However, by leveraging cognitive learning strategies and properly scaffolding and adapting instruction to the student's individual needs, we can minimize the degree to which their working memory capacity limits their learning, thereby maximizing the transfer of new information and the retention of existing information in long-term memory.

Design Constraints

The brain's information-processing pipeline is designed to be incredibly efficient. However, even the most efficient designs have limitations. Design is all about balancing trade-offs to achieve the best possible outcome in the face of constraints. To understand the constraints and the rationale behind a design, it can be helpful to attempt some naive critiques.

Critique: *Why is long-term memory needed? Why can't the brain just hold everything in working memory forever through rehearsal?*

Rationale: Rehearsal requires a lot of effort. It is very taxing on the brain. When the brain engages in rehearsal, it's like a muscle that is lifting a weight.

Just like a muscle has a limit to the amount of weight it can hold, the brain has a limit to the amount of new information it can hold in working memory via rehearsal. Most people can only hold about 7 digits (or more generally 4 chunks of coherently grouped items) simultaneously and only for about 20 seconds (Miller, 1956; Cowan, 2001; Brown, 1958). And that assumes they aren't needing to perform any mental manipulation of those items – if they do, then fewer items can be held due to competition for limited processing resources (Wright, 1981).

Long-term memory solves this problem by providing a place where the brain can store lots of information for a long time without requiring much effort.

Critique: *Why doesn't the brain just store everything it encounters in long-term memory? That way, it would never forget anything.*

Rationale: When it comes to information storage, more is not always better. In order for it to be worthwhile to store a piece of information, the benefit must offset the cost. Creating connections between neurons is costly in the sense that it requires biological resources – the connections are physical growths between cells, which means they have to be actively constructed and maintained by the body.

To illustrate with a concrete example, suppose that you want to buy a biography book that will help you understand somebody's background and their impact on society. One book contains 300 pages, costs $20, and covers formative experiences in their childhood, their career arc, and occasional anecdotes to illustrate key points and themes. Another book contains 10,000 pages, costs $1,000, covers all of the information in the first book, and also includes a description of every single meal the person ate throughout their life. Unless you have a specific, intense interest in this person's dietary habits (which you probably don't), it's easy to see that the first option is superior.

Case Study: Information Flow During a Computation

To illustrate how information flows through these memory banks when solving a math problem, let's analyze what happens as we compute 4^3 using typical arithmetic strategies while writing down some intermediate steps. (Remember that exponentiation is just repeated multiplication: 4^3 means to take three 4's and multiply them together, that is, $4^3 = 4 \times 4 \times 4 = 64$.)

First, let's get a sense of how each memory bank will help us solve the problem:

1. **Sensory memory** will capture visual data that lets us read the problem or any intermediate work that we've written down, thereby allowing the written information to be loaded into working memory. It will also filter out any distractions (e.g. background noise) as we solve the problem.

2. **Working memory** will hold the relevant pieces of the problem, request additional information from long-term memory, and apply that information to incrementally transform the pieces of the problem into the solution. Our problem-solving narrative will take place within the working memory.

3. **Long-term memory** will, upon request from working memory, produce definitions, facts, and procedures that we learned previously. It is like an internal "reference book" that we can use to look up additional information that would be helpful while solving the current problem.

It's worth re-emphasizing that the problem-solving narrative will take place within the working memory. Sensory and long-term memory will supply working memory with information, which working memory will combine, transform, and use to guide our behavior to solve the problem. As researchers elaborate (Roth & Courtney, 2007):

> *"Working memory (WM) is the active maintenance of currently relevant information so that it is available for use. A crucial component of WM is the ability to update the contents when new information becomes more relevant than previously maintained information. New information can come from different sources, including from sensory stimuli (SS) or from long-term memory (LTM).*
> *...*
> *In order for information in working memory to guide behavior optimally ... it must reflect the most relevant information according to the current context and goals. Since the context and the goals change frequently it is necessary to update the contents of WM selectively with the most relevant information while protecting the current contents of WM from interference by irrelevant information.*
> *...*
> *There are ... many ways in which WM can be changed, including through the manipulation of information being maintained (Cohen et al., 1997; D'Esposito, Postle, Ballard and Lease, 1999), the addition or removal of items being maintained (Andres, Van der Linden and Parmentier, 2004), or the replacement of one item with another (Roth, Serences, and Courtney, 2006)."*

Now, let's walk through the specific steps needed to solve the problem while observing what happens in each memory bank.

Sensory Memory (SM)	**Working Memory (WM)**	**Long-Term Memory (LTM)**
View problem: 4^3 Send relevant info to WM: 4 exponent 3		
	Rehearsing: 4 exponent 3 Request definition of exponent from LTM.	
		Retrieve definition: "A exponent B means A multiplied by itself B
	Rehearsing: 4 exponent 3, "A exponent B means A multiplied	

	by itself B times” Apply “A exponent B means A multiplied by itself B times” to 4 exponent 3 to get 4 × 4 × 4. Clear out all other WM. Rehearsing: 4 × 4 × 4 Request procedure of repeated multiplication from LTM.	times”
	Rehearsing: 4 × 4 × 4, “multiply in any order, but left-to-right by default” Apply “multiply in any order, but left-to-right by default” to 4 × 4 × 4 to get (4 × 4) × 4. Clear out all other WM. Rehearsing: (4 × 4) × 4 Request 4 × 4 from LTM.	Retrieve procedure: “multiply in any order, but left-to-right by default”
	Rehearsing: (4 × 4) × 4, 4 × 4 = 16 Apply 4 × 4 = 16 to (4 × 4) × 4, resulting in 16 × 4. Clear out all other WM. Rehearsing: 16 × 4 Request 16 × 4 from LTM.	Retrieve fact: 4 × 4 = 16
	Rehearsing: 16 × 4, “multiply place values separately and add results” (Write 16 × 4 on paper for later reloading.) Apply “multiply place values separately and add results” to 16 × 4. Tens place value is 10 so multiply 10 × 4. Clear out 16 × 4 from WM. Rehearsing: 10 × 4, “multiply place values separately and add results”	Unable to retrieve fact 16 × 4. Automatic redirect to retrieve procedure: “multiply place values separately and add results”

	Request 10×4 from LTM.	
		Retrieve fact: $10 \times 4 = 40$
	Rehearsing: $10 \times 4 = 40$, “multiply place values separately and add results” (Write 40 on paper for later reloading) Reload WM.	
View written work: 16×4 40		
Send relevant info to WM: 16×4, “one number written down”		
	Rehearsing: 16×4, “multiply place values separately and add results”, “one number written down” Apply “multiply place values separately and add results” and “one number written down” to 16×4. Next is ones place; ones place value is 6 so multiply 6×4. Clear out all other WM. Rehearsing: 6×4 Request 6×4 from LTM.	
		Retrieve fact: $6 \times 4 = 24$
	Rehearsing: $6 \times 4 = 24$ (Write 24 on paper for later reloading) Reload WM.	
View written work: 16×4 $40 + 24$		
Send relevant info to WM: $40 + 24$		
	Rehearsing: $40 + 24$ Request $40 + 24$ from LTM.	

		Unable to retrieve 40 + 24. Automatic redirect to procedure: "add digits"
	Rehearsing: 40 + 24, "add digits" Apply "add digits" to 40 + 24. First digit will be 4 + 2. Clear all other WM. Rehearsing: 4 + 2, "add digits" Request 4 + 2 from LTM.	
		Retrieve fact: 4 + 2 = 6
	Rehearsing: 4 + 2 = 6, "add digits" (Write 6 on paper for later reloading) Reload WM.	
View written work: 16 × 4 40 + 24 6 Send relevant info to WM: 40 + 24, "one number written down"		
	Rehearsing: 40 + 24, "add digits", "one number written down" Apply "add digits" and "one number written down" to 40 + 24. Second digit will be 0 + 4. Clear out all other WM. Rehearsing: 0 + 4 Request 0 + 4 from LTM.	
		Retrieve fact: 0 + 4 = 4
	Rehearsing: 0 + 4 = 4 (Write 4 on paper for later reloading) Reload WM.	
View written work: 16 × 4 40 + 24 64		

Send relevant info to WM: 64	Rehearsing: 64 Answer is 64.	

What learning resulted from this computation? Remember that learning occurs when the wiring of long-term memory is changed in a positive way that increases a student's ability to perform a task. This can involve any combination of wiring up new information, wiring up connections between existing pieces of information, reorganizing existing wiring so that the information can be retrieved more efficiently, etc.

With this in mind, let's take inventory of the processes that occurred within long-term memory in the example above:

- *Retrieval of definitions:*
  "A exponent B means A multiplied by itself B times"

- *Retrieval of facts:*
  $4 \times 4 = 16$
  $10 \times 4 = 40$
  $6 \times 4 = 24$
  $4 + 2 = 6$
  $0 + 4 = 4$

- *Redirects to procedures (and retrieval/execution of those procedures):*
  $4 \times 4 \times 4$ → "multiply in any order, but left-to-right by default"
  16×4 → "multiply place values separately and add results"
  $40 + 24$ → "add digits"

All of these pieces of information will become further consolidated in long-term memory, and there will be additional wiring connecting these component skills as part of a larger procedure for computing exponents.

Additionally, the fact $4^3 = 64$ will also begin consolidating in long-term memory (though it will soon be forgotten unless it is repeatedly reviewed into the future). Indeed, many people who frequently perform mental math with exponents know their cubes from 1^3 to 6^3 by heart and can simply retrieve their values as opposed to computing them.

Neuroscience of Working Memory

Recall that when the brain thinks about objects, concepts, associations, etc, it represents these things by activating different patterns of neurons with electrical impulses. Loosely speaking, the brain's working memory capacity represents the amount of effort that it can devote to activating these neural patterns and persistently maintaining their simultaneous activation. The cognitive load of a task represents the amount of this effort that the brain would need to put forth in order to complete the task.

As summarized by D'Esposito (2007):

> *"...[T]he neuroscientific data presented in this paper are consistent with most or all neural populations being able to retain information that can be accessed and kept active over several seconds, via persistent neural activity in the service of goal-directed behaviour.*
>
> *The observed persistent neural activity during delay tasks may reflect active rehearsal mechanisms. Active rehearsal is hypothesized to consist of the repetitive selection of relevant representations or recurrent direction of attention to those items.*
> *...*
> *Research thus far suggests that working memory can be viewed as neither a unitary nor a dedicated system. A network of brain regions, including the PFC [prefrontal cortex], is critical for the active maintenance of internal representations that are necessary for goal-directed behaviour. Thus, working memory is not localized to a single brain region but probably is an emergent property of the functional interactions between the PFC and the rest of the brain."*

Long-term learning is represented by the creation of strategic electrical wiring between neurons. Whenever a neuron is activated with electrical impulses, the impulses naturally travel through its outward connections to reach other neurons, potentially causing those other neurons to activate as well. By creating strategic connections between neurons, the brain can more easily, quickly, accurately, and reliably activate more intricate patterns of neurons.

Talcott (2021) summarizes this process as follows:

> *"Individual neurons can be thought of as rather simple biological batteries, each maintaining a gradient of biochemical ions across its cell membrane, which results in a small, local electrical charge – or potential.*
> *...*
> *Incoming signals from neighbouring brain cells are communicated to the neuron's dendrites and act to continuously modify the magnitude of the neuron's electrical charge. When the sum of*

> *signals from other neurons drives the electrical gradient to and then past a critical voltage, an electrical signal – the action potential – is generated and propagated along the neuron's axon. This signal ultimately modulates the activity of other neurons to which it connects.*
>
> *This process of synaptic transmission comprises the release of neurotransmitter chemicals at the junction between two cells – the synapse. Neurotransmitters released in response to the action potential in the pre-synaptic cell bind to receptors on the dendrites of the post-synaptic cell. The effects of such neurotransmitter binding serve to modify the electrical potential in the cell, either exciting it toward generating an action potential or inhibiting it from doing so.*
>
> *...*
>
> *Cognition – our thinking, reasoning and learning processes – are derived from activity in neural networks within the brain ... Neurodevelopment is a lifelong process involving the modification of the structural and functional properties of the brain ... One of the most striking aspects of the post-natal neurodevelopmental period in early childhood is in this near continuous refinement of neural connectivity, including both the strengthening of productive synapses and elimination of those that are less robust or redundant...*
>
> *Structural and functional connectivity provides mechanisms for implementing adaptation of the brain in response to an individual's experience of the world. As children are born with nearly a full complement of brain cells, adaptation of responses to environmental change – the underlying basis of learning for any organism – is accomplished mainly through modifying neural connectivity. Connectivity increases in parallel with children's advancement of their cognitive capacities and learning achievement.*
>
> *Adapted from a theory first articulated by Donald Hebb in the 1940, one well-supported principle regarding the relationships between brain structure and function in the developing brain is 'what fires together, wires together' ... When reinforced through repetition (experience), this coupling increases the probability of their activity being coincident in the future. This feedback process also works in reverse, such that connections that are not actively reinforced can be eliminated through a competitive elimination process, which favours the survival of more functionally adaptive networks at the expense of less efficient or redundant competitor networks through development...*
>
> *These mechanisms of synaptic plasticity are widely considered to be a predominant way through which information is coded and retained in brain networks ... Learning and memory (a cognitive demonstration of learning through recall of material to which an individual has been exposed) are therefore both expressed in the brain and related at the neural level to modification of connectivity within neural networks in response to repeated patterns of environmental stimuli and their associations."*

Wiring induces a "domino effect" by which entire patterns of neurons are automatically activated as a result of initially activating a much smaller number of neurons in the pattern. However, when the brain is initially learning something, the corresponding neural pattern has not been "wired up" yet, which means that the brain has to devote effort to activating each neuron in the pattern. In other words, because the dominos have not been set up yet, each one has to be toppled in a separate stroke of effort. This imposes severe limitations on how much new information the brain can hold simultaneously in working memory.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research, 169*, 323-338.

  Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol., 59*, 193-224.

  Nairne, J. S., & Neath, I. (2012). Sensory and working memory. *Handbook of Psychology, Second Edition, 4.*

  Importance: *In-depth overviews of different memory types and their interactions, at both cognitive (mental) and physiological (neural) levels.*

- D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 761-772.

  Tolcott, J. B. The neurodevelopmental underpinnings of children's learning: Connectivity is key.

  Importance: *When the brain thinks about objects, concepts, associations, etc, it represents these things by activating different patterns of neurons with electrical impulses. Loosely speaking, the brain's working memory capacity represents the amount of effort that it can devote to activating these neural patterns and persistently maintaining their simultaneous activation, and the cognitive load of a task represents the amount of this effort that the brain would need to put forth in order to complete the task.*

  Long-term learning is represented by the creation of strategic electrical wiring between neurons. Wiring induces a "domino effect" by which entire patterns of neurons are automatically activated as a result of initially activating a much smaller number of neurons in the pattern. However, when the brain is initially learning something, the corresponding neural pattern has not been "wired up" yet, which means that the brain has to devote effort to activating each neuron in the pattern. This imposes severe limitations on the number of new pieces of information that the brain can hold simultaneously in working memory.

- Roth, J. K., & Courtney, S. M. (2007). Neural system for updating object working memory from different sources: sensory stimuli or long-term memory. *Neuroimage, 38*(3), 617-630.

  Importance: *The brain's information-processing pipeline is centered around working memory, which pulls relevant information from sensory and long-term memory into an area where it can be combined, transformed, and used to guide perceptions and behavior.*

- Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in cognitive sciences, 17*(3), 111-120.

  Importance: *Forgetting refers to the absence of expression of previously properly acquired memory in a situation that normally would cause such expression. This can reflect actual memory loss or a failure to retrieve existing memory. The phenomenon of forgetting in long-term memory has been widely researched, but the lower-level mechanisms underlying the phenomenon are not yet well understood.*

Chapter 4. Core Technology: the Knowledge Graph

Summary: *Math Academy utilizes a knowledge graph, an interconnected structure of thousands of topics from 4th grade through university-level mathematics, to organize its curriculum and facilitate algorithmic decision-making. The knowledge graph allows Math Academy to place each student at the edge of their individual "knowledge frontier," fill in any gaps in foundational knowledge, leverage mastery learning to efficiently extend student knowledge, provide spaced reviews and remedial reviews when necessary, and capitalize on "encompassing" relationships to achieve turbo-boosted learning speed.*

Understanding the Knowledge Graph

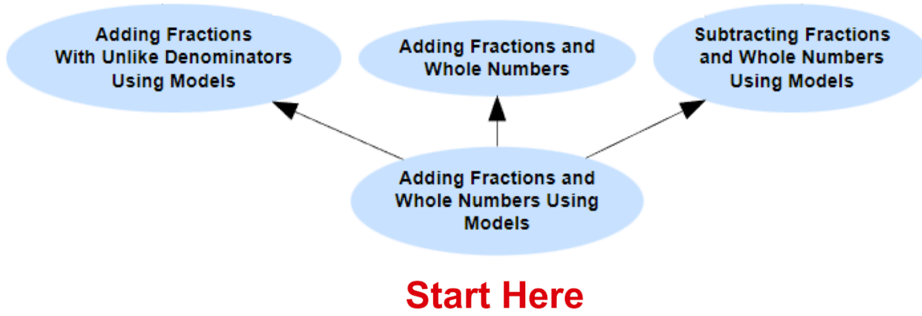
| Linking Topics and Prerequisites

To understand how Math Academy leverages specific cognitive learning strategies, it is helpful to have high-level understanding of our **knowledge graph**, which organizes our curriculum in a way that enables algorithmic decision-making.

Here, the word *"graph"* is a term that readers may be unfamiliar with. Usually, the word *"graph"* refers to a chart illustrating the relationship between two variables, such as a bar chart or a line chart. But in our context, the word *"graph"* refers to a diagram consisting of objects and arrows between them. This terminology is common in the mathematical field of graph theory.

Our knowledge graph contains multiple thousands of interlinked **topics**. Each linkage between topics indicates a relationship between them, such as one topic being a **prerequisite** for another topic. (There are lots of different kinds of relationships, but for now, we'll just focus on prerequisites.)

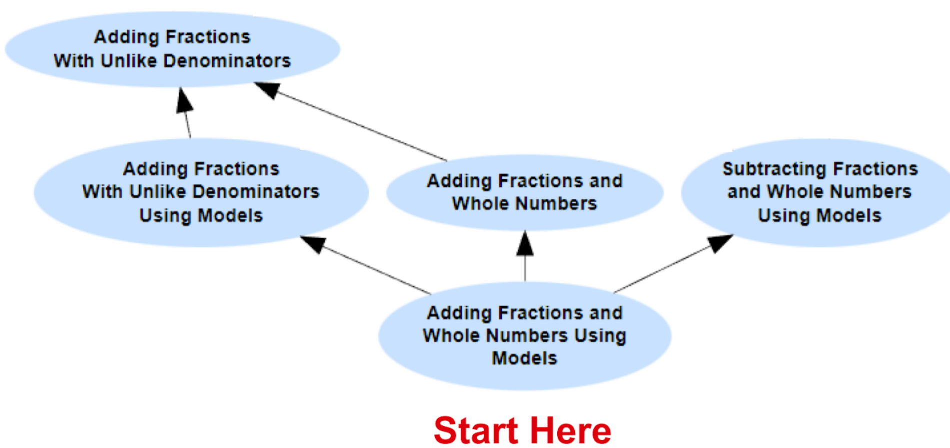
For instance, below is a simple example of a knowledge graph that shows a topic *Adding Fractions and Whole Numbers Using Models* (bottom) that is the prerequisite for three other topics (top). After a student learns the topic on the bottom, they will be ready to learn any topic that it points to. In other words, the arrows point along potential "learning paths" that the student can follow.



However, if multiple arrows point to a higher topic, then that means the higher topic has multiple prerequisites that the student needs to learn beforehand.

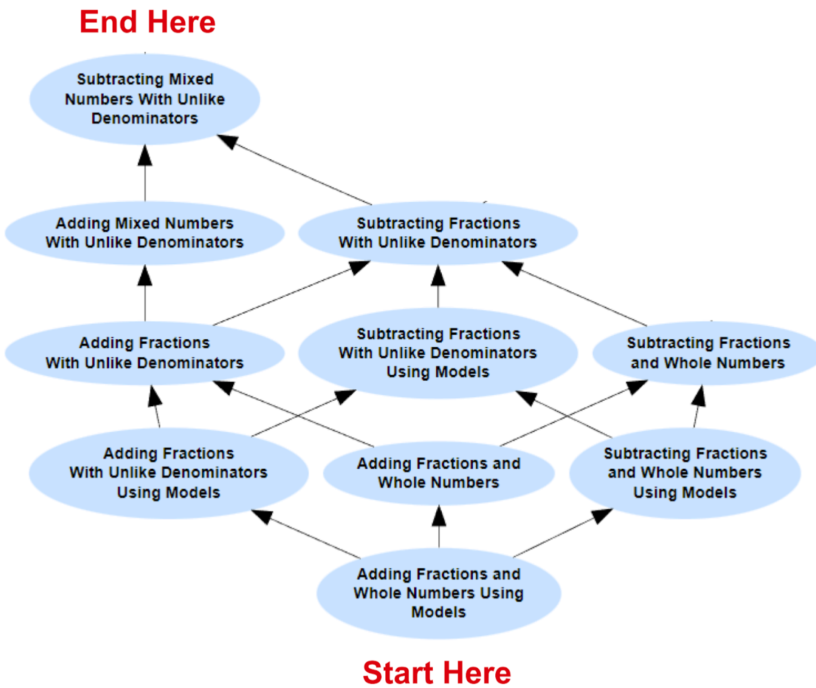
To illustrate, the topic *Adding Fractions With Unlike Denominators* has been added to the top of the knowledge graph. As indicated by the arrows pointing to it, it has two prerequisites that the student needs to learn beforehand:

1. *Adding Fractions With Unlike Denominators Using Models*
2. *Adding Fractions and Whole Numbers*

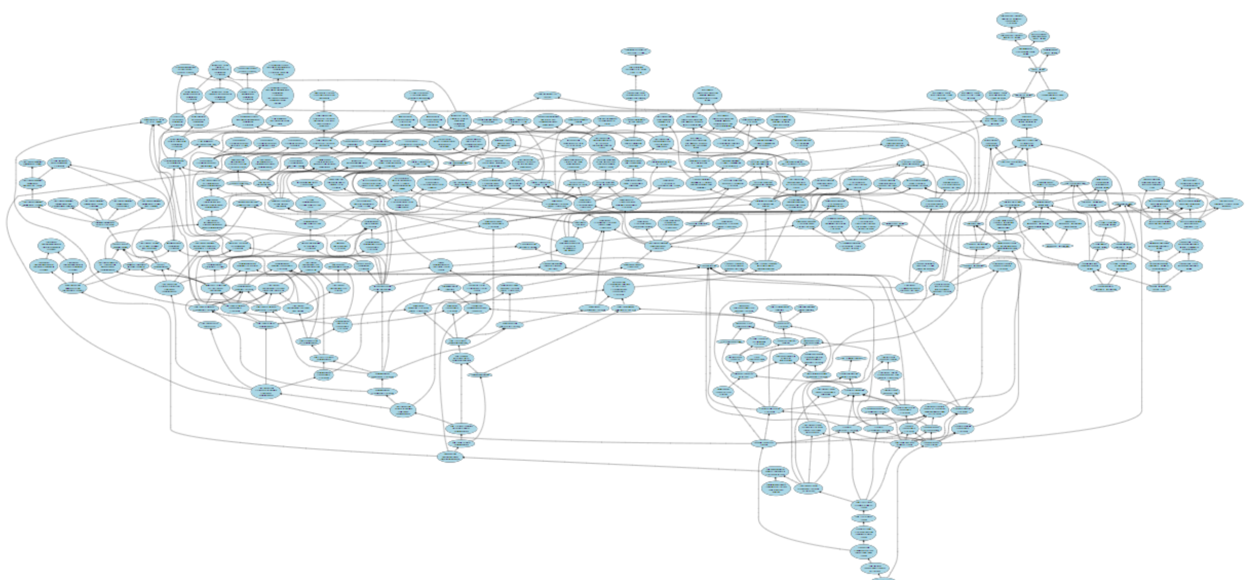


| Zooming Out

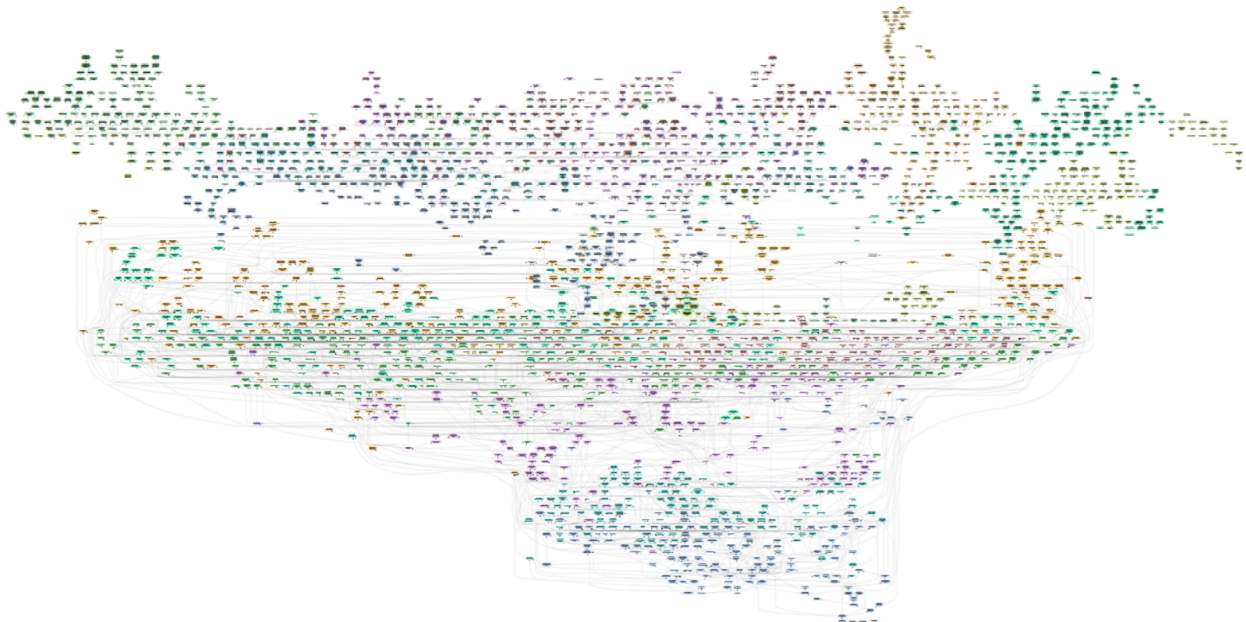
Zooming out more, we see that knowledge graphs can encode a lot of complicated information that would otherwise be hard to describe and reason about.



Zooming out *even more*, below is the knowledge graph for an entire course consisting of about 300 topics.



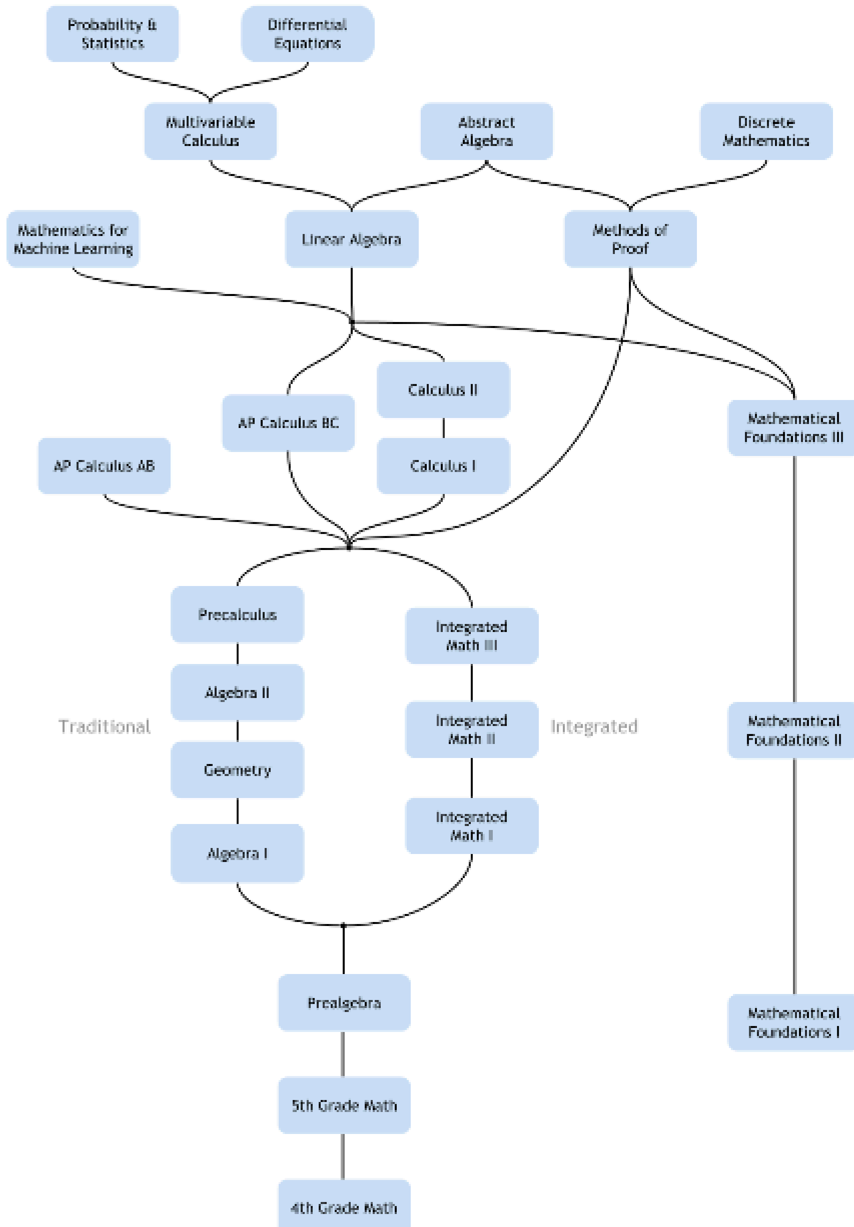
Fully zoomed out, Math Academy's entire curriculum consists of *multiple thousands* of topics spanning 4th Grade through university-level math. All these topics are connected up together in the knowledge graph. In this view, a course is simply a section of our knowledge graph. (In the visualization below, different colors represent different courses.)



The knowledge graph above contains the following courses: *4th Grade Math, 5th Grade Math, Pre-Algebra, Algebra I, Geometry, Algebra II, Pre-Calculus, Calculus I, Calculus II, Linear Algebra, Multivariable Calculus, *Differential Equations, *Probability & Statistics, *Discrete Mathematics, *Abstract Algebra.* (As of October 2023, this accounts for most, but not all, of the content in our system – courses not listed have large overlap with the preceding list. Asterisks indicate that a course is still under development.)

| Course Graph

On school and university websites, it is common to see courses arranged into a **course graph**, which can be interpreted as a highly-compressed version of a knowledge graph where a single entity represents hundreds of topics. Math Academy's course graph is shown below:

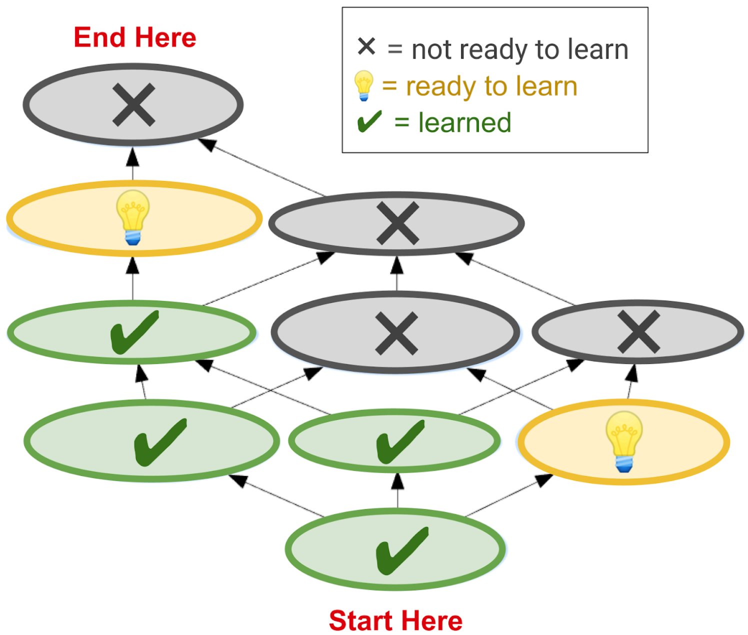


However, it is important to realize that each course is ultimately just a set of topics in the knowledge graph. The knowledge graph is the ultimate source of truth; a course graph simply summarizes and communicates information about the high-level structure of a knowledge graph so that humans can understand it.

Using the Knowledge Graph

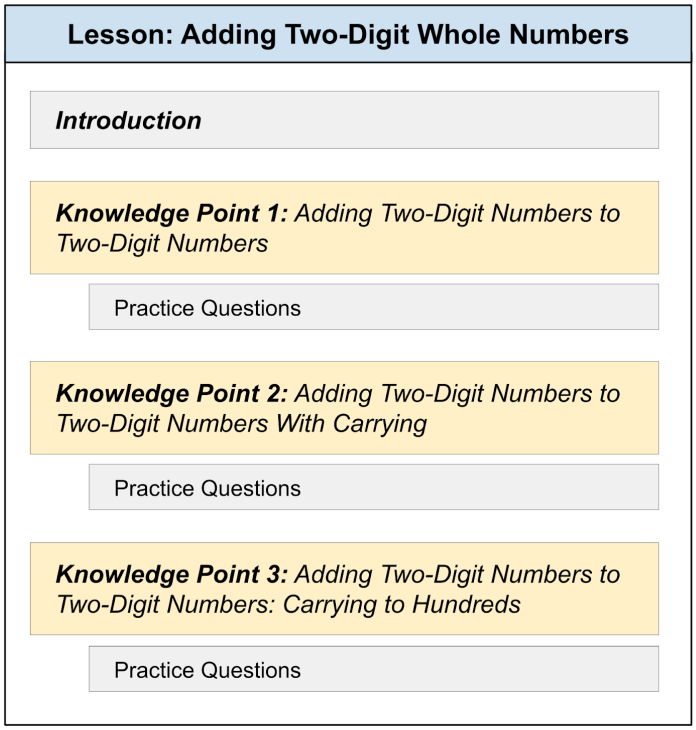
| Scaffolded Mastery Learning

Math Academy's knowledge graph enables us to implement **mastery learning**, in which students demonstrate proficiency on **prerequisites** before moving on to more advanced topics.

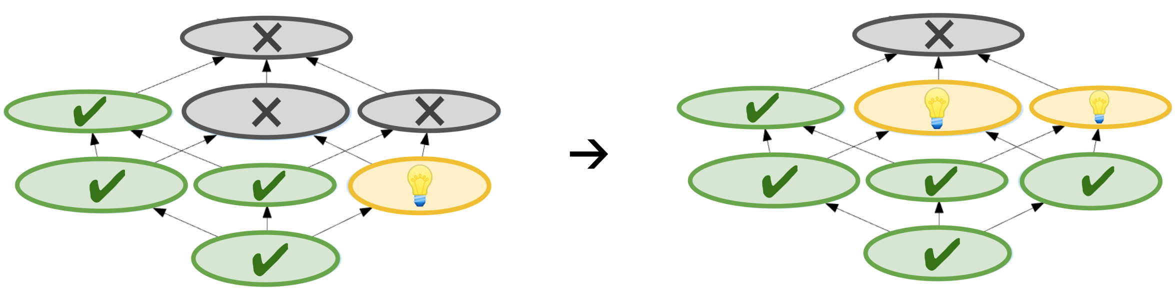


Each topic involves a **lesson** that is broken down into several key pieces of learning called **knowledge points**. Each knowledge point contains a **worked example** and asks **questions** similar to the worked example.

Knowledge points build on each other to help **scaffold** students through the lesson: the first knowledge point covers the most basic idea or skill of the lesson, and later knowledge points gently introduce more advanced cases.



To demonstrate mastery of a topic, a student must answer sufficiently many questions correctly in each successive knowledge point in the lesson. Once this is accomplished, more advanced topics become available for the student to work on.



| Additional Linkages

> Key Prerequisites Enable Targeted Remediation

Each knowledge point is linked to one or more **key prerequisite** topics that represent the prerequisite knowledge that is most directly being used in that knowledge point. If a student ever fails a lesson twice at the same knowledge point, we automatically provide remedial reviews on the key prerequisites. This helps the student strengthen their foundations in the areas where they are most in need of additional practice, so that they are better prepared to pass the lesson the next time around.

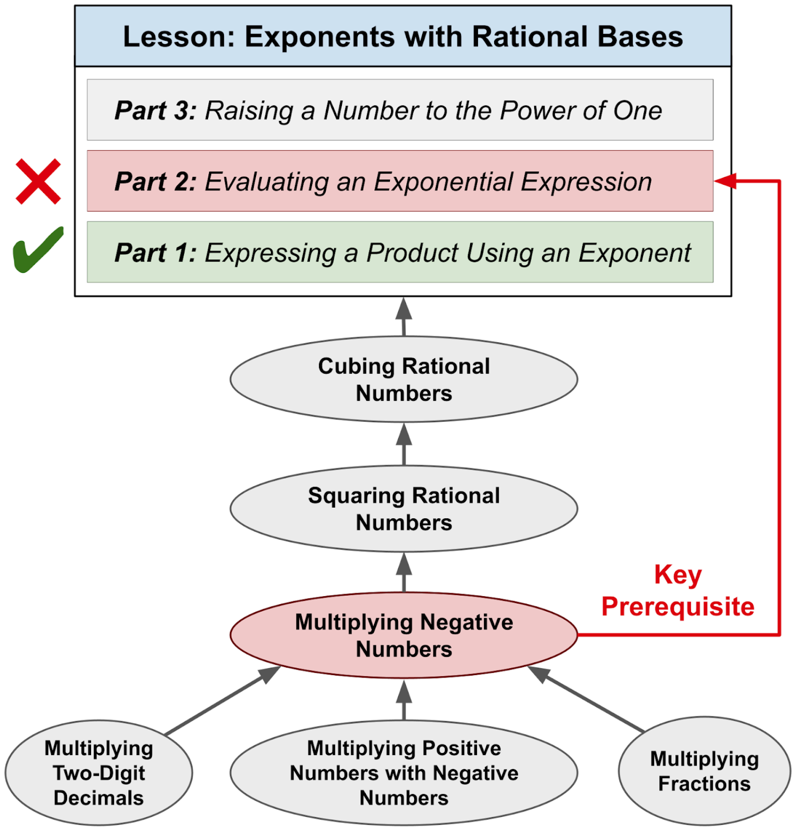
As a concrete example, suppose that while re-attempting the lesson *Exponents with Rational Bases*, a student

- manages to pass *Part 1: Expressing a Product Using an Exponent*, e.g. expressing $4 \times 4 \times 4$ as 4^3, but

- gets stuck again at *Part 2: Evaluating an Exponential Expression*, e.g. computing $(-4)^3 = (-4) \times (-4) \times (-4)$.

In this situation, the student has demonstrated that they understand the concept of an exponent, but they are struggling to use multiplication to compute the result.

Although multiplication occurs several steps back in the sequence of prerequisites, we have linked *Part 2: Evaluating an Exponential Expression* to the key prerequisite topic *Multiplying Negative Numbers*, which allows us to automatically trigger a targeted remedial review on *Multiplying Negative Numbers*.



> Encompassings Enable Turbo-Boosted Learning Speed

Our knowledge graph also stores **encompassing** relationships between topics. Advanced mathematical problems implicitly practice or "encompass" many simpler skills. Using sophisticated algorithms that capitalize on these encompassings, Math Academy enables students to spend most of their time learning new material while simultaneously making sure they keep getting practice on things they've previously learned. This results in turbo-boosted learning speed.

How does this work? The main idea is that whenever a student is due to review some previously-learned material, we serve them the smallest possible set of learning tasks that encompasses all the due review. The student receives all the review they need, in the most concentrated form possible.

To illustrate, consider the following multiplication problem, in which we multiply the two-digit number 39 by the one-digit number 6:

$$
\begin{array}{r}
{\scriptstyle 5} \\
3\,9 \\
\times \qquad 6 \\
\hline
2\,3\,4
\end{array}
$$

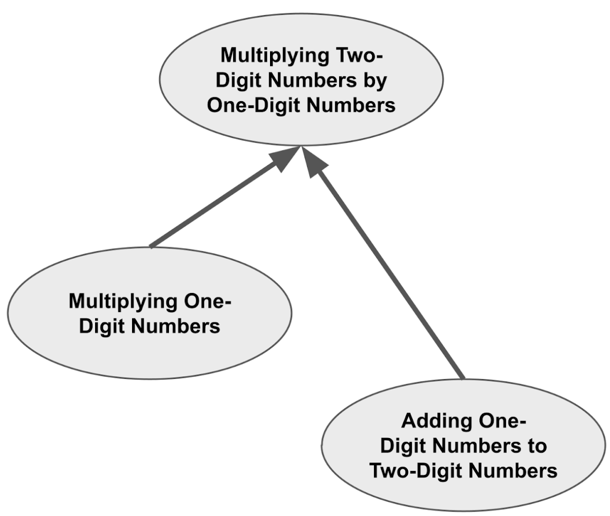
In order to perform the multiplication above, we have to multiply one-digit numbers and add a one-digit number to a two-digit number:

- First, we multiply 6 × 9 = 54. We carry the 5 and write the 4 at the bottom.

- Then, we multiply 6 × 3 = 18 and add 18 + 5 = 23. We write 23 at the bottom.

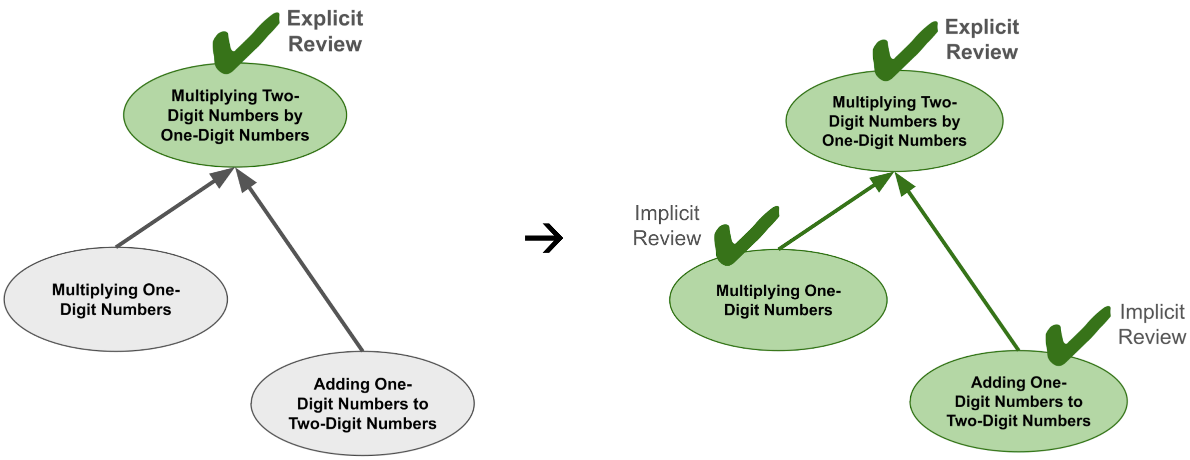
In other words, *Multiplying a Two-Digit Number by a One-Digit Number* **encompasses** *Multiplying One-Digit Numbers* and *Adding a One-Digit Number to a Two-Digit Number*.

We can visualize this using an **encompassing graph** as shown below. The encompassing graph is similar to a prerequisite graph, except the arrows indicate that a simpler topic is encompassed by a more advanced topic. (Encompassed topics are usually prerequisites, but prerequisites are often not fully encompassed.)



Now, suppose that a student is due for reviews on all three of these topics. Because of the encompassings, the only review that they will actually have to do is *Multiplying a Two-Digit*

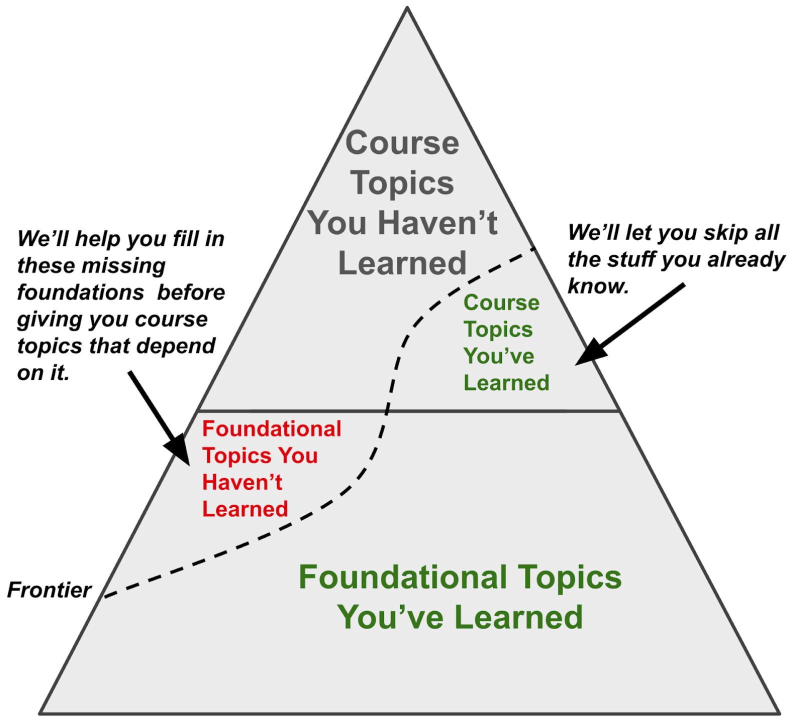
Number by a One-Digit Number. When they complete this review, it will implicitly provide repetitions on the topics that it encompasses because the student has effectively practiced those skills as well.



| Diagnostic Exams

When a student joins Math Academy, they take an adaptive diagnostic exam that leverages the knowledge graph to quickly identify their **knowledge frontier**. The knowledge frontier is the boundary between what they know and what they don't know, and it indicates what topics they are ready to learn. Following the diagnostic, whenever a student is served new lessons, those lessons always cover topics that are on the student's knowledge frontier.

In addition to assessing knowledge of course content, our diagnostic exams also assess knowledge of lower-grade **foundations** that students need to know in order to succeed in the course (i.e. they are prerequisites for the course). It is common for incoming students to be excited about a course but lack some foundational knowledge – and our knowledge graph enables us to identify and fill in any missing foundational knowledge while simultaneously allowing students to learn course topics that don't rely on that missing foundational knowledge.


Course Topics You Haven't Learned
We'll help you fill in these missing foundations before giving you course topics that depend on it.
We'll let you skip all the stuff you already know.
Course Topics You've Learned
Foundational Topics You Haven't Learned
Frontier
Foundational Topics You've Learned

Chapter 5. Accountability and Incentives

Summary: *Students and teachers are often not aligned with the goal of maximizing learning, which means that in the absence of accountability and incentives, classrooms are pulled towards a state of mediocrity. Accountability and incentives are typically absent in education, which leads to a "tragedy of the commons" situation where students pass courses (often with high grades) despite severely lacking knowledge of the content. However, Math Academy is properly held accountable and incentivized to maximize student learning.*

Accountability and Incentives are Necessary but Absent in Education

According to K. Anders Ericsson (1993, with Krampe & Tesch-Romer), one of the most influential researchers in the field of human expertise and performance:

> *"...[D]eliberate practice requires effort and is not inherently enjoyable. Individuals are motivated to practice because practice improves performance."*

In other words, maximal learning does not happen naturally as a result of maximizing other things like enjoyment, comfort, convenience, and ease of practice. In fact, maximal learning is at odds with some of these things. Sacrifices must be made.

At the risk of stating the obvious: if you want to maximize learning, then you should not make decisions on the basis of anything other than how those decisions affect measurable learning. However, what may not be so obvious is that students and teachers are often not aligned with the goal of maximizing learning.

Students often just want to get a good enough grade to avoid angering their parents, or to get into college (or get a scholarship to college) – and in college, they often just want to do well enough to get their degree and either get a job or be accepted to graduate school. From the perspective of such students, the goal is to earn grades that are good enough to keep moving along their desired career path, while minimizing the amount of extra effort. Earning sufficient grades with minimal effort is totally different from maximizing learning.

Likewise, while teachers generally want their students to learn, they also receive substantial pressure from parents and administrators to make the learning process feel comfortable and enjoyable, and check boxes on people's intuitions (however mistaken) about learning, while simultaneously ensuring that students don't fall behind on any standardized tests. A teacher's goal is often for their students to perform well enough not to raise eyebrows from parents and administrators, while minimizing the amount of griping from students (and parents) about how much effort is required.

These forces pull classrooms towards a state of mediocrity: students need to learn some baseline amount that is deemed "enough" for their grade level, but there is no need to learn more than that, even if it is possible (and extremely advantageous) to learn much more in the same amount of time.

The pull towards mediocrity is not unique to education. However, other industries do a better job of *counteracting* it by leveraging accountability mechanisms and incentives to motivate people to maximize performance. For instance, in professional athletics, coaches are held accountable for winning (their continued employment depends on it) and they are often incentivized with massive financial bonuses for achievements like qualifying for tournaments and winning championships. The same is true for players. Along the chain of command from team owners to coaches to players, there is also a chain of accountability and incentives.

While it's true that college rankings can be viewed as some kind of incentive structure, it's important to realize that learning is not the basis of such rankings. The rankings may incentivize other things, but not learning. As MIT researchers elaborate (Subirana, Bagiati, & Sarma, 2017):

> *"Taking a look at major University ranking methodologies one can easily observe they consistently lack any objective measure of what content knowledge and skills students retain from college education in the long term.*
> *...*
> *In general, college academics taught in the classroom don't seem to be recognized explicitly by public market indicators. As an example, MIT was ranked number one in the world by US News Report in the latest ranking available, however taking a closer look at the ranking methodology one can see it does not include any metric of what students retained from the classroom. In fact, all major market ranking methodologies consistently lack any objective measure of student college academic retention ([MIT office of the provost 2012])."*

Likewise, while it's true that teacher credentialing can be viewed as some sort of accountability mechanism, it's important to realize that accountability for *learning* in particular is lacking. As discussed in chapter 2, most teacher credentialing programs do not cover, much less assess, prospective teachers on their knowledge of the science of learning and ability to leverage effective practice techniques to maximize student learning.

It's also worth noting that university professors generally aren't required to earn teaching credentials, and they're not even incentivized to teach as their primary concern – they are primarily measured in terms of research output, not teaching. Yet, they are also given more autonomy in designing their courses, and as a result, college courses tend to be more instructor-centered than student-centered (as compared to K-12 courses). A typical university professor gives some lectures, assigns weekly problem sets, and then gives a mid-term and a final exam that are curved so that no matter how much learning did or did not occur, the result is always a normal distribution and a shrug.

We re-emphasize some quotes from chapter 2:

> *"A recent textbook analysis (Pomerance, Greenberg, & Walsh, 2016) took the six key learning strategies from this report by Pashler and colleagues, and found that very few teacher-training textbooks cover any of these six principles – and none cover them all, suggesting that these strategies are not systematically making their way into the classroom."* – Weinstein, Madan, & Sumeracki (2018)

> *"The preparation of virtually every college teacher consists of in-depth study in an academic discipline: chemistry professors study advanced chemistry, historians study historical methods and periods, and so on. Very little, if any, of our formal training addresses topics like adult learning, memory, or transfer of learning. … [I]ronically (and embarrassingly), it would be difficult to design an educational model that is more at odds with the findings of current research about human cognition than the one being used today at most colleges and universities."* – Halpern & Hakel (2003)

What Happens in the Absence of Accountability and Incentives

| Tragedy of the Commons

As discussed in chapter 1, Bloom & Sosniak (1981) noted that teachers typically focus on a "cross section" of many students covering a small subset of curriculum over a short period of time.

> *"Although the curriculum in a particular subject may extend over a period of ten or more years, each teacher has the child only for a term, year, or course. And the teacher is responsible only for what happens during that period of time."*

As a result, maintenance and improvement of students' mathematical knowledge is a responsibility shared by a group of many teachers.

However, it is widely known that in the absence of accountability and incentives that promote collective interests, people will focus on behaviors that benefit themselves as individuals, and pay less attention to how their actions affect the group as a whole. As a result, when a group is given responsibility for the maintenance and improvement of a shared resource, the resource will typically deteriorate. While some individuals may care for the resource properly, they are typically unable or unwilling to pick up the slack of those who do not. This kind of deterioration of a shared or "common" resource is known as the **tragedy of the commons**.

A concrete example of the tragedy of the commons is littering. In the absence of accountability and incentives, public spaces will become filled with trash. Even people who dispose of their trash properly will generally not be motivated to pick up the trash of others. To prevent a public space from becoming filled with trash, it is necessary to create accountability mechanisms, such as fines for littering, and incentives, such as paid jobs to incentivize some people to periodically clean the space. But if the accountability and incentives are not implemented properly (e.g. the fine for littering is too low or unenforceable, or the paid jobs do not hire enough people or do not hold them accountable for actually cleaning the entire space), then the space will still become filled with trash.

The tragedy of the commons takes place in education in a similar way. Instead of "littering," the tragic action is allowing students to pass courses despite severely lacking knowledge of the content. A teacher who "picks up other people's trash" is a teacher who holds students accountable for learning the material in their course, including any prerequisite material that they are missing.

When there is a lot of "trash," i.e. students are severely lacking prerequisite knowledge, a teacher who "picks up other people's trash" puts forth a ton of effort supporting students through remedial assignments/assessments and help sessions, while simultaneously holding the line on expectations and enduring griping from students who experience a rude awakening about how much extra work they have to put in to shore up their missing foundations. Few teachers do this, just as few people pick up other people's trash. Instead, when faced with a situation like this, the typical teacher will just run the class as usual, curve (or otherwise inflate) the grades, and leave the problem for the next year's teacher to deal with (or not deal with).

While littering fines and paid janitorial jobs often provide the necessary accountability and incentives to keep spaces clean, teachers typically do not face penalties for allowing students to pass courses despite severely lacking knowledge of the content, and teachers are given no financial incentive for working hard to remedy these kinds of problematic situations that are created by other teachers. As a result, it is common for students to pass courses despite severely lacking knowledge of the content.

| Grades Can't Be Trusted

> Evidence for Grade Inflation

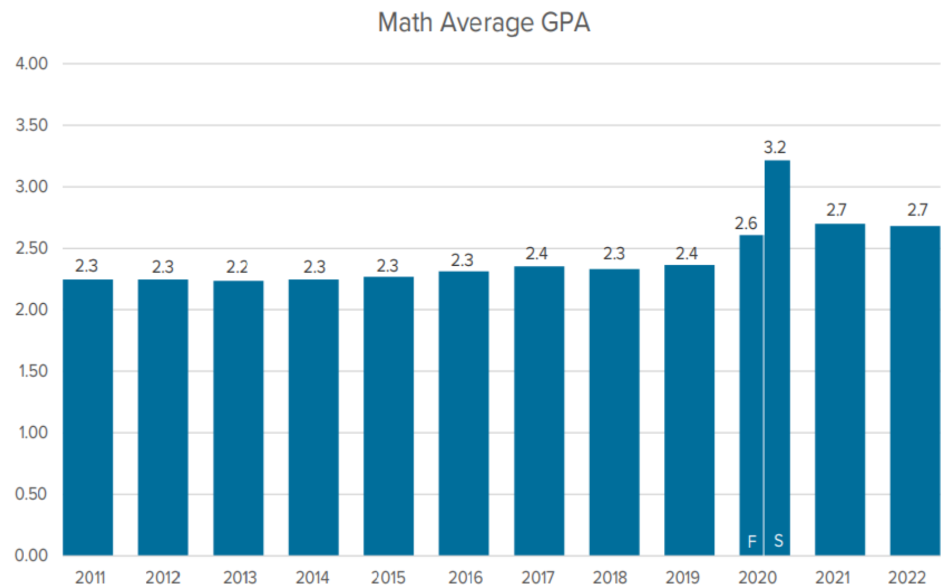
One of the most obvious examples of students passing courses (often with high grades) despite severely lacking knowledge of the content is the co-occurrence of extreme learning loss and extreme grade inflation during the COVID-19 pandemic.

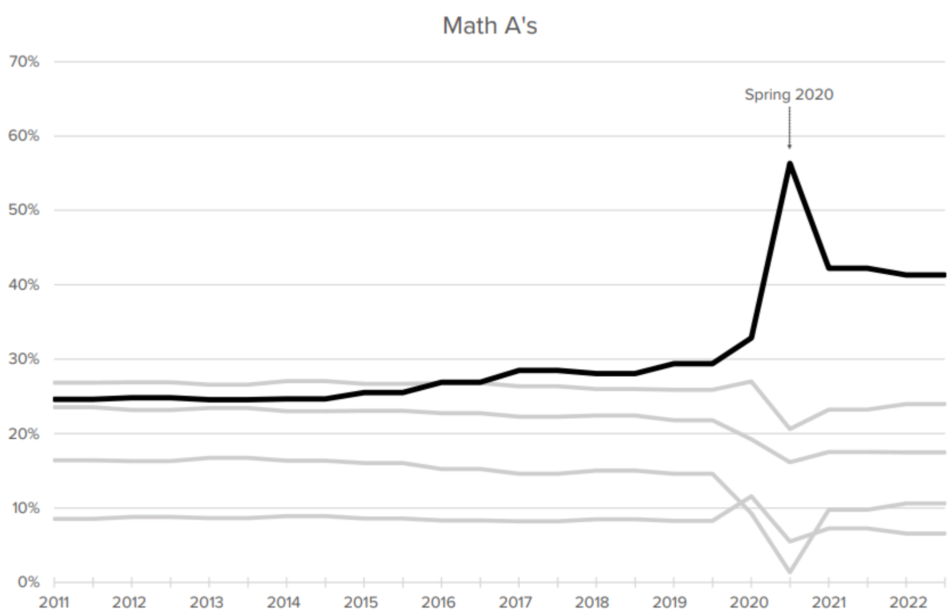
Researchers have found that the learning loss experienced by students during COVID-19 was even more extreme than that experienced by evacuees during Hurricane Katrina, one of the deadliest hurricanes to hit the United States (Kuhfeld, Soland, & Lewis, 2022):

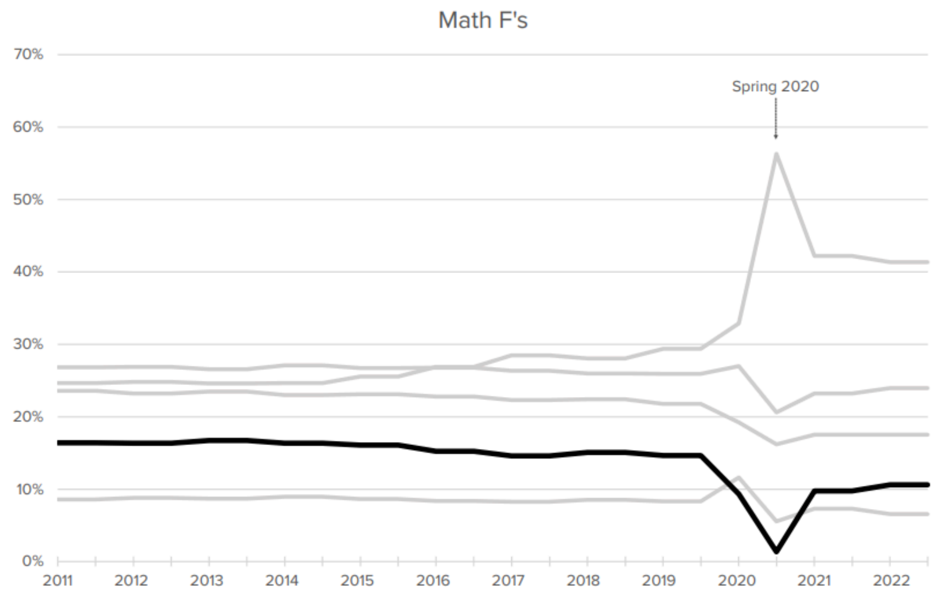
> *"Using test scores from 5.4 million U.S. students in grades 3-8, we tracked changes in math and reading achievement across the first two years of the pandemic. Average fall 2021 math test scores in grades 3-8 were .20-27 standard deviations (SDs) lower relative to same-grade peers in fall 2019 … These drops are significantly larger than estimated impacts from other large school disruptions, such as after Hurricane Katrina (Sacerdote [2012] when reported math scores dropped .17 SDs in one year for New Orleans evacuees)."*

Based on the magnitude of this learning loss, one would reasonably expect that student grades would have dropped during the pandemic. But instead, the opposite happened: grades skyrocketed and remained elevated even after most schools returned to normal in-person instruction after the pandemic. As researchers at the CALDER Center (Center for Analysis of Longitudinal Data in Education Research) discovered when analyzing educational data from the state of Washington (Goldhaber & Young, 2023):

> *"…[A]lmost no students received an F grade in the spring of 2020. The share of F grades dropped from … 9.3% to 1.4% in math courses … between the fall and spring semesters of 2020. The distribution of grades higher than F mostly increased for A grades, with the share of A's jumping from 32.9% to 56.3% in math … The average GPA in math jumped from 2.6 to 3.2 … The figures also suggest that English and science grades largely returned to pre-pandemic levels by 2021-22, but math grades did not. Indeed, the math GPA in 2021-22 was 2.7, 0.4 points higher than it was in 2018-19."*


Math Average GPA
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
2.3
2.3
2.2
2.3
2.3
2.3
2.4
2.3
2.4
2.6
3.2
2.7
2.7
F
S
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022


Math A's
70%
60%
50%
40%
30%
20%
10%
0%
Spring 2020
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022



Figures reproduced from Goldhaber & Young (2023) with permission.

However, standardized test scores have not increased commensurately:

> *"To better understand what these shifts in grading might mean, we perform some simple regressions to descriptively assess the extent to which the relationship between grades and test scores has changed over time. ... [A] student who got an 'A' in Algebra 1 was predicted to be in the 73rd percentile of the test distribution in 2015-16, the 68th percentile in 2018-19, and the 58th percentile in 2021-22."*

This phenomenon is not limited to the United States. It is widespread. For instance, a similar situation is described in an analysis of student grades in Italy (Doz, 2021):

> *"...[T]he results showed a statistically significant difference in pre- and post-COVID-19 quarantine grades. End-of-year grades were higher than those before the COVID-19 confinement. Furthermore, the results indicated that more than half of the students in the sample achieved a higher grade at the end of the school year. ... The findings suggest that greater caution should be paid in interpreting students' grades pre- and post-COVID-19 confinement, since it cannot be excluded that such students' achievements are inflated. Excessively high students' grades that do not represent their actual knowledge and competencies could give educators and legislators misleading and even false information about the quality of distance learning and students' knowledge."*

Indeed, grade inflation has been happening for a while, that is, COVID-19 amplified an existing trend. As researchers from ACT, Inc. describe, high school grade point averages (HSGPA) have increased while standardized test scores – not just aptitude-oriented tests like the SAT, but also achievement-oriented tests like the ACT, the NAEP, and even end-of-course exams – have not (Sanchez & Moore, 2022).

> *"...[A] mismatch between HSGPA and test scores suggests grade inflation is most likely present. HSGPA across time has been compared to ACT® (Bejar & Blew, 1981; Bellott, 1981) and SAT scores (Godfrey, 2011), NAEP data (U.S. Department of Education, National Assessment of Educational Progress [NAEP], Long-Term Trend Reading Assessments, 2020), and end-of-course exams (Gershenson, 2018). Consistently, analyses have shown that HSGPA has steadily increased over the last several decades, but standardized assessment scores have remained stagnant or have fallen (Bejar & Blew, 1981; Gershenson, 2020)."*

As elaborated by Gershenson (2018):

> *"...[R]ising high school grade point averages (GPAs) have been accompanied by stagnant SAT, ACT, and NAEP scores, strongly suggesting lowered classroom standards. And in higher education, As are now the most common grade awarded, despite constituting just 15 percent of grades in the early 1960s."*
>
> *"While many students are awarded good grades, few earn top marks on the statewide end-of-course exams for those classes. ... In fact, more than one-third of the students who received Bs from their teachers in Algebra 1 failed to score 'proficient' on the EOC exam."*
>
> *"...[E]arning a good grade in a course is no guarantee that a student has learned what the state expects her to have learned in that course. Results show that even students who earn the best grades often fail to demonstrate mastery of key skills and knowledge when measured on the state test. Recall that just 21 percent of A students and 3 percent of B students attain the 'superior' designation on the EOC, and more than one-third of B students don't reach proficiency at all."*

> Why Grade Inflation is a Problem

Gershenson (2018) mentions that grade inflation can create a "vicious cycle" of students being set up for failure in future courses:

> *"That's clearly a problem since receiving an A or B in a course signals academic success to most students and their families. When students earn passing grades despite not mastering the academic material, a vicious cycle can follow, whereby they're set up for failure via unmerited promotion to the next course or grade level."*
>
> *"...[G]rade inflation results in promoting students to subsequent grades and later accepting them to postsecondary institutions for which they are academically ill-prepared. Consequently, they struggle and risk dropping out.*
> *...*
> *[G]rade inflation may have the political consequence of encouraging families to believe everything is going well at school, even when a school is troubled and needs reform. It is easy for parents to ignore systemic mediocrity when their children's grades seem strong."*

This concern is echoed by the Goldhaber & Young (2023):

> *"Public opinion surveys point to a discrepancy between what parents believe about their student's level of achievement, i.e., that students have recovered academically, and what test results like NAEP suggest about their achievement (Esquivel, 2022; Kane & Reardon, 2023; Vázquez Tonnes, 2023).*
>
> *...*
>
> *Algebra 1 – the course for which we noted the greatest weakening in the relationship between test scores and grades – is seen as a gateway to more advanced math concepts (Snipes & Finkelstein, 2015). ... Schools use grades in classes such as Algebra 1 to determine whether students need extra support, remediation, or even if they must repeat a course before moving on. If this signal is no longer accurately conveying a student's level of achievement, school systems risk under-supporting students who need help.*
>
> *Likewise, families and students use grades as a signal of how a student is doing in school; the expectation is that if a student is having academic trouble, that trouble will show up in their grades. Decisions such as whether to enroll a child in after school tutoring or summer school may rest on a belief that the grades on a report card accurately reflect a student's levels of achievement. As we noted above, many parents are under the impression that their children are not suffering from learning loss due to the pandemic; however, test scores indicate otherwise. It is possible that without a grade that signals trouble, parents may not choose to get needed extra academic support."*

In short, parents typically think that an "A" indicates mastery of grade-level standards, but it often doesn't. If a student's school says that they're doing fine in math, then it does not automatically follow that the student is keeping college and career doors open that depend on mathematical proficiency. Different schools sometimes have their own interpretations of what it means for their students to be doing fine in math, and that doesn't always match up with grade-level standards, much less what is expected by colleges and careers.

This is a problem because it sets students up for failure later in life when it matters most. Every year, countless first-year college students decide to major in aerospace engineering or astrophysics or some other math-heavy subject, only to have that dream crushed when they realize they can't even pass an entry-level math course like Calc II (not even with the help of a tutor). These problems can be remedied when students are young, before their knowledge deficits grow too large – but problems can only be fixed after they are detected, and grades are no longer a reliable tool for detecting these problems. Inflated grades signal to students and parents that all career doors remain open, when in fact, many are in the process of being locked shut.

| Resistance to Objective Measurement

> Radical Constructivism Rejects the Idea of Measuring Learning Objectively

As discussed above, there is overwhelming evidence that grades have increased while standardized test scores have not. However, because remedying grade inflation and its downstream effects requires lots of extra effort from all parties involved (including teachers, students, parents, administrators), there is opposing pressure to reject the idea that grade inflation is occurring. Given the evidence, the only way to argue against the existence of grade inflation is to argue against the very idea of measuring learning objectively.

As prominent psychologists John Anderson, Lynne Reder, and Herbert Simon describe (1998), this is indeed a tenet of an educational philosophy known as **radical constructivism**:

> *"The denial of the possibility of objective evaluation is perhaps the most radical and far-reaching of the constructivist claims. ... D. Charney documents that empiricism has become a four-letter word in deconstructionist writings. D. H. Jonassen describes the issue from the perspective of a radical constructivist:*
>
> *'If you believe, as radical constructivists do, that no objective reality is uniformly interpretable by all learners, then assessing the acquisition of such a reality is not possible.'"*

Take it from Ernst von Glasersfeld (1984) himself, who is widely regarded as the philosopher who first formulated radical constructivism:

> *"Radical constructivism, thus, is radical because it breaks with convention and develops a theory of knowledge in which knowledge does not reflect an 'objective' ontological reality, but exclusively an ordering and organization of a world constituted by our experience."*

To concretely understand the radical constructivist position in the context of grade inflation, recall that in the absence of accountability and incentives, public spaces will become filled with trash. Logically, the existence of excessive trash in a public space provides an argument for increasing accountability and incentives surrounding littering and janitorial work. However, a radical constructivist will resist this conclusion on the grounds that "one person's trash is another person's treasure" and therefore it is impossible to objectively measure the amount of trash on the ground.

Clearly, this counterargument is ridiculous and nobody would actually espouse it in the context of trash. People who enter a space can see, and agree, about how much trash is on the ground. However, the counterargument persists in the context of education because "seeing the trash for oneself" often requires a combination of expertise in the subject matter – which most people do not have, especially in the context of mathematics. And even those who do see the trash often turn a blind eye to it out of convenience because they don't want to put in the extra effort to fix the situation, especially when their efforts will be met with griping from others who do not see the trash.

As Anderson, Reder, and Simon (1998) elaborate, the radical constructivists' rejection of objective reality leads to other problematic conclusions:

> *"Another sign of the radical constructivists' discomfort with evaluation manifests itself in the motto that the teacher is the novice and the student the expert. The idea is that every student gathers equal value from every learning experience. The teacher's task is to come to understand and value what the student has learned. As J. Confrey writes:*
>
> *'Seldom are students' responses careless or capricious. We must seek out their systematic qualities which are typically grounded in the conceptions of the student. ... [F]requently when students' responses deviate from our expectations, they possess the seeds of alternative approaches which can be compelling, historically supported and legitimate if we are willing to challenge our own assumptions.'*
>
> *Or, as Cobb, Wood, and Yackel write:*
>
> *'The approach respects that students are the best judges of what they find problematical and encourages them to construct solutions that they find acceptable given their current ways of knowing.'*
>
> *If the student is supposed to move, in the course of the learning experiences, from a lower to a higher level of competence, why are the student's judgments of the acceptability of solutions considered valid? While the teacher is valued who can appreciate children's individuality, see their insights, and motivate them to do their best and to value learning, definite educational goals must be set. More generally, if the 'student as judge' attitude were to dominate education, when instruction had failed and when it had succeeded, when it was moving forward and when backward, would no longer be clear.*
>
> *Understanding why the student, at a particular stage, is doing what he or she is doing is one thing. Helping the student understand how to move from processes that are 'satisfactory' in a limited range of tasks to processes that are more effective over a wider range is another matter. As L. B. Resnick argues, many concepts that children naturally come to (for example, that motion implies force) are not what the culture expects of education and in these cases 'education must follow a different path: still constructivist in the sense that simple telling will not work, but much less dependent on untutored discovery and exploration.'"*

> Radical Constructivism is a Present Force in Education

Radical constructivism might seem so outlandish that it is hard to imagine anyone seriously supporting it. However, it is indeed a present force in education. For instance, one document that circulated among educators during the 2021-22 school year (the year that most schools returned to in-person instruction after the COVID-19 pandemic) is *Where Is Manuel? A Rejection of 'Learning Loss'*, which, in a refusal to accept the reality that some demographics were more affected by pandemic-induced learning losses than others, outright rejects the idea that learning loss occurred during the pandemic.

The document makes a number of outlandish claims, some of are factually incorrect, and others of which are so vague that they can neither be proven or falsified (which effectively renders them meaningless):

> *"It is important to note that we believe that learning takes place everywhere and always. ... Funding and attention to 'fix learning loss' disregards our essential and suggested actions [to move toward antiracist mathematics education for all students].*
>
> *...*
>
> *This farce embodies the assumption that learning didn't happen, or that it didn't happen enough. This assumption is an insult to educators and families alike. ... When teachers could not connect, students continued their learning and growing with and within their families and communities. Some of this learning was closely matched to traditional school standards, and some of this learning was not as aligned to school standards but went deeper and was more authentic than anything that could have been learned through a computer screen or even in a school building. This learning may be different, but it is not any less important and should not be treated as if it is wrong or insufficient.*
>
> *Let's revisit Manuel. With persistence, the educators would have discovered that Manuel's days away from class were rich with learning experiences. Instead of attending class remotely, he went with his father to work and helped with his younger siblings and animals on the family farm. ... Manuel and his siblings did some work assigned by their teachers, but they were more motivated and engaged in the learning that was acquired outside of school. Manuel has not lost learning. The flexibility of remote learning has allowed him to supplement his studies from school with a rich mathematical understanding of the world.*
>
> *...*
>
> *Resist the thinking that students like Manuel are behind, and instead remind yourself that they are right where they should be after a pandemic. Resist deficit thinking and do not send deficit messages to the students like Manuel, and others who did attend daily, and instead look for what knowledge they gained and how they grew. ... Resist making the assumption that the learning students like Manuel experienced was not enough, and instead assume their experiences contributed to their present and future success in ways that are just as good, if not better, than what could have been learned through school."*

The organization producing this document, TODOS: Mathematics for ALL, is not just a fringe group. Between 2020-23, its leadership has included members of the Riverside and Santa Clara County Offices of Education and as well as professors from numerous universities including UCLA, UT Austin, The Ohio State University, University of Arizona, San Francisco University, University of Alberta, University of Missouri, Iowa State University, East Carolina University, University of New Mexico, Texas State University, and Utah Valley University. Furthermore, TODOS is a member of the Conference Board of the Mathematical Sciences (CBMS), which means that it is recognized by the International Mathematical Union (IMU) as one of the 19 national mathematical societies for the United States. For reference, IMU awards some of the highest honors in the mathematical profession, including the Fields Medal, which is widely considered to be the mathematical equivalent of the Nobel Prize.

TODOS has released numerous other documents espousing similar viewpoints. For instance, in *The Mo(ve)ment to Prioritize Antiracist Mathematics: Planning for This and Every School Year* (2020), released shortly before *Where is Manuel*, TODOS stated the following:

> *"Following school closure due to COVID-19, we have noted a resurgence of deficit views of students when they are described as 'behind' or 'unable to catch up since they missed so much school.' We believe this description of students is harmful. It frames students as individually responsible for a loss of learning and detracts from the broader issues of students and families surviving through a pandemic. Mathematics learning is a messy web of interconnected concepts. So we assert that instead of being distracted by framing students as lacking skills, we use the fall to start anew from an asset-based perspective. We urge policymakers, school district administrators, teachers, curriculum developers, and software developers to avoid playing into the fear-inciting discourses of students falling behind and ranking them by perceived ability.*
>
> *To take it a step further, in this moment we must rethink what counts as valid mathematical knowledge. … [W]e must expand our understanding of what it means to be good at mathematics, make space for alternative ways of knowing and doing mathematics based in the community, and acknowledge the brilliance, both in mathematics and beyond, of BIPOC [Black, Indigenous, and People of Color] in our classrooms."*

Likewise, in the joint position statement between TODOS and the National Council of Supervisors of Mathematics (NCSM) from 2016:

> *"…[Deficit views of historically marginalized children can arise from] the continuous labeling of children's readiness to learn mathematics via standardized tests and other institutional tools that position and sanction specific forms of mathematics knowledge. … A social justice priority in mathematics education is to openly challenge deficit thinking and the institutional tools and practices that perpetuate static views about children and their mathematics competencies.*
> *…*
> *Second, deficit thinking implies that students "lack" knowledge and experiences expected by the dominant group. Deficit thinking ignores, dismisses, or casts as barriers mathematical knowledge and experiences children engage with outside of school every day. A social justice approach to mathematics education assumes students bring knowledge and experiences from their homes and communities that can be leveraged as resources for mathematics teaching and learning (Civil, 2007; Gonzalez et al., 2005; Leonard & Martin, 2013; Turner et al., 2012).*
> *…*
> *Mathematics achievement, often measured by standardized tests, has been used as a gatekeeping tool to sort and rank students by race, class, and gender starting in elementary school (Davis & Martin, 2008; Ellis, 2008; Spielhagen, 2011)."*

Note that while we have discussed radical constructivism at a high level in this chapter, our general critique cuts deeper and will continue in chapter 8, where we emphasize that effective practice requires direct instruction as opposed to unguided instruction.

Math Academy is Held Accountable for Student Learning

Math Academy's existence depends on its ability to make students learn. If a student gets stuck and can't make progress in our system, then we're out of a job. If the learning on our system doesn't show up in students' grades and test scores outside of our system, then we're out of a job. We are properly incentivized to maximize student learning – real learning, not just the perception of it.

For this reason, we have no choice but to hold the line on what it means to truly learn something. When a student learns a topic on our system, they have to demonstrate that they really understand it. They have to solve real problems – successfully – and not just the simplest cases.

Perhaps surprisingly, this turns out to be one of our competitive advantages: we hold students accountable for learning, and they hold us accountable for providing material that is easy to learn from. More generally, we hold our users accountable for getting value out of our product, and our users hold us accountable for creating a valuable product.

This differentiates us from other free and ultra-low-cost online learning platforms whose dependence on a massive user base forces them to employ ineffective learning strategies that do not repel unserious students. Such platforms often cover only the simplest cases of each topic, and allow students to move on to more advanced topics despite poor performance on prerequisite topics. They are like teachers who go through the motions and check boxes, whereas Math Academy is like a tutor whose livelihood depends on the actual learning outcomes of its students.

Granted, we do sometimes have to deliver unfortunate news to incoming students and their parents who mistakenly believe, on the basis of their good grades, that they are comprehensively proficient in the mathematical subjects that they have taken at school. We can't promise that students and parents will be happy with their diagnostic results. It's not uncommon for our diagnostic test to expose that a student needs to relearn some things that they supposedly "already know" (but don't actually know) from school. But we can promise that, if our diagnostic test reveals any gaps in a student's knowledge, then we will automatically add those gaps to their learning plan to get them back on track.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review, 100*(3), 363.

  Importance: *Deliberate practice requires effort and is not inherently enjoyable. Individuals are motivated to practice because practice improves performance. In other words, maximal learning does not happen naturally as a result of maximizing other things like enjoyment, comfort, convenience, and ease of practice. In fact, maximal learning is at odds with some of these things. Sacrifices must be made.*

- Subirana, B., Bagiati, A., & Sarma, S. (2017). *On the Forgetting of College Academics: at" Ebbinghaus Speed"?.* Center for Brains, Minds and Machines (CBMM).

  Importance: *While it's true that college rankings can be viewed as some kind of incentive structure, it's learning is not the basis of such rankings. Major university ranking methodologies consistently lack any objective measure of what content knowledge and skills students retain from college education in the long term. The rankings may incentivize other things, but not learning.*

- Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher, 51*(7), 500-506.

  Importance: *The learning loss experienced by students during COVID-19 was even more extreme than that experienced by evacuees during Hurricane Katrina, one of the deadliest hurricanes to hit the United States.*

- Goldhaber, D., & Young, M. G. (2023). Course Grades as a Signal of Student Achievement: Evidence on Grade Inflation Before and After COVID-19. CALDER Research Brief No. 35.

  Doz, D. (2021). Students' mathematics achievements: A comparison between pre-and post-COVID-19 pandemic. *Education and Self Development, 16*(4), 36-47.

Importance: *During the COVID-19 pandemic, grades skyrocketed and remained elevated even after most schools returned to normal in-person instruction after the pandemic. This is a widespread phenomenon occurring beyond the United States. Grade inflation prevents learning issues from being detected, and schools and families risk under-supporting students who need help.*

- Sanchez, E. I., & Moore, R. (2022). Grade Inflation Continues to Grow in the Past Decade. Research Report. ACT, Inc.

  Gershenson, S. (2018). Grade Inflation in High Schools (2005-2016). *Thomas B. Fordham Institute.*

  Importance: *Grade inflation has been happening for a while. High school grade point averages (HSGPA) have increased while standardized test scores – not just aptitude-oriented tests like the SAT, but also achievement-oriented tests like the ACT, the NAEP, and even end-of-course exams – have not. Grade inflation can create a "vicious cycle" of students being set up for failure in future courses.*

- Anderson, J. R., Reder, L. M., Simon, H. A., Ericsson, K. A., & Glaser, R. (1998). Radical constructivism and cognitive psychology. *Brookings papers on education policy,* (1), 227-278.

  Von Glasersfeld, E. (1984). An introduction to radical constructivism. *The invented reality, 1740*, 28.

  TODOS: Mathematics for All (2020). *Where Is Manuel? A Rejection of 'Learning Loss'*.

  del Rosario Zavala, Maria, Ma Bernadette Andres-Salgarino, Zandra de Araujo, Amber G. Candela, Gladys Krause, and Erin Sylves (2020). The Mo(ve)ment to Prioritize Antiracist Mathematics: Planning for This and Every School Year. *TODOS: Mathematics for All.*

  National Council of Supervisors of Mathematics and TODOS: Mathematics for ALL. (2016). Mathematics education through the lens of social justice: Acknowledgment, actions, and accountability. *Joint Position Paper.*

  Importance: *Given the evidence, the only way to argue against the existence of grade inflation is to argue against the very idea of measuring learning objectively. Indeed, this outlandish position is taken by proponents of the radical constructivist philosophy of education, such as TODOS: Mathematics for ALL.*

II. ADDRESSING CRITICAL MISCONCEPTIONS

Chapter 6. The Persistence of Neuromyths

Summary: *Scientifically inaccurate beliefs about the brain endure even among professional educators. They persist because they sound scientific and are convenient to believe, especially if they provide a level of comfort, such as reinforcing one's preconceptions or offering false hope of a "quick fix" to someone who is facing a problem.*

Neuromyths are Common Misconceptions about the Brain

Despite the cognitive processes in the brain being fairly well understood at a high level, there are countless myths that persist despite decisive evidence to the contrary – even among professionals in the field of education. As Grospietsch & Lins (2021) emphasize:

> *"Numerous empirical studies reveal widespread endorsement of such misconceptions on the topic of learning and the brain both among the public at large and among pre-service and in-service teachers (e.g., Dekker et al., 2012; Ferrero et al., 2016). Even school principals, award-winning teachers and university instructors widely endorse neuromyths like 'we only use 10% of our brains', 'learning differences due to hemispheric use', or the 'existence of learning styles' (Horvath et al., 2018; Zhang et al., 2019)."*

As Pasquinelli (2012) notes, even high-ranking government officials are not immune to these myths:

> *"In 1998, the state of Florida passed a bill for day-care centers to play classical music to children. The same year, the Georgia governor asked for $105,000 for the production and distribution of classical music to newborns. He did so because he had read that listening to Mozart's music can boost IQ scores. Too good to be true."*

These misconceptions are so common that they have acquired a special name: **neuromyths** (Grospietsch & Lins, 2021).

> *"The term 'neuromyth' was coined by the neurosurgeon Alan Crockard in the 1980s to describe scientifically inaccurate understandings of the brain in medical culture (Howard-Jones, 2010). The Organisation for Economic Co-operation and Development (Organisation for Economic Co-operation and Development [OECD], 2002) defines-neuromyths as "misconception[s] generated by a misunderstanding, a misreading, or a misquoting of facts scientifically established (by brain research) to make a case for use of brain research in education and other contexts" (p. 111)."*

Why Neuromyths Persist

Neuromyths can often be characterized as the oversimplification, misinterpretation, and/or misapplication of a nuanced complex scientific finding.

> *"Grospietsch (2019) came to define neuromyths as misconceptions based on a kernel of 'truth', meaning that they take a scientific term or research finding (= neurofact) as a starting point for their argumentation, which morphs into a no-longer-scientifically-accurate implication for teaching and learning (= neuromyth) through a series of erroneous conclusions and logical fallacies."*

Despite being at best useless, and at worst detrimental, neuromyths persist because they sound scientific and are convenient to believe, especially if they provide a level of comfort, such as reinforcing one's preconceptions or offering false hope of a "quick fix" to someone who is facing a problem. As Pasquinelli (2012) explains:

> *"...[W]hy do neuromyths persist independently of their falsity and poor applicative value? It is likely that under the urge of application, educators are susceptible of turning toward easy-fixes that are presented in a respectable, scientific jargon and are loosely inspired by neuroscience (Hirsh-Pasek & Bruer, 2007).*
>
> *...*
>
> *Certain neuromyths for instance seem to fulfill a "soothing" function. ... Bangerter & Heath (2004) have shown that the interest in the Mozart effect ... positively correlates with lower teachers' salaries and low national tests scores per pupil spending. ... [Those] that are more in need of good education are easy prey to scientific legends about learning*
>
> *...*
>
> *Another feature of the human mind that might favor neuro- and other myths is confirmation bias, that is, the tendency to seek or interpret fresh information in a way that confirms previous beliefs (Nickerson, 1998).*
>
> *...*
>
> *Finally, it is largely accepted that both adults and children behave as if they had intuitions about laws of physics, biology, and psychology (diSessa, 1993). These intuitions are in many cases misconceptions that interfere with scientific instruction while preceding and influencing the acquisition of knowledge (Bloom & Weisberg, 2007)."*

To understand how learning can be optimized in an adaptive learning platform like Math Academy, it is important to first remedy any misconceptions about learning that could otherwise turn into sources of confusion.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Grospietsch, F., & Lins, I. (2021, July). Review on the prevalence and persistence of neuromyths in education–Where we stand and what Is still needed. In *Frontiers in Education* (Vol. 6, p. 665752). Frontiers Media SA.

  Betts, K., Miller, M., Tokuhama-Espinosa, T., Shewokis, P. A., Anderson, A., Borja, C., ... & Dekker, S. (2019). International Report: Neuromyths and Evidence-Based Practices in Higher Education. *Online Learning Consortium.*

  Pasquinelli, E. (2012). Neuromyths: Why do they exist and persist?. *Mind, Brain, and Education, 6*(2), 89-96.

  Importance: *Despite the cognitive processes in the brain being fairly well understood at a high level, there are countless myths that persist despite decisive evidence to the contrary – even among professionals in the field of education. These "neuromyths" can often be characterized as the oversimplification, misinterpretation, and/or misapplication of a nuanced complex scientific finding. Despite being at best useless, and at worst detrimental, neuromyths persist because they sound scientific and are convenient to believe, especially if they provide a level of comfort, such as reinforcing one's preconceptions or offering false hope of a "quick fix" to someone who is facing a problem.*

Chapter 7. Myths & Realities about Individual Differences

Summary: *Different people generally have different working memory capacities and learn at different rates, but people do not actually learn better in their preferred "learning style." Instead, different people need the same form of practice but in different amounts. Additionally, not everybody can learn every level of math, but most people can learn the basics. In practice, however, few people actually reach their full mathematical potential because they get knocked off course early on by factors such as missing foundations, ineffective practice habits, inability or unwillingness to engage in additional practice, or lack of motivation.*

People Differ in Learning Speed, Not Learning Style

Myth: *Everybody has the same working memory capacity and learns at the same rate, but different people learn differently depending on their preferred learning style.*

Reality: The exact opposite is true. Different people generally have different working memory capacities and learn at different rates. While people may have preferred learning "styles" (e.g. visual vs verbal), they do not actually learn better when given information in their preferred style. The myth is that different people need the same amount of practice but in different forms – whereas the reality is that different people need the same form of practice but in different amounts.

| Learning Style Preferences are Irrelevant

One of the most widespread – and most widely debunked – neuromyths is that people learn better when they receive information in their preferred "learning style." To quote the authors of one of the largest and most comprehensive studies on the persistence of neuromyths (Betts et al., 2019):

> *"Learning styles is one of the most widespread myths in education (Pashler, McDaniel, Rohrer & Bjork, 2008; Reiner & Willingham, 2010; Roher & Pashler, 2012). Despite repeated testing of hypotheses relating to learning styles, there is no evidence to date showing that individuals learn*

> *better when they receive information in their preferred learning styles (Newtown & Miah, 2017; Newtown & Miah, 2017).*
> *...*
> *In 2006, a learning styles challenge was put forth by a team of underwriters offering $1,000 and then moving it up to $5,000 to provide scientific evidence supporting this myth (Wallace, 2014). To date, there has not been a payout."*

As Grospietsch & Lins (2021) elaborate:

> *"According to Grospietsch and Mayer (2021b), the kernel of truth behind this neuromyth is that people differ in the mode in which they prefer to receive information (visually or verbally; e.g., Höffler et al., 2017).*
>
> *The first erroneous conclusion that can be drawn from this kernel of truth is that there are auditory, visual, haptic and intellectual learning styles, as Vester (1975) suggested in the German context.*
>
> *The next erroneous conclusion drawn is that people learn better when they obtain information in accordance with their preferred learning style.*
>
> *Finally, the third erroneous yet widely disseminated conclusion is that teachers must diagnose their students' learning styles and take them into account in instruction. ... [T]here is no empirical evidence confirming the effectiveness of considering students' learning styles in instruction (Willingham et al., 2015)."*

As Kirschner & Hendrick sum it up (2024, pp.108):

> *"These so-called learning styles have been exposed as nonsense in research time after time. There are no 'image thinkers' or 'language thinkers'. Everyone thinks with both systems and everyone benefits from using both. The more often you use the two systems together, the stronger the trace in your memory and the better you will remember and thus learn."*

| Working Memory Capacity (WMC) Differences are Relevant

However, one aspect of the brain that has been widely documented not only to vary between people, but also to affect people's general cognitive performance, is **working memory capacity** (**WMC**). As Conway et al. (2007) describe:

> *"A fundamental characteristic of WM [working memory] is that it has a limited capacity, which constrains cognitive performance, such that individuals with greater capacity typically perform better than individuals with lesser capacity on a range of cognitive tasks.*
>
> *For example, older children have greater capacity than younger children, healthy adults have greater capacity than patients with frontal-lobe damage or disease, younger adults have greater capacity than elderly adults, and in all such cases, those individuals with greater WM capacity*

> *out-perform individuals with lesser capacity in several important cognitive domains, including complex learning, reading and listening comprehension, and reasoning.*
>
> *In short, we know that variation in WM capacity exists and that this variation is important to everyday cognitive performance."*

These differences in working memory capacity have been characterized not only at a psychological level, but also at the physical level of brain activity measures. Vogel & Machizawa (2004) have found that brain activity reaches a plateau when people attempt to perform tasks that meet or exceed their WMC, and people with high WMC reach this plateau much later than people with low WMC:

> *"Here, we provide electrophysiological evidence for lateralized activity in humans that reflects the encoding and maintenance of items in visual memory. The amplitude of this activity is strongly modulated by the number of objects being held in the memory at the time, but approaches a limit asymptotically for arrays that meet or exceed storage capacity.*
>
> *Indeed, the precise limit is determined by each individual's memory capacity, such that the activity from low-capacity individuals reaches this plateau much sooner than that from high-capacity individuals. Consequently, this measure provides a strong neurophysiological predictor of an individual's capacity, allowing the demonstration of a direct relationship between neural activity and memory capacity.*
> *...*
> *That is, simply by measuring the amplitude increase across memory array sizes, we could accurately predict an individual's memory capacity."*

Engström, Landtblom, & Karlsson (2013) have explained why this happens: the higher one's WMC, the less neural activity their brain requires to perform the task – in other words, the task is less taxing on their brain.

> *"Low- and high-capacity participants showed an increase in activity as a function of increasing demands but differed in that high-capacity participants started from a lower level."*

> WMC Impacts Perceived Effort

It comes as no surprise, then, that people with higher WMC will generally perceive a given task to be easier than people with lower WMC. Indeed, this has been demonstrated experimentally in a study that measured how difficult people found it to identify spoken words in the presence of background noise (Rudner et al., 2012):

> *"...[P]articipants were asked to rate effort at SNRs [signal-to-noise ratios, i.e. difficulty levels] individually adapted to their speech recognition performance. Thus, individual differences in speech recognition ability were taken into account in the ratings of perceived effort; even so WM capacity explained variance in perceived effort between conditions.*

> ...
> *[T]he difference in perceived listening effort in modulated and steady state noise at relatively favorable SNRs is a function of WM capacity. ... [P]ersons with greater WM capacity find listening in noise less effortful than persons with lower WM capacity across all three levels and noise types.*
> ...
> *[R]atings of listening effort may be an indicator of the relative degree of engagement of explicit processing resources in WM. Thus, a relation between WM and rated effort may indicate that persons with greater WM capacity are using fewer explicit processing resources"*

> WMC Impacts Abstraction Ability

Similarly, it has also been shown that high WMC facilitates **abstraction**, that is, seeing "the forest for the trees" by learning underlying rules as opposed to memorizing example-specific details (McDaniel et al., 2014). This is unsurprising, given that understanding large-scale patterns requires balancing many concepts simultaneously in WM.

> *"...[A]fter training (on a function-learning task), participants either displayed an extrapolation profile reflecting acquisition of the trained cue-criterion associations (exemplar learners) or abstraction of the function rule (rule learners; Studies 1a and 1b).*
> ...
> *Studies 1c and 2 examined the persistence of these learning tendencies on several categorization tasks. Study 1c showed that rule learners were more likely than exemplar learners (indexed a priori by extrapolation profiles) to resist using idiosyncratic features (exemplar similarity) in generalization (transfer) of the trained category. Study 2 showed that the rule learners but not the exemplar learners performed well on a novel categorization task (transfer) after training on an abstract coherent category.*
> ...
> *[W]orking memory capacity (as measured by Ospan following Wiley et al., 2011) was a significant and unique predictor of the tendency to rely on rule versus exemplar processes in the function learning task, such that higher working memory capacity was related to reliance on rule learning.*
>
> *For a number of reasons, greater working memory capacity could facilitate abstracting the function rule during learning, including the ability to maintain and compare several stimuli concurrently (Craig & Lewandowsky, 2012), to partition the training stimuli into two linear segments and switch back and forth between them during learning (Erickson, 2008; Sewell & Lewandowsky, 2012), and to reject or ignore initial biases (e.g., a positive linear) in order to discern the given function (cf., Wiley et al., 2011).*
>
> *Thus, learners enjoying greater working memory capacity might be more inclined to engage processes that would support rule learning (relating several training trials, partitioning training trials, ignoring initial biases) than would learners with more limited working memory capacity."*

Abstracting underlying rules improves one's ability to extrapolate knowledge to new contexts, a skill that is widely assessed in academic settings. Indeed, individual differences in abstraction ability have been shown to impact educational outcomes (McDaniel et al., 2018):

> *"Students may do well answering exam questions that are similar to examples presented in class. Yet, some of these students perform poorly on exam questions that require applying instructed concepts to a new problem whereas others fare better on such questions.*
>
> *Our hypothesis is that these performance differences reflect, in part, individual differences in learners' tendencies to focus on acquiring the particular exemplars and responses associated with the training exemplars (exemplar learners) versus attempting to abstract underlying regularities reflected in particular exemplars (abstraction learners). ... [W]e differentiated students on this dimension, and then tracked their final exam performances in introductory chemistry courses.*
>
> *Abstraction learners demonstrated advantages over exemplar learners for transfer questions but not for retention questions. The results converge on the idea that individual differences displayed in how learners acquire and represent concepts persist from laboratory concept learning to learning complex concepts in science courses."*

> WMC Impacts Learning Speed

As one might infer from the impact of WMC on perceived effort and abstraction ability, WMC has also been shown to impact **speed of learning**, that is, the rate at which one's ability to perform a task improves over the course of exposure, instruction, and practice on the task (though the impact of WMC on task performance is diminished after the task is learned to a sufficient level of performance).

Multiple studies have linked individual differences in speed of learning and WMC in the context of categorization tasks (see McDaniel et al., 2014 for a summary):

> *"...[A]cross several types of categorization tasks, Craig and Lewandowsky (2012) and Lewandowsky (2011) reported significant correlations between speed of learning and working memory capacity. In the present study, we found a similar general association between speed of learning in the function task and working memory capacity as indexed by Ospan alone.*
> *...*
> *Learning the function rule presumably requires maintaining and comparing stimuli across trials ("comparative hypothesizing", Klayman, 1988) and possibly partitioning the stimuli into subsets for the different slopes and switching back and forth across these partitioned segments during training (Lewandowsky et al., 2002; Sewall & Lewandowsky, 2012), and these processes require working memory capacity (both from a theoretical perspective, Craig & Lewandowsky, 2012; and based on empirical findings, Sewall & Lewandowsky, 2012). Consequently, for participants attempting to abstract the function rule, higher working memory capacity (as indexed by Ospan scores), would facilitate learning."*
>
> *"The implication is that for the rule learners, those with higher working memory capacity were able to more effectively support the processing needed to determine the functional relation among the training points, thereby supporting faster learning."*

Another study reported that reducing WMC slowed learning during a puzzle (Reber & Kotovsky, 1997):

> *"Participants solving the Balls and Boxes puzzle for the first time were slowed in proportion to the level of working memory (WM) reduction resulting from a concurrent secondary task. On a second and still challenging solution of the same puzzle, performance was greatly improved, and the same WM load did not impair problem-solving efficiency. Thus, the effect of WM capacity reduction was selective for the first solution of the puzzle, indicating that learning to solve the puzzle, a vital part of the first solution, is slowed by the secondary WM-loading task."*

The impact of WMC on learning speed is not limited to puzzles in academic laboratories – it extends to real-life contexts of academics and professional expertise. For instance, in a study of piano players, WMC was a significant predictor of performance even for experts who had logged thousands of hours of practice – that is, high-WMC pianists attained the same level of performance with fewer hours of practice, or a greater level of performance with the same hours of practice, compared to low-WMC pianists (Meinz & Hambrick, 2010).

> *"In evaluating participants having a wide range of piano-playing skill (novice to expert), we found that deliberate practice accounted for nearly half of the total variance in piano sight-reading performance. However, there was an incremental positive effect of WMC, and there was no evidence that deliberate practice reduced this effect. Evidence indicates that WMC is highly general, stable, and heritable, and thus our results call into question the view that expert performance is solely a reflection of deliberate practice."*

To be clear, the variation in ability was explained primarily by the amount of effective practice, but WMC was indeed a significant secondary factor. As Kulasegaram, Grierson, & Norman (2013) summarize:

> *"Although all studies support extensive DP [deliberate practice] as a factor in explaining expertise, much research suggests individual cognitive differences, such as WM capacity, predict expert performance after controlling for DP. The extent to which this occurs may be influenced by the nature of the task under study and the cognitive processes used by experts. The importance of WM capacity is greater for tasks that are non-routine or functionally complex."*

At the other end of the spectrum, Swanson & Siegel (2011) found that students with learning disabilities generally have lower WMC:

> *"We argue that in the domain of reading and/or math, individuals with LD have smaller general working-memory capacity than their normal achieving counterparts and this capacity deficit is not entirely specific to their academic disability (i.e., reading or math). ... We find that in situations that place high demands on processing, individuals with LD have deficits related to controlled attentional processes (e.g., maintaining task relevant information in the face of distraction or interference) when compared to their chronological aged-matched counterparts.*
>
> *...*
>
> *One conclusion from the experimental literature is that individual differences in WM (of which executive processing is a component) are directly related to achievement (e.g., reading comprehension) in individuals with average or above average intelligence (e.g., Daneman & Carpenter, 1980). Thus, children or adults with normal IQs have difficulty (or efficiency varies) in*

> *executive processing and that such difficulties are not restricted to those with depressed intelligence*
>
> *...*
>
> *Our conclusions from approximately two decades of research are that WM deficits are fundamental problems of children and adults with LD. Further, these WM problems are related to difficulties in reading and mathematics, and perhaps writing. Although WM is obviously not the only skill that contributes to academic difficulties [e.g., vocabulary and syntactical skills are also important (Siegal and Ryan, 1988)], WM does play a significance role in accounting for individual differences in academic performance."*

| Lack of Evidence for WMC Training

While it is possible to train and improve on tasks that are typically used to measure WMC, evidence is currently lacking that these task-specific performance improvements actually represent an increase in WMC that can be transferred to more general contexts. As described by Redick et al. (2015):

> *"Despite the promising results of initial research studies, the current review of all of the available evidence of working memory training efficacy is less optimistic. Our conclusion is that working memory training produces limited benefits in terms of specific gains on short-term and working memory tasks that are very similar to the training programs, but no advantage for academic and achievement-based reading and arithmetic outcomes.*
>
> *...*
>
> *Previous work has shown that manipulations can increase a person's score on a working memory measure (e.g., re-taking a test, motivation, strategy instruction), but this improvement in the individual's working memory score may not reflect a true change in underlying working memory ability. For example, Ericsson et al. (1980) demonstrated a subject who, through mnemonic strategies, was able to increase his serial recall of digits to 79 in a row, though when tested on memory span measures that did not include digits, his scores were in the normal range (7 ± 2).*
>
> *...*
>
> *The bulk of the evidence from studies with rigorous methodology provide little evidence for the efficacy of working memory training in improving academic and achievement outcomes such as reading, spelling, and math. The observation of positive near transfer to working memory and lack of academic or achievement test far transfer corresponds with previous meta-analyses (Melby-Lervåg & Hulme, 2013; Rapport et al., 2013), and indicates that contrary to popular belief, the evidence for the educational benefit of working memory training is lacking."*

However, as Anderson (1987) points out, training domain-specific skills can *effectively* turn long-term memory into an extension of working memory:

> *"Chase and Ericsson (1982) showed that experience in a domain can increase capacity for that domain. Their analysis implied that what was happening is that storage of new information in long-term memory, became so reliable that long-term became an effective extension of short-term memory."*

For emphasis, we quote Chase and Ericsson (1982) directly:

> *"The major theoretical point we wanted to make here is that one important component of skilled performance is the rapid access to a sizable set of knowledge structures that have been stored in directly retrievable locations in long-term memory. We have argued that these ingredients produce an effective increase in the working memory capacity for that knowledge base."*

It comes as no surprise, then, that Redick et al. (2015) recommend that students focus on training subject-specific skills directly:

> *"We recommend that in contrast to unstructured, unguided, general interventions such as cognitive training and videogame training, more research should be focused on training specific skills and abilities that are likely to exhibit near transfer to very similar academically relevant outcomes – for example, training specific language skills in children with text comprehension difficulties (Clarke, Snowling, Truelove, & Hulme, 2010), or computer-assisted instruction of reading and math skills (Rabiner, Murray, Skinner, & Malone, 2010)."*

These recommendations are echoed (Anderson et al., 1998) by K. Anders Ericsson, one of the most influential researchers in the field of human expertise and performance:

> *"...[M]odern educators have trained many generalizable abilities such as creativity, general problem-solving methods, and critical thinking. However, decades of laboratory studies and theoretical analyses of the structure of human cognition have raised doubts on the possibility of training general skills and processes directly, independent of specific knowledge and tasks.*
>
> *For example, research on thinking and problem solving show that successful performance depends on special knowledge and acquired skills, and studies of learning and skill acquisition show that improvements in performance are primarily limited to activities in the specific domain."*

The recommendations are also echoed by researchers Amanda VanDerHeyden and Robin S. Codding (2020), who have extensive experience researching academic intervention in mathematics:

> *"The evidence summarized and analyzed in meta-analytic studies illustrates that (a) although cognitive measures correlate with mathematics achievement, these measures do not correlate with student responsiveness to intervention; (b) using cognitive assessment tools does not provide the information necessary to improve academic skill weaknesses; and (c) cognitive interventions do very little to improve academic performance outcomes (Burns, 2016).*
> *...*
> *[Jacob and Parkinson (2015)] concluded that there are very few rigorous intervention studies examining the causal link between executive function interventions and academic outcomes. ... [T]hese existing studies showed improvements on measures of executive function but no improvements on academic achievement. Thus, the notion that executive function training can bring about gains in mathematics proficiency is not consistent with existing evidence. The evidence serves as a reminder that the most effective way to address a math skill deficit is to directly remediate math skills rather than trying to improve working memory or executive functioning as a means to address math skill deficits."*

| Different Students Need Different Amounts of Practice

The takeaway from all of this is that an adaptive learning system should focus on subject-specific learning tasks and adapt to a student's observed learning speed, not their preferred learning style. Each student needs to be given enough practice to achieve mastery on each learning task – and that amount of practice may vary depending on the particular student and the particular learning task.

While this may seem like a disappointing truth for students who generally need more practice than others, we re-emphasize a study quoted earlier in this chapter, which showed that the impact of WMC on task performance was diminished after the task was learned to a sufficient level of performance (Reber & Kotovsky, 1997).

> *"Participants solving the Balls and Boxes puzzle for the first time were slowed in proportion to the level of working memory (WM) reduction resulting from a concurrent secondary task.* ***On a second and still challenging solution of the same puzzle, performance was greatly improved, and the same WM load did not impair problem-solving efficiency.*** *Thus, the effect of WM capacity reduction was selective for the first solution of the puzzle, indicating that learning to solve the puzzle, a vital part of the first solution, is slowed by the secondary WM-loading task."*

More generally, as Unsworth & Engle (2005) explain:

> *"...[I]ndividual differences in WM capacity occur in tasks requiring some form of control, with little difference appearing on tasks that required relatively automatic processing."*

In this view, extra practice should not be viewed as limiting the progress of students who are slower to learn, but rather as empowering them to develop greater automaticity and lessen the impact of the cognitive difference responsible for their slower learning, thereby allowing continued learning on more advanced material.

We emphasize that this is fully compatible with, and in fact a necessary part of maintaining a **growth mindset**. Nobody's current level of knowledge is "fixed" or set in stone, and in order to support every student and maximize their learning, it's necessary to provide some students with more practice than others. The whole goal of adapting the amount of practice to individual differences in student learning speeds is to support maximum student growth. In fact, in the absence of such adaptivity student growth would certainly be stunted:

- If a student is catching on slowly, and you don't give them enough practice and instead move them on to the next thing before they are able to do the current thing, then you'll soon push them so far out of their depth that they'll just be struggling all the time and not actually learning anything, thereby stunting their growth.

- Likewise, if a student picks up on something really quickly and you make them practice it for way longer than they need to instead of allowing them to move onward to more advanced material, that's also stunting their growth.

To maximize each individual student's growth on each individual skill that they're learning, Math Academy gives each student enough practice to achieve mastery and allows them to move on to more advanced skills immediately after mastering the prerequisites.

Your Mathematical Potential Has a Limit, but it's Likely Higher Than You Think

Myth: *Everybody can learn every level of math.*

Reality: Most people can learn basic math like arithmetic and some algebra – but beyond that, higher levels of math become increasingly abstract and technical, and fewer people have the cognitive resources to learn it quickly enough to make a career out of it, much less get to that point relatively early in their lives.

| Levels of Math

One problem with this myth is that most people do not understand just how deep the levels of mathematics can go, and how cognitively taxing it is to learn the deepest levels. Arithmetic is a completely different ballpark from graduate-level math (and beyond). Most people consider calculus to be "really advanced math," but calculus is not even halfway to the level at which expert mathematicians operate.

For reference, we offer a loose formulation of the levels of mathematics below:

1. *Arithmetic* – seldom considered "hard math"

2. *Algebra* – often considered "hard math" by people who disliked math in school

3. *Calculus* – considered "hard math" by the general public

4. *Real Analysis, Abstract Algebra, Partial Differential Equations, etc.* – considered "hard math" by most college students majoring in math

5. *Algebraic Topology, Differential Geometry, etc.* – considered "hard math" by most graduate students doing PhDs in math, as well as many research professors in math

6. *The math underlying solutions to the most famous problems in modern mathematics, e.g. Ricci Flow with Surgery which underlies the proof of Poincaré Conjecture* – considered "hard math" by the world's top mathematicians

To put these levels in perspective, it can be helpful to draw an analogy to athletics:

- Learning arithmetic is like basic ambulatory movement: almost everyone can do it.

- Learning high school calculus is like being able to run ten miles without stopping: it takes time and effort to work up to it, but by training effectively and consistently, many people can accomplish it.

- Learning research-level mathematics is like qualifying for the 100-meter dash at the Olympics: it requires a certain biological predisposition coupled with the commitment of thousands of hours to the most grueling forms of training.

The reason why this is harder to accept in the context of mathematics than in the context of athletics is that we cannot observe the makeup and functioning of our brains as clearly as we can our bodies. But, as elaborated earlier in this chapter, individual differences in brains do exist (e.g. working memory capacity) and are relevant to key mathematical skills (e.g. abstraction ability).

| The Abstraction Ceiling

To help lend some concreteness to something as abstract as "abstraction ability," it may help to hear the famed Douglas Hofstadter (2012) recount his firsthand experience of running up against an **"abstraction ceiling"** in his own brain while pursuing a PhD in mathematics:

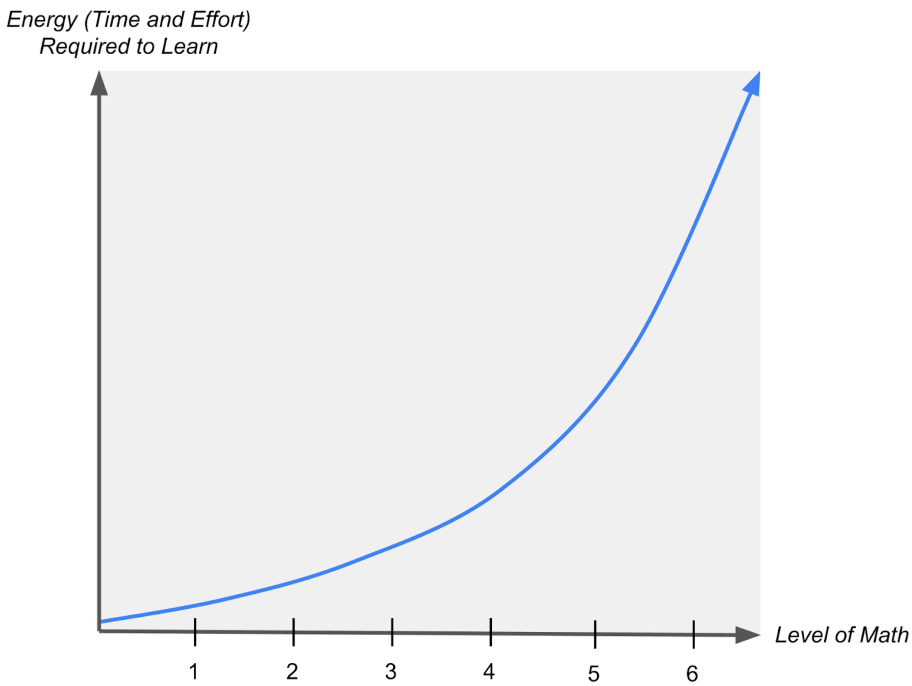
> *"I am a 'mathematical person', that's for sure, having grown up profoundly in love with math and having thought about things mathematical for essentially all of my life (all the way up to today),*

> *but in my early twenties there came a point where I suddenly realized that I simply was incapable of thinking clearly at a sufficiently abstract level to be able to make major contributions to contemporary mathematics.*
> *...*
> *I had never suspected for an instant that there was such a thing as an 'abstraction ceiling' in my head. I always took it for granted that my ability to absorb abstract ideas in math would continue to increase as I acquired more knowledge and more experience with math, just as it had in high school and in college.*
> *...*
> *I found out a couple of years later, when I was in math graduate school, that I simply was not able to absorb ideas that were crucial for becoming a high-quality professional mathematician. Or rather, if I was able to absorb them, it was only at a snail's pace, and even then, my understanding was always blurry and vague, and I constantly had to go back and review and refresh my feeble understandings. Things at that rarefied level of abstraction ... simply didn't stick in my head in the same way that the more concrete topics in undergraduate math had ... It was like being very high on a mountain where the atmosphere grows so thin that one suddenly is having trouble breathing and even walking any longer.*
> *...*
> *To put it in terms of another down-home analogy, I was like a kid who is a big baseball star in high school and who is consequently convinced beyond a shadow of a doubt that they are destined to go on and become a huge major-league star, but who, a few years down the pike, winds up instead being merely a reasonably good player on some minor league team in some random podunk town, and never even gets to play one single game in the majors. ... Sure, they have oodles of baseball talent compared to most other people – there's no doubt that they are highly gifted in baseball, maybe 1 in 1000 or even 1 in 10000 – but their gifts are still way, way below those of even an average major leaguer, not to mention major-league superstars!*
> *...*
> *On the other hand, I think that most people are probably capable of understanding such things as addition and multiplication of fractions, how to solve linear and quadratic equations, some Euclidean geometry, and maybe a tiny bit about functions and some inklings of what calculus is about."*

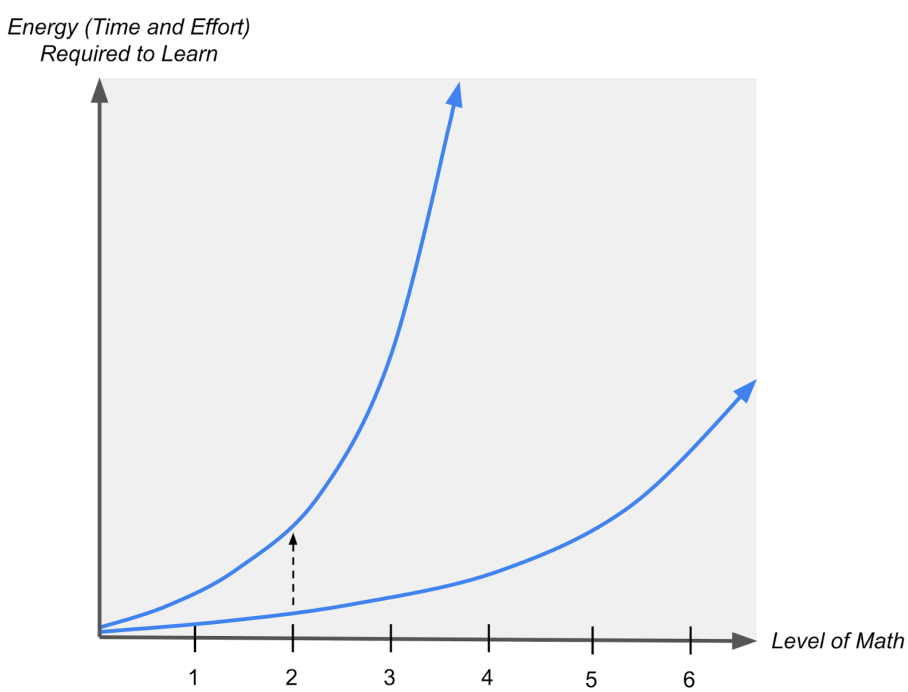
As Hofstadter describes, the abstraction ceiling is not a "hard" threshold, a level at which one is suddenly incapable of learning math, but rather a "soft" threshold, a level at which the amount of time and effort required to learn math begins to skyrocket until learning more advanced math is effectively no longer a productive use of one's time. That level is different for everyone. For Hofstadter, it was graduate-level math; for another person, it might be earlier or later (but almost certainly earlier).

| Learning Energy vs Level of Math

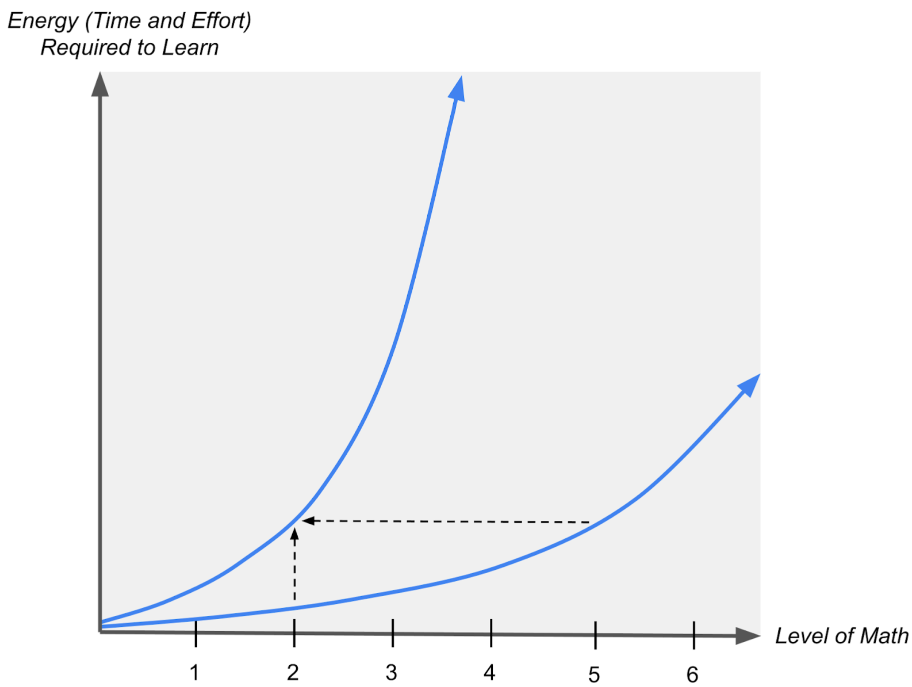
The central insight is that the further you go in math, the more energy it requires to learn the next level up. Whether they realize it or not, everybody who learns math is on an exponential curve of energy (time and effort) versus the level of math. (A key feature of exponential curves is that they can look fairly flat at the beginning, but appear to skyrocket later on, despite there being a constant "multiplier" to get from one point to the next.)



As we described earlier in this chapter, people with lower working memory capacities generally perceive cognitive skills to require more effort and more practice to master. It is as if there is a "multiplier" on the amount of energy required.



Another key feature of exponential curves is that vertical scaling is equivalent to horizontal translation. For instance, if we take the curve 2^x and multiply it by 8 (representing a person who requires 8 times more energy to learn math), then we have the curve $8 \cdot 2^x$ which is also equivalent to 2^{3+x}, a horizontal shift 3 levels to the left.



At the beginning (at the left), the two graphs are not that different, but as we look further to the right (progressively more levels of math), they quickly separate, and one graph skyrockets much earlier than the other. Everybody is on this exponential curve of energy (time and effort required to learn) versus level of abstraction, but everyone's curve is shifted horizontally depending on their cognitive ability and degree of motivation/interest. For some people, math doesn't get hard until graduate-level Algebraic Topology; for others, it becomes hard as early as high school algebra.

These graphs demonstrate the importance of learning efficiency: by increasing learning efficiency, Math Academy is able to *divide* the energy required to learn, *flattening* the graphs above, which is the same as shifting them to the right. When students enjoy high-efficiency learning conditions, they not only make faster progress, but also reach higher levels of math than they would otherwise.

Students get off the math train and stop taking math classes once it begins to feel too inefficient. Once the progress-to-work ratio gets too low, they lose interest and focus on other endeavors where their progress-to-work ratio is higher. Efficiency keeps the progress-to-work ratio as high as possible, keeping students on the math learning train for as far as possible.

| All Can Learn Some, Many Can Learn More, but Few Can Learn All

> Nature or Nurture? Both Matter

This characterization is compatible with the usual findings of studies into the effects of nature and nurture on skill acquisition: both matter. Talent – the top speed at which one can acquire skills in a particular domain – matters, and so does hard work. As Kirschner & Hendrick describe (2024, pp.142):

> *"If you are trying to convince students that innate ability accounts for 0% of success and effort accounts for 100% of success then you are misleading them. Instead of saying to students 'talent doesn't matter, only effort matters', what we should be saying to students is 'yes, talent and natural ability play a big part in success but effort matters on the margins, and the marginal gains can go on to yield significant gains'."*

Lack of talent does not necessarily mean "you can't do this," but it does mean that someone lacking talent will need to work much harder, possibly to an infeasible extent, as compared to someone with talent. But because the human lifespan is so limited and human talents can be so diverse, "you need to work much harder than others to accomplish this" effectively means the same thing as "you probably won't do this because you'll find more efficient, productive, and fulfilling uses of your time doing other things."

As prominent psychologist Dean Keith Simonton summarizes (2007):

> *"...[T]he concept of talent does not require the existence of 'innate constraints to the attainment of elite achievement'. On the contrary, genetic endowment may merely influence the rate at which domain-specific expertise is acquired without imposing any upper or lower bounds on attainment. Thus, empirical research indicates that outstanding creative individuals require less time to master the requisite knowledge and skill than do less creative individuals (Simonton, 2000).*
>
> *In addition, talent may affect the magnitude of performance for individuals with the same acquired level of expertise. Talented persons may 'get more bang for the buck' out of a given quantity of declarative and procedural knowledge. But, again, this enhancement effect does not amount to the imposition of any 'innate constraints'."*

Elsewhere, Simonton elaborates (2013, pp.17-26) further on the importance of both nature and nurture to the development of expert performance. It has been well established that expert performance is contingent upon favorable sociocultural conditions, family and education circumstances, and massive amounts of deliberate practice:

> *"Environmental factors play a major role in the development of greatness. Furthermore, these factors are extremely diverse. They include the larger sociocultural conditions ... as well as more proximate circumstances, such as family background and education ... In addition, it has been well established in a wide range of achievement domains that greatness is contingent on what has been called 'deliberate practice' (e.g., Ericsson, Krampe, & Tesch-Römer, 1993; Krampe & Ericsson, 1996).*
>
> *...*
>
> *If any of these [essential environmental factors] attain levels of zero, and thus become totally unsupportive of greatness, then greatness will fail to materialize. A violinist who never practices will never become a virtuoso violinist, and probably not even a decent amateur player. The same holds for sociocultural factors."*

However, many critical traits underlying expert performance have been shown to have a significant genetic component:

> *"Reasonably precise heritability coefficients have been estimated for many critical intellectual, dispositional, and physical variables (Bouchard, 2004; Bouchard, Lykken, McGue, Segal, & Tellegen, 1990). ... [O]ften genetics accounts for at least one-third of the variation, and sometimes more than half.*
>
> *...*
>
> *This is not to say that there do not exist abilities or traits that lack significant heritabilities ... It's just that the latter represent the exception rather than the rule. Certainly most major cognitive abilities are inherited to a very substantial degree, and heritabilities are moderately high for all dispositional variables associated with the attainment of greatness."*

While the development of expert performance depends on favorable environmental conditions and massive amounts of deliberate practice, the *speed* of development can be accelerated (or decelerated) by genetic factors:

> *"...[I]t is far more fruitful to define innate talent in terms of expertise acquisition (Simonton, 2008b). This definition starts by viewing talent as a set of cognitive abilities, dispositional traits, and (where necessary) physical attributes ... [T]his variable set in whole or in part must either (a) accelerate the rate at which domain-specific expertise is acquired (i.e., "better faster" effect) or (b) enhance domain-specific performance for a given amount of acquired expertise (i.e., "more bang for the buck" effect). ... Nature is what facilitates and accentuates nurture.*
>
> *...*
>
> *This definition allows us to explain four facts that would otherwise be inexplicable. First, individuals vary immensely in how long it takes to acquire the expertise requisite for greatness (Simonton, 2000). ... Second, those who take less time to acquire expertise are actually better off than those who take more time (Simonton, 2000). ... Third, greatness is positively associated with broad interests, hobbies, and even versatility (e.g., Root-Bernstein, Bernstein, & Garnier, 1995; Root-Bernstein et al., 2008; Simonton, 1976; Sulloway, 1996). ... [I]t would seem impossible for*

> *anyone to become a great polymath if every domain always required a full decade to acquire sufficient expertise. Fourth and last, empirical research in both behavioral genetics and differential psychology has conclusively identified sets of abilities and traits that feature both substantial heritability coefficients and sizeable predictive validities (Bouchard & Lykken, 1999; Simonton, 2008b)."*

> Speed of Skill Acquisition Matters Because Time is Limited

Humans are subject to many real-world constraints like limited lifespans and the need to learn a marketable skill quickly enough to get a job that affords basic life amenities. Additionally, things are always competing for our attention: whenever something feels hard or uninteresting, there are many other opportunities to do things that we might find easier and at least as interesting. Consequently, we tend to be pulled in other directions once we enter a range where developing further expertise in a domain becomes overwhelmingly arduous. We switch to other things that we (often, correctly) feel are a better use of our limited time.

As prominent psychologist Robert Sternberg recounts (2014):

> *"Most people who want to become experts – whether as violinists, skiers, physicists, or whatever – do not make it. They drop out along the way. They try and discover that, for whatever reason, it is not the way for them to go. I know, because as soon as I made the transition from high school to college, I found that I could not realistically compete as a cellist in the much stiffer competition I found in college compared with high school. Eventually I, like many others, decided that my time would be better spent elsewhere."*

There are compounding factors, too: when something becomes hard and we stop doing as well, we often like it less and lose interest/motivation, which makes it even more difficult. When we are young children in school, our teachers and/or parents might force us to continue investing effort into learning specific subjects like math even when we would prefer not to, but in high school and beyond, parents and teachers are less involved in monitoring how much and how effectively we practice. As we face greater responsibilities from life in general, we are met with so many incentives to get good grades that we might shy away from taking classes that require an outsized amount of effort. All of these factors converge to pull us towards an "off ramp" when mathematics gets hard for us.

So, for all practical purposes, it is completely untrue that everyone can learn every level of math and become a research mathematician – just like it is completely untrue that anyone can become an Olympic sprinter, a professional basketball player, a world-famous comedian, a Grammy-winning singer, etc. But at the same time, almost everybody can learn basic arithmetic – just like almost everybody can learn to run, shoot a free throw, tell a joke, or hum a tune. And with proper training, most people can learn some algebra – just like most people can run a 5k, shoot a three-pointer, amuse an audience, or sing a soothing tune.

| Why the "All Can Learn All" Myth Persists

There are at least two reasons why the "all can learn all" myth persists. First, the reality that "all can learn some, many can learn more, but few can learn all" can feel unfair and uncomfortable – especially in the context of mathematics, since the students who are best at math tend to be viewed as the smartest.

Second, people often overweight the importance of learning advanced math (and other technical subjects) to general success in life. In reality, lots of jobs, even those that are well-respected and lucrative, don't require advanced math. For instance, how many doctors, lawyers, members of Congress, and even university presidents can and do use calculus in their work? Few, if any. For professions like those, advanced math is not essential. (While it may be true that a larger-than-expected minority of people in such careers may have learned advanced math at some point, knowledge of advanced math itself is typically not one of the relevant factors contributing to one's ability to secure and maintain such a career.)

What's important is that everyone gain basic math skills, and people with quantitative talent who are interested in math and want to go into professions that use advanced math – not just aspiring mathematicians but also aspiring physicists, bioinformaticians, rocket scientists, machine learning engineers, etc. – don't take the "off ramp" too early and miss out on the opportunity to build a career around something they enjoy and are good at. We will address this idea more thoroughly in the next myth.

As a final part of debunking the current myth, it's important to realize that even professional educators and coaches who train students are susceptible to promote the myth that anyone can do anything with a bit of hard work.

Trainers, like parents, often don't want to tell their children that they're not gifted/talented enough in an area to build a career out of it – which is understandable because not only does it feel mean, but it may not even be true: new and unexpected gifts/talents can emerge as a child develops.

However, for trainers in particular, there are also other incentives at play that are important to be aware of. For a trainer, there is no upside to telling a student's parents that their child is not gifted/talented enough in an area to build a career out of it. It just makes parents and students really upset (even though professional trainers are probably correct more often than not) and, in the context of private training, often leads to loss of business. So, trainers are incentivized to avoid this treacherous territory and instead take one of the following false positions:

1. Gifts/talents are meaningless and anyone can do anything with a bit of hard work.

2. Gifts/talents are important, but there is no way to know whether a gift/talent might emerge as a young student develops, so there is no use basing any sort of decision around them.

In debunking the current myth, we demonstrated that the first position cannot be held rationally. However, believing in it can feel so empowering, and in turn so convincing, that it can be an effective position for a trainer to take to keep students and parents happy.

Likewise, in the second position, while it is true that even professional trainers don't know with absolute certainty how their students (especially young students) will develop over time, they do generally have *some* or even *a lot* of information about whether a student has a gift/talent – and if not, then how likely it is that the gift/talent might emerge later in the child's development. Leveraging that information can be a critical part of helping a child enter an area where they have both the gifts/talents and the level of interest to eventually build a career that they enjoy.

| Struggle Does Not Imply Inability

Myth: *If you do poorly in a math class, it means you are incapable of learning that level of math.*

Reality: If you do poorly in a math class, it doesn't necessarily mean that you are incapable of learning that level of math. There are a number of reasons that could be the root cause of your struggle. While it's true that everyone's mathematical potential has a limit, in practice the ceilings we hit rarely represent our true "abstraction ceiling" as described by Hofstadter. All sorts of factors can artificially lower our ceilings, such as missing foundations, ineffective practice habits, inability or unwillingness to engage in additional practice, or lack of motivation.

> Struggle Can Be Caused by Missing Foundations

When people age, they accumulate biological damage that eventually reaches a tipping point and leads to a cascade of catastrophic health issues. The same thing happens to students learning mathematics.

Students accumulate weaknesses and knowledge gaps as they progress through math – even a grade of B+ or A- means that there are things in the course that the student never completely grasped, much less mastered. Additionally, gaps can be created if a student takes a course that is not comprehensive and does not cover some topics that are assumed to be prior knowledge in higher-level courses. Once a student has accumulated a critical number of gaps (and by the way, a gap begets more gaps), then the student is doomed to struggle unless proper remediation is enacted to fill in those gaps.

Math Academy automatically takes steps to detect and remediate each individual student's gaps in knowledge. However, remediation is extremely difficult to accomplish outside the context of an adaptive, automated learning system. It rarely happens in the classroom – teachers just don't have the bandwidth to spend enough time with each student to figure out exactly which pieces of foundational knowledge are missing. And while remediation can often be performed by a skilled tutor, it generally requires many tutoring sessions over a long period of time, continuing indefinitely into the future to prevent new gaps from forming, which makes it prohibitively expensive for most families.

Students usually stop taking math classes once they amass a critical number of knowledge gaps. The usual sequence of events starts with students trying to imitate procedures cookbook-style, without really understanding what's going on, because they can't intuitively grasp any of the new material that they're being taught. Soon after that, they find themselves unable to solve any problems that involve critical thinking or many steps.

It's similar to how professional athletes usually retire not because they're too old, but because they've accumulated too many injuries. As Indiana Jones once put it: "it's not the years, it's the mileage." Or as math writer/cartoonist Ben Orlin humorously described, it's the "law of the broken futon": a single missing part can, over time, warp an entire futon and render it unusable.

Students will almost assuredly accumulate these deficits in traditional classrooms. It's only the most gifted and motivated students who are able and willing to identify and "self-repair" their gaps on their own.

- In traditional classrooms, students often get stuck on foundational topics but are required to complete homework on more advanced topics, leading them to "scrape by" without really understanding the subject matter.

- Students also do not review material learned in previous years, and often do not even review material from the course that they're in unless they are preparing for a test. This leads them to quickly forget what they've learned, requiring re-learning scratch if and when those topics show up again in the future.

- Often, traditional courses are not even comprehensive! It's not uncommon for instructors to run out of time before the end of the year and skip sections of the textbook.

Math Academy, however, remedies these issues so that students never end up with knowledge holes.

- By practicing mastery learning, we ensure that students are not required to complete more advanced topics until they have demonstrated proficiency on the prerequisites. This way, students are always ready to properly absorb the new concepts being learned.

- We also engage in spaced repetition, a systematic way of reviewing previously learned material at appropriate intervals to retain knowledge. This way, students don't forget what they've learned.

- Our courses are fully comprehensive. When developing our courses, we look at all the major textbooks to ensure that we cover all of their combined content. Any topic that you could reasonably expect to find in some version of a course, you'll find in our system.

Even if students enter Math Academy with knowledge holes, we automatically take steps to detect and repair them. Our diagnostic exams not only assess course content, but also lower-grade foundations, so that we can identify and fill in every individual student's gaps in foundational knowledge.

> Struggle Can Be Caused by Ineffective Practice

As we explained when summarizing the science of learning, effective learning feels like a workout with a personal trainer. It should center around deliberate practice, a type of active learning in which individualized training activities are specially chosen to improve specific aspects of performance through repetition and successive refinement.

We will cover active learning and deliberate practice in more depth in later chapters, but below are some key points:

- Effective learning is active, not passive. It is not effective to attempt to learn by passively watching videos, attending lectures, reading books, or re-reading notes.

- Deliberate practice requires repeatedly practicing skills that are beyond one's repertoire. However, this tends to be more effortful and less enjoyable, which can mislead non-experts to practice within their level of comfort.

- Classroom activities that are enjoyable, collaborative, and non-repetitive (such as group discussions and freeform/unstructured project-based or discovery learning) can sometimes be useful for increasing student motivation and softening the discomfort associated with deliberate practice – but they are only supplements, not substitutes, for deliberate practice.

- Deliberate practice must be a part of a consistent routine. The power of deliberate practice comes from compounding of incremental improvements over a longer period of time. It is not a "quick fix" like cramming before an exam.

On Math Academy, students spend the entirety of their time engaged in deliberate practice by solving problems (and receiving feedback) on new topics and topics most in need of review. We intersperse active problem-solving with instruction so that students receive minimum effective doses of information right before they use it to actively solve problems and receive feedback.

> Struggle Can Be Caused by Insufficient Practice

Struggle can be caused by needing more practice than other students (or, equivalently, the pace of the class might be too fast). This is not necessarily a catastrophic issue in itself because it can usually be remedied by engaging in further practice. However, it can cause problems if coupled with other factors such as the following:

- The instructional material is not highly scaffolded.

- Few practice problems are available.

- Exam problems are substantially different from homework problems.

- The additional practice required exceeds the amount of effort that you are willing to put forth to learn the material.

Math Academy remedies all but the last of these issues:

- Our content is about 10x more finely scaffolded than what you'd find elsewhere.

- If a student struggles during a task, we give more questions – that is, more chances to learn and demonstrate their learning.

- We have quick, frequent quizzes where questions are similar to (but not the same as) those learned during lessons.

- We even tailor the speed of the spaced repetition process to every separate student on every separate topic to ensure that students are getting just enough review to retain information over the long term.

> Struggle Can Be Caused by Lack of Motivation

Properly motivated students are usually driven by one or more of the following factors:

- *They are intrinsically interested in the material.* Some students truly love math and see beauty in the way various mathematical ideas fit together and give rise to new perspectives.

- *The material is highly relevant to their future goals.* For instance, an aspiring rocket scientist might not love math but might be motivated to learn it because of how useful it is for getting rockets into space. Likewise, an aspiring doctor might not love math but might be required to evidence a baseline level of mathematical knowledge when applying to medical school. Even students who do not have specific future goals might feel strongly about keeping potential career doors open which would otherwise be shut by not learning enough math.

- *They enjoy competing in mathematical exams and science fairs.* Some students have neutral feelings about math, but find that they are good at it, and that they enjoy learning more advanced mathematics to provide a competitive edge in exams and science fairs.

- *Their parents have motivated them with a meaningful extrinsic reward.* Sometimes, a student may not fall into any categories above, but their parents (often rightfully so) want them to fully take advantage of any opportunities to learn math while they are still in school. For some students, this may mean learning the basic math they need to get by in life after school; for other students, this may mean learning more advanced math to open a wide variety of career doors. If a student is highly interested in other activities like reading novels, playing video games, or even something as simple as going out for dessert, offering them extrinsic rewards in return for meeting checkpoints in their math learning can often provide sufficient motivation to keep them from "checking out" during learning.

If a student is not driven by any of the motivational factors above, they may "check out" or otherwise struggle due to a lack of interest in learning the material.

| Analogy to Lifespans

The key takeaway from this section is that your mathematical potential has a limit, but it's likely higher than you think. If this idea still feels unclear, then it may help to draw an analogy to human lifespans.

Your lifestyle will affect the length of your life, but even if you live perfectly healthily, there is no guarantee that you will become a supercentenarian (110+ years old). A tiny fraction of people will live that long, but probably not you, even if you do everything right. However, it is still true that you can vastly extend your lifespan if you live healthily, as compared to if you live unhealthily.

The same is true in athletics. Even if you practice effectively for longer than anyone else, there is no guarantee that you will become a hall-of-fame athlete. However, it is still true that you will become vastly more skilled in your chosen sport than if you practiced ineffectively or didn't put in as much time. Quite likely, you will become more skilled than you or anyone else thought was possible.

Mathematics is no different. Even if you devote your life to effective study, there is no guarantee that you will become a world-class mathematician. But by putting a serious effort into effective study, you will learn far more math and open far more career doors than you would otherwise.

| Student Bite Size vs Curriculum Portion Size

Thinking deeply about the role of instruction in supporting learners, it's possible to arrive at the following misconception.

> The Misconception: If Instruction is Done Perfectly, Won't All Students Learn at the Same Rate?

Higher math is heavily g-loaded, which creates a cognitive barrier for many students. The goal of instructional scaffolding, guidance, and review is to help boost students over that barrier.

But if the purpose of scaffolding, guidance, and review is to help students overcome cognitive blockers, then wouldn't a theoretical learning environment with infinite scaffolding, guidance, and review completely factor out cognitive differences, causing students to learn at the same rate?

Sure, the speed at which students learn (and remember what they've learned) varies from student to student. It has been shown that some students learn faster and remember longer, while other students learn slower and forget more quickly (e.g., Kyllonen & Tirre, 1988; Zerr et al., 2018; McDermott & Zerr, 2019).

But perhaps these studies are simply reflective of unfavorable learning conditions, and people would learn at the same rate in an optimally favorable learning environment?

> The Resolution: Under Favorable Learning Conditions, Student Bite Size Equals Curriculum Portion Size

Continuing to think deeply about this thought experiment, one will eventually realize that infinite scaffolding, guidance, and review is *not* synonymous with optimally favorable learning conditions.

Sure, students will eat meals of information at similar bite rates when each spoonful fed to them is infinitesimally small. However, "eating at the same rate" would be a ceiling effect.

Faster learners would be capable of learning faster, but the curriculum would be too granular relative to their generalization ability and/or provide too much review relative to their forgetting rate, thereby creating a ceiling effect that prevents fast learners from learning at their top speed.

A maximally-favorable learning environment would require that the curriculum's granularity is equal to the student's bite size and the rate of review is equal to the student's rate of forgetting.

The amended metaphor: *Students eat meals of information at similar bite rates when each spoonful fed to them is sized appropriately relative to the size of their mouth.*

This type of learning environment would maximize each student's individual potential, free of ceiling effects. Critically, however, students would not progress through it at the same rate: equal bite rates does *not* imply equal rates of food volume intake.

> Consistency with Observations

This framing of favorable learning conditions ("student bite size equals curriculum portion size") is consistent with the phenomenon that math becomes hard for different students at different levels. The following factors affect students differentially as they move up the levels of math:

- *Combinatorial explosion in the problem space* – lowers the "bite size" more for students with lower generalization ability, or, equivalently, reduces the perceived granularity of the curriculum. (This may be a contributing factor in cases when, e.g., students do fine in math but struggle in physics.)

- *Large body of knowledge to maintain* – increases the amount of review more for students with higher forgetting rates. Also reduces effective "bite size" since an increasing portion of each bite will consist of reviewing fuzzy prerequisite material.

It is also consistent with the concept of soft and hard ceilings on the highest level of math that one can reach:

- Say we have a student with low generalization ability and high forgetting rate. Then a favorable curriculum takes more time to work through (as compared to a favorable curriculum for an average student) due to increased granularity and review, and that multiplier increases as they go up the levels of math.

- At some point "it requires lots of practice to learn" becomes synonymous with "can't learn" – first in a soft sense of "the benefits of engaging in this much practice do not outweigh the opportunity costs of neglecting to develop my skills in other domains that I find easier," and then in a hard sense of "the amount of practice required exceeds the sum of waking hours over the remainder of my life."

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Grospietsch, F., & Lins, I. (2021, July). Review on the prevalence and persistence of neuromyths in education–Where we stand and what Is still needed. In *Frontiers in Education* (Vol. 6, p. 665752). Frontiers Media SA.

  Betts, K., Miller, M., Tokuhama-Espinosa, T., Shewokis, P. A., Anderson, A., Borja, C., ... & Dekker, S. (2019). International Report: Neuromyths and Evidence-Based Practices in Higher Education. *Online Learning Consortium.*

  Importance: *One of the most widespread – and most widely debunked – neuromyths is that people learn better when they receive information in their preferred "learning style."*

- Conway, A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. (2007). Variation in Working Memory: An Introduction. In Conway, A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. (Eds.), *Variation in working memory* (pp.3-17). Oxford University Press.

  Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748-751.

  Engström, M., Landtblom, A. M., & Karlsson, T. (2013). Brain and effort: brain activation and effort-related working memory in healthy participants and patients with working memory deficits. *Frontiers in human neuroscience, 7*, 140.

  Importance: *One aspect of the brain that has been widely documented not only to vary between people, but also to affect people's general cognitive performance, is working memory capacity (WMC). These differences in working memory capacity have been characterized not only at a psychological level, but also at the physical level of brain activity measures.*

- Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of* *the American Academy of Audiology, 23*(08), 577-589.

Importance: *People with higher WMC will generally perceive a given task to be easier than people with lower WMC.*

- McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General, 143*(2), 668.

  McDaniel, M. A., Cahill, M. J., Frey, R. F., Rauch, M., Doele, J., Ruvolo, D., & Daschbach, M. M. (2018). Individual differences in learning exemplars versus abstracting rules: Associations with exam performance in college science. *Journal of Applied Research in Memory and Cognition, 7*(2), 241-251.

  Importance: *It has also been shown that high WMC facilitates abstraction, that is, seeing "the forest for the trees" by learning underlying rules as opposed to memorizing example-specific details. Individual differences in abstraction ability have been shown to impact educational outcomes.*

- McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General, 143*(2), 668.

  Reber, P. J., & Kotovsky, K. (1997). Implicit learning in problem solving: The role of working memory capacity. *Journal of Experimental Psychology: General, 126*(2), 178.

  Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological science, 21*(7), 914-919.

  Kulasegaram, K. M., Grierson, L. E., & Norman, G. R. (2013). The roles of deliberate practice and innate ability in developing expertise: evidence and implications. *Medical education, 47*(10), 979-989.

  Swanson, H. L., & Siegel, L. (2011). Learning disabilities as a working memory deficit. *Experimental Psychology, 49*(1), 5-28.

  Importance: *WMC has also been shown to impact speed of learning, that is, the rate at which one's ability to* *perform a task improves over the course of exposure, instruction, and practice on the task. This effect extends beyond laboratory settings into real-life contexts of academics and professional expertise.*

- Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review, 27*(4), 617-633.

  Reber, P. J., & Kotovsky, K. (1997). Implicit learning in problem solving: The role of working memory capacity. *Journal of Experimental Psychology: General, 126*(2), 178.

  Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & cognition, 33*(2), 213-220.

  Importance: *While it is possible to train and improve on tasks that are typically used to measure WMC, evidence is currently lacking that these task-specific performance improvements actually represent an increase in WMC that can be transferred to more general contexts. However, the impact of WMC on task performance is diminished after the task is learned to a sufficient level of performance, and is minimal for tasks that have been learned to the level of automatic processing.*

- Hofstadter, D., & Carter, K. (2012). Some Reflections on Mathematics from a Mathematical Non-mathematician. *Mathematics in School, 41*(5), 2-4.

  Importance: *The further one goes in math, the more energy it requires to learn the next level up. This leads to an "abstraction ceiling" – not a "hard" threshold, a level at which one is suddenly incapable of learning math, but rather a "soft" threshold, a level at which the amount of time and effort required to learn math begins to skyrocket until learning more advanced math is effectively no longer a productive use of one's time.*

- Simonton, D. K. (2007). Talent and expertise: The empirical evidence for genetic endowment. *High Ability Studies, 18*(1), 83-84.

  Simonton, D. K. (2013). If innate talent doesn't exist, where do the data disappear?. *The complexity of greatness: Beyond talent or practice*, 17-26.

  Importance: *It has been well established that expert performance is contingent upon favorable sociocultural conditions, family and education circumstances, and massive amounts of deliberate practice. However, many critical traits underlying expert performance have been shown to have a significant genetic component. While the development of expert performance depends on favorable environmental conditions and massive amounts* *of deliberate practice, the speed of development can be accelerated (or decelerated) by genetic factors. Lack of talent does not necessarily mean "you can't do this," but it does mean that someone lacking talent will need to work much harder, possibly to an infeasible extent, as compared to someone with talent.*

Chapter 8. Myths & Realities about Effective Practice

Summary: *The most effective learning techniques require substantial cognitive effort from students and typically do not emulate what experts do in the professional workplace. Direct instruction is necessary to maximize student learning, whereas unguided instruction and group projects are typically very inefficient. Effortful processes like testing, repetition, and computation are essential parts of effective learning, and competition is often helpful.*

Effective Practice Does Not Emulate the Professional Workplace

Myth: *Effective methods of practice emulate what experts do in the professional workplace.*

Reality: A well-known phenomenon in cognitive psychology is that instructional techniques that promote the most learning in experts, promote the least learning in beginners, and vice versa. This is called the expertise reversal effect (first introduced by Sweller et al., 2003). As Kirschner & Hendrick summarize (2024, pp.67):

> *"As the novice is not a miniature expert, it's important to realize that what may work very well for an expert (e.g. discovery learning, problem-based learning [in the sense of working in groups to solve an open-ended problem], inquiry learning) usually doesn't work well or is even harmful and counterproductive for the novice (and vice versa)."*

Additionally, in the professional workplace, employees engage in activities that maximize group output, which is totally different – and in some ways, opposite – from maximizing individual learning.

| Direct Instruction is Needed

> Definition and Importance

It is true that many highly skilled professionals spend a lot of time solving open-ended problems, and in the process, discovering new knowledge as opposed to obtaining it through **direct instruction**. However, this does not mean that beginners should do the same. The expertise reversal effect suggests the opposite – that beginners (i.e. students) learn most effectively through direct instruction.

Direct instruction is intuitively obvious. If a coach is trying to get a student to become a great chess player or pianist, they don't tell the student "go play around and come back with something insightful." Rather, the coach explicitly demonstrates a skill and then provides corrective feedback to the student as they practice the skill. As Kirschner & Hendrick describe (2024, pp.68):

> *"While an expert can be given a problem to be solved after having been taught a certain technique or principle, a novice should be given a more structured approach to using that principle for solving the same problem, for example in the form of a worked example."*

Indeed, this is backed up by decades of research. As prominent psychologists Richard Clark, Paul Kirschner, and John Sweller summarize (2012):

> *"Decades of research clearly demonstrate that for novices (comprising virtually all students), direct, explicit instruction is more effective and more efficient than partial guidance. So, when teaching new content and skills to novices, teachers are more effective when they provide explicit guidance accompanied by practice and feedback, not when they require students to discover many aspects of what they must learn.*
>
> *...*
>
> *We also have a good deal more experimental evidence [since the 1960s] as to what constitutes effective instruction: controlled experiments almost uniformly indicate that when dealing with novel information, learners should be explicitly shown all relevant information, including what to do and how to do it. We wonder why many teacher educators who are committed to scholarship and research ignore the evidence and continue to encourage minimal guidance when they train new teachers.*
>
> *After a half century of advocacy associated with instruction using minimal guidance, it appears that there is no body of sound research that supports using the technique with anyone other than the most expert students. Evidence from controlled, experimental (a.k.a. "gold standard") studies almost uniformly supports full and explicit instructional guidance rather than partial or minimal guidance for novice to intermediate learners. These findings and their associated theories suggest teachers should provide their students with clear, explicit instruction rather than merely assisting students in attempting to discover knowledge themselves."*

> Unguided Instruction has a History of Pseudoscience

Clark, Kirschner, & Sweller (2012) explain that unguided instruction persists by cloaking itself in a different disguise each time it is debunked:

> *"Richard Mayer (a cognitive scientist at the University of California, Santa Barbara) examined evidence from studies conducted from 1950 to the late 1980s comparing pure discovery learning (defined as unguided, problem-based instruction) with guided forms of instruction. He suggested that in each decade since the mid-1950s, after empirical studies provided solid evidence that the then-popular unguided approach did not work, a similar approach soon popped up under a different name with the cycle repeating itself.*
>
> *Each new set of advocates for unguided approaches seemed unaware of, or uninterested in, previous evidence that unguided approaches had not been validated. This pattern produced discovery learning, which gave way to experiential learning, which gave way to problem-based and inquiry learning, which has recently given way to constructivist instructional techniques."*

As they elaborate elsewhere (Kirschner, Sweller, & Clark, 2006), these unguided approaches are often based on modeling the activities of professionals:

> *"Examples of applications of these differently named but essentially pedagogically equivalent approaches include science instruction in which students are placed in inquiry learning contexts and asked to discover the fundamental and well-known principles of science by modeling the investigatory activities of professional researchers (Van Joolingen, de Jong, Lazonder, Savelsbergh, & Manlove, 2005)."*

They also explain that the current formulation, constructivist instruction, uses scientific camouflage but is not actually scientific itself:

> *"Turning again to Mayer's review of the literature, many educators confuse 'constructivism,' which is a theory of how one learns and sees the world, with a prescription for how to teach.*
>
> *In the field of cognitive science, constructivism is a widely accepted theory of learning; it claims that learners must construct mental representations of the world by engaging in active cognitive processing. Many educators (especially teacher education professors in colleges of education) have latched on to this notion of students having to 'construct' their own knowledge, and have assumed that the best way to promote such construction is to have students try to discover new knowledge or solve new problems without explicit guidance from the teacher.*
>
> *Unfortunately, this assumption is both widespread and incorrect. Mayer calls it the 'constructivist teaching fallacy.' … Learning requires the construction of knowledge. Withholding information from students does not facilitate the construction of knowledge."*

In his critical review, Mayer (2004) had plenty more to say:

> *"The research in this brief review shows that the formula constructivism = hands-on activity is a formula for educational disaster.*

> *...*
> *Like some zombie that keeps returning from its grave, pure discovery continues to have its advocates.*
> *...*
> *Pure discovery did not work in the 1960s, it did not work in the 1970s, and it did not work in the 1980s, so after these three strikes, there is little reason to believe that pure discovery will somehow work today.*
> *...*
> *[T]he issue addressed in this article is not whether constructivism is a good idea for education. but rather whether the educational implications attributed to constructivism are good ideas. In the case of discovery methods, the implications attributed to constructivism are not good ideas.*
> *...*
> *The debate about discovery has been replayed many times in education, but each time, the research evidence has favored a guided approach to learning."*

These interpretations are echoed throughout the literature. As other prominent psychologists John Anderson, Lynne Reder, and Herbert Simon state (1998):

> *"A consensus exists within cognitive psychology that people do not record experience passively but interpret new information with the help of prior knowledge and experience. ... However, denying that information is recorded passively does not imply that students must discover their knowledge by themselves, without explicit instruction, as claimed by radical constructivists."*
>
> *"Radical constructivism emphasizes discovery learning, learning in complex situations, and learning in social contexts, while strongly distrusting systematic evaluation of educational outcomes. ... [C]ertain of its devotees exhibit an antiscience bias that, should it prevail, would devote any hope for progress in education.*
> *...*
> *Little positive evidence exists for discovery learning and it is often inferior. Discovery learning, even successful in enabling the acquisition of the desired construct, may require a great deal of valuable time that could have been spent practicing the construct (which is an active process, too) if it had been learned from instruction. Because most learning only takes place after the construct has been discovered, when the search is lengthy or unsuccessful, motivation commonly lags. As D. P. Ausubel wrote in 1968, summarizing the findings from the research on discovery learning:*
>
> *'Actual examination of the research literature allegedly supportive of learning by discovery reveals that valid evidence of this nature is virtually nonexistent. It appears that the various enthusiasts of the discovery method have been supporting each other research-wise by taking in each other's laundry, so to speak, that is, by citing each other's opinions and assertions as evidence and by generalizing wildly from equivocal and even negative findings.'"*

> Unguided Instruction is Logically and Scientifically Inconsistent

Anderson, Reder, & Simon (1998) also explain that opponents of direct instruction are, ultimately, opponents of extensive practice – a position that is clearly problematic:

> *"Some argue that direct instruction leads to 'routinization' of knowledge and drives out understanding:*

> *'The more explicit I am about the behavior I wish my students to display, the more likely it is that they will display the behavior without recourse to the understanding which the behavior is meant to indicate; that is, the more likely they will take the form for the substance.'*
>
> *An extension of this argument is that excessive practice will also drive out understanding. This criticism of practice (called 'drill and kill,' as if this phrase constituted empirical evaluation) is prominent in constructivist writings. Nothing flies more in the face of the last 20 years of research than the assertion that practice is bad.*
>
> *All evidence, from the laboratory and from extensive case studies of professionals, indicates that real competence only comes with extensive practice. By denying the critical role of practice, one is denying children the very thing they need to achieve competence. ... the grain of truth in the drill-and-kill criticisms [is that]: Students need to be engaged when they are studying."*

Likewise, there are critical issues with the idea of learning primarily from complex situations:

> *"First, a learner who is having difficulty with many of the components can easily be overwhelmed by the processing demands of the complex task. Second, to the extent that many components are well mastered, the student will waste a great deal of time repeating those mastered components to get an opportunity to practice the few components that need additional practice.*
>
> *A large body of research in psychology shows that part training is often more effective when the part component is independent, or nearly so, of the larger task. ... Practicing one's skills periodically in full context is important to motivation and to learning to practice, but not a reason to make this the principal mechanism of learning."*

Along these lines, Clark, Kirschner, & Sweller (2012) further explain that, in addition to being supported by a mountain of experimental evidence, the superiority of direct instruction follows intuitively from modern understandings of working and long-term memory:

> *"These two facts – that working memory is very limited when dealing with novel information, but that it is not limited when dealing with organized information stored in long-term memory – explain why partially or minimally guided instruction typically is ineffective for novices, but can be effective for experts. When given a problem to solve, novices' only resource is their very constrained working memory. But experts have both their working memory and all the relevant knowledge and skill stored in long-term memory."*

As Sweller, Clark, and Kirschner (2010) elaborate elsewhere:

> *"Recent 'reform' curricula both ignore the absence of supporting data and completely misunderstand the role of problem solving in cognition. If, the argument goes, we are not really teaching people mathematics but rather are teaching them some form of general problem solving, then mathematical content can be reduced in importance. According to this argument, we can teach students how to solve problems in general, and that will make them good mathematicians able to discover novel solutions irrespective of the content.*
>
> *We believe this argument ignores all the empirical evidence about mathematics learning. Although some mathematicians, in the absence of adequate instruction, may have learned to solve*

> *mathematics problems by discovering solutions without explicit guidance, this approach was never the most effective or efficient way to learn mathematics.*
> *...*
> *[L]ong-term memory, a critical component of human cognitive architecture, is not used to store random, isolated facts but rather to store huge complexes of closely integrated information that results in problem-solving skill. That skill is knowledge domain-specific, not domain-general. An experienced problem solver in any domain has constructed and stored huge numbers of schemas in long-term memory that allow problems in that domain to be categorized according to their solution moves.*
>
> *In short, the research suggests that we can teach aspiring mathematicians to be effective problem solvers only by providing them with a large store of domain-specific schemas. Mathematical problem-solving skill is acquired through a large number of specific mathematical problem-solving strategies relevant to particular problems. There are no separate, general problem-solving strategies that can be learned.*
> *...*
> *Whereas a lack of empirical evidence supporting the teaching of general problem-solving strategies in mathematics is telling, there is ample empirical evidence of the validity of the worked-example effect. A large number of randomized controlled experiments demonstrate this effect (e.g., Schwonke et al., 2009; Sweller & Cooper, 1985). For novice mathematics learners, the evidence is overwhelming that studying worked examples rather than solving the equivalent problems facilitates learning.*
>
> *Studying worked examples is a form of direct, explicit instruction that is vital in all curriculum areas, especially areas that many students find difficult and that are critical to modern societies. Mathematics is such a discipline. Minimal instructional guidance in mathematics leads to minimal learning (Kirschner, Sweller, & Clark, 2006)."*

> Unguided Instruction Leads to Major Issues in Practice

Clark, Kirschner, & Sweller (2012) also describe what *actually* happens in classrooms that do not use direct instruction:

> *"In real classrooms, several problems occur when different kinds of minimally guided instruction are used.*
>
> *First, often only the brightest and most well-prepared students make the discovery.*
>
> *Second, many students, as noted above, simply become frustrated. Some may disengage, others may copy whatever the brightest students are doing – either way, they are not actually discovering anything.*
>
> *Third, some students believe they have discovered the correct information or solution, but they are mistaken and so they learn a misconception that can interfere with later learning and problem solving. Even after being shown the right answer, a student is likely to recall his or her discovery – not the correction.*
>
> *Fourth, even in the unlikely event that a problem or project is devised that all students succeed in completing, minimally guided instruction is much less efficient than explicit guidance. What can*

> *be taught directly in a 25-minute demonstration and discussion, followed by 15 minutes of independent practice with corrective feedback by a teacher, may take several class periods to learn via minimally guided projects and/or problem solving."*

These issues are also backed up by numerous studies:

> *"Hardiman, Pollatsek, and Weil (1986) and Brown and Campione (1994) noted that when students learn science in classrooms with pure-discovery methods and minimal feedback, they often become lost and frustrated, and their confusion can lead to misconceptions. Others (e.g., Carlson, Lundy, & Schneider, 1992; Schauble, 1990) found that because false starts are common in such learning situations, unguided discovery is most often inefficient."*

To emphasize, these issues are so problematic that they can actually result in *negative* educational progress:

> *"Not only is unguided instruction normally less effective; there is also evidence that it may have negative results when students acquire misconceptions or incomplete or disorganized knowledge."*

But despite these issues, the students who learn least in unguided settings still tend to prefer it because it feels less effortful:

> *"...[W]hen learners are asked to select between a more-guided or less-guided version of the same course, less-skilled learners who choose the less-guided approach tend to like it even though they learn less from it. It appears that guided instruction helps less-skilled learners by providing task-specific learning strategies. However, these strategies require learners to engage in explicit, attention-driven effort and so tend not to be liked, even though they are helpful to learning."*

Of course, experienced, effective teachers are well acquainted with these issues and (rightfully so) brush off any recommendations to use unguided learning:

> *"...[M]any experienced educators are reluctant to implement – because they require learners to engage in cognitive activities that are highly unlikely to result in effective learning. As a consequence, the most effective teachers may either ignore the recommendations or, at best, pay lip service to them (e.g., Aulls, 2002)."*

This sentiment is sharply echoed by Mayer (2004):

> *"...[T]he contribution of psychology is to help move educational reform efforts from the fuzzy and unproductive world of educational ideology – which sometimes hides under the banner of various versions of constructivism – to the sharp and productive world of theory-based research on how people learn."*

To top it all off, as Kirschner, Sweller, & Clark (2006) summarize, even on the rare occasion that a student does manage to learn in an unguided setting, their learning tends to be shallower than it would have been in a strongly guided setting:

> *"Moreno (2004) concluded that there is a growing body of research showing that students learn more deeply from strongly guided learning than from discovery. Similar conclusions were reported by Chall (2000), McKeough, Lupart, and Marini (1995), Schauble (1990), and Singley and Anderson (1989).*
>
> *Klahr and Nigam (2004), in a very important study, not only tested whether science learners learned more via a discovery versus direct instruction route but also, once learning had occurred, whether the quality of learning differed. Specifically, they tested whether those who had learned through discovery were better able to transfer their learning to new contexts. The findings were unambiguous. Direct instruction involving considerable guidance, including examples, resulted in vastly more learning than discovery. Those relatively few students who learned via discovery showed no signs of superior quality of learning."*

As Kirschner & Hendrick summarize (2024, pp.76):

> *"...[I]f you want your students to learn to solve problems, they first need both the declarative and procedural knowledge within the subject area of the problem in question. This is also true if you want to teach them to communicate, discuss, write, or whatever twenty-first century skill people are talking about. You can't communicate about something, write about something, discuss or argue about something, etc., without first knowing about that something and then also knowing the rules (i.e. the procedures) for doing it."*

| Many Hands Make Light Work... and Light Learning

Professionals often work in groups because it gives them an economic advantage. Real-world projects are often extremely complex and require a massive amount of highly skilled labor across a wide variety of disciplines. The amount of work necessary to bring the project to fruition might exceed what one person can put forth over their entire lifetime, and the number of skill domains covered by the work might be more than any one person can hope to master in a single lifetime. This problem is solved by constructing a team where each member is highly skilled in one or more of the relevant domains, and there are enough members to complete the workload in a feasible amount of time.

The goal of division of labor in the professional workplace is to maximize the output of a team. On the surface, it might seem like a tempting strategy to apply in the classroom: won't maximizing the output of a classroom effectively maximize the learning of individual students? But the answer is a resounding no. Division of labor is division of learning, which means that it actually *minimizes* the learning of individual students.

To *maximize* the learning of individual students, it is necessary to actively engage every individual student on every single piece of material to be learned. Division of labor is the complete opposite of that, since each student actively learns only the material that corresponds to their individual responsibility in the division of labor. The rest of the project, they observe only passively, if at all. At best, each student only learns a tiny fraction of the material. At worst, one student ends up doing all the work while the rest of the group learns nothing.

As Anderson, Reder, & Simon (1998) summarize:

> *"Some of the learning contexts recommended in radical constructivist writings involve tasks that can be solved by a single problem solver, but the movement more and more is to convert these to group learning situations. ... While a person must learn to deal with the social aspects of jobs, all skills required for these jobs do not need to be trained in a social context. ... Training independent parts of a task separately is preferable, because fewer cognitive resources will be required for performance, thereby reserving adequate capacity for learning.*
> *...*
> *A review by the National Research Council (NRC) Committee on Techniques for the Enhancement of Human Performance noted that ... relatively few studies 'have successfully demonstrated advantages for cooperative versus individual learning,' and that a number of detrimental effects arising from cooperative learning have been identified – the 'free rider,' the 'sucker' [reducing effort to avoid being taken advantage of by free riders], the 'status differential' [low-ability team members lose social status and reduce effort] and 'ganging up' [directing group effort towards circumventing the intended efforts of the task] effects [Salomon & Globerson, 1989].*
>
> *The NRC review of cooperative learning notes a substantial number of reports of no-differences, but, unfortunately, a huge number of practitioner-oriented articles about cooperative learning gloss over difficulties with the approach and treat it as an academic panacea. It is applied too liberally without the requisite structuring or scripting to make it effective. ... A reported practice among some students is to divide the labor across classes so that one member of a group does all of the work for a project in one class, while another carries the burden for a different class. Clearly these are not the intended outcomes of cooperative learning but will occur if thoughtful implementation and scripting of the learning situation are not evident."*

Granted, fun, collaborative group activities can sometimes be useful for increasing student motivation and softening the discomfort associated with intense, individualized deliberate practice. However, they do not directly move the needle on student performance – rather, they "grease the wheels" and reduce psychological friction during the process of deliberate practice. Performance improvements come directly from deliberate practice.

Effective Practice Requires Effort

| There is No Such Thing as Effortless Learning

Myth: *There exist effective methods of practice that require low or no effort.*

Reality: Talent development takes work – not just a little work, but a lot of work. There is absolutely no confusion about this in the talent development community. Can you imagine asking an athletic coach to help you become a star player using training methods that don't tire you out and make you sweat?

A common theme in the science of learning is that effective learning feels like a workout with a personal trainer. It should center around deliberate practice, a type of active learning in which individualized training activities are specially chosen to improve specific aspects of performance through repetition and successive refinement. These practice activities are done entirely for the purpose of pushing one's limits and improving performance; consequently, they tend to be more effortful and less enjoyable.

Unfortunately, many types of training methods are ineffective, but require little effort, and can therefore seem attractive to even the most well-intentioned, hardworking students because they create an **illusion of competence** (e.g. Karpicke, Butler, & Roediger, 2009; also called an **illusion of comprehension** in earlier works reviewed in e.g. Bjork & Bjork, 2023). Examples include looking at notes, rereading course materials, and highlighting. In a review of scientific studies on various methods of practice, low-effort methods like these were found to have the lowest utility in terms of promoting learning, retention, and application of knowledge (Dunlosky et al., 2013):

Utility	Techniques
High	Practice testing Distributed Practice
Moderate	Interleaved Practice Elaborative interrogation Self-explanation
Low	Summarization Highlighting

	The keyword mnemonic Imagery use for text learning Rereading

On the other hand, the two highest-utility methods – practice testing and distributed practice – are particularly effortful. The benefits of practice testing come from effortful retrieval of information, and the benefits of distributed practice come from spreading out practice sessions to allow for some amount of forgetting to set in between them (which thereby increases the level of effort required during subsequent practice sessions). As Brown, Roediger, & McDaniel (2014, pp.48) summarize:

> *"Spacing out your practice feels less productive for the very reason that some forgetting has set in and you've got to work harder to recall the concepts. It doesn't feel like you're on top of it. What you don't sense in the moment is that this added effort is making the learning stronger."*

And as Kang (2016) describes, these two high-effort methods are *even more* effective (and, of course, *even more* effortful) when combined:

> *"Testing or spaced practice, each on its own, confers considerable advantages for learning. But, even better, the two strategies can be combined to amplify the benefits: Reviewing previously studied material can be accomplished through testing (often followed by corrective feedback) instead of rereading."*

To be clear, this is not to say that passively reading or re-reading material should be completely avoided. It is useful to familiarize oneself with instructional material before engaging in effortful practice, and it is also useful to revisit that material if one runs into issues while attempting to carry out the effortful practice. However, it is not until effortful practice that true learning actually occurs.

Familiarizing oneself with instructional material is similar to warming up before a workout: the warmup does not actually lead to improvements in strength or endurance, but it does help maximize performance and avoid injury during the workout. No matter what skill is being trained, improving performance is always an effortful process.

As Qadir & Imran (2018) summarize, learning is all about creating **desirable difficulties**:

> *"While we intuitively dislike difficulties and thus try to avoid them, many difficulties (but not all) have a positive effect on learning. The well-known cognitive psychologist Bob Bjork coined the term '**desirable difficulties**' for such difficulties that have a positive effect on learning.*
>
> *...*

> *Learning – i.e., actual learning that requires the ability to remember and transfer concepts in the long term – requires effort ... Research has shown that while retrieval is harder with spaced learning and interleaving, resulting in the feeling that the learning is less accomplished, the resulting learning is actually deeper and will lead to easier retrieval in the future."*

| To Oppose Effortful Practice is to Oppose Talent Development

Myth: *Testing, repetition, computation, and competition detract from learning.*

Reality: In the world of talent development, nobody is confused about the importance of these methods. Can you imagine telling an athletic coach that things like competitive tryouts, repetitious drills, exhausting physical conditioning, and assigning playing time based on performance during scrimmage and competitive games against other teams, detract from developing athletic talent?

> Testing and Repetition are Necessary

As we covered while debunking the previous myth, practice *testing* and distributed practice (also known as spaced *repetition*) are widely understood by researchers to be two of the most effective practice techniques. We have also discussed the importance of deliberate practice, individualized training activities specially chosen to improve specific aspects of a student's performance through *repetition* and successive refinement, which has been shown to be one of the most prominent underlying factors responsible for individual differences in performance, even among highly talented elite performers.

It is not possible to rationally argue that one can maximize learning without engaging in testing and repetition. If someone attempts to argue that position, what they are really saying is that they disagree with the premise of maximizing learning. And that is fine – plenty of people would prefer for their education to maximize other things like fun and entertainment while, as a secondary concern, meeting some low bar for shallowly learning some surface-level basic skills. But that is a completely different and opposite thing from talent development.

What's more, in a subject as hierarchical as math, where each advanced skill requires many simpler skills to be applied in complex ways, avoiding testing and repetition can lead to major struggle. To learn a complex skill, a student must first be fluent with the simpler component skills – and to comfortably perform the complex skill, a student must be fully automatic with the simpler component skills. If a student does not develop fluency and eventual automaticity on each skill, they will be doomed to struggle on the more advanced skills of which those simpler skills are components. Testing and repetition are the two learning strategies that most directly build fluency and eventual automaticity. (To be clear, repetition does not mean giving students excessive practice past the point of mastery, but rather, giving students *enough* practice to achieve mastery before moving them on to more advanced skills.)

> Computation is Necessary

There are several reasons why practicing computation is a necessary part of learning math.

1. In the absence of computation, it's easy to lose touch with the concrete meaning of various symbols, procedures, and ideas. Computation keeps learners aware of what these things mean in terms of concrete numbers. In fact, the whole point of an abstract idea is to streamline and unify existing knowledge of concrete examples. Computational examples are to mathematics as experiences are to life.

2. Is someone a talented basketball player if they can talk about the strategy of the game but cannot actually make any shots? No. The same applies to someone who can talk about mathematical ideas but is unable to perform computations.

3. It is impossible to gain a full, holistic understanding of a subject without knowing the component skills. If someone can't shoot a basketball, how can they possibly understand how different shots compare in terms of difficulty, and what plays might open up good shots? The same is true in mathematics.

4. Computation often helps build conceptual understanding. Math is full of ideas that cannot be properly understood without experience carrying out computations. (One of the clearest examples of this is the concept of the discriminant of a quadratic equation: if a student has experience computing solutions to quadratic equations using the quadratic formula, then they will find it much easier to observe that the b^2-4ac term, known as the discriminant, controls the number of solutions.)

Math resources that don't give proper emphasis to computation end up having to water down their curriculum and cherry-pick problems, giving students the easiest possible problem-solving cases that don't require too much in the way of foundational skills. That can be exciting for students because they get enough conceptual understanding to feel like they have learned the material in proper depth even though they actually haven't.

This is fine if a student is just curious about math and wants to learn a bit without putting in too much time and effort – but if a student is serious about learning math well enough to make a career out of it, then watered-down courses won't give them what they need. That's where Math Academy comes in: we teach math as if we were training a professional athlete or musician, or anyone looking to acquire a skill to the highest degree possible, and we've designed the curriculum to go toe-to-toe vs any similar course you would find in the top universities and the most popular textbooks in the world.

Computation and Silly Mistakes

Opponents of computation will sometimes claim that it unnecessarily slows down the learning process when students are given further practice in response to silly mistakes. However, in the context of a skill hierarchy like mathematics, it's not good enough to be "almost" able to execute a skill properly. If a gymnast is "almost" able to land a backflip, then that's great progress, but at the same time, they're not ready to try any combination moves of which a backflip is a component. Even if it's a silly mistake keeping the gymnast from landing the backflip, they still have to rectify it before layering on more advanced skills.

More generally, when students are not made to clean up their silly mistakes on low-level skills, they eventually hit a wall where no matter how hard they try, they are unable to reliably perform advanced skills due to the compounding probability of silly mistakes in the component skills.

Additionally, many students frequently claim that they made a silly mistake when in fact their mistake was indicative of a deeper conceptual misunderstanding. Sometimes this claim is in good faith (i.e., they honestly believe they made a silly mistake), other times it's in bad faith (i.e., they're trying to exploit the grading system to get credit they don't deserve), but regardless, it's something that needs to prevented.

> Competition Can be Helpful and is Unavoidable in the Big Picture

While competition is not inherent to the learning process, appropriately structured competition does not necessarily detract from it either, and in many cases, can incentivize learners to increase the quantity and quality of practice to maximize their level of achievement. For instance, many Math Academy students are highly motivated by weekly leaderboards to maintain a consistent practice schedule.

Of course, this is not to say that every student must compete in order to learn productively. Some students prefer not to participate in Math Academy's weekly leaderboards, and that is totally fine. Similarly, this is not to say every form of competition promotes learning. The key phrase is *appropriately structured* competition. It is easy to imagine disaster scenarios arising from inappropriately structured competition (e.g. the student with the highest score gets an A and all others fail the class).

That said, if a student is serious about developing their talent to a high enough level to build a career around it, then competition is a reality that they must eventually face. In talent development, anyone who seriously attempts to reach any level of success in a sport, instrument, etc, knows that they have to work really hard and compete against other people (who are also working really hard) for limited positions. Mathematics is no exception. There is a limited number of professorships available for mathematics and related disciplines, and outside academia, there is a limited number of positions available for jobs that involve solving hard problems using advanced mathematics.

It's worth emphasizing that while competition gets a bad rap, its purpose is positive: generally speaking, the purpose of competition is to assign responsibilities to the people most capable of performing them and motivate those people to continue working hard and improving. The bad rap tends to be vocalized by people who are not aligned with this process – for instance, people who confuse their enjoyment of a job with their capability or value to society in performing it, or people who wait until the last minute to begin developing a talent and then experience a rude awakening when they realize that their level of capability is far behind that of other people.

> Why the Myth Persists in Education (But Not in Talent Development)

Why does this myth persist in the practice of education, whereas there is no confusion in the field of talent development? One key factor is that in talent development, the optimization problem is clear: an individual's performance is to be maximized, so the methods used during practice are those that most efficiently convert effort into performance improvements. On the other hand, in education, there are many other factors (especially bureaucratic ones) that constrain and cloud the optimization problem. The end result is that teachers are incentivized to use easy, fun, low-accountability, hard-to-measure practice techniques that keep students, parents, and administrators off their back. Unfortunately, these practice techniques tend to be ineffective.

For instance, consider the idea of testing. In talent development, all parties involved are proponents of testing. If a child is training to play a sport at a high level, such as becoming an Olympic sprinter, then the child, their parents, and their coach will all want to see regular measurements of the child's 100-meter dash time. If that time is going down, then practice is working and everybody is happy. If the time is not going down, then it signals that something needs to be adjusted in the practice routine and nobody is happy until the problem is solved. The act of measuring performance is critical because it tells everyone whether the child is making progress towards achieving their goal.

In education, however, many people are against testing. Typically, parents want their children to get high grades and learn a bit without feeling too stressed, their children want to minimize the amount of work they have to do to satisfy (or, perhaps, not anger) their parents, and administrators want parents to be happy and test scores to be sufficiently high. Teachers are squeezed by pressure on both sides – getting as many kids as possible over some threshold test score, while assigning as high grades and as little work as possible.

In this position, it is easy to dislike testing – if testing were to go away, then it would be easy to satisfy all parties involved by centering the class around discussions and fun activities. Students wouldn't have to work too hard, they would learn a little bit, they would receive good grades on the basis of participation, parents would be happy that their children are getting high grades and learning a bit without feeling too stressed, and administrators would be happy that parents are happy. It is only natural for those in this position to oppose testing and instead argue for the existence and importance of subjective forms of learning that cannot be objectively measured, even though such forms of learning are unscientific by definition.

> Why Talent Development is Important in Math

Practitioners of talent development tend to be found in hierarchical skill domains like sports and music, where each advanced skill requires many simpler skills to be applied in complex ways. This is because it's hard to climb up the skill hierarchy without intentionally trying to do so.

To learn an advanced skill, you must be able to comfortably execute its prerequisite skills, and the prerequisite skills underlying those, and so on. Getting to the point of comfortable execution on any skill takes lots of practice over time – and even after you get there, you have to continue practicing to maintain your ability.

None of this happens naturally. If you don't carefully manage the process, then you struggle. Nobody gets to be really good at a sport or instrument without taking their talent development seriously and intentionally trying to maximize their learning.

Conveniently, most students aren't expected to achieve a high level of success in sports or music, so they can get away with de-prioritizing talent development. If every student in gym class were expected to be able to do a backflip by the end of the year, things would have to change – but the expectations are so low that meeting them does not require talent development.

When it comes to math, however, things become problematic. Like sports and music, math is an extremely hierarchical skill domain, so achieving a high level of success requires a dedication to talent development. However, unlike sports and music, most students are expected to achieve a relatively high level of success in math: many years of courses increasing in difficulty, culminating in at least algebra, typically pre-calculus, often calculus, and sometimes even higher than that.

As a result, in math, de-prioritizing talent development leads to major issues. When students do the mathematical equivalent of playing kickball during class, and then are expected to do the mathematical equivalent of a backflip at the end of the year, it's easy to see how struggle and general negative feelings can arise.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Clark, R., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator, 36*(1), 5-11.

  Sweller, J., Clark, R., & Kirschner, P. (2010). Teaching general problem-solving skills is not a substitute for, or a viable addition to, teaching mathematics. *Notices of the American Mathematical Society, 57*(10), 1303-1304.

  Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist, 41*(2), 75-86.

  Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning?. *American psychologist, 59*(1), 14.

  Anderson, J. R., Reder, L. M., Simon, H. A., Ericsson, K. A., & Glaser, R. (1998). Radical constructivism and cognitive psychology. *Brookings papers on education policy,* (1), 227-278.

  Importance: *Decades of research support the fact that students learn most effectively through direct instruction. This also follows intuitively from modern understandings of working and long-term memory. The alternative, unguided instruction, can be so problematic as to actually result in negative educational progress.*

- Anderson, J. R., Reder, L. M., Simon, H. A., Ericsson, K. A., & Glaser, R. (1998). Radical constructivism and cognitive psychology. *Brookings papers on education policy,* (1), 227-278.

  Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International journal of Educational research, 13*(1), 89-99.

  Importance: *Group learning is generally inefficient, and is typically applied without the requisite structuring or scripting to make it effective, causing it to become detrimental instead.*

- Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own?. *Memory, 17*(4), 471-479.

  Bjork, E. L., & Bjork, R. A. (2023). Introducing Desirable Difficulties Into Practice and Instruction: Obstacles and Opportunities. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 111-21). Society for the Teaching of Psychology.

  Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest, 14*(1), 4-58.

  Kang, S. H. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 12-19.

  Qadir, J., & Imran, M. A. (2018). Learning 101: The untaught basics. *IEEE Potentials, 37*(3), 33-38.

***Importance:** Many types of training methods are ineffective, but require little effort, and can therefore seem attractive to even the most well-intentioned, hardworking students because they create an illusion of competence. Examples include looking at notes, rereading course materials, and highlighting. On the other hand, the most effective methods – such as practice testing and distributive practice – are particularly effortful, and are even more effective (and, of course, even more effortful) when combined. Learning is all about creating desirable difficulties.*

Chapter 9. Myths & Realities about Mathematical Acceleration

Summary: *Students often become mathematically accelerated by working on Math Academy, and there are many misconceptions surrounding educational acceleration. Acceleration does not lead to adverse psychological consequences in capable students; rather, whether a student is ready for advanced mathematics depends solely on whether they have mastered the prerequisites. Acceleration does not imply shallowness of learning; rather, students undergoing acceleration generally learn – in a shorter time – as much as they would otherwise in a non-accelerated environment over a proportionally longer period of time. Accelerated students do not run out of courses to take and are often able to place out of college math courses even beyond what is tested on placement exams. Lastly, for students who have the potential to capitalize on it, acceleration is the greatest educational life hack: the resulting skills and opportunities can rocket students into some of the most interesting, meaningful, and lucrative careers, and the early start can lead to greater career success.*

Acceleration is Often Misunderstood

On a system like Math Academy, where students can learn math multiple times as efficiently as in a traditional classroom, students who continue working at a normal "school workload" pace of an hour or more per weekday throughout the year will learn multiple years of math in one year – in a fully comprehensive curriculum, without skipping any content.

While this may seem unexpected and even shocking to those unfamiliar with academic acceleration (the practice of allowing students to learn academic material at a younger age and/or faster rate than is typical), this is a normal and expected consequence of increased learning efficiency: if a student's learning becomes 4x more efficient, and they continue putting forth the same amount of time into learning, then they will learn 4x as much material.

The benefits of mathematical acceleration are as numerous as the misconceptions surrounding it. As lamented by researcher James Borland (1989, pp.185):

> *"Acceleration is one of the most curious phenomena in the field of education. I can think of no other issue in which there is such a gulf between what research has revealed and what most practitioners believe. The research on acceleration is so uniformly positive, the benefits of*

> *appropriate acceleration so unequivocal, that it is difficult to see how an educator could oppose it."*

The purpose of this chapter is to clear up misconceptions and, at the same time, communicate the benefits of mathematical acceleration.

Developmental Appropriateness

| Advanced Study is Appropriate Once Prerequisites Have Been Mastered

Myth: *Learning math early is not appropriate for students' social/emotional and cognitive/academic development.*

Reality: Educational acceleration does not lead to adverse psychological consequences in capable students. For instance, according to a study titled *Academic Acceleration in Gifted Youth and Fruitless Concerns Regarding Psychological Well-Being: A 35-Year Longitudinal Study* that followed thousands of accelerated students throughout their lives over the course of 35 years (Bernstein, Lubinski, & Benbow, 2021):

> *"The amount of educational acceleration did not covary with psychological well-being. Further, the psychological well-being of participants in both studies was above the average of national probability samples. Concerns about long-term social/emotional effects of acceleration for high-potential students appear to be unwarranted, as has been demonstrated for short-term effects.*
> *...*
> *These findings are consistent with research on the effects of academic acceleration on psychological well-being. That is, there is little evidence that academic acceleration has negative consequences on the psychological well-being of intellectually talented youth (Assouline et al., 2015; Benbow & Stanley, 1996; Colangelo et al., 2004; Gross, 2006; Robinson, 2004).*
> *...*
> *These findings do not support the frequently expressed concerns about the possible long-term social and emotional costs of acceleration by counselors, parents, and administrators. ... Those who were accelerated had few regrets for doing so. Indeed, if anything, they tended to wish that they had accelerated more."*

Whether a student is ready for advanced mathematics depends solely on whether they have mastered the prerequisites. If a student has mastered prerequisites, then it is appropriate for them to continue learning advanced math early, and not appropriate to stunt their development by holding them back. As the study authors note:

> *"Many fear negative possibilities of moving a gifted child to a more advanced class. Yet it also is important to consider the negative possibilities of holding children back in classes aiming to teach*

> *subject matter that they have already mastered (Benbow & Stanley, 1996; Gross, 2006; Stanley, 2000). Choosing not to accelerate is as much of a decision as choosing to do so …*
>
> *This is particularly important given the extensive empirical literature showing positive effects of acceleration on academic achievement (Kulik & Kulik, 1984, 1992; Lubinski, 2016; Rogers, 2004; Steenbergen-Hu et al., 2016) and creativity (Park et al., 2013; Wai et al., 2010). … Presenting students with an educational curriculum at the depth and pace with which they assimilate new knowledge is beneficial. Other studies have shown that academic acceleration tends to enhance professional and creative achievements before age 50 (Park et al., 2013; Wai et al., 2010)."*

Numerous other studies on the long-term effects of educational acceleration have drawn similar conclusions. As Wai (2015) summarizes:

> *"…[F]or many decades there has been a large body of empirical work supporting educational acceleration for talented youths (Colangelo & Davis, 2003; Lubinski & Benbow, 2000; VanTassel-Baska, 1998). Although neglecting this evidence seems increasingly harder to do (Ceci, 2000; Stanley, 2000), putting research into practice has been challenging due to social and political forces surrounding educational policy and implementation (Benbow & Stanley, 1996; Gallagher, 2004; Stanley, 2000).*
> *…*
> *The educational implications of these studies are quite clear. They collectively show that the various forms of educational acceleration have a positive impact. The key is appropriate developmental placement (Lubinski & Benbow, 2000) both academically and socially. … Educational acceleration is essentially appropriate pacing and placement that ensures advanced students are engaged in learning for life. Every student deserves to learn something new each day (Stanley, 2000). The evidence clearly supports allowing students who desire to be accelerated to do so, and does not support holding them back.*
> *…*
> *[T]he long-term studies reviewed here show that adults who had been accelerated in school achieved greater educational and occupational success and were satisfied with their choices and the impact of those choices in other areas of their lives."*

| Why the Myth of Developmental Inappropriateness Persists

This myth of acceleration being developmentally inappropriate may be perpetuated in part by convenience. In schools, each grade typically progresses through the math curriculum in lockstep, which means that accelerated students would need to be placed in above-grade courses. This can lead to major logistical challenges.

For instance, if above-grade course is not offered by the school (which would certainly be the case for accelerated 5th graders in elementary schools, 8th graders in middle schools, and 12th graders in high schools), then either

- the students would need to take the class at another school (which introduces transportation, scheduling, and administrative issues) or

- the school would need to hire a teacher who is capable of teaching the higher-grade material (and it's hard enough for schools to hire teachers who are capable of teaching grade-level mathematics).

And even if the above-grade course is offered by the school, there may be schedule conflicts with grade-level courses that mathematically accelerated students still need to take. (Course schedules are typically optimized to minimize conflicts within grade levels, but not across grade levels.)

Besides logistical issues, there are other factors that can disincentivize acceleration and lead the myth to be perpetuated out of convenience. As Steenbergen-Hu, Makel, & Olszewski-Kubilius (2016) describe:

> *"[E]ducation administrators may have perverse incentives to avoid acceleration. For example, although acceleration can often actually save schools money because students spend fewer years in school, it can also 'cost' schools money. Because school funding is often allocated based on headcounts and accelerated students spend fewer years in school, schools receive fewer dollars overall, or in the case of dual enrollment, may have to spend some of those dollars outside the district.*
>
> *Similarly, in states that offer open enrollment, students could leave a district for one where their needs are better met. Moreover, in the age of accountability via test score performance, keeping students who could be accelerated with their same-age peers can boost average test scores, regardless of whether the students are learning."*

Even in schools that do offer acceleration, typically only a small portion of students per grade are accelerated. Given how many logistical challenges and other disincentivizing factors there are, how few students are typically accelerated, and how easy it is to imagine a young student struggling socially when they are placed in a class with older students away from age-level friends, it is not surprising that the myth persists.

On Math Academy, however, all of these issues and concerns vanish. Students can accelerate their mathematical learning – even to the point of learning highly advanced university-level math – entirely through our system from the comfort of their home or grade-level classroom.

Depth of Learning

| Accelerated Students Learn More Material, Just as Deeply

Myth: *Mathematically accelerated students become accelerated by rushing through watered-down courses, leading to shallower learning.*

Reality: It is well documented in the literature of academic acceleration studies that students undergoing acceleration generally learn – in a shorter time – as much as they would otherwise in a non-accelerated environment over a proportionally longer period of time.

For instance, Kulik & Kulik's well-known review (1984) of 26 academic acceleration studies found that talented students who were accelerated by one year (i.e. they learned two years' worth of material in one year) performed as well as students one year older who were equivalently talented but not accelerated:

> *"First, talented youngsters who were accelerated into higher grades performed as well as the talented, older pupils already in those grades. Second, in the subjects in which they were accelerated, talented accelerates showed almost a year's advancement over talented same-age nonaccelerates."*

As Kulik & Kulik (1984) noted, *"most [other] reviewers of the controlled studies have reached favorable conclusions about the effects of acceleration."* Furthermore, many of these conclusions were expressed with a level vehemence that is rare to find in academic literature, except out of frustration when a result so clearly supported by science is ignored by the education system for no reason other than the inertia of tradition:

- *"In her 1958 review, Goldberg pointed out that it was hard to find a single research study showing acceleration to be harmful and that many studies proved acceleration to be a satisfactory method of challenging able students."*
- *"A 1964 review by Gowan and Demos concluded simply that 'accelerated students do better than non-accelerated students matched for ability' (p. 194)."*
- *"Gold (1965) echoed their [Gowan and Demos's] sentiments and added, 'No paradox is more striking than the inconsistency between research findings on acceleration and the failure of our society to reduce the time spent by superior students in formal education' (p. 238)."*
- *"Perhaps what is needed," Gallagher suggested in 1969, "is some social psychologist to explore why this procedure [of academic acceleration] is generally ignored in the face of such overwhelmingly favorable results" (p. 541).*

- *"Dillon in 1973 also lamented the lack of interest in acceleration and offered a social psychological explanation: 'Apparently the cultural values favoring a standard period of dependency and formal education are stronger than the social or individual need for achievement and independence. This is an instance of the more general case one remarks throughout education: When research findings clash with cultural values, the values are more likely to prevail, (p. 717).'"*

- *"In a review of research on acceleration in mathematics, Begle (1976) concluded that accelerated students scored higher than comparable controls in almost all comparisons and almost never scored lower. The accelerated students also did better than average, nonaccelerated, older students, and when they did not do as well as talented older students, they did not lag far behind."*

This review (Kulik & Kulik, 1984), considered together with about a dozen more recent others, gave rise to the following conclusion in the second-order review titled *What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement* (Steenbergen-Hu, Makel, & Olszewski-Kubilius, 2016):

> *"...[T]he conversation needs to evolve beyond whether such interventions [of academic acceleration] can ever work. There is not an absence of evidence, nor is there evidence of absence of benefit. The preponderance of existing evidence accumulated over the past century suggests that academic acceleration ... can greatly improve K–12 students' academic achievement."*

| How We Ensure Comprehensive, Deep Learning

Even beyond the academic literature, Math Academy uses specific techniques to ensure that even the most accelerated students are learning comprehensively and deeply:

- Math Academy's courses are more comprehensive than typical courses in traditional classrooms, and

- completing a Math Academy course requires a student to demonstrate a degree of knowledge greater than that necessary to pass the course in a traditional classroom.

To ensure that our courses are fully comprehensive, we perform curriculum comparisons to ensure that our courses provide a superset of content covered by major textbooks. That is to say: given a major textbook, our corresponding course not only covers the content found within the textbook, but also covers additional content found in other major textbooks. Our courses are the "real thing," and they cover all the content that one could reasonably expect to find in any major textbook or standard class syllabus.

Yes, this means that we have invested a lot of time, effort, and money developing an absolute mountain of content. Our courses are the product of nearly a decade of work by a team of more than 10 PhD mathematicians. We have over 2500 fully-scaffolded lessons each consisting of an introduction, 3-4 fully-worked examples called Knowledge Points or KPs, and 10+ questions (with full solutions) within each KP. In total, we have a bank of over 150,000 questions. These counts continue to increase.

On the other hand, typical courses in traditional classrooms are seldom comprehensive. Textbooks usually aim to cover a vast array of everything that an instructor *might* want to teach (because if they miss any material that an instructor wants to cover, they'll lose the instructor to a different textbook). But instructors generally pick and choose from that material. While the basics of most math courses are generally agreed upon, what's covered beyond the basics can vary from one instructor to another depending on various factors such as the following:

- *the class schedule* – due to many schedule interruptions throughout the year (e.g. standardized testing, field trips, school assemblies), teachers have less class time than one would imagine, so they have to prioritize and streamline what they're going to teach in order to avoid running out of time before the end of the year.

- *the instructor's interests* – different classes often cover different offshoots from the core material and may go deeper in some areas than others, depending on what the instructor finds most interesting or is most comfortable teaching.

Additionally, unlike traditional classrooms where students can pass their courses despite not having mastered all of the material covered, Math Academy is a mastery-based learning system in which students do not move forward until they demonstrate mastery of topics. As a result, a student who completes a course on Math Academy must have demonstrated mastery on 100% of the topics in the course – whereas a student who gets an A, B, C, or D in a course in a traditional classroom may have only mastered 90%, 80%, 70%, or 60% of the material.

Continuity of Courses

| Accelerated Students Don't Run out of Math Courses

Myth: *If a student takes math classes early, they will run out of math classes to take.*

Reality: While many people think calculus is the "end of the road" for math, it is but an entry-level requirement for university-level math courses. There are even more university-level math courses above calculus than there are high school courses below calculus, and many of these university-level math courses are available on Math Academy.

After a single-variable calculus course (like AP Calculus BC), most serious students who study quantitative majors like math, physics, engineering, and economics have to take core "engineering math" courses including Linear Algebra, Multivariable Calculus, Differential Equations, and Probability & Statistics (the advanced calculus-based version, not the simpler algebra-based version like AP Statistics). Beyond those core "engineering math" courses, different majors include plenty of specialized courses that branch off in various ways.

There are so many university-level math courses that a student could not fit them all into a standard 4-year undergraduate course load even if they overloaded their schedule every year – however, the more of these courses a student is able to take, the more academic opportunities and career doors are open to them in the future.

| Advanced Students Can Place Out of College Courses Beyond Placement Tests

Myth: *There's no use in learning math past calculus in high school because you'll have to take it again in college (since advanced placement courses and college math placement tests only go up through calculus).*

Reality: When the most advanced students place out of classes, it is not through transfer credit or placement exams. Generally, they are placing out of courses that are beyond what's tested on the placement exam.

They do this by not only learning the material beforehand, but also taking the initiative to schedule a meeting with an undergraduate advisor or coordinator for the math department. Some schools have a policy of arranging undergraduate for-credit exams, while others may have a less formal process, such as arranging a meeting with a professor who will determine the student's placement by discussing mathematics with them, getting a sense of their background and knowledge, and maybe having them solve some problems at the board.

If you want to learn math ahead of time and place into more advanced courses, there are a couple pitfalls to watch out for:

1. If you learn material ahead of time, but not comprehensively, then you might not be able to evidence enough knowledge to place out of it. Or, if you manage to place out of a

course without having learned the material comprehensively, you might end up way out of your depth in the more advanced course that you end up taking.

2. If you learn material ahead of time, but do not continually review that material, then you will likely become rusty and unable to evidence enough knowledge to place out of it.

In order to avoid these pitfalls, you need to learn material comprehensively and continually review it after learning it. Math Academy does both of these things.

Relevance to Students' Futures

| Learning Math Early Reduces Risk and Opens Doors to Opportunities

Myth: *Learning math early can be impressive, but it's just a party trick. It doesn't have much real impact on a student's future, especially if they're going into something other than engineering.*

Reality: You know how, when you take a language class, there's often a couple kids who speak the language at home and think the class is super easy? You can do that with math. Learning math ahead of time basically guarantees an A and guards against all sorts of risks such as the teacher not knowing the content very well or otherwise not being able to explain it well. This is especially helpful at university, when lectures are often unsuitable for a first introduction to a topic.

Of course, the natural objection is "won't you be bored in class?" – but if you do super well in advanced classes, especially at university, then that opens all kinds of doors to recommendations for internships, research projects with professors, etc. Even if you aren't a genius, you appear to be one in everyone else's eyes, and consequently you get a ticket to those opportunities reserved for top students. Students who receive and capitalize on these opportunities can launch themselves into some of the most interesting, meaningful, and lucrative careers that are notoriously difficult to break into.

Learning math early also gives students the opportunity to delve into a wide variety of specialized fields that are usually reserved for graduates with strong mathematical foundations. This fast-tracks students towards discovering their passions, developing valuable skills in those domains, and making professional contributions early in their careers, which ultimately leads to higher levels of career accomplishment. As described by the authors of a 40-year longitudinal study of thousands of mathematically precocious students (Park, Lubinski, & Benbow, 2013):

> *"The relationship between age at career onset and adult productivity, particularly in science, technology, engineering, and mathematics (STEM) fields, has been the focus of several researchers throughout the last century (Dennis, 1956; Lehman, 1946, 1953; Simonton, 1988, 1997; Zuckerman, 1977), and a consistent finding is that earlier career onset is related to greater productivity and accomplishments over the course of a career. All other things being equal, an earlier career start from acceleration will allow an individual to devote more time in early adulthood to creative production, and this will result in an increased level of accomplishment over the course of one's career.*
>
> *...*
>
> *[In this study] Mathematically precocious students who grade skipped were more likely to pursue advanced degrees and secure STEM accomplishments, reached these outcomes earlier, and accrued more citations and highly cited publications in STEM fields than their matched and retained intellectual peers."*

And while it's true that students don't need to know much beyond algebra to get a job in fields like computer science, medicine, etc. – the people in such fields who *do* also know advanced math are extra valuable and in demand because they can work on projects that combine domain expertise and math.

Higher-Grade Math is Typically More Productive than Grade-Level Competition Problems

Myth: *If a student learns their grade-level math and wants to do more math, it is more productive to have them work on extremely challenging competition math problems at their current grade level than to continue learning more advanced math that they would normally learn in higher grade levels.*

Reality: When a middle or high school teacher has a bright math student, and the teacher directs them towards competition math, it's usually not because that's the best option for the student. Rather, it's the best option for the teacher. It gives the student something to do while creating minimal additional work for the teacher.

Competition math problems generally don't require students to learn new fields of math. Rather, the difficulty comes from students needing to find clever tricks and insights to arrive at solutions using the mathematical tools that they have already learned. A student can wrestle with a competition problem for long periods of time, and all the teacher needs to do is give a hint once in a while and check the student's work once they claim to have solved the problem.

But if you look at the kinds of math that most quantitative professionals (like rocket scientists and AI developers) use on a daily basis, those competition math tricks show up rarely, if ever. What *does* show up everywhere is university-level math subjects like linear algebra, multivariable calculus, differential equations, and (calculus-based) probability and statistics. Given that most students who enjoy math end up applying math in some other field (as opposed to becoming pure mathematicians), it would be more productive for them to get a broad view of math as early as possible so that they can sooner apply it to projects in their field(s) of interest.

Of course, the countering view is that "students should go 'deep' with the math that they've already learned – they'll learn the other math subjects when they're ready." But, in practice, the second part of that claim is not true. There are so many other math subjects that even most math majors only learn a tiny slice of all the math that's out there.

Students generally can't learn other math subjects "on the job" after graduation, either – if you're trying to solve cutting-edge problems that nobody has solved before, then there is no "known path" that can tell you what additional math you need. And to even realize that a field of math can help you solve your problem, you generally need to have learned a substantial amount of that field in the first place.

In practice, the only way for students to "learn the other math subjects when they're ready" is to learn as much math as possible during school.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Bernstein, B. O., Lubinski, D., & Benbow, C. P. (2021). Academic acceleration in gifted youth and fruitless concerns regarding psychological well-being: A 35-year longitudinal study. *Journal of Educational Psychology, 113*(4), 830.

  Importance: *Concerns about long-term social/emotional effects of acceleration for high-potential students appear to be unwarranted, as has been demonstrated for short-term effects. Those who were accelerated had few regrets for doing so. Indeed, if anything, they tended to wish that they had accelerated more.*

- Wai, J. (2015). Long-term effects of educational acceleration. *A nation empowered: Evidence trumps the excuses holding back America's brightest students, 2*, 73-83.

  Importance: *For many decades there has been a large body of empirical work supporting educational acceleration for talented youths. Educational acceleration is essentially appropriate pacing and placement that ensures advanced students are engaged in learning for life. Adults who had been accelerated in school achieved greater educational and occupational success and were satisfied with their choices and the impact of those choices in other areas of their lives. The evidence clearly supports allowing students who desire to be accelerated to do so, and does not support holding them back.*

- Kulik, J. A., & Kulik, C. L. C. (1984). Effects of accelerated instruction on students. *Review of educational research, 54*(3), 409-425.

  Importance: *Students undergoing acceleration generally learn – in a shorter time – as much as they would otherwise in a non-accelerated environment over a proportionally longer period of time. Specifically, talented students who were accelerated by one year (i.e. they learned two years' worth of material in one year) performed as well as students one year older who were equivalently talented but not accelerated.*

- Park, G., Lubinski, D., & Benbow, C. P. (2013). When less is more: Effects of grade skipping on adult STEM productivity among mathematically precocious adolescents. *Journal of*

Educational Psychology, 105(1), 176.

Importance: *All other things being equal, an earlier career start from acceleration will allow an individual to devote more time in early adulthood to creative production, and this will result in an increased level of accomplishment over the course of one's career. Mathematically precocious students who grade skipped were more likely to pursue advanced degrees and secure STEM accomplishments, reached these outcomes earlier, and accrued more citations and highly cited publications in STEM fields than their matched and retained intellectual peers.*

III. COGNITIVE LEARNING STRATEGIES

Chapter 10. Active Learning

Summary: *It is a decisive finding in the literature that students learn better when they are actively engaged in learning exercises as opposed to passively consuming educational content. True active learning requires every individual student to be actively engaged on every piece of the material to be learned.*

Definition and Importance

It is a common misconception that the fastest way to learn math is by watching videos, attending lectures, reading books, or re-reading notes. This is false. As hundreds of studies have shown, passively consuming educational content leads to significantly worse educational outcomes than **active learning**, where students are actively performing learning exercises (Freeman et al., 2014).

In a passive learning scenario like watching a video, students may believe they are effectively learning if they can understand the video and follow along. However, following along and understanding the video's contents isn't learning, even if students claim that it is. As discussed in chapter 3, learning is a positive change in long-term memory. In order for students to have learned something, they need to be able to consistently reproduce that information and use it to solve problems. None of these things happen when students watch a video, even if they understand it perfectly. The same reasoning applies to attending lectures, reading books, re-reading notes, and all other passive learning techniques.

The superiority of active learning is so robust across subjects and experimental methodologies that a highly-cited meta-analysis states, verbatim (Freeman et al., 2014):

> *"...[C]alls to increase the number of students receiving STEM degrees could be answered, at least in part, by abandoning traditional lecturing in favor of active learning.*
> *...*
> *Given our results, it is reasonable to raise concerns about the continued use of traditional lecturing as a control in future experiments."*

According to Nobel laureate Carl Wieman, lecturing is the educational equivalent of bloodletting (Westervelt, 2016):

> *"You let some blood out and go away and they get well. Was it bloodletting that did it, or something else? … You give people lectures, and [some students] go away and learn the stuff. But it wasn't that they learned it from lecture – they learned it from homework, from assignments. When we measure how little people learn from an actual lecture, it's just really small.*
> *…*
> *The quality of teaching is not something that university administrators are rewarded for, and correspondingly know or care about … It's like you've got a hospital and you're not bothering to check if your doctors are using antibiotics or bloodletting."*

Like lecturing, re-reading does not count as active learning either. As Brown, Roediger, & McDaniel (2014, pp.10) describe:

> *"The finding that rereading textbooks is often labor in vain ought to send a chill up the spines of educators and learners, because it's the number one study strategy of most people – including more than 80 percent of college students in some surveys – and is central in what we tell ourselves to do during the hours we dedicate to learning.*
>
> *Rereading has three strikes against it. It is time consuming. It doesn't result in durable memory. And it often involves a kind of unwitting self-deception, as growing familiarity with the text comes to feel like mastery of the content. The hours immersed in rereading can seem like due diligence, but the amount of study time is no measure of mastery."*

It's important to realize that true active learning means every individual student is engaged in activity, not just the class as a whole. For instance, although a class-wide discussion might seem like active learning on the surface, it does not immediately follow that each student is active. Often, it is only a proportionally small number of enthusiastic, vocal students who participate in all parts of the discussion and can be considered truly active. Even if the instructor cold-calls on students who have not been participating, most students will only pay attention enough that they won't look foolish or be embarrassed if called upon.

Moreover, true active learning requires every individual student to be actively engaged on every piece of material to be learned. Divide-and-conquer group projects do not count as fully active learning because each student is only actively learning the material that corresponds to their individual responsibility in the division of labor. The rest of the project, they observe only passively, if at all.

Math Academy's approach to math education is entirely centered around true active learning – on every single topic in their course, students are solving problems within minutes of starting the lesson (following a minimum effective dose of initial explanation). They spend the vast majority of their time engaged in active problem-solving.

Case Studies

| Case Study 1: Why Active Learning is Obvious

To get a clear, concrete picture of what active learning entails and why it is so beneficial, it's helpful to go through a case study in what is perhaps the most familiar setting for active learning: learning a new sport.

Suppose that you want to learn how to play tennis. You go to your local tennis club, where there is a coach who used to play tennis professionally. They offer personal lessons for a pricey $100/hour, but you really want to learn from the best, as efficiently as possible, so you fork over the money for a lesson the following week.

The next week, you show up for your lesson. The coach greets you and begins the hour-long session. It proceeds as follows:

> *[5 minutes] Coach talks about the beauty of tennis and why it's a great sport to learn.*
>
> *[5 minutes] Coach demonstrates a tennis stance and explains the specific components of the stance: bent knees, forward lean, racket in front of you, etc.*
>
> *[5 minutes] Coach demonstrates the ideal place to stand when receiving a volley: near the baseline, in the middle of the court, so that you're back far enough that your opponent can't hit the ball behind you, but you're close enough to the net to launch forward towards any shorter volleys.*
>
> *[10 minutes] Coach demonstrates a forehand swing, explains how the force should come from the legs and the twisting of the body (rather than the arm) and emphasizes the importance of "follow through" on the swing.*
>
> *[20 minutes] Coach demonstrates a backhand swing, breaks down the components, and shares stories about historic moments in tennis when a player had no time to position themself for a forehand and therefore had to rely on their backhand to win the game. Coach demonstrates a one-arm backhand, an advanced move that looks particularly cool.*
>
> *[15 minutes] Coach demonstrates serving and shows off some lightning-fast, precisely targeted serves that seem impossible to return. Again, these are advanced moves that look really cool.*

When the session ends, the coach asks if you want to schedule another session the following week.

What do you do? Are you a happy customer? Do you want to schedule another session? Heck no! The coach just waxed philosophical and showed you moves the whole time. You didn't actually learn anything. You might as well have just watched tennis on TV. You signed up for a tennis lesson to become a better tennis player – not to watch the coach hit the ball. You just wasted $100 on a complete waste of time and you want your money back.

Of course, this situation is unlikely to occur in real life athletic training because coaches know that continued employment depends on their ability to make students learn. They are held accountable for improving the performance of their students. They need to get real, demonstrable results, and get them fast – and if they can't, then they're going to lose a client and develop a reputation as a grifter who tricks people into paying a lot of money for a service that just doesn't work.

In real life athletic training, a coach is going to have their students actively performing moves within the first couple minutes of the session. Sure, the coach might take a minute to demonstrate and break down a new move as the student watches, but for the next 10+ minutes after that, the student is going to be actively practicing that new move. The coach will observe the student and point out areas where they need to correct their form to be more effective – and as the student gets better at the new move, they will experience a real, demonstrable improvement in their athletic performance. Maybe they'll be able to hit the ball faster or more precisely. Maybe they'll be able to return a tricky volley that originally kept going past them at the beginning of the session. Whatever the improvement, it will be tangible and reproducible.

It's worth emphasizing: in a personal coaching session, when does the learning occur? It's not when you pay the coach the money. It's not when you watch the coach demonstrate a move. It's when you actually start doing things that you weren't able to do before. It's when you attempt a move, the coach corrects your form, and you attempt the move again with better results. The learning is the incremental gain in your ability to perform a skill. If you're not getting those gains, you're not learning.

The same reasoning applies if you're getting a lesson on piano or guitar. You're not just absorbing information – you're developing skills. Mathematics, too, is skill-based. Learning how to solve a new type of equation is totally different from, say, learning some new history about the life of Napoleon. At the core, the keys to effective training in mathematics are the same as the keys to effective training in athletics or music.

| Case Study 2: Most Students Don't Even Pay Attention During Lectures

As he details in an interview, Peter Reinhardt, co-founder and CEO of the customer data platform Segment, learned the hard way that most students aren't even paying attention during lecture (Y Combinator, 2018).

According to Reinhardt, Segment's original product was actually a classroom lecture tool:

> *"The idea was to give students this button to push to say, "I'm confused." The professor would get this graph over time of how confused the students were. We thought it was a really cool idea. We were college students at the time and we had a bunch of professors who were excited about it at MIT and elsewhere. ... [W]e talked to like 20 other professors and they were all excited about it."*

However, despite all the excitement from professors, when testing the tool in an actual classroom at Boston University, it was a total disaster – not because the tool didn't work, but because the students weren't even paying attention to what the professor was saying.

Instead, most students were either partially or fully engrossed with scrolling through Facebook. The lecture was, effectively, just a charade: the students weren't even passively learning; they were just sitting there doing other stuff.

> *"It was just a total disaster. All the students opened their laptops and they all went straight to Facebook.*
>
> *The way we discovered this is we were standing in the black of the classroom and just counting laptop screens. We'd be looking over the shoulders of the students and going one, two, three and we discovered at the beginning of class, about 60% of the students were on Facebook and by the end about 80% were on Facebook. Oh my God.*
> *...*
> *Standing in the back of a BU classroom. It was an anthropology class. And I remember arriving at the 60% and the 80% and we went up and apologized to the professor and walked out."*

The situation seemed so hopeless that – despite having gone through a months-long startup incubator program, building out a highly sophisticated product consisting of hundreds of thousands of lines of code, and raising over half a million dollars in investments literally a week before – they completely abandoned the product and called up their investors to return the investment money.

> *"We went through the whole YC [Y Combinator startup incubator] with this idea. Built it out. Hundreds of thousands of lines of code. Super-complicated classroom lecture tool product. It had presentation view and people could ask questions. It was very complicated. We actually even raised money at Demo Day with this idea. About [$]600K.*
> *...*

> *We had just gotten wires for these checks, for this money, literally a week before. We called back all the investors and we were like, 'Well, it turns out this is a terrible idea. So what do you want us to do with the money?' Almost all of them said, well, we invested for the team so go find something else."*

As fate would have it, most of the investors were willing to retain the investment if the team pivoted to solving a more promising problem – so the team pivoted to building a web analytics tool that would help other developers avoid this kind of catastrophe by getting a better understanding of how users were behaving on their apps.

> *"We realized, we should have been able to figure out some of this analysis by not just standing in the back of the classroom. Like, we should have been able to see some of this digitally. You couldn't see it in the analytics metrics at all. We decided, hey, let's build an analytics tool."*

| Case Study 3: How Active Learning Saved MIT's Physics Classes

In the early 2000s, MIT solved issues with its physics courses by switching from passive to active learning. The problem, according to John Belcher, the MIT professor who spearheaded the effort, was that a staggering number of undergraduates were failing their freshman (first-year) physics course, a general education requirement (Dori & Belcher, 2005):

> *"Teaching freshman courses in a large lecture hall with over 300 students ... is based on the assumption that the instructor can 'pour out' knowledge from his or her vast reservoir into the empty glasses of the students' minds.*
> *...*
> *If this were true, students at MIT would not fail these large required classes. The high failure rates in these courses at MIT, approaching 15%, and the low attendance in lectures at the end of the term, less than 50%, suggest that there is a basic flaw in this model of instruction."*

Keep in mind that MIT is one of the most selective universities in the world, known especially for admitting students with extremely high mathematical capabilities and passion for quantitative subjects like physics. With a yearly cohort of about 1000 students, a 15% failure rate means that, roughly, a staggering 150 MIT students were failing those courses each year. If this many highly-qualified MIT admits are failing a first-year general education requirement, then clearly the problem lies in the delivery of the course, not the ability of the students.

The problem, as Belcher and colleagues characterized it, was that the physics courses were centered around passive learning. The solution, then, would be to switch the courses over to active learning instead (Dori & Belcher, 2005).

> *"The thinking required while attending a lecture is low-level comprehension of factual knowledge that goes from the ear to the writing hand (Towns & Grant, 1997). Johnson et al. (1998) pointed out that students' attention to what the instructor was saying decreased as the lecture proceeded. ... As Bybee and Ben-Zvi (1998) indicated, science educators have focused primarily on content and secondarily on instruction, leaving assessment and implementations to others or completely ignoring them.*
>
> *Until the early 1990's, most physics instructors were largely unaware of the outcomes of research in physics education (Laws, Rosborough & Poodry, 1999). During the past 15 years, a number of physics curricula have been developed that utilize educational research outcomes. ... The common thread in all these curricula is that they emphasize elements of active learning and conceptual understanding that build on making predictions, and observing and discussing the outcomes with peers. Hake (1998) showed that the learning gains in undergraduate physics are almost double when active learning is involved."*

The movement of these MIT physics courses from passive lectures to active learning classrooms came to be known as the Technology Enhanced Active-Learning (TEAL) project, where TEAL classes operated as follows:

> *"A typical [TEAL] class is comprised of mini lectures scattered throughout the class, separated by periods in which students are engaged in hands-on desktop experiments, visualizations, problem solving, and peer discussion."*

Indeed, Belcher and colleagues reported astounding results, with active learning reducing the failure rate by nearly two-thirds:

> *"The failure rates in the two experimental groups were less than 5% in the small- and large-scale experimental groups, respectively, compared with 13% in the traditional control group (Spring 2002)."*

In a cohort of 1000 students, this would mean that, of the 130 students who would fail the passive learning class, only 50 would still fail the active learning class, and the other 70 would be rescued from failure and end up passing the class.

| Case Study 4: If You're Active Half the Time, That's Still Not Enough

In a study (Deakin & Cobley, 2003, pp.115-136) of figure skaters who had been practicing for a similar number of years, the proportion of active practice (relative to passive practice) was a defining attribute separating the elite and non-elite skaters. The elite skaters spent a greater proportion of their practice time actively practicing some of the trickiest, most taxing moves (jumps & spins), and even when resting from those taxing activities, they were more likely to continue actively practicing less taxing movements like footwork and arm positions.

The authors note specific percentage breakdowns, which we have organized into a table to illustrate how each group of skaters would use 100 minutes of practice time.

> *"...[T]he elite group spent an average of 14% of their total on-ice practice time on rest; the competitive group, 31%; and the test skaters, 46%. ... The elite and competitive skaters spent 68% and 59% of their sessions practicing jumps whereas the test group was engaged in those activities for only 48% of their on-ice time.*
> *...*
> *Not only did the elite group practice jumps and spins for a higher proportion of on-ice session, but they also rested less and used the remaining 18% of their on-ice time to practice other elements of their programs, such as footwork and arm positions."*

Group	**Active Minutes** (= **Taxing** + **Non-Taxing**)	**Passive Minutes**	**Active Minutes per Passive Minute**
Elite	86 (= 68 + 18)	14	6.1
Competitive Non-Elite	69 (= 59 + 10)	31	2.2
Non-Competitive	54 (= 48 + 6)	46	1.2

In the table, we see that the elite skaters allocated their practice time far more efficiently: during practice, the elite skaters were over 6 times more active than passive, while non-competitive skaters were nearly as passive as they were active.

The key takeaway is that, while some amount of active learning is certainly better than no active learning, the best outcomes are achieved by fully maximizing the amount of productive active learning. (Of course, some passive instruction will generally be needed to demonstrate to a learner what it is that they need to practice, but that passive instruction should be kept to a minimum effective dose before launching into more extensive active learning.)

If we ballpark-estimate the proportion of time that a Math Academy student spends on active learning, we get a similar proportion as the elite skaters. On average, a typical Math Academy lesson might consist of 3 worked examples, each followed by about 3 practice questions (give or take depending on how well the student does), and the lesson followed by several explicit reviews (about 4 questions each, again give or take depending on performance) spaced into the future. So, for every 3 worked examples that a student reads, they will be actively doing about $3 \times 3 + 3 \times 4 = 21$ practice problems, putting the net ratio at about 7 active practice problems per passive worked example.

Neuroscience of Active Learning

The effects of active learning can be seen quite literally in the brain: in brain imaging studies, active learning consistently leads to more neural activation than passive learning.

For instance, in a study of students actively writing letters versus passively viewing them, the active writing produced higher brain activity in the sensori-motor network and beyond (Kersey & James, 2013):

> *"Self-generated production of cursive letters during learning led to the recruitment of a sensori-motor network known to also be active during letter perception and reading, however, passive observation of a letter being formed did not. This finding adds to the growing literature suggesting that self-generated writing is important for setting up reading networks in the developing brain.*
> *...*
> *Further, when we directly compared active to passive learning of cursive letters, greater recruitment of the bilateral insula and claustrum was shown during the perception of actively learned letters than passively learned letters ... [This would suggest that] children were better able to phonologically process letters that they learned by writing than those that they learned by observing an experimenter write ... [and that] writing practice has led to more similar neural representation between printed letters and those letters learned by writing."*

Not only does active performance produce more physical brain activity than passive viewing, as described above, but researchers have also found that prior active performance can lead to higher brain activity even during passive viewing later on, in a sense "carrying over" to make the passive viewing more active within the brain.

Specifically, Calvo-Merino et al. (2006) demonstrated that when someone views another person performing an action, the viewer experiences higher activation in motor areas if they have frequently performed that action themself:

> *"We found greater premotor, parietal, and cerebellar activity when dancers viewed moves from their own motor repertoire, compared to opposite-gender moves that they frequently saw but did not perform."*

The same researchers elaborated more in an earlier paper (Calvo-Merino et al., 2005):

> *"Comparing the brain activity when dancers watched their own dance style versus the other style therefore reveals the influence of motor expertise on action observation.*
>
> *We found greater bilateral activations in premotor cortex and intraparietal sulcus, right superior parietal lobe and left posterior superior temporal sulcus when expert dancers viewed movements that they had been trained to perform compared to movements they had not.*

> *Our results show that this 'mirror system' integrates observed actions of others with an individual's personal motor repertoire, and suggest that the human brain understands actions by motor simulation."*

Persistence of Misconceptions

Why do misconceptions about active and passive learning persist, despite clear intuition and decisive evidence supporting active learning? Several reasons are obvious:

1. Passive learning is more convenient for students and teachers alike. Teachers don't have to spend time and effort implementing learning activities, and students don't have to spend time and effort engaging those activities. Most teachers are happy to lecture about the beautiful intricacies of their field of study and believe that their students are learning, and most students are happy to lean back, relax, pay half attention (if that), and believe that they are learning. (In general, it is always tempting to believe that which is most convenient.)

2. It's easy to mistakenly believe that you have learned a concept well enough to reason and solve problems when you are not actually made to attempt those things. (For the same reason, many people mistakenly believe that they can outrun a bear.)

3. Some teachers resist active learning methods like cold-calling out of fear that it will make students uncomfortable – even though research has shown that cold-calling not only heightens engagement but also increases voluntary participation and comfort over time (Dallimore, Hertenstein, & Platt, 2013).

A fourth, less-obvious reason was discovered by a study (Deslauriers et al., 2019) on Harvard physics classes, which not only measured educational outcomes in active versus passive learning settings, but also measured students' *perceptions* of their learning. As quoted in the study:

> *"Compared with students in traditional lectures, students in active classes perceived that they learned less, while in reality they learned more.*
>
> *Students rated the quality of instruction in passive lectures more highly, and they expressed a preference to have 'all of their physics classes taught this way,' even though their scores on independent tests of learning were lower than those in actively taught classrooms.*
> *...*
> *When students experienced confusion and increased cognitive effort associated with active learning, they perceived this disfluency as a signal of poor learning, while in fact the opposite is true."*

In other words, active learning produced more learning by increasing cognitive activation, but students *mistakenly* interpreted that extra cognitive effort as an indication that they were not learning as well, when in fact the opposite is true. Active learning creates a **desirable difficulty** that makes class feel more challenging but improves learning. Passive learning, on the other hand, promotes an **illusion of competence** in which students (and their teachers) overestimate their knowledge because they are not made to exercise it.

That said, it's dubious whether students and teachers – at their core – truly believe these misconceptions, given that their behavior quickly changes to an active learning model of working through practice questions with direct and immediate feedback when they are held accountable for demonstrating learning, such as when preparing for a standardized test like an AP exam.

In a similar way, one might question whether students – at their core – actually dislike active learning. Tharayil et al. (2018) note that students' perceptions of active learning have not been consistent across studies and are often positive:

> *"Although much of the published literature suggests that students often respond positively to active learning strategies (Arce 1994; Armbruster et al. 2009; Carlson and Winquist 2011; Hoffman 2001; Leckie 2001; Oakley et al. 2007; O'Brocta and Swigart 2013; Reddy 2000; Richardson and Birge 1995), there are counterbalancing studies which show mixed student responses (Bacon et al. 1999; Brent and Felder 2009; Goodwin et al. 1991; Hall et al. 2002; Kvam 2000; Rangachari 1991; Wilke 2003) or negative student responses (Lake 2001; Yadav et al. 2011)."*

People often do not look forward to workouts, yet they don't mind it once they actually begin exercising, and then they feel proud of their efforts afterwards. If active learning is similar to physical activity, then students may prefer passive to active learning simply because it's easier (a typical human behavior), but they may feel much more engaged during active learning (whereas passive learning is pretty boring for students), and they may feel better about themselves after doing actual work and knowing that they made real progress.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415.

  Importance: *Hundreds of studies have shown that passively consuming educational content leads to significantly worse educational outcomes than active learning, where students are actively performing learning exercises.*

- Dori, Y. J., & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts?. *The journal of the learning sciences, 14*(2), 243-279.

  Importance: *In the early 2000s, MIT solved issues with its physics courses by switching from passive to active learning, which reduced the failure rate by nearly two-thirds.*

- Kersey, A. J., & James, K. H. (2013). Brain activation patterns resulting from learning letter forms through active self-production and passive observation in young children. *Frontiers in psychology, 4*, 567.

  Importance: *Active performance produces more physical brain activity than passive viewing.*

- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current biology, 16*(19), 1905-1910.

  Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. Cerebral cortex, 15(8), 1243-1249.

Importance: *Prior active performance can lead to higher brain activity even during passive viewing later on, in a sense "carrying over" to make the passive viewing more active within the brain.*

- Deakin, J. M., & Cobley, S. (2003). A search for deliberate practice. *Expert performance in sports*, 115-36.

  Importance: *In a study of figure skaters who had been practicing for a similar number of years, the proportion of active practice (relative to passive practice) was a defining attribute separating the elite and non-elite skaters.*

- Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*(39), 19251-19257.

  Importance: *Active learning produces more learning by increasing cognitive activation, but students often mistakenly interpret extra cognitive effort (such as productive struggle and occasional confusion) as an indication that they are not learning as well, when in fact the opposite is true.*

Chapter 11. Deliberate Practice

Summary: *Deliberate practice is the most effective form of active learning. It consists of individualized training activities specially chosen to improve specific aspects of a student's performance through repetition and successive refinement. It is the opposite of mindless repetition. The amount of deliberate practice has been shown to be one of the most prominent underlying factors responsible for individual differences in performance across numerous fields, even among highly talented elite performers. Deliberate practice demands effort and intensity, and may be discomforting, but its long-term commitment compounds incremental improvements, leading to expertise.*

Definition and Importance

| Deliberate vs Non-Deliberate Practice

While active learning leads to significantly better educational outcomes than passive learning, not all active learning strategies are created equal. The most effective type of active learning is **deliberate practice**, which consists of individualized training activities specially chosen to improve specific aspects of a student's performance through repetition and successive refinement.

Deliberate practice is the opposite of mindless repetition, and it has been shown to be one of the most prominent underlying factors responsible for individual differences in performance, even among highly talented elite performers (Ericsson, Krampe, & Tesch-Romer, 1993). K. Anders Ericsson, first author of that study and one of the most influential researchers in the field of human expertise and performance, elaborates further on what it means to engage in deliberate practice (Ericsson, 2006):

> *"The core assumption of deliberate practice (Ericsson, 1996, 2002, 2004; Ericsson et al., 1993) is that expert performance is acquired gradually and that effective improvement of performance requires the opportunity to find suitable training tasks that the performer can master sequentially – typically the design of training tasks and monitoring of the attained performance is done by a teacher or a coach.*
>
> *Deliberate practice presents performers with tasks that are initially outside their current realm of reliable performance, yet can be mastered within hours of practice by concentrating on critical aspects and by gradually refining performance through repetitions after feedback.*

> *Hence, the requirement for concentration sets deliberate practice apart from both mindless, routine performance and playful engagement, as the latter two types of activities would, if anything, merely strengthen the current mediating cognitive mechanisms rather than modify them to allow increases in the level of performance."*

Ericsson offers (2003, pp.72-73) a concrete and familiar example illustrating the distinction between deliberate and non-deliberate practice in the context of music:

> *"As children, many people may have spent a lot of time practicing the piano with modest improvements, or known other people who did. When parents forced them to practice, many piano students would simply play the same piece repeatedly without full concentration on specific aspects of their performance. Under those circumstances the existing performance level becomes only more stable and 'practice' makes it permanent. The relation between current level of performance and the number of hours of 'practice' is weak for this type of beginner (Lehmann, 1997).*
>
> *Successful practice requires identifying specific goals for how to change the performance. ... Most deliberate practice by music students is solitary as they attempt to master specific assignments, often new pieces of music selected by their teachers to be of an appropriate difficulty level. Musicians will encounter difficult passages while mastering a new piece of music. To achieve mastery, the musician first identifies the source of the problem, often by playing the passage in a slow tempo. ... With focused repetitions the pianist will generally reach mastery.*
>
> *Sometimes the pianist will still experience difficulties and work on specific exercises that eventually lead to desired changes. In music, there is a large body of training techniques that have been designed to help musicians develop control over performance and attain the desired speed and dexterity. The use of techniques designed to overcome weaknesses and increase control exemplifies the essence of deliberate practice."*

Below is another example, also offered by Ericsson and colleagues, illustrating deliberate practice the context of athletics (Plant et al., 2005):

> *"...[M]any people know recreational golf and tennis players whose performance has not improved in spite of 20–30 years of active participation. The mere act of regularly engaging in an activity for years and even decades does not appear to lead to improvements in performance, once an acceptable level of performance has been attained (Ericsson, 2002).*
>
> *For example, if someone misses a backhand volley during a tennis game, there may be a long time before the same person gets another chance at that same type of shot. When the chance finally comes, they are not prepared and are likely to miss a similar shot again. In contrast, a tennis coach can give tennis players repeated opportunities to hit backhand volleys that are progressively more challenging and eventually integrated into representative match play.*
>
> *However, unlike recreational play, such deliberate practice requires high levels of concentration with few outside distractions and is not typically spontaneous but carefully scheduled (Ericsson, 1996, 2002). A tennis player who takes advantage of this instruction and then engages in particular practice activities recommended by the teacher for a couple of hours in deeply focused manner (deliberate practice), may improve specific aspects of his or her game more than he or she otherwise might experience after many years of recreational play."*

| Deliberate Practice is Effective, Non-Deliberate Practice is Not

It's important to realize that the effects of deliberate practice hold across a wide variety of domains, not just music and athletics. As summarized by Reeves (2014):

> *"Since the original presentation of deliberate practice, the theory has been tested and applied to a number of domains. Ericsson originally speculated that deliberate practice would be of particular value in domains such as: chess, sports, mathematics, and sciences (Ericsson 1993). He has personally investigated a number of these domains and their relationship to deliberate practice. Domains in which he found significant evidence of deliberate practice being a predictor of elite level of expertise include: problem-solving, dart throwing, rhythmic gymnastics, golf, education, nursing, medical expertise, interpreting, and golf (Ericsson 1993, Ericsson 2000a, Ericsson 2000b, Ericsson 2007a, Ericsson 2007b, Ericsson 2007c, Ericsson 2008a, Ericsson 2008b).*
>
> *Additionally, other academicians have taken this theory and successfully applied it to other domains as well. Most recently, deliberate practice was found to be an effective tool for enhancing microsurgical skills in surgeons (El Tecle 2013) as well as hysteroscopy skills in obstetrics and gynecology residents (Rackow 2012). Outside of medicine, deliberate practice has been shown to be significant in accelerating a wide variety of other skills, including: knowledge development (Pachman 2013), critical thinking skills (Cahill 2012), team sports (Helson 1998), chess (Charness 2005), and advanced writing skills (Kellogg 2009) among others."*

The effectiveness of deliberate practice, and the ineffectiveness of non-deliberate practice, is so strong that metrics of professional experience that combine the two (such as "years of experience") have been found to only weakly predict actual performance – whereas, on its own, the amount of purely deliberate practice is a much stronger predictor. As summarized by Ericsson (2008):

> *"Traditionally, professional expertise has been judged by length of experience, reputation, and perceived mastery of knowledge and skill. Unfortunately, recent research demonstrates only a weak relationship between these indicators of expertise and actual, observed performance. In fact, observed performance does not necessarily correlate with greater professional experience.*
>
> *Expert performance can, however, be traced to active engagement in deliberate practice (DP), where training (often designed and arranged by their teachers and coaches) is focused on improving particular tasks. DP also involves the provision of immediate feedback, time for problem-solving and evaluation, and opportunities for repeated performance to refine behavior."*

Along these lines, Lehtinen et al. (2017) emphasize that in the context of academics in particular, quantity of study time is not by itself a strong predictor of academic improvement – rather, the *quality* of study time is the critical determinant.

> *"It is important to note that it is 'deliberate' practice that matters, not just any practice. For example, Plant, Ericsson, Hill, and Asberg (2005) found that improvement in performance in higher education did not significantly correlate with the amount of time spent studying. It did, however, relate to concentrated learning aimed at specific performance goals."*

Furthermore, Debatin et al. (2023) note that high-quality deliberate practice requires complete individualization, an aspect that is sometimes overlooked even by academics in the field:

> *"...[Some] authors have neglected the most important characteristic of deliberate practice: individualization of practice. Many of the analyzed effect sizes derived from measures that did not assess individualized practice and, therefore, should not have been included in meta-analyses of deliberate practice.*
> *...*
> *In our study of 178 chess players, we found that at a high level of individualization and quality of practice, the effect size of structured practice was more than three times higher than that found at the average level."*

Intuitively, the specific aspects of performance that one student is most in need of refining will generally be different for another student, meaning that the most effective exercises on which to spend practice time will differ from student to student.

Effort is Required

| Deliberate Practice Feels Like Exercising with a Personal Trainer

On Math Academy, students spend the entirety of their time engaged in deliberate practice by solving problems (and receiving feedback) on new topics and topics most in need of review. We intersperse active problem-solving with instruction so that students receive minimum effective doses of information right before they use it to actively solve problems and receive feedback.

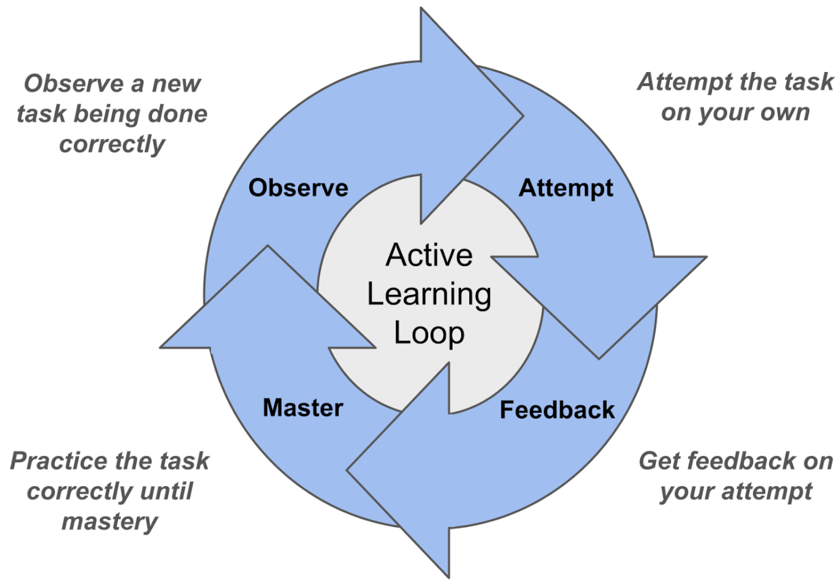
In this way, learning on Math Academy feels like exercising with a personal trainer:

1. The trainer quickly demonstrates an exercise, which you observe.

2. You attempt the exercise, and the trainer corrects anything that is wrong with your form.

3. You continue practicing the exercise, receiving feedback from the trainer, until you are able to complete it comfortably with proper form.

4. The trainer introduces you to a more challenging exercise, and you go back to step 1.



Like exercise with a personal trainer, learning with Math Academy requires effort. Building neural connections takes work, just like building muscle. We will challenge you, but we will not ask you to do anything that you're unprepared for.

| Cycle of Strain and Adaptation

As Ericsson, Krampe, & Tesch-Romer (1993) describe, deliberate practice requires intense, near-maximal-effort training. The goal is to push the limit of one's performance capacity forward during each practice session.

> *"Deliberate practice aimed at improving strength and endurance in sports clearly shows the importance of near maximal effort during practice and the resulting fatigue. Physical activity and exercise produce no benefit unless they are sufficiently intense ... elite athletes train at much higher intensities to improve their performance."*

This creates a continual cycle of strain and adaptation. As Ericsson (2006) elaborates:

> *"Measurable increases in physical fitness do not simply result from wishful thinking. Instead people have to engage in intense aerobic exercise that pushes them well beyond the level of comfortable physical activity if they are to improve their aerobic fitness (Ericsson, 2003a; Ericsson et al., 1993; Robergs & Roberts, 1997).*
> *...*
> *When the human body is put under exceptional strain, a range of dormant genes in the DNA are expressed and extraordinary physiological processes are activated. Over time the cells of the body,*

> *including the brain (see Hill & Schneider, Chapter 37) will reorganize in response to the induced metabolic demands of the activity by, for example, increases in the number of capillaries supplying blood to muscles and changes in metabolism of the muscle fibers themselves.*
>
> *These adaptations will eventually allow the individual to execute the given level of activity without greatly straining the physiological systems. To gain further beneficial increases in adaptation, the athletes need to increase or change their weekly training activities to induce new and perhaps different types of strain on the key physiological systems."*

Even in contexts outside of sports, these adaptations can be detected as physical changes in the brain:

> *"...[A]thletic training involves pushing the associated physiological systems outside the comfort zone to stimulate physiological growth and adaptation (Ericsson, 2001, 2002, 2003a, 2003c, 2003d). Furthermore, recent reviews (Gaser & Schlaug, 2003; Hill & Schneider, Chapter 37; Kolb & Whishaw, 1998) show that the function and structure of the brain is far more adaptable than previously thought possible.*
>
> *Especially, early and extended training has shown to change the cortical mapping of musicians (Elbert, Pantev, Wienbruch, Rockstoh, & Taub, 1995), the development of white matter in the brain (Bengtsson et al., 2005), the development of "turn out" of ballet dancers, the development of perfect pitch, and flexibility of fingers (Ericsson & Lehmann, 1996).*
>
> *In sum, elite performers search continuously for optimal training activities, with the most effective duration and intensity, that will appropriately strain the targeted physiological system to induce further adaptation without causing overuse and injury."*

Discomfort is Required

Deliberate practice requires repeatedly practicing skills that are *beyond* one's repertoire. However, this tends to be more effortful and less enjoyable, which can mislead non-experts to practice within their level of comfort.

For instance, this was observed as a factor differentiating intermediate and expert Gaeilic football players (Coughlan et al., 2014):

> *"Expert and intermediate level Gaelic football players executed two types of kicks during an acquisition phase and pre-, post-, and retention tests. During acquisition, participants self-selected how they practiced and rated the characteristics of deliberate practice for effort and enjoyment.*
>
> *The expert group predominantly practiced the skill they were weaker at and improved its performance across pre-, post- and retention tests. Participants in the expert group also rated their practice as more effortful and less enjoyable compared to those in the intermediate group.*

> *In contrast, participants in the intermediate group predominantly practiced the skill they were stronger at and improved their performance from pretest to posttest but not on the retention test."*

Likewise, as described by Ericsson (2006) in the context of singing:

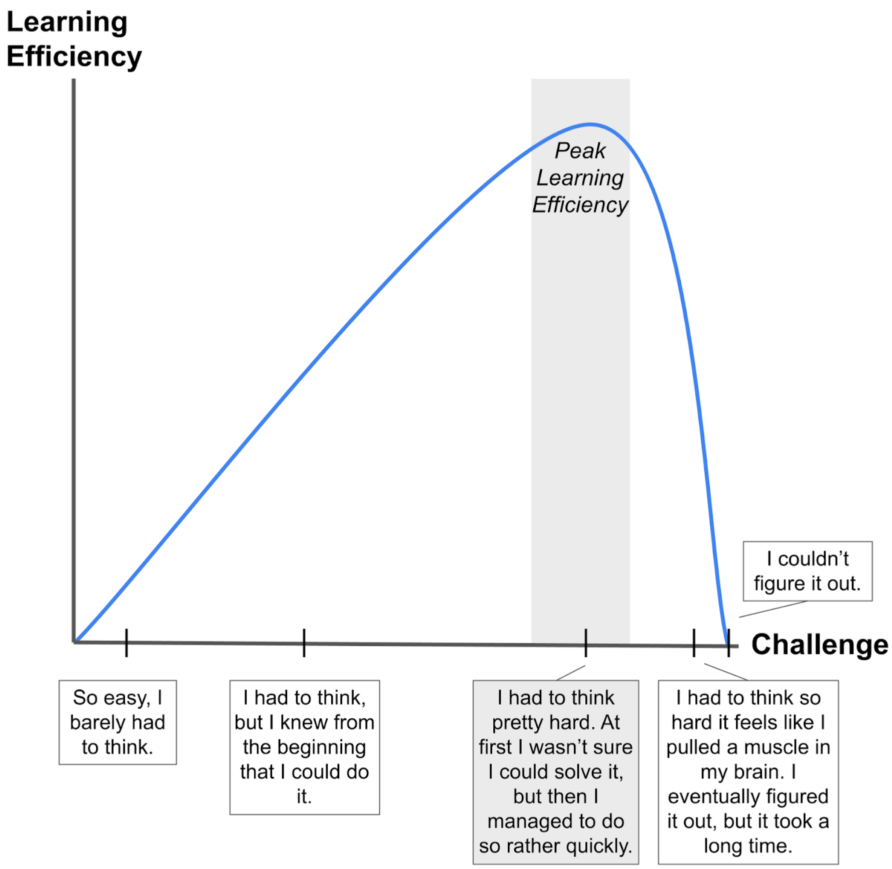
> *"In a recent study of singers Grape, Sandgren, Hansson, Ericsson, and Theorell (2003) revealed reliable differences of skill in the level of physiological and psychological indicators of concentration and effort during a singing lesson.*
>
> *Whereas the amateur singers experienced the lesson as self-actualization and an enjoyable release of tension, the professional singers increased their concentration and focused on improving their performance during the lesson."*

And as Lehtinen et al. (2017) elaborate:

> *"The 'art of deliberate practice' obviously includes the ability and willingness to conduct highly concentrated activities which might be, to some degree, aversive in nature. For example, maximal capacity training in running is demanding and situationally unpleasant even for world-class runners, but it is undeniably a necessary part of running training.*
> *...*
> *However, less experienced individuals like novices tend to focus their practice on more pleasant levels of effort. For example, unexperienced musicians often practice pieces (or parts of pieces) which they have already mastered. They try to avoid errors and failures and they do not challenge their own learning. It is in response to this that trainers/mentors/guides etc. are most valuable.*
> *...*
> *In the realm of mathematics education, a distinction should also be made between routine practice with existing skills and the types of deliberate practice that push students to develop their emerging skills and knowledge structures. ... In geometry learning, Pachman, Sweller, and Kalyuga (2013) ... found that more knowledgeable students even tended to choose achievable rather than difficult problems if they had the opportunity to choose. Training with these geometrical tasks resulted in minimal performance improvements. Only when a deliberate practice model was applied and these more knowledgeable students were presented with designer-selected difficult problems to solve did their skills improve."*

It's a common misconception that maximum-efficiency learning should feel maximally scaffolded, perfectly smooth and easy the whole way through. While this is more true than not, it misses an important nuance: maximum-efficiency learning should feel just-enough scaffolded that the learning tasks are challenging yet still achievable.

This is more obvious in the context of athletics: maximum-efficiency training involves pushing athletes to the brink of their capabilities. At the beginning of a training session, an athlete undergoing maximal-efficiency training will probably not be confident in their ability to successfully perform the training tasks, but they will end up doing so.



When you're developing skills at peak efficiency, you are maximizing the difficulty of your training tasks subject to the constraint that you end up successfully overcoming those difficulties in a timely manner. A noteworthy corollary is that you are also *minimizing* your confidence in your ability to complete the training tasks (again subject to the constraint that you end up successfully completing them in a timely manner).

In that view, confidence is more of a "hindsight" thing than an "in-the-moment" thing. If you feel confident while engaging in maximum-efficiency learning, it's not because the task in front of you seems easy relative to your abilities, but because you've been in situations before where tasks felt challenging relative to your abilities but you've always managed to come out successful.

One can also gain confidence by looking at progress over time. While the amount of progress over any single training session may feel small, consistent deliberate practice will lead to large performance gains over longer periods of time. When a student looks at training activities from months ago (e.g., math problems that felt hard at the time), and these activities now feel much easier than the student remembered, this can provide a large confidence boost.

Long-Term Compounding

| Expertise is the Product of Incremental Improvements Over Time

Lehtinen et al. (2017) are careful to note that a single round of deliberate practice will not result in instant expertise – rather, it is the compounding of these incremental improvements over a longer period of time that lead someone to become an expert:

> *"The formation of expert-like practice activities is not a single event, but a long process in itself. The acquisition of high level competence in complex domains such as mathematics is a laborious process that needs deliberate practice during a number of years."*

Anderson, Reder, & Simon (1998) elaborate further:

> *"...[U]nderstanding of a domain does not come in one fell swoop of insight but is built up bit by bit over time.*
>
> *For example, to say that a student has understood a concept such as fractions means that the student can use that knowledge flexibly in many situations. Thus, the student can figure out how much pizza each of three children will get if they have to share half a pizza; the student will recognize that, when thirty-five people must be transported by busses that each hold twenty people, two buses are required, not one-and-three-quarters; the student can explain why one inverts a fraction to divide by it; and so on. A child does not suddenly acquire the ability to do all of this.*
>
> *The belief in moments of transformation in education is undoubtedly linked to the old belief in developmental psychology that children transit abruptly between stages. ... Instead, as R. S. Siegler documents with great care, development is always gradual and continuous. The same is true of education."*

Consequently, as Ericsson, Krampe, & Tesch-Romer (1993) emphasize, long-term motivation and commitment are essential:

> *"...[D]eliberate practice requires effort and is not inherently enjoyable. Individuals are motivated to practice because practice improves performance. ... Thus, an understanding of the long-term consequences of deliberate practice is important."*

| Motivational Supplements are Not Substitutes for Deliberate Practice

To this end, classroom activities that are enjoyable, collaborative, and non-repetitive (such as group discussions and freeform/unstructured project-based or discovery learning) can sometimes be useful for increasing student motivation and softening the discomfort associated with deliberate practice.

However, it's important to realize that these activities are only *supplements*, not *substitutes*, for deliberate practice. Unlike deliberate practice, they do not directly move the needle on student performance – rather, they "grease the wheels" and reduce psychological friction during the process of deliberate practice. Performance improvements come directly from deliberate practice, but occasional motivational activities can inspire students to continue engaging in deliberate practice over the long term even when it feels difficult and uncomfortable.

Again, this is perhaps most obvious in the contexts of music and athletics:

- Musicians often enjoy fiddling around on their instruments and jamming with friends in a freestyle, creative way. These are fun activities that can enhance motivation and sometimes produce creative ideas that can be integrated into their defining style as artists. But elite musicians know that an even more central component of their practice routine is consistently pushing themselves beyond their repertoire, using intensely focused effort to gain new skills and improve specific areas of weakness through repetition and successive refinement.

- Athletes often enjoy the camaraderie of team bonding activities, which might include "trick shot" competitions, group discussions about team goals and individual expectations, and exchanging stories and perspectives over team dinners. Again, these are fun activities that can help teammates fuel their passion for the game and feel connected to one another. However, elite athletes know that at the end of the day, their performance on the field comes from routinely pushing their physiological and mental limits every day during practice, where they focus intensely on gaining new skills and improving specific areas of weakness through repetition and successive refinement.

As is said about famous basketball player Kobe Bryant (Cacciola, 2020):

> *"At the team's pre-Olympic training camp the following summer, Bryant was the first player to arrive. In fact, he beat most members of the coaching staff – and was getting in workouts at 5:30 a.m. … The foundation for all of Bryant's feats – the 81-point game, the scoring titles, the series-clinching jump shots, the three championships he had already won with the Lakers – was*

> *his work ethic and desire. The spectacular was rooted in the mundane, in the monotony of hard labor."*

The overall takeaway from this chapter is that by engaging in deliberate practice on Math Academy, you will gain the ability to reason coherently and solve problems in levels of math that you were previously unable to comprehend. But as any personal trainer will tell you: if you want to achieve your goals, you have to put in the work. Excellence is the product of effective training over a long period of time, and effective training requires intense effort focused in areas beyond your repertoire.

Misinterpretations of Deliberate Practice

Because "deliberate practice" has effectively become synonymous with "maximally effective practice," people will sometimes refer to a form of practice as "deliberate practice" simply because they personally believe it to be maximally effective. Consequently, whenever a form of practice is claimed to be "deliberate practice," the claim should not be taken at face value. After thorough investigation, it is not uncommon to find that someone is cutting corners on one of the two requirements of deliberate practice – "mindful" and "repetition" – and then resisting objective, quantifiable measurement of their performance that would expose the ineffectiveness of their practice. This is not always intentional – it may be an honest mistake – but regardless, it is something to watch out for.

To emphasize:

- Deliberate practice is not mindless repetition. If you're doing the same thing over and over again, then you're doing deliberate practice wrong. Deliberate practice is about making performance-improving adjustments on every single repetition. Any individual adjustment is small and yields a small improvement in performance, but when you compound these small changes over a massive number of cycles, you end up with massive changes and massive gains in performance. None of this happens if you're mindlessly doing the same thing over and over again without making adjustments.

- Likewise, even if you're mindful during practice, you can't skimp on repetition and still call it "deliberate practice." Deliberate practice necessitates a high volume of action-feedback-adjustment cycles in every single training session. Otherwise, the compounding doesn't happen. Any activity that throttles the number of these cycles cannot be described as deliberate practice.

Many heated debates in math education stem from these misinterpretations of deliberate practice.

- Mindless repetition, doing the same thing over and over again without making performance-improving adjustments, is not deliberate practice.
- Likewise, any activity that throttles the volume of action-feedback-adjustment cycles is not deliberate practice (e.g., excessively challenging problems or think-pair-share).

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review, 100*(3), 363.

  Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. *The Cambridge handbook of expertise and expert performance, 38*(685-705), 2-2.

  Importance: *Deliberate practice consists of individualized training activities specially chosen to improve specific aspects of a student's performance through repetition and successive refinement. It is one of the most prominent underlying factors responsible for individual differences in performance, even among highly talented elite performers. However, it requires effort and is not inherently enjoyable. Individuals are motivated to practice because practice improves performance. Thus, an understanding of the long-term consequences of deliberate practice is important.*

- Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Academic emergency medicine, 15*(11), 988-994.

  Importance: *The effectiveness of deliberate practice, and the ineffectiveness of non-deliberate practice, is so strong that metrics of professional experience that combine the two (such as "years of experience") have been found to only weakly predict actual performance – whereas, on its own, the amount of purely deliberate practice is a much stronger predictor.*

- Lehtinen, E., Hannula-Sormunen, M., McMullen, J., & Gruber, H. (2017). Cultivating mathematical skills: From drill-and-practice to deliberate practice. *ZDM, 49*, 625-636.

  Importance: *It is "deliberate" practice that matters, not just any practice. Deliberate practice requires highly concentrated activities which might be, to some degree, aversive in nature. However, less experienced individuals like novices tend to focus their practice on more pleasant levels of effort. They practice things that they have already mastered, they try to avoid errors and failures, and they do not challenge their own learning.*

For instance, in geometry learning, Pachman, Sweller, and Kalyuga (2013) found that more knowledgeable students even tended to choose achievable rather than difficult problems if they had the opportunity to choose. Training with these geometrical tasks resulted in minimal performance improvements. Only when a deliberate practice model was applied and these more knowledgeable students were presented with designer-selected difficult problems to solve did their skills improve.

Likewise, Plant, Ericsson, Hill, and Asberg (2005) found that improvement in performance in higher education did not significantly correlate with the amount of time spent studying. It did, however, relate to concentrated learning aimed at specific performance goals.

The formation of expert-like practice activities is not a single event, but a long process in itself. The acquisition of high level competence in complex domains such as mathematics is a laborious process that needs deliberate practice during a number of years.

Additional Resources

- Parrish, S. The Ultimate Deliberate Practice Guide: How to Be the Best. *FS.blog.*
- Clear, J. The Beginner's Guide to Deliberate Practice. *JamesClear.com.*
- Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard business review, 85*(7/8), 114.

Chapter 12. Mastery Learning

Summary: *By organizing its curriculum into a knowledge graph that keeps track of prerequisite relationships between topics, Math Academy is able to implement mastery learning, a strategy in which students demonstrate proficiency on prerequisites before advancing. While even loose approximations of mastery learning have been shown to produce massive gains in student learning, mastery learning faces limited adoption due to clashing with traditional teaching methods and placing increased demands on educators. Math Academy implements true mastery learning at a fully granular level, which requires fully individualized instruction and is only attainable through one-on-one tutoring.*

Mastery Learning is Underused

One of our main paradigms is **mastery learning**, also proposed by Bloom (1968), in which students must demonstrate proficiency on prerequisite topics before moving on to more advanced topics. True mastery learning at a fully granular level requires fully individualized instruction, which is only attainable through one-on-one tutoring.

There are methods by which a single teacher can loosely approximate mastery learning, such as Bloom's Learning For Mastery (LFM) strategy and Keller's Personalized System of Instruction (PSI). As Kulik, Kulik, & Bangert-Drowns (1990) summarize:

> *"In both LFM and PSI courses, material to be learned is divided into short units, and students take formative tests on each unit of material (Bloom, 1968; Keller, 1968). ... Lessons in LFM courses are teacher presented, and students move through these courses at a uniform, teacher-controlled pace. Lessons in PSI courses are presented largely through written materials, and students move through these lessons at their own rates."*

However, as Bloom (1984) discovered when characterizing the two-sigma problem, a single teacher practicing mastery learning with 30 students could only produce a one-sigma effect size as compared to the two-sigma effect size of individual tutoring. And while numerous studies reproduced the finding that even loose approximations of mastery learning (managed manually by a single teacher) produce substantial learning gains, most studies were unable to reproduce gains as strong as one sigma (the average effect size was about 0.5 standard deviations) (Kulik, Kulik, & Bangert-Drowns, 1990):

> *"The data show that mastery learning programs have positive effects on student achievement. On the average, such programs raise final examination scores by about 0.5 standard deviations, or from the 50th to the 70th percentile, in colleges, high schools, and the upper grades of elementary schools. Although PSI and LFM strategies differ on several points and the two teaching methods have been studied in distinct ways, studies of PSI and LFM report similar results. PSI raised examination scores by an average of 0.48 standard deviations; LFM raised examination scores by an average of 0.59 standard deviations."*

Unfortunately, despite producing well-documented learning gains in classrooms, even loose approximations of mastery learning were not widely adopted as they faced opposition for deviating from traditional convention and requiring more effort from teachers and administrators (Sherman, 1992). (It's true that a minority of teachers now attempt some degree of differentiated instruction, but this is not the same as true mastery learning, which holds all students to the same standard and is completely individualized.)

As lamented by John Gilmour Sherman (1992), who was a co-creator, researcher, and practitioner of Keller's Personalized System of Instruction (PSI):

> *"Some PSI courses have been prohibited in spite of their success. I know of several colleagues who were given 'cease and desist' orders. Some are names prominent in the literature, their courses effective, according to objective data.*
>
> *I experienced this also. Avoiding a frontal attack, the chairman of the Psychology Department at Georgetown declared by fiat that something on the order of 50% of class time must be devoted to lecturing. By reducing the possibility of self-pacing to zero, this effectively eliminated PSI courses.*
>
> *He issued this order on the grounds that in the context of lecturing 'it is the dash of intellects in the classroom that informs the student.' No data were presented on this point! The spectacle of purporting to defend scholarship while deciding the merits of instructional methods by assertion is silly.*
>
> *The troubling aspect of all these cases was that data played no part in the decisions. It is disturbing when one has to wonder whether research on the education process makes any difference."*

As Buskist, Cush, & DeGrandpre (1991) elaborate, mastery learning methods like PSI were shot down because they threatened the traditional educational establishment:

> *"The first and most important task of any institution is self-preservation. Once in place, its primary goal must be to hold its ground or, if possible, advance. If this goal is not met, then its demise is eminent. PSI poses certain implications that threaten the preservation of the educational establishment and its guardians (department heads, deans, and academic vice-presidents). According to Keller,*
>
> *'Suppose that a system such as PSI were to be given official approval for adoption throughout the educational scale from top to bottom... With every student given individual treatment, when would formal education start?... What would happen to the classroom hour, the college quarter,*

> *the semester, or the academic year?... Who would win the scholarships and prizes? Who would make Phi Beta Kappa? Who would be the class valedictorian?... What changes would be made in the payment of tuition when the period of course attendance varied? How would a course of study be defined?'*
>
> *In other words, the major impediment to educational reform is the educational system itself. That is perhaps why all major efforts at educational reform in this century have been directed at renovating curricula and not at changing how teachers teach (see, e.g., Skinner, 1984; McGovern, 1990).*
>
> *Revamping curricula requires no revamping of the educational establishment. The curriculum changes, but that is all. Courses are still taught by lecture within a term's time. Grade distributions still approximate the normal curve and students enroll in upper division courses without first mastering more fundamental material. Students still take about four years, give or take a term, to finish what higher education demands of them.*
>
> *The Keller Plan runs contrary to this strategy. It is a bold attempt to change how we teach, despite what we teach. In an entirely PSI-based college, students might finish in two years, maybe sooner, and learn a good deal more. Imagine what would happen if the entire educational system were PSI-based: huge numbers of people, most still in the throes of puberty, might be graduating from college-an unsettling thought for many educators. Indeed, PSI represents a threat to the educational system and its guardians.*
>
> *PSI simply does not fit well into our modern educational system. The longstanding tradition of teaching by lecture has accumulated inertia that has proven difficult to dislodge. In the interests of self-preservation, the educational establishment has backed reforms favoring what teachers teach instead of how teachers teach."*

Implementing Mastery Learning

The traditional convention is to march students through a linear sequence of topics according to a predetermined schedule (like that shown below). Any students who get lost are continually asked to learn new topics despite not having mastered the prerequisites. As a result, those students spend class time learning little to nothing and developing a general distaste for math.

Unit 3: Adding & Subtracting Fractions

This Week

Module 3.1: Adding and Subtracting Fractions and Whole Numbers

- Adding Fractions and Whole Numbers Using Models
- Adding Fractions and Whole Numbers
- Subtracting Fractions and Whole Numbers Using Models
- Subtracting Fractions and Whole Numbers

Next Week

Module 3.2: Adding and Subtracting Fractions

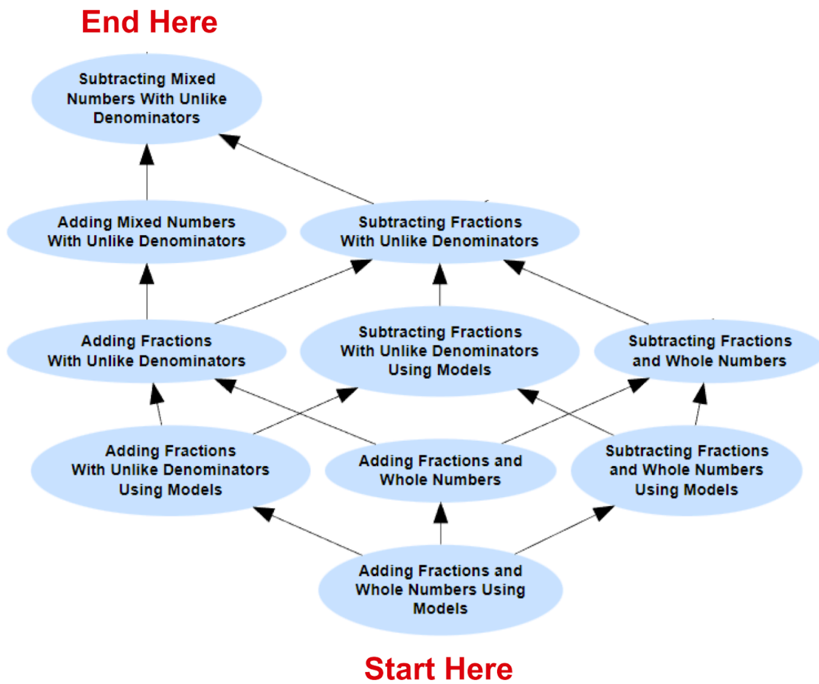
- Adding Fractions With Unlike Denominators Using Models
- Subtracting Fractions With Unlike Denominators Using Models
- Adding Fractions With Unlike Denominators
- Subtracting Fractions With Unlike Denominators

Week After Next

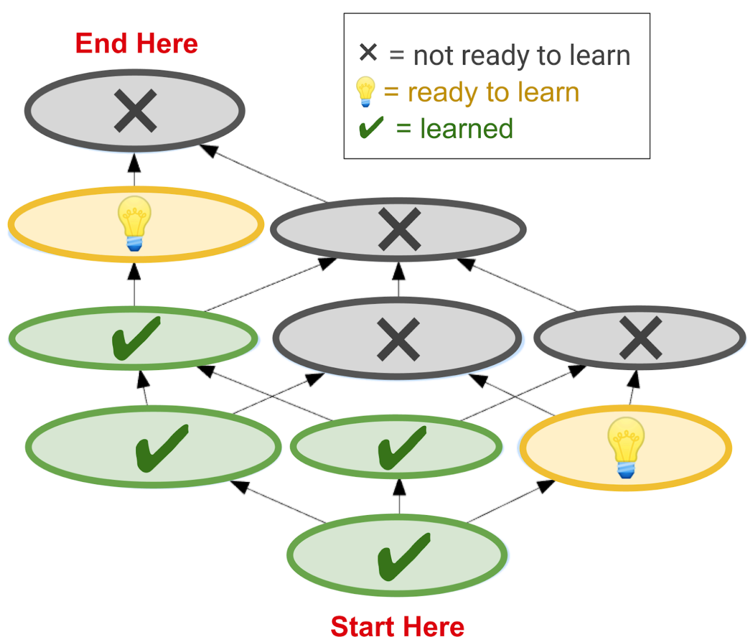
Module 3.3: Adding and Subtracting Mixed Numbers

- Adding Mixed Numbers With Unlike Denominators
- Subtracting Mixed Numbers With Unlike Denominators

Math Academy, however, implements true mastery learning at a fully granular level. We accomplish this by organizing topics into a **knowledge graph** that shows all the topics and **prerequisite relationships** between them. In the knowledge graph below, arrows point from simpler prerequisite topics to more advanced "post"-requisite topics.



By overlaying a student's progress on the knowledge graph, we can identify the topics that they are ready to learn – that is, the topics for which they have demonstrated a sufficient level of proficiency on the prerequisites. We only serve students lessons on these topics. If a student gets stuck on a topic, they can try again another day, but in the meantime, they are allowed to learn other topics that don't depend on the problematic one.



It is infeasible for a single teacher, who can only teach one topic at a time, to manually support true mastery learning across a class full of students who all have different learning profiles. As researchers have discovered, knowledge profiles vary immensely even across students in the same grade (Pedersen, et al., 2023):

> *"our results suggest that nearly 38% and 49% of students in grade four and eight classrooms may either struggle to understand 'grade-level' content or have already mastered the content, respectively."*

However, Math Academy's fully-automated interactive lessons support de-synchronized learning where different students are simultaneously taught different topics. With a software system that emulates the decisions of an expert tutor, we are able to provide fully individualized instruction at scale and achieve true mastery learning at a fully granular level.

Knowledge Frontier as Zone of Proximal Development

Mastery learning is closely related to Vygotsky's **Zone of Proximal Development**, which refers to the range of tasks that a student is able to perform while supported, but cannot do on their own. Students maximize their learning when they are completing tasks within this range.

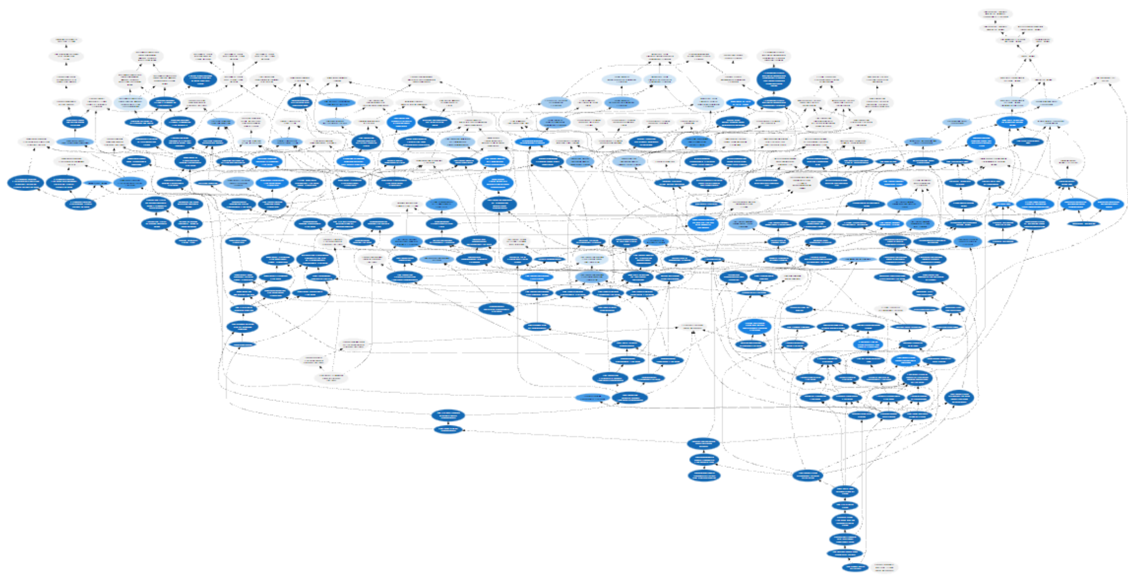
In the context of an adaptive learning system, a student's zone of proximal development coincides with their **knowledge frontier** or **edge of mastery**, the set of new topics for which they have mastered the prerequisites. Selecting new learning tasks from a student's knowledge frontier can lead to drastic improvements in learning.

For instance, another learning platform has reported (Zou et al., 2019) that even when a teacher has access to student performance data chooses a new topic that they believe is appropriate for the class, a student is about 3-4x as likely to be successful in mastering that topic if it lies along their knowledge frontier (as opposed to residing beyond their frontier). This led the authors to conclude that

> *"...[S]imply providing teachers with data is not always sufficient for good instructional decision making. ... Optimizing learning outcomes requires correct teaching decisions that lead students on the right path, based on the student's ZPD [Zone of Proximal Development]."*

A student's knowledge frontier can be visualized as the edge of their **knowledge profile**, which, loosely speaking, represents how "developed" their mathematical brain is. Every time they learn a new math topic, it's as if they grow a new brain cell and connect it to existing brain cells. Initially, this new brain cell is weak and requires frequent nurturing, but over time it becomes strong and requires less frequent care.

For instance, a knowledge profile for a second-semester calculus student is visualized below. Learned topics are shaded (with darker shading indicating that more successful practice has been completed), and arrows between topics represent prerequisite relationships.

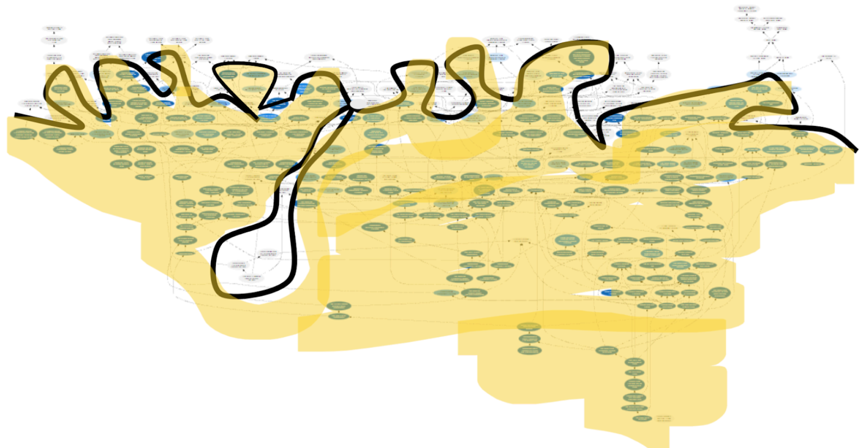


(Note that this visualization only shows a "subsystem" within the student's full mathematical brain – there are several hundred topics in the calculus course, but there are thousands of topics in Math Academy's entire mathematical curriculum, which spans elementary school through university-level math.)

The **knowledge frontier** is the edge of the knowledge profile. It separates what the student knows from what they don't know.

- The student knows all the simpler topics below their knowledge frontier. That is, they know all the prerequisites, the prerequisites of the prerequisites, and so on.

- However, they do not know any of the advanced topics at or above their knowledge frontier.



When a student first starts on the Math Academy system, we begin with a **placement diagnostic**, a **dynamic assessment** that quickly estimates their knowledge frontier. Following the diagnostic, whenever a student is served new lessons, those lessons always cover topics that are on the student's estimated knowledge frontier, and the estimated knowledge frontier quickly becomes more accurate as the student completes those lessons.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6), 4-16.

  Importance: *The average tutored student performed better than 98% of the students in a traditional class, an effect size of two sigmas (standard deviations). However, a single teacher practicing mastery learning with 30 students could only produce a one-sigma effect size as compared to the two-sigma effect size of individual tutoring.*

- Kulik, C. L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of educational research*, 60(2), 265-299.

  Importance: *Numerous studies reproduced the finding that even loose approximations of mastery learning (managed manually by a single teacher) produce substantial learning gains, though generally not as high as one sigma (the average effect size was about 0.5 standard deviations).*

- Sherman, J. G. (1992). Reflections on PSI: Good news and bad. *Journal of Applied Behavior Analysis*, 25(1), 59.

  Buskist, W., Cush, D., & DeGrandpre, R. J. (1991). The life and times of PSI. *Journal of Behavioral Education, 1*, 215-234.

  Importance: *Despite producing well-documented learning gains in classrooms, Keller's Personalized System of Instruction (a method by which a single teacher can loosely approximate mastery learning) was not widely adopted as it faced opposition for deviating from traditional convention and requiring more effort from teachers and administrators.*

- Pedersen, B., Makel, M. C., Rambo-Hernandez, K. E., Peters, S. J., & Plucker, J. (2023). Most mathematics classrooms contain wide-ranging achievement levels. *Gifted Child Quarterly,*

67(3), 220-234.

Importance: *Knowledge profiles vary immensely even across students in the same grade: nearly 38% and 49% of students in grade four and eight classrooms may either struggle to understand "grade-level" content or have already mastered the content, respectively.*

- Zou, X., Ma, W., Ma, Z., & Baker, R. S. (2019). Towards helping teachers select optimal content for students. In *Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part II 20* (pp. 413-417). Springer International Publishing.

  Importance: *Another learning platform has reported that even when a teacher has access to student performance data and chooses a new topic that they believe is appropriate for the class, a student is about 3-4x as likely to be successful in mastering that topic if it lies along their knowledge frontier (as opposed to residing beyond the frontier).*

Chapter 13. Minimizing Cognitive Load

Summary: *When the cognitive load of a learning task exceeds a student's working memory capacity, the student experiences cognitive overload and is not able to complete the task. Math Academy avoids cognitive overload by finely scaffolding content with numerous small steps: each lesson is broken up into several "knowledge points" of increasing difficulty, each containing a worked example and requiring the student to demonstrate mastery on practice problems before proceeding to the next knowledge point. Our content is about 10x more finely scaffolded than what you'd find elsewhere. This makes learning accessible to all students regardless of their working memory capacity. Scaffolding is gradually removed as students progress, ensuring sustained learning without dependence on supports.*

The Learning Staircase

Learning is like climbing a staircase. Each step is a learning task – the higher the step, the more advanced the topic is. At the top of the staircase is **higher-order thinking** such as critical thinking and problem solving. However, different students have different stair-climbing abilities, and many students never make it to the top because they get stuck at individual stairs that are too tall for them to climb.

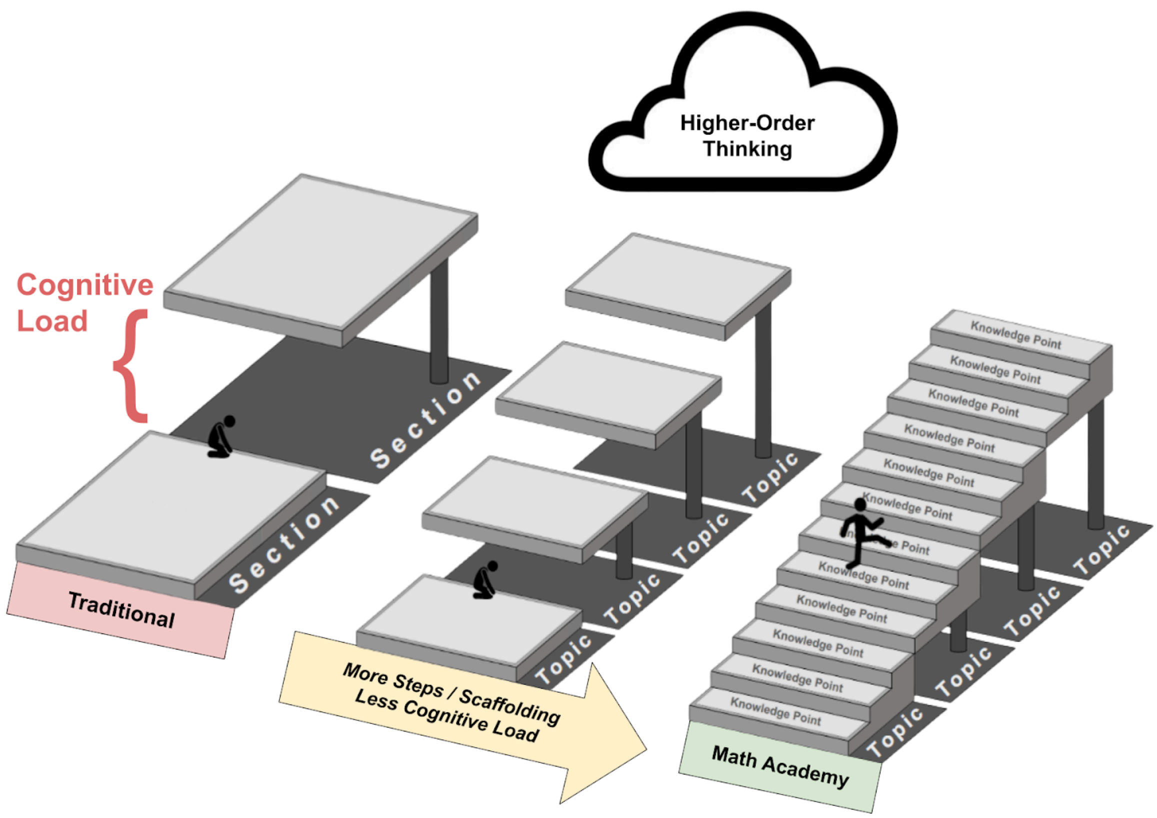
Math Academy's solution is to split individual stairs into even smaller stairs so that all students can climb them. The smaller we make the individual stairs, the more students can climb all the way to the top.

For instance, a typical calculus textbook might consist of 100 steps (10 chapters × 10 sections in each chapter). But in our calculus course, we have about 1000 steps (~300 topics × 3-4 **knowledge points** or stages of increasing difficulty per topic). In other words, our content is about 10x more finely scaffolded than what you'd find elsewhere.

In technical terms, we are **minimizing cognitive load**. Cognitive load refers to the amount of **working memory** that is required to complete a task. Working memory consists of limited-capacity, limited-duration short-term memory storage along with capabilities for organizing, manipulating, and generally "working" with the information stored in short-term memory.

In the staircase analogy, the height of each step represents cognitive load. Different students have different working memory capacities, and if the cognitive load of a learning task exceeds a student's working memory capacity, then the student will not be able to complete the task due to **cognitive overload**.

Cognitive overload has massive negative ramifications for students: not only has working memory capacity been shown to predict performance in mathematical problem solving (Swanson & Beebe-Frankenberger, 2004), but perhaps shockingly, it has also been shown to be a better predictor than IQ when predicting a young student's future academic success (Alloway & Alloway, 2010). By minimizing the cognitive load and avoiding cognitive overload, we make learning accessible to many students for whom it would otherwise be insurmountable.

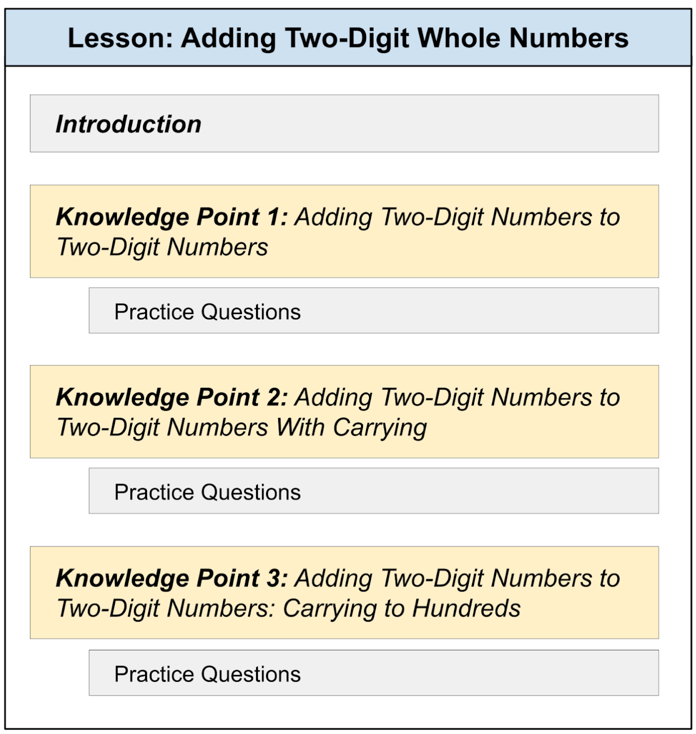


Math Academy maintains this high level of scaffolding even when teaching higher-order thinking. In our multi-part problems, students explore challenging, complex problem contexts one part at a time, and each part leverages an individual skill that they have previously learned in an earlier topic. This way, we fully "split up the staircase" as students climb from practicing individual skills in isolation to combining skills in novel higher-order problem contexts.

Micro-Scaffolding

Even within individual knowledge points, we take additional measures to minimize cognitive load. Each knowledge point starts with a demonstration or **worked example**, which has been shown by numerous studies (see Sweller, 2006 for a review) to reduce cognitive load and help students develop a baseline mental framework or schema when their level of understanding is initially low.

After the worked example, students solve problems that are similar to the worked example, only progressing onto the next knowledge point once they have demonstrated mastery of the previous one. This way, we avoid asking students to solve problems that overload their working memory capacity.

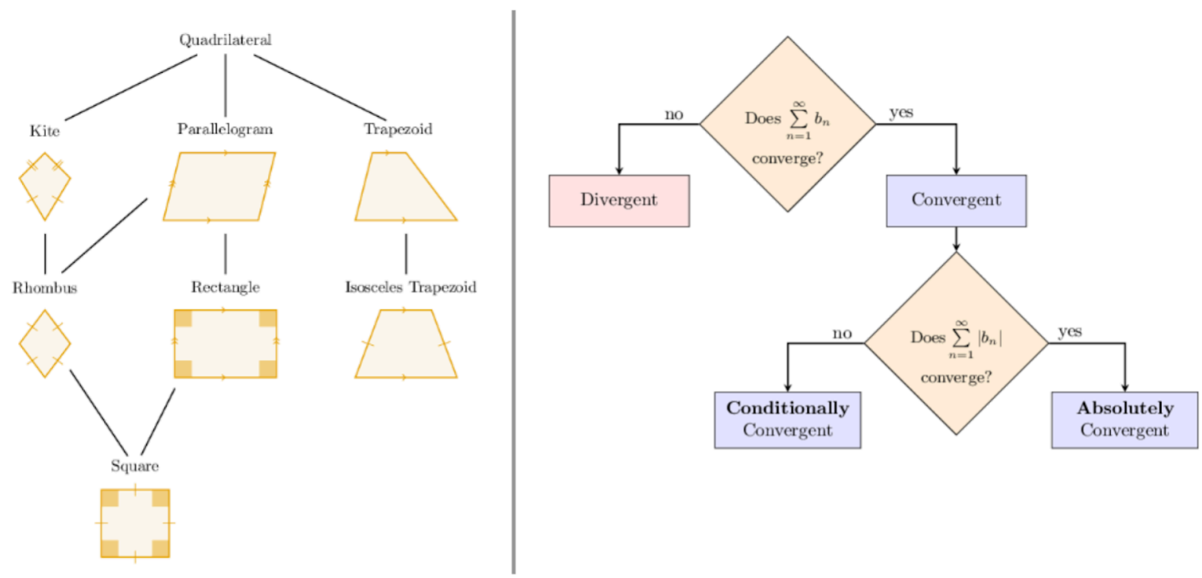


In the explanations of worked examples and practice questions, we leverage **subgoal labeling** by grouping steps into meaningful units. This minimizes the number of chunks of information that students need to store in their working memory, thereby reducing cognitive load. Additionally, subgoal labeling has been shown to help students grasp the structure of the problem, thereby enabling the learning to transfer to novel problems in the same category (Catrambone, 1995).

We also leverage **dual-coding theory** by including visualizations and diagrams when possible to help students develop **mental images**. In addition to helping students make connections that they can use to recall information and consolidate information into chunks, this also helps students avoid cognitive overload by distributing cognitive load more evenly between two subsystems within the working memory system: the **phonological loop**, which stores verbal information, and the **visuo-spatial sketchpad**, which stores visual imagery (Baddeley, 1983).

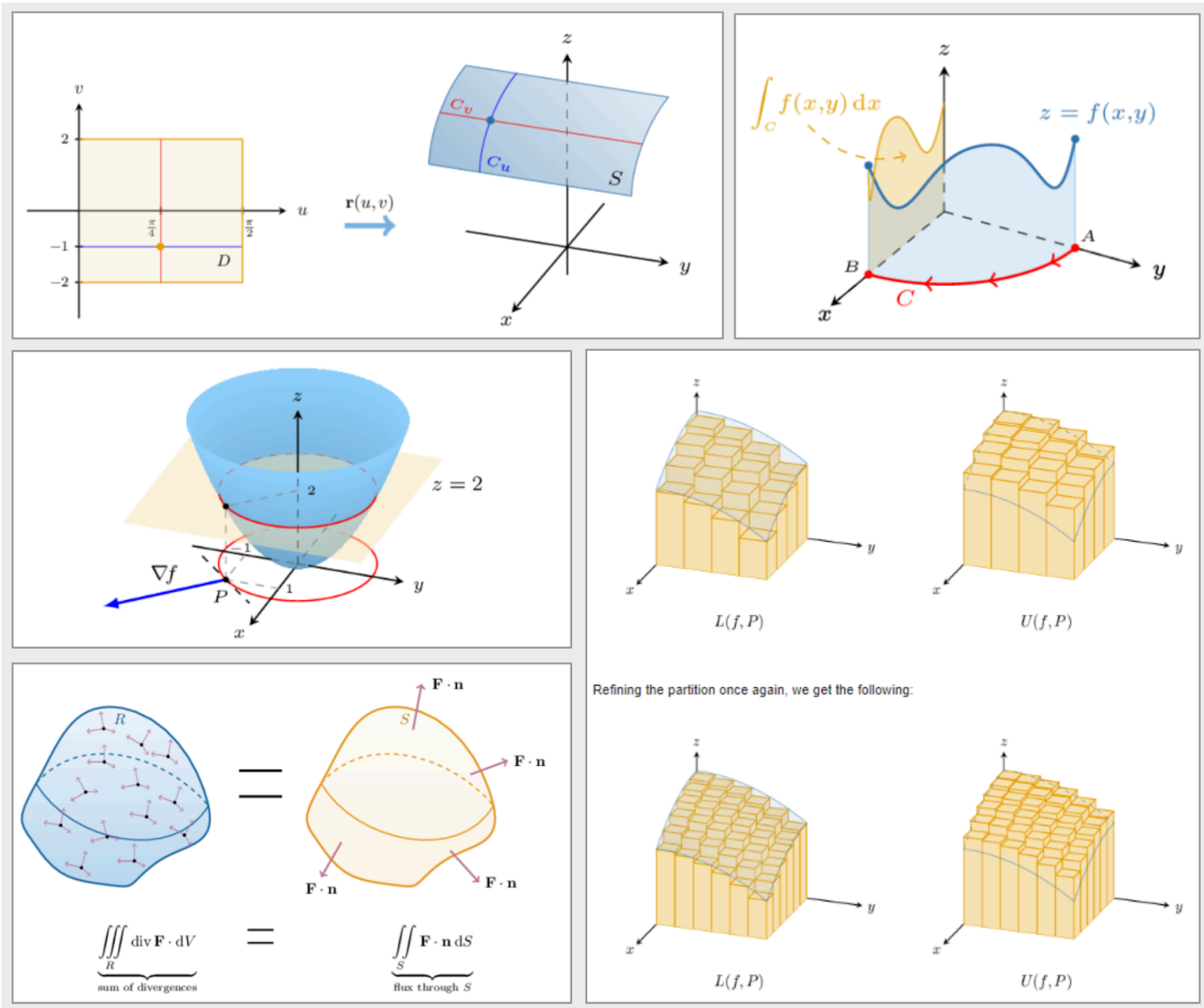
It's worth noting that unlike most other educational programs, Math Academy makes heavy use of visualizations and diagrams throughout the entire math curriculum – not just in elementary mathematics, but all the way through university-level subjects.

For instance, just as we use flowcharts to help students classify shapes in elementary mathematics, we also use flowcharts to help students classify series in Calculus:

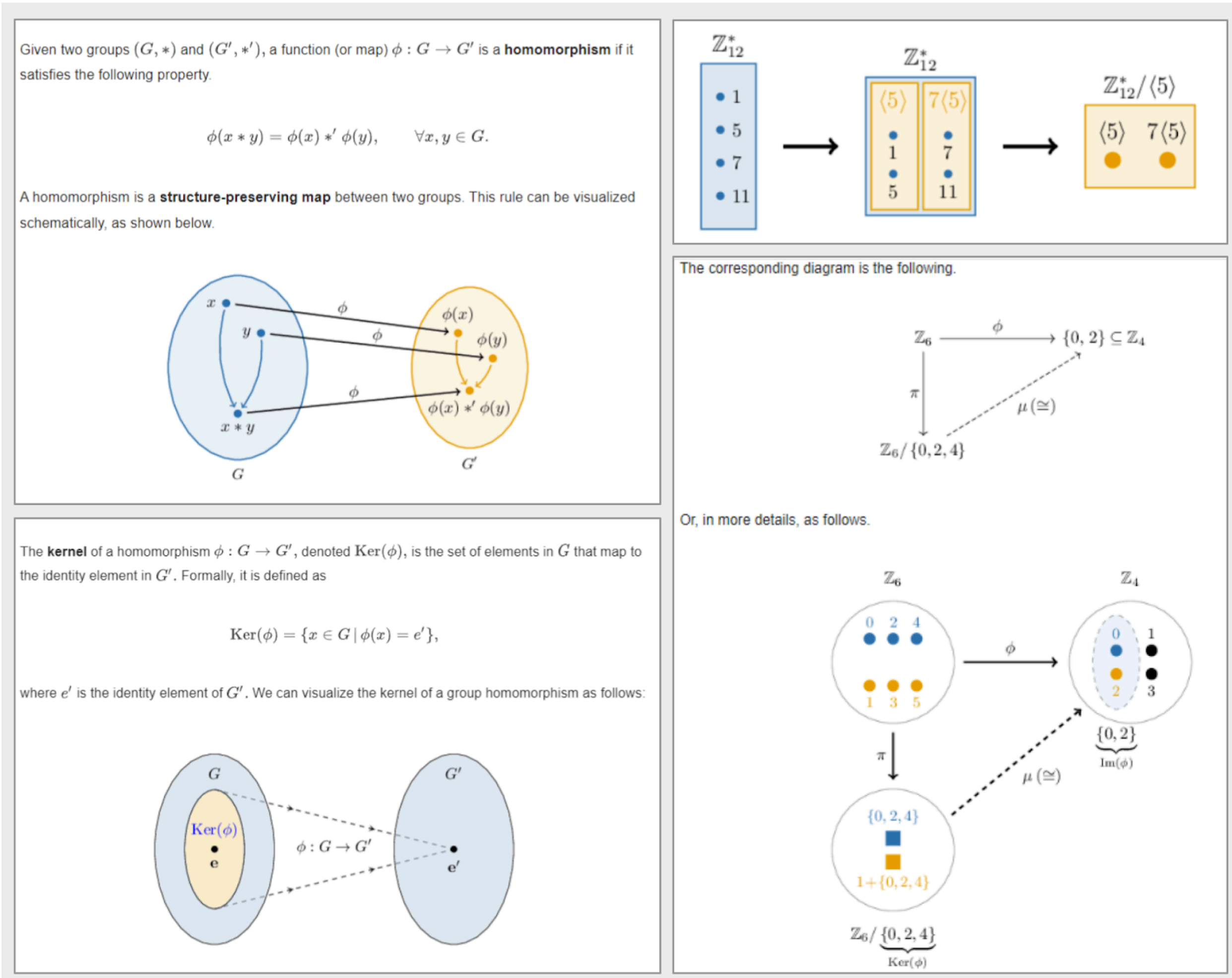


The visualization doesn't stop at Calculus. It continues all the way through more advanced university-level courses like Multivariable Calculus – even Abstract Algebra, an upper-level math-major course about the "structure" of abstract mathematical objects, whose textbooks and lectures are usually associated with dense, dry, image-less strings of symbols.

Sample Images – Multivariable Calculus



Sample Images – Abstract Algebra



The Expertise Reversal Effect

While it's important to use scaffolding to minimize cognitive load when students are learning new material, it's also important to gradually strip away the scaffolding as they become comfortable with that material so that the scaffolding does not become a crutch. This phenomenon is known as the **expertise reversal effect**: the instructional techniques that promote the most learning in beginners, promote the least learning in experts, and vice versa.

On Math Academy, after a student completes an initial lesson on a topic, we gradually strip away scaffolding during later reviews. While we scaffold lessons by having students solve questions that are similar to worked examples (one worked example at a time), we mix up review problems so that it is not obvious which worked example is the best reference. This encourages students to solve review problems without referring to the worked examples – and while they can go back to the lesson and dig up a similar example for reference if they get really stuck on a review problem, even students who do this must reason about the structure of their problem to match it to helpful reference material.

We also continually quiz our students on the material that they have learned – and during quizzes, no scaffolding is provided. Quizzes are quick and frequent, but each quiz covers a wide variety of previously learned material. Additionally, quizzes are timed, and students are unable to refer back to lessons for reference.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of educational psychology, 96*(3), 471.

  Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of experimental child psychology, 106*(1), 20-29.

  Importance: *Cognitive overload has massive negative ramifications for students: not only has working memory capacity been shown to predict performance in mathematical problem solving, but perhaps shockingly, it has also been shown to be a better predictor than IQ when predicting a young student's future academic success.*

Chapter 14. Developing Automaticity

Summary: *Automaticity is the ability to perform low-level skills without conscious effort. Analogous to a basketball player effortlessly dribbling while strategizing, automaticity allows individuals to avoid spending limited cognitive resources on low-level tasks and instead devote those cognitive resources to higher-order reasoning. In this way, automaticity is the gateway to expertise, creativity, and general academic success. However, insufficient automaticity, particularly in basic skills, inflates the cognitive load of tasks, making it exceedingly difficult for students to learn and perform.*

Importance of Automaticity

| Autonomicity Frees Up Working Memory

An essential yet often-overlooked part of minimizing cognitive load is developing **automaticity** on basic skills – that is, the ability to execute low-level skills without having to devote conscious effort towards them. Automaticity is necessary because it frees up limited working memory to execute multiple lower-level skills in parallel and perform higher-level reasoning about the lower-level skills.

As a familiar example, think about all the skills that a basketball player has to execute in parallel: they have to run around, dribble the basketball, and think about strategic plays, all at the same time. If they had to consciously think about the mechanics of running and dribbling, they would not be able to do both at the same time, and they would not have enough brainspace to think about strategy.

This extends to academics as well. As described by Hattie & Yates (2013, pp.53-58):

> *"You cannot comprehend a 'big picture' if your mind's energies are hijacked by low-level processing. Continuity is broken. The goal shifts from understanding the total context to understanding the immediate word before you. ... If you read connected text (such as sentences) at any pace under 60 wpm, then understanding what you read becomes almost impossible.*
> *...*
> *Many [students] arrive at school with a lack of automaticity within their basic sound-symbol functioning. With a minimal level of phonics training, they may be able to fully identify letters, verbalise sound symbol relationships, and read isolated words through sheer effort. But, if the pace*

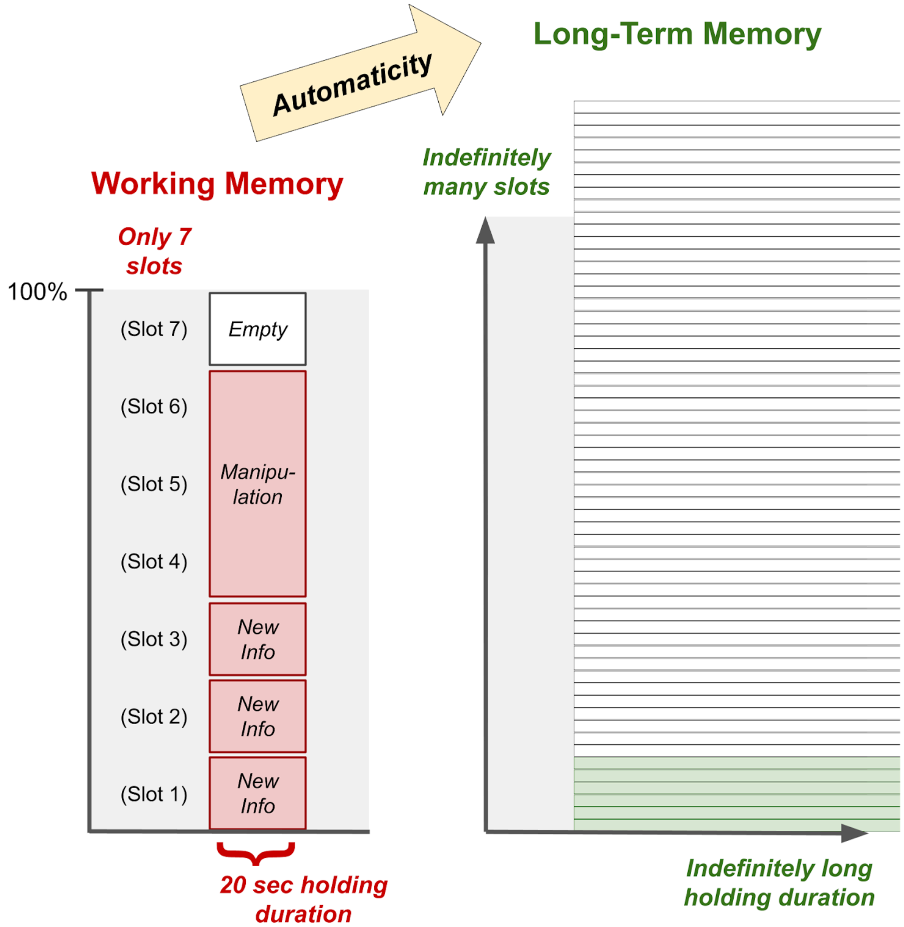
> *of processing is not brought up to speed, through intensive self-directed practice, reading for understanding will remain beyond grasp.*
> ...
> *A well-replicated finding is that students who present with difficulties in mathematics by the end of the junior primary years show deficits in their ability to access number facts with automaticity. Such deficits stymie further development in this area, often with additional adverse consequences such as students experiencing lack of confidence, lack of enjoyment, and feelings of helplessness."*

| Working Memory is Limited, but Long-Term Memory is Not

Unfortunately, working memory has such limited capacity that most people can only hold a handful of pieces of new information simultaneously in their heads (spanning about 7 digits, or more generally 4 chunks of coherently grouped items), and only for about 20 seconds as the memory degrades from decay or interference (Miller, 1956; Cowan, 2001; Brown, 1958; Ricker, Vergauwe, & Cowan, 2016). And that assumes they aren't needing to perform any mental manipulation of those items – if they do, then fewer items can be held due to competition for limited processing resources (Wright, 1981). This severe limitation of the working memory when processing novel information is known as the **narrow limits of change principle** (Sweller, Ayres, & Kalyuga, 2011).

An intuitive analogy by which to understand the limits of working memory is to think about how your hands place a constraint on your ability to hold and manipulate physical objects. You can probably hold your phone, wallet, keys, pencil, notebook, and water bottle all at the same time – but you can't hold much more than that, and if you want to perform any activities like sending a text, writing in your notebook, or uncapping your water bottle, you probably need to put down several items.

In the same way, your working memory only has about 7 slots for new information, and once those slots are filled, if you want to hold more information or manipulate the information that you are already holding, you have to clear out some slots to make room.



In particular, you can't solve a problem if you can't fit all its pieces in your working memory. This means that if a student doesn't achieve automaticity on lower-level skills, it doesn't even matter how well the teacher scaffolds a new skill – they won't be able to do it. And even for tasks within a student's cognitive capacity, it has been shown that a heavy cognitive load drastically increases the likelihood of errors (Ayres, 2001).

When you develop automaticity on a skill or piece of information, however, you can use it without it occupying a slot in your working memory. Instead, the skill is stored in your **long-term memory**, where indefinitely many things can be held for indefinitely long without requiring cognitive effort.

As Anderson (1987) summarizes, automaticity can effectively turn long-term memory into an extension of short-term memory:

> *"Chase and Ericsson (1982) showed that experience in a domain can increase capacity for that domain. Their analysis implied that what was happening is that storage of new information in long-term memory, became so reliable that long-term became an effective extension of short-term memory."*

For emphasis, we quote Chase and Ericsson (1982) directly:

> *"The major theoretical point we wanted to make here is that one important component of skilled performance is the rapid access to a sizable set of knowledge structures that have been stored in directly retrievable locations in long-term memory. We have argued that these ingredients produce an effective increase in the working memory capacity for that knowledge base."*

| Expertise Requires Automaticity

Automaticity is the mental capacity that differentiates experts from beginners, a phenomenon that has been thoroughly studied in various contexts including the game of chess. As summarized by Ross (2006):

> *"...[A] typical grandmaster has access to roughly 50,000 to 100,000 chunks of chess information [Gobet & Simon, 1998]. A grandmaster can retrieve any of these chunks from memory simply by looking at a chess position, in the same way that most native English speakers can recite the poem 'Mary had a little lamb' after hearing just the first few words."*

As elaborated by Gobet & Simon (1998):

> *"...[S]kill in playing chess depends both on (a) recognizing familiar chunks in chess positions while playing games, and (b) exploring possible moves and evaluating their consequences. ... Expert memory, in turn includes slowly acquired structures in long-term memory (retrieval structures, templates) that augment short-term memory with slots (variable places) that can be filled rapidly with information about the current position."*

Indeed, as Benjamin Bloom noted (1986) while identifying automaticity as a key theme in his own research on talent development, automaticity was described as the "hands and feet of genius" as early as the 19th century:

> *"Our talent development studies support the 1899 research of Bryan and Harter who were concerned with the development of automaticity in expert Morse Code telegraphers. They most eloquently described the benefits of automaticity as an outcome of the learning process.*
>
> *'The learner must come to do with one stroke of attention what now requires half a dozen, and presently in one still more inclusive stroke, what now requires thirty-six. He must systematize the work to be done and must acquire a system of automatic habits corresponding to the system of tasks. When he has done this he is master of the situation in his [occupational or professional]*

> *field. … Finally, his whole array of habits is swiftly obedient to serve in the solution of new problems. Automatism is not genius, but it is the hands and feet of genius.'"*

It's important to realize that automaticity goes beyond simple familiarity. If you truly "know" something, then you should be able to access and leverage that information both quickly and accurately. If you can't, then you're just "familiar" with it. And when learning hierarchical bodies of knowledge – whether it be math, chess, a sport, or an instrument – it's important to truly know things, not just be familiar with them. Why? Because you can't build on familiarity. That's what the term "shaky foundations" refers to. You can only build on a solid foundation of knowledge.

To help students develop automaticity (and, consequently, expertise in mathematics), Math Academy requires students to practice each skill until they have reached a sufficient level of mastery. Students start at their edge of knowledge (not their edge of "familiarity") and are not pushed forward along learning paths until they have mastered the prerequisite skills. Additionally, to help consolidate skills into long-term memory after mastery, skills are continually reviewed into the future through a systematic method called spaced repetition (which is described later in this document).

Case Study: Computing Exponents With vs Without Automaticity on Multiplication and Addition Facts

To convey the importance of automaticity, it helps to walk through a case study in which we observe a problem being solved by students who have different levels of automaticity in their underlying skills. As we will see, a student's overall learning experience can vary drastically depending on their level of automaticity.

Suppose that we have three different students – Otto, Rica, and Finn – whose names are chosen to represent their respective levels of automaticity.

- *Otto* has developed *full automaticity* on multiplication facts and procedures.

- *Rica* doesn't know her multiplication facts – she *recalculates* them from scratch. She is able to carry out multiplication procedures, but she isn't fully comfortable with them and has to proceed slowly, writing every single step down.

- *Finn*, likewise, doesn't know his multiplication facts – but he doesn't know his addition facts either, so he uses *finger-counting* for everything. He is not at all comfortable with multiplication procedures.

These students are each given a lesson on cubes of numbers. After an explanation of what it means to cube a number, and a demonstration with a worked example, they're each given a problem to practice on their own: compute 4^3. Let's observe the thought processes (both reasoning and emotions) as each of these students solves the problem.

Otto is so comfortable with his multiplication and addition facts that he solves the problem in 10 seconds in his head. He feels it was easy, is excited to try another, and can't wait for harder problems like cubing negative numbers, decimal numbers, and fractions.

- *$4^3 = 4 \times 4 \times 4$. I know $4 \times 4 = 16$, easy, and then $16 \times 4 =$ … well that's $10 \times 4 = 40$ and $6 \times 4 = 24$, together making $40 + 24 = 64$. Done, easy! What's next?*

Rica solves the problem in 2 minutes, but her answer is not correct. She takes another 2 minutes to correct the mistake but gets tired and wants to take a break before moving on to the next problem. She's not looking forward to harder problems.

- *$4^3 = 4 \times 4 \times 4$. What's 4×4? I don't know, let's compute it.*
  - *4×4 is the same as $4 + 4 + 4 + 4$, which is … well, $4 + 4 = 8$, plus 4 is 12, plus 4 is 16.*
- *Where was I? Oh right, $4 \times 4 = 16$ and then $16 \times 4 =$ … ugh, gotta go through that multiplication procedure.*
  - *Put the 16 on top, then $\times$ 4 on bottom, and now we carry out the procedure. First $4 \times 6 = 6 + 6 + 6 + 6$, count that up to get $6 + 6 = 12$, plus 6 is 18, plus 6 is 22. Write down 2, carry another 2. Then $4 \times 1 = 4$, add the carried 2, write down 6.*
- *Done. Result is 62. Oh wait, the teacher says that's close but not quite right. Fine, let's try this again.*
- (Rica repeats the entire procedure above and this time gets a result of 64.)
- *Great, teacher says that 64 is right. I know there are more problems to do but that one was kind of hard and I'm tired. Teacher, can I take a break and do the next one later?*

Finn takes 10 minutes to solve the problem, but his answer is not correct. He tries again for another 10 minutes but makes a different mistake. The teacher has to sit with him for another 10 minutes to carry him through the problem. By the time Finn is done with the problem, it has almost been a full class period. He is totally exhausted and overwhelmed and dreads doing the rest of the homework.

- *$4^3 = 4 \times 4 \times 4$. What's 4×4? I don't know, let's compute it.*
    - *4×4 is the same as $4 + 4 + 4 + 4$, which is … ugh, gotta count all this up. This is annoying.*
        - *Start at 4, then 4 more is 5, 6, 7, 8.*
        - *Start at 8, then 4 more is 9, 10, 11, 12.*
        - *Start at 12, then 4 more is 12, 13, 14, 15.*
    - *Phew, that took a while, but now I have $4 + 4 + 4 + 4 = 15$. Why was I doing that, again? Oh right, I was really doing $4 \times 4 = 15$.*
- *Wait, we're not even done yet. I did $4 \times 4 = 15$, but that was because I wanted to do $4 \times 4 \times 4$. Okay so now I need to do 15×4. Ew, that's going to be even harder. I don't like this. But fine, let's do it.*
    - *15×4 is the same as $15 + 15 + 15 + 15$, and those are big numbers so I need to line it up on paper.*
        - *Put 15 at the top, then another 15 below, then another 15, then another 15.*
        - *Let's add the right column:*
            - *Start at 5, then 5 more is 6, 7, 8, 9, 10.*
            - *Start at 10, then 5 more is 10, 11, 12, 13, 14.*
            - *Start at 14, then 5 more is 15, 16, 17, 18, 19.*
        - *Write down 9, carry the 1, then add down the left column: start at 1, then 1 more is 2, then 3, then 4, then 5. Write down the 5, we have 59.*
- *Answer is 59. Glad that's over. That took forever. Oh wait, the teacher says that's wrong. Noooo… do I have to do this whole thing over again?! This is way too much work.*

- (Finn repeats the entire procedure above and this time gets a result of 66, which is still incorrect. He is getting very noticeably frustrated and his teacher sits down with him to go through his work. They find and fix several errors together and arrive at the correct result of 64.)

- *I can't do any more of this today. I'm too tired. I hate math, and my teacher gives me way too much work. And the next problem looks harder, and there are even more on the homework! This is terrible. Class is almost over so I'm just going to zone out until the bell rings.*

This case study demonstrates that the more automaticity a student has on their lower-level skills,

- the easier they will find it to acquire new higher-level skills,

- the more quickly and independently they will be able to execute those skills,

- the better they will feel about the learning process as a whole, and

- the more excited they will be to continue learning more advanced material.

Students who develop automaticity will feel empowered, while students who do not will feel overwhelmed and defeated.

Automaticity, Creativity, and Higher-Level Thinking

| Automaticity is Necessary for Creativity

The relationship between automaticity and creativity is commonly misunderstood. Some people think that automaticity and creativity are opposite and competing forces: supposedly, because automaticity requires repeated practice, it turns students into mindless robots, whereas to leverage the power of human creativity, one needs to break free from that robotic mindset. This line of reasoning might sound alluring – and even convenient, since students often don't enjoy the repeated practice that's required to develop automaticity – but there's one problem: it's completely false.

In reality, automaticity is a necessary component of creativity. The whole purpose of automaticity is to *reduce* the amount of bandwidth that the brain must allocate to robotic tasks, thereby freeing up cognitive resources to engage in higher-level thinking. If a student does not develop automaticity, then they will have to consciously think about every low-level action that they perform, which will exhaust their cognitive capacity and leave no room for high-level creative thinking.

As a concrete example, consider what is typically considered one of the most creative activities: writing. Effective writing requires a frictionless pipeline from ideas in one's mind to words on paper. If a writer had to consciously think about spelling, grammar, word definitions, transitions between sentences, when to make a new paragraph, etc, they would become bogged down in low-level robotic tasks and would have no mental bandwidth to think about high-level creative details like vivid imagery, logical cohesiveness, and emotions evoked by various phrases and ideas.

Indeed, the importance of automaticity is documented by researchers in the field of writing education (Kellogg & Whiteford, 2009):

> *"Serious, effective composition is at once a severe test of memory, language, and thinking ability … it depends on the author's ability to manage the burdensome demands made on working memory by the task of written composition.*
> *…*
> *[T]he necessary coordination and control cannot succeed without reducing the relative demands that planning, generation, and reviewing make on working memory. The writer cannot flexibly and adaptively coordinate planning, generating, and reviewing when the needs of any single process consume too many available resources. The writer cannot be mindful of the whole while struggling with the parts."*

What's more, this view is supported by an overwhelming amount of research over at least the past half-century:

> *"Empirical support for the importance of working memory resources, especially executive attention, in the development of advanced writing skills is strong. First, a measurement of overall working memory capacity in college students correlates with their writing performance (Ransdell & Levy, 1996). Vanderberg and Swanson (2007) extended such findings by discovering that it is individual differences in central executive capacity that reliably accounts for variability in writing skills among 10th graders in high school. Controlled executive attention, rather than the storage of representations, is most critical in explaining individual differences in skill. Converging experimental results show that distracting executive attention with a concurrent task of remembering six digits disrupts both the quality and fluency of text composition (Ransdell, Levy, & Kellogg, 2002).*
>
> *The advancement of writing skills from beginner to advanced levels depends on the availability of adequate working memory resources and the capacity to allocate them appropriately to planning, sentence generation, and reviewing. McCutchen (1996) reviewed a large body of evidence in support of this view. For example, children's fluency in generating written text is limited until they master the mechanical skills of handwriting and spelling (Graham, Berninger, Abbott, & Whitaker,*

> *1997). Learning the mechanics of writing to the point that they are automatic during primary school years is necessary to free the components of working memory for planning, generating, and reviewing. Mastery of handwriting and spelling is also a necessary condition for writers to begin to develop the control of cognition, emotion, and behavior that is needed to sustain the production of texts as adolescents (Graham & Harris, 2000).*
>
> *Revision is constrained or even nonexistent in developing writers because of working memory limitations. Revision requires detecting a problem, diagnosing its cause, and finding an appropriate way to correct it (Flower et al., 1986). If revision fails because of working memory limitations, as opposed to knowledge of what revision entails, then providing cues to detect problems in the text should benefit revision, because writers can then devote resources solely to diagnosis and solution. Cuing in fact does improve the revision of even college students (Hacker, Plumb, Butterfield, Quathamer, & Heineken, 1994).*
>
> *As Beal (1996) observed, very young writers have trouble even seeing the literal meaning of their texts. The beginning author focuses on his or her thoughts not on how the text itself reads. Maintaining the author's ideas in working memory requires much, if not all, of the available storage and processing capacity of working memory in during childhood and early adolescence. This prevents the student from reading the text carefully and maintaining a clear representation of what it actually says that is independent of what the author intended to say."*

| Automaticity is Necessary for Higher-Level Thinking

The same reasoning applies to mathematics. In order to operate at higher levels of mathematical thinking and abstract thought, it's necessary to have developed automaticity at the lower levels. Consider the following realization from a skeptic-turned-convert principal (Brown, Roediger, & McDaniel, 2014, pp.44-45):

> *"What about Principal Roger Chamberlain's initial concerns about practice quizzing at Columbia Middle School – that it might be nothing more than a glorified path to rote learning? When we asked this question after the study was completed, he paused for a moment to gather his thoughts.*
>
> *'What I've really gained a comfort level with is this: for kids to be able to evaluate, synthesize, and apply a concept in different settings, they're going to be much more efficient at getting there when they have the base of knowledge and the retention, so they're not wasting time trying to go back and figure out what that word might mean or what that concept was about. It allows them to go to a higher level.'"*

To put it bluntly, according to Lehtinen et al. (2017):

> *"Fluency in basic arithmetic tasks and number combination skills has proved to be crucial for later mathematical learning and weaknesses in automatization of these skills is characteristic of mathematically disabled children."*

Allen-Lyall (2018) elaborates further:

> *"Internalized facts allow for efficient mental computations that make easier multi-step problem solving or recognizing and making connections between mathematical concepts, such as multiplication and division, ratio comparison, fraction equivalencies, or exploration of object relationships in the world of geometry (Chapin & Johnson, 2006; National Research Council, 2005).*
>
> *...*
>
> *When one internalizes multiplication facts, less brainpower is required to perform tasks that require more complex or successive arithmetic manipulations (Geary, 1999; Geary, Saults, Liu, & Hoard, 2000). Flexible thinking and conceptual leaps between mathematical concepts are possible when products are not computed using successive addition or determined by visual inspection of tables or charts (Royer, 2003). The relationship between factors and products becomes a point of departure into more challenging mathematics. Beginning every new mathematical step forward with a return to multiplication as repeated addition or reliance upon visual assistance may interrupt intuitive mathematical thinking (Goswami, 2008).*
>
> *Fluid mental computations are thwarted by the needs of working memory necessarily allocated to ascertaining the product of two factors or, conversely, the factors of a particular product. Memorizing facts reduces cognitive load, allowing for working memory to better allocate resources when processing number relationships required by more complex mathematics (Goswami, 2008; LeFevre, DeStefano, Coleman & Shanahan, 2005)."*

| Automaticity is a Gatekeeper to Mathematical Literacy and Academic Success

In a broader scope, Allen-Lyall (2018) also explains how automaticity on math facts is a gatekeeper to mathematical literacy, which in turn impacts future academic and career prospects:

> *"Extending beyond successful school mathematics performance, broader options for college study and employment opportunity become increasingly likely when one feels confident in one's mathematical thinking and is able to demonstrate solid achievement (Atweh & Clarkson, 2001; Marsh & Hau, 2004; Valero, 2004; Williams & Williams, 2010).*
>
> *For myriad reasons, facts acquisition becomes an educational gatekeeper to true mathematical literacy. Consequently, helping children to be successful with this seemingly small element of early mathematics learning truly matters in a world rife with challenges requiring the mathematical communication of ideas between and within fields (D'Ambrosio & D'Ambrosio, 1994; Thomas, 2001)."*

As other researchers have discovered, the impact on academic achievement begins immediately: students who are slow on their basic math facts begin falling behind their faster peers as soon as multi-digit arithmetic (Joy Cumming & Elkins, 2010).

> *"Profiles of children based on latency performance on the fact bundles were clustered. The slowest cluster reported use of counting strategies on many bundles; the fastest cluster reported use of retrieval or efficient-thinking strategies. Cluster group was the best predictor of performance on multidigit tasks. Addition fact accuracy contributed only for tasks without carrying, and grade level was not significant. Analysis by error type showed most errors on the multidigit sums were*

> *due to fact inaccuracy, not algorithmic errors. The implication is that the cognitive demands caused by inefficient solutions of basic facts made the multidigit sums inaccessible."*

In retrospect, beliefs that paint a false dichotomy between automaticity and creativity are not only factually incorrect, but amusingly ironic. Such beliefs suggest that de-emphasizing repetition promotes creativity as a skill for life success – when in reality, it causes students to perpetually spend mental bandwidth on low-level tasks that they could have (through repetition) learned to do automatically, thereby limiting their capacity for higher-level and creative mathematical thinking, as well as their future academic and career prospects.

Neuroscience of Automaticity

At a physical level in the brain, automaticity involves developing strategic neural connections that reduce the amount of effort that the brain has to expend to activate patterns of neurons.

Researchers have observed this in functional magnetic resonance imaging (fMRI) brain scans of participants performing tasks with and without automaticity (Shamloo & Helie, 2016). When a participant is at wakeful rest, not focusing on a task that demands their attention, there is a baseline level of activity in a network of connected regions known as the default mode network (DMN). The DMN represents background thinking processes, and people who have developed automaticity can perform tasks without disrupting those processes:

> *"The DMN is a network of connected regions that is active when participants are not engaged in an external task and inhibited when focusing on an attentionally demanding task … at the automatic stage (unlike early stages of categorization), participants do not need to disrupt their background thinking process after stimulus presentation: Participants can continue day dreaming, and nonetheless perform the task well."*

When an external task requires lots of focus, it inhibits the DMN: brain activity in the DMN is reduced because the brain has to redirect lots of effort towards supporting activity in task-specific regions. But when the brain develops automaticity on the task, it increases connectivity between the DMN and task-specific regions, and performing the task does not inhibit the DMN as much:

> *"…[S]ome DMN regions are deactivated in initial training but not after automaticity has developed. There is also a significant decrease in DMN deactivation after extensive practice.*
>
> *…*
>
> *The results show increased functional connectivity with both DMN and non-DMN regions after the development of automaticity, and a decrease in functional connectivity between the medial prefrontal cortex and ventromedial orbitofrontal cortex. Together, these results further support the hypothesis of a strategy shift in automatic categorization and bridge the cognitive and*

> *neuroscientific conceptions of automaticity in showing that the reduced need for cognitive resources in automatic processing is accompanied by a disinhibition of the DMN and stronger functional connectivity between DMN and task-related brain regions."*

In other words, automaticity is achieved by the formation of neural connections that promote more efficient neural processing, and the end result is that those connections reduce the amount of effort that the brain has to expend to do the task, thereby freeing up the brain to simultaneously allocate more effort to background thinking processes.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63*(2), 81.

  Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences, 24*(1), 87-114.

  Importance: *Human working memory has such limited capacity that most people can only hold a handful of pieces of new information simultaneously in their heads: about 7 digits, or more generally 4 chunks of coherently grouped items.*

- Ayres, P. L. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology, 26*(2), 227-248.

  Importance: *Even for tasks within a student's cognitive capacity, a heavy cognitive load drastically increases the likelihood of errors.*

- Ross, P. E. (2006). The expert mind. *Scientific American, 295*(2), 64-71.

  Importance: *A typical grandmaster has access to roughly 50,000 to 100,000 chunks of chess information. A grandmaster can retrieve any of these chunks from memory simply by looking at a chess position, in the same way that most native English speakers can recite the poem "Mary had a little lamb" after hearing just the first few words.*

- Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6*(3), 225-255.

  Importance: *Skill in playing chess depends both on (a) recognizing familiar chunks in chess positions while playing games, and (b) exploring possible moves and evaluating their consequences. Expert memory, in turn* *includes slowly acquired structures in long-term memory (retrieval structures, templates) that augment short-term memory with slots (variable places) that can be filled rapidly with information about the current position.*

- Bloom, B. S. (1986). Automaticity: "The Hands and Feet of Genius." *Educational leadership, 43*(5), 70-77.

  Importance: *Automaticity was a key theme in Benjamin Bloom's research on talent development and was described as the "hands and feet of genius" as early as the 19th century.*

- Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250-266.

  Importance: *In the field of writing, the importance of automaticity is supported by an overwhelming amount of research over at least the past half-century. Serious, effective composition places burdensome demands on working memory. In order to free the components of working memory for planning, generating, and reviewing, students must learn the mechanics of writing to the point that they are automatic during primary school years. The writer cannot be mindful of the whole while struggling with the parts.*

- Allen-Lyall, B. (2018). Helping students to automatize multiplication facts: A pilot study. *International Electronic Journal of Elementary Education, 10*(4), 391-396.

  Importance: *Facts acquisition is an educational gatekeeper to true mathematical literacy. Internalized facts allow for efficient mental computations that make easier multi-step problem solving or recognizing and making connections between mathematical concepts. Memorizing facts reduces cognitive load, allowing for working memory to better allocate resources when performing tasks that require more complex or successive manipulations.*

- Joy Cumming, J., & Elkins, J. (1999). Lack of automaticity in the basic addition facts as a characteristic of arithmetic learning problems and instructional needs. *Mathematical Cognition, 5*(2), 149-180.

  Importance: *The impact of automaticity on academic achievement begins immediately: students who are slow on their basic math facts begin falling behind their faster peers as soon as multi-digit arithmetic.*

- Shamloo, F., & Helie, S. (2016). Changes in default mode network as automaticity develops in a categorization task. *Behavioural Brain Research, 313*, 324-333.

  Importance: *Automaticity is achieved by the formation of neural connections that promote more efficient neural processing, and the end result is that those connections reduce the amount of effort that the brain has to expend to do the task, thereby freeing up the brain to simultaneously allocate more effort to background thinking processes.*

Chapter 15. Layering

Summary: *Layering is the act of continually building on top of existing knowledge – that is, continually acquiring new knowledge that exercises prerequisite or component knowledge. This causes existing knowledge to become more ingrained, organized, and deeply understood, thereby increasing the structural integrity of a student's knowledge base and making it easier to assimilate new knowledge. To reap the benefits of layering, Math Academy moves students forward to new topics immediately after they demonstrate mastery of prerequisites, and employs a highly-connected curriculum where new topics exercise and build on earlier topics.*

Facilitation and Structural Integrity

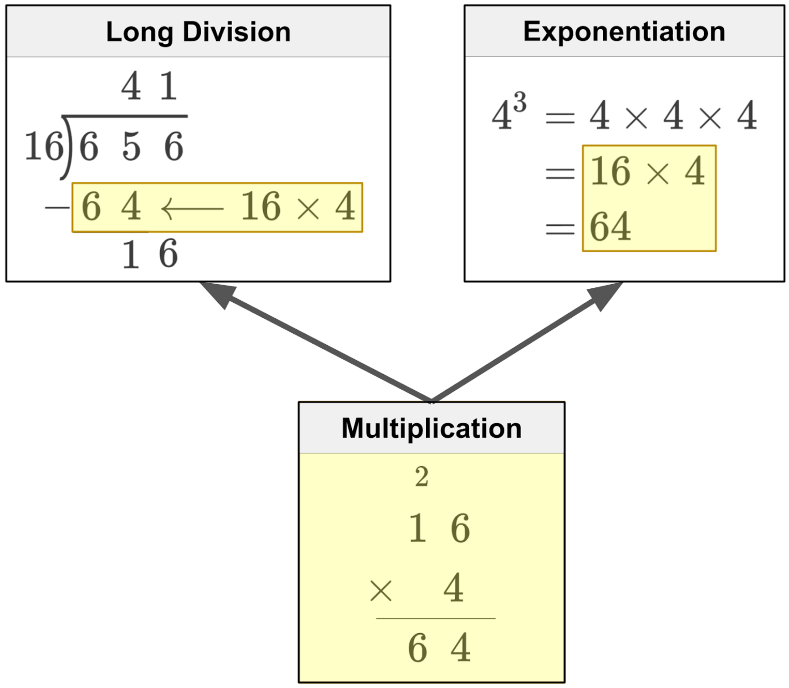
| Facilitation

As students learn progressively more advanced material, they reinforce and deepen their foundational knowledge. In academic literature, this is known as **facilitation**: when a new task exercises knowledge learned in a prior task, learning can be facilitated in two ways:

- (*Retroactive Facilitation*) The new task can restore memory of prior knowledge to the same extent as identical repetition of the prior task, leading to long-lasting retention (Ausubel, Robbins, & Blake, 1957; Arzi, Ben-Zvi, & Ganiel, 1985).

- (*Proactive Facilitation*) Knowledge acquired during the prior task can improve the acquisition of knowledge that is specific to the new task (Arzi, Ben-Zvi, & Ganiel, 1985).

As a concrete example, consider that multiplication is a component skill in both long division and exponentiation. When you learn long division, you also practice the more basic skill of multiplication, which not only reinforces your knowledge of multiplication (*retroactive facilitation*), but also makes it easier for you to learn how to compute exponents (*proactive facilitation*).



To take advantage of facilitation, it is necessary to continually **layer** on top of existing knowledge – that is, continually acquire new knowledge that exercises prerequisite or component knowledge. In general, the more connections (neural, cognitive, social, and experiential) there are to a piece of knowledge, the more ingrained, organized, and deeply understood it is (Cross, 1999), and the easier it is to recall via spreading activation through connections. The most efficient way to increase the number of connections to existing knowledge is to continue layering on top of it.

| Structural Integrity

Layering produces **structural integrity**, a well-known engineering concept that also applies to knowledge (the underlying structure of one's knowledge is known as their **schema**). When advanced features are built on top of a system, they sometimes fail in ways that reveal previously-unknown foundational weaknesses in the underlying structure. This forces engineers to fortify the underlying structure so that the system can accommodate new elements without compromising its integrity.

Fortifying the underlying structure often requires improving its organization and elegance, which, when applied to student schemas, produces deep understanding and insight. When the structural integrity of a system is increased, it also becomes easier to add more advanced features *in general*. In the same way, when the structural integrity of a student's schema is increased, it becomes easier to assimilate new knowledge *in general*.

How We Layer

To reap the benefits of layering, Math Academy employs two key features:

1. Moving students forward to new topics immediately after they demonstrate mastery of prerequisites.

2. A highly-connected curriculum where new topics exercise and build on earlier topics.

After a student completes a lesson on Math Academy, new lessons are immediately unlocked. The student will later review what they learned in the lesson that they completed, but they are not "held back" to practice already-learned topics any more than is necessary. This stands in contrast to traditional classrooms, where students are often tethered to the pace of the class and prevented from learning more advanced concepts that come later in the class schedule or in a higher grade level, even though they have already mastered the prerequisites.

Additionally, Math Academy's curriculum is intentionally structured so that earlier topics are applied and reinforced in higher-level topics. We have

- advanced application topics that transition students from purely mathematical framings to contexts involving word problems,

- topics that explicitly teach non-obvious connections between other topics, and

- multi-part problems that pull together many earlier topics to explore a challenging, complex problem context one part at a time.

A high degree of connectivity also arises naturally from our principle that *any lesson should cover all types of problems that a student could reasonably be expected to solve if they truly know the topic.*

It's worth noting that unlike Math Academy, many other educational resources violate this principle and consequently lose out on the benefits of layering because their advanced topics do not actually exercise and build on earlier topics. For instance, some watered-down calculus courses steer clear of any kinds of problems that involve extensive algebra, only covering the simplest cases possible. As a result, students in those courses are not only unable to solve standard problems outside the tiny sandbox of the course, but they also do not fortify their foundations, which can lead them to forget lots of lower-level math despite taking a higher-level math course.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Ausubel, D. P., Robbins, L. C., & Blake Jr, E. (1957). Retroactive inhibition and facilitation in the learning of school materials. *Journal of Educational Psychology, 48*(6), 334.

  Importance: *When a new task exercises knowledge from a previous task, the new task can improve memory of that knowledge as much as identical repetition of the original task.*

- Arzi, H. J., Ben-Zvi, R., & Ganiel, U. (1985). Proactive and retroactive facilitation of long-term retention by curriculum continuity. *American educational research journal, 22*(3), 369-388.

  Importance: *Layering improves retention of prior knowledge and acquisition of new knowledge. Consequently, a program composed of a hierarchical sequence of learning units is superior to a discontinuous array of discrete courses.*

- Cross, K. P. (1999). Learning is about making connections. *The Cross Papers; 3*.

  Importance: *The more connections there are to a piece of knowledge, the more ingrained, organized, and deeply understood it is, and the easier it is to recall.*

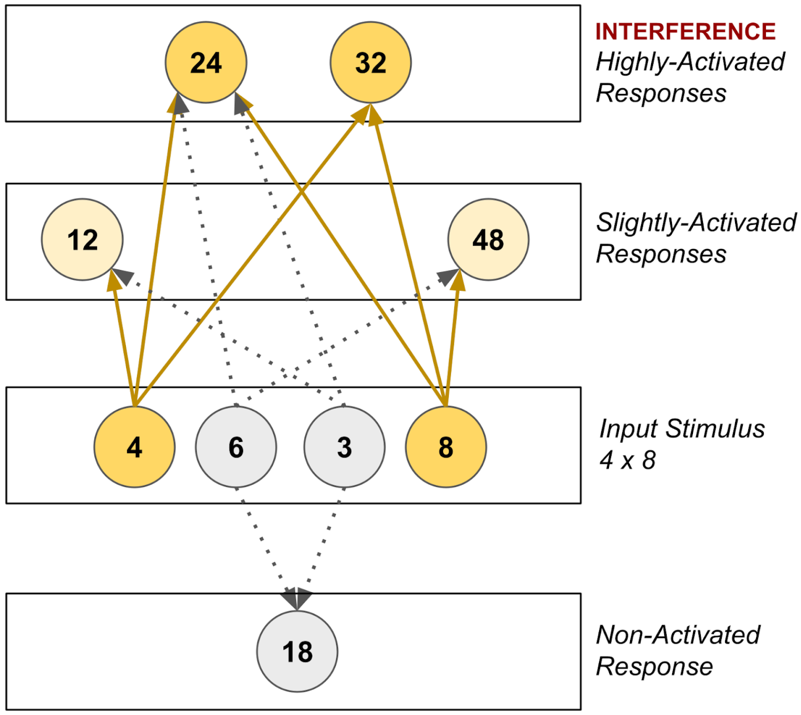
Chapter 16. Non-Interference

Summary: *Associative interference occurs when related knowledge interferes with recall. It is more likely to occur when highly related pieces of knowledge are learned simultaneously or in close succession. However, Math Academy mitigates the effects of interference by teaching dissimilar concepts simultaneously and spacing out related pieces of knowledge over time. This reduces confusion, enhances recall, and facilitates efficient, simultaneous learning of multiple topics, promoting smooth, rapid progress in courses while maintaining varied and engaging learning experiences.*

Associative Interference

Associative interference describes the phenomenon that conceptually related pieces of knowledge can interfere with each other's recall. For instance, it is easy to mistake a leopard for a cheetah.

The same happens in math. Multiple studies have shown that well over half, and potentially as high as 90%, of multiplication mistakes are caused by interference (see Campbell, 1987 for a summary). For instance, when recalling 4×8, related facts like $\mathbf{4} \times 6 = \mathbf{24}$ and $3 \times \mathbf{8} = \mathbf{24}$ interfere with the spreading activation during the recall process and increase the likelihood of the error $4 \times 8 = 24$. (This phenomenon occurs throughout math – multiplication facts are just a convenient setting for academic studies.)



Non-Interference

While it is not possible for a teacher to change the structure of knowledge to make different pieces of information less related, further research in **interference theory** has revealed a factor that can be controlled by a teacher to reduce the impact of interference: time spacing. In a study by Underwood & Ekstrand (1967):

> *"2 groups learned A-B for 32 trials, learned A-C to one perfect recitation 3 days later, and recalled A-C after 24 hr. 2 other groups learned both lists in immediate succession followed by 24-hr, recall of A-C. 1 group from each schedule had 6 A-B pairs retained in A-C. The results showed that the 3-day separation of A-B and A-C markedly reduced proactive inhibition ..."*

In other words, interference is more likely to occur when highly related pieces of knowledge are learned simultaneously or in close succession – but by spacing out these related pieces of knowledge over time, a teacher can mitigate the effects of interference. We call this strategy **non-interference**.

Unfortunately, traditional classrooms ignore the benefits of non-interference and instead operate in a way that *exacerbates* the problem. The typical math curriculum is divided into units of related material and taught in subsequent lessons. This promotes confusion, impedes recall, and places a severe bottleneck on how many topics can be successfully taught simultaneously, thereby creating lots of friction and massively slowing down the learning process.

Math Academy, however, practices non-interference by teaching new concepts alongside dissimilar material. Students are allowed to choose from an array of diverse, non-overlapping learning tasks. As they complete tasks, their knowledge graphs are updated and the system chooses new topics to guide them most efficiently through the course. By utilizing non-interference, Math Academy reduces confusion, improves recall, and successfully teaches many topics simultaneously, thereby enabling students to make smooth, fast progress through their courses.

What's more, non-interference also helps keep Math Academy's learning tasks varied and exciting for students. Learning can feel like a grind when you are made to focus on the same types of concepts and problems for a long time, just like exercising can feel like a grind if the entire workout consists of a single exercise (especially if it's one of your least favorite exercises). So, just like a personal trainer packs a wide variety of exercises into each workout to maintain motivation, Math Academy packs a wide variety of topics into each learning session to keep things exciting.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Campbell, J. I. (1987). The role of associative interference in learning and retrieving arithmetic facts. *Cognitive processes in mathematics*, 107-122.

  Importance: *Multiple studies have shown that well over half, and potentially as high as 90%, of multiplication mistakes are caused by interference. (For instance, when recalling 4 × 8, related facts like 4 × 6 = 24 and 3 × 8 = 24 interfere and increase the likelihood of the error 4 × 8 = 24.)*

- Underwood, B. J., & Ekstrand, B. R. (1967). Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology, 74*(4p1), 574.

  Importance: *Interference is more likely to occur when highly related pieces of knowledge are learned simultaneously or in close succession. By spacing out these related pieces of knowledge over time, a teacher can mitigate the effects of interference.*

Chapter 17. Spaced Repetition (Distributed Practice)

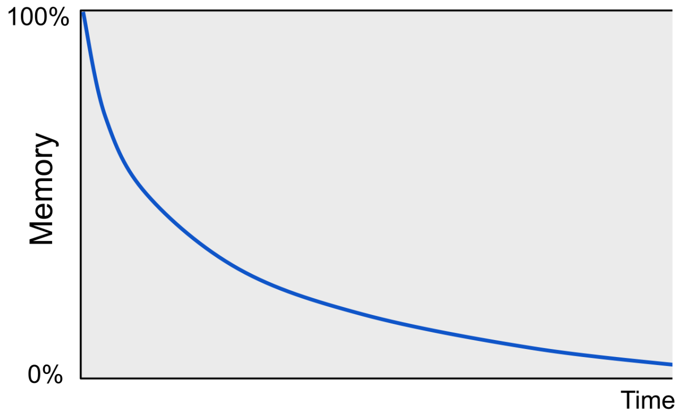
Summary: *When reviews are spaced out or distributed over multiple sessions (as opposed to being crammed or massed into a single session), memory is not only restored, but also further consolidated into long-term storage, which slows its decay. This is known as the spacing effect. A profound consequence of the spacing effect is that the more reviews are completed (with appropriate spacing), the longer the memory will be retained, and the longer one can wait until the next review is needed. This observation gives rise to a systematic method for reviewing previously-learned material called spaced repetition (or distributed practice). A repetition is a successful review at the appropriate time. Spaced repetition is complicated in hierarchical bodies of knowledge, like mathematics, because repetitions on advanced topics should "trickle down" to update the repetition schedules of simpler topics that are implicitly practiced (while being discounted appropriately since these repetitions are often too early to count for full credit towards the next repetition). However, Math Academy has developed a proprietary model of Fractional Implicit Repetition (FIRe) that not only accounts for implicit "trickle-down" repetitions but also minimizes the number of reviews by choosing reviews whose implicit repetitions "knock out" other due reviews (like dominos), and calibrates the speed of the spaced repetition process to each individual student on each individual topic (student ability and topic difficulty are competing factors).*

Retaining Knowledge Indefinitely

| The Spacing Effect

Learning new topics is only half of the puzzle. The other half is remembering what you've learned. In order to retain knowledge, you must periodically review it – otherwise, in the absence of review, it will decay.

A common way to visualize memory decay is through a **forgetting curve**, first studied by psychologist Hermann Ebbinghaus in the late 19th century:

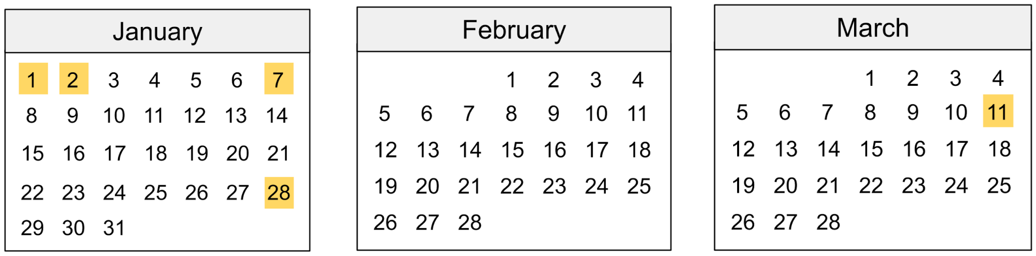


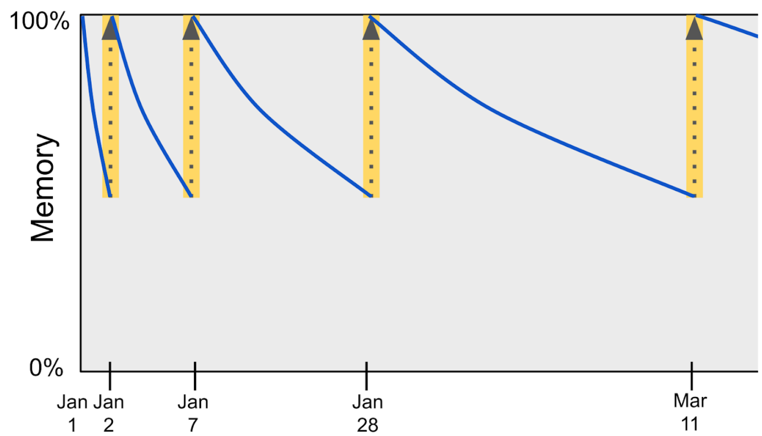
Ebbinghaus (1885) discovered that when reviews are spaced out or *distributed* over multiple sessions (as opposed to being crammed or *massed* into a single session), memory is not only restored, but also further **consolidated** into long-term storage, which slows its decay. This is now known as the **spacing effect**.

| Spaced Repetition

A profound consequence of the spacing effect is that the more reviews are completed (with appropriate spacing), the longer the memory will be retained, and the longer one can wait until the next review is needed. This observation gives rise to a systematic method for reviewing previously-learned material called **spaced repetition** (or **distributed practice**). A **repetition** is a successful review at the appropriate time.

Here is an example that illustrates the process and power of spaced repetition. Suppose you learn a new word. Initially, you might only remember that word for a day. But if you quiz yourself on its meaning tomorrow, then you might remember it until the end of the week. And if you quiz yourself again at the end of the week, then you might remember for several weeks. If you stick to this spaced repetition schedule, then you'll eventually be able to go many years between repetitions (Bahrick et al., 1993).





The main challenge of spaced repetition is choosing the optimal amount of time between repetitions. If you wait too long, you will forget the word and move backwards in your spaced repetition schedule. But if you perform the next repetition too early, your memory won't strengthen as much and you won't move forward as quickly.

| Building Intuition About Spaced Repetition

As Qadir & Imran (2018) describe, spaced repetition can be understood intuitively by way of analogy to muscle-building:

> *"...[M]assed learning can give temporary fluency, just like a body builder can pump muscles temporarily by cramming exercises. However, growth only occurs with a spaced exercise routine (in which exercise and rest follow each other cyclically). Similarly, long-term learning also requires spaced practice and does not result from cramming."*

As Brown, Roediger, & McDaniel (2014, pp.9-10, 81-82, 100-101) elaborate:

> *"It's widely believed by teachers, trainers, and coaches that the most effective way to master a new skill is to give it dogged, single-minded focus, practicing over and over until you've got it down.*

> *Our faith in this runs deep, because most of us see fast gains during the learning phase of massed practice. What's apparent from the research is that gains achieved during massed practice are transitory and melt away quickly.*
>
> *...*
>
> *Massed practice gives us the warm sensation of mastery because we're looping information through short- term memory without having to reconstruct the learning from long-term memory. But just as with rereading as a study strategy, the fluency gained through massed practice is transitory, and our sense of mastery is illusory. It's the effortful process of reconstructing the knowledge that triggers reconsolidation and deeper learning.*
>
> *...*
>
> *When you recall learning from short- term memory, as in rapid- fire practice, little mental effort is required, and little long-term benefit accrues. But when you recall it after some time has elapsed and your grasp of it has become a little rusty, you have to make an effort to reconstruct it. This effortful retrieval both strengthens the memory but also makes the learning pliable again, leading to its* ***reconsolidation****. Reconsolidation helps update your memories with new information and connect them to more recent learning."*

The process of **reconsolidation** can be likened (pp.73-74) to the process of composing an essay through many iterations:

> *"An apt analogy for how the brain consolidates new learning may be the experience of composing an essay. The first draft is rangy, imprecise. You discover what you want to say by trying to write it. After a couple of revisions you have sharpened the piece and cut away some of the extraneous points. You put it aside to let it ferment. When you pick it up again a day or two later, what you want to say has become clearer in your mind. Perhaps you now perceive that there are three main points you are making. You connect them to examples and supporting information familiar to your audience. You rearrange and draw together the elements of your argument to make it more effective and elegant.*
>
> *Similarly, the process of learning something often starts out feeling disorganized and unwieldy; the most important aspects are not always salient. Consolidation helps organize and solidify learning, and, notably, so does retrieval after a lapse of some time, because the act of retrieving a memory from long- term storage can both strengthen the memory traces and at the same time make them modifiable again, enabling them, for example, to connect to more recent learning. This process is called reconsolidation. This is how [spaced] retrieval practice modifies and strengthens learning."*

| Consensus Among Researchers

It's worth noting that the spacing effect is still an active area of research. As Hartwig, Rohrer, & Dedrick (2022) describe, there may be other factors at play besides reconsolidation – but while the exact mechanism(s) underlying the spacing effect may still be debated, the result and utility of the spacing effect is fully agreed upon by researchers:

> *"Researchers have proposed numerous theoretical explanations for the spacing effect (for reviews, see Benjamin & Tullis, 2010; Delaney et al., 2010; Dempster, 1989). According to various theories, the spacing effect may derive from mechanisms such as encoding variability (i.e., contextual variation provides richer encoding when two learning episodes are spaced apart), deficient*

> *processing (i.e., processing of material during a second learning episode is diminished if close in time to the first episode), consolidation (i.e., a second learning episode benefits from any memory consolidation that occurs in the interim), or study-phase retrieval (i.e., spacing promotes effortful retrieval during a second learning episode). However, no single mechanism has accounted for the entire body of spacing-related findings, and it is possible that a combination of mechanisms may best explain the effect (Delaney et al., 2010).*
>
> *Regardless of mechanism, spacing effects are robust – occurring across various materials, procedures, and learner characteristics (Dunlosky et al., 2013). Most important for the present study, spacing effects have been demonstrated in numerous classroom-based randomized studies (e.g., Seabrook et al., 2005; Sobel et al., 2011; for a review, see Dunlosky et al., 2013). Moreover, classroom studies have found spacing effects with math learning (Barzagar Nazari & Ebersbach, 2019; Hopkins et al., 2016; Lyle et al., 2020; Schutte et al., 2015). In short, considerable data show that spaced math practice improves scores on delayed tests. ... [T]he literature is clear that practice should be spaced across many class sessions if students are to retain the information long-term (Rawson et al., 2013; Rawson et al., 2018)."*

As Rohrer (2009) states:

> *"...[T]he spacing effect is arguably one of the largest and most robust findings in learning research, and it appears to have few constraints."*

Indeed, according to researcher Kang (2016), hundreds of studies have demonstrated that spaced repetition produces superior long-term retention. As a memorable example, he describes one of the earliest spaced repetition studies, whose findings have been backed up by 254 follow-up studies over the past century:

> *"In one early study, to illustrate a specific instance, college students were asked to learn the Athenian Oath (Gordon, 1925). One group of students heard the oath read 6 times in a row; another group heard the oath 3 times on 1 day and 3 more times 3 days later.*
>
> *On the immediate test, the group that received massed repetition recalled slightly more than the group that received spaced repetition. But on the delayed test 4 weeks later, the spaced group clearly outperformed the massed group.*
>
> *While massed practice might appear [slightly] more effective than spaced practice in the short term, spaced practice produces durable long-term learning."*

The benefits of spacing are so pronounced and conclusive that they have even attracted attention from the field of advertising, where the spacing effect has been reproduced in numerous studies of consumer memory of brands (Schmidt & Eisend, 2015).

Spaced Repetition is Underused

Unfortunately, as with mastery learning, spaced repetition deviates from traditional convention in education and consequently remains rarely used in classrooms. As Kang (2016) laments:

> *"Despite over a century of research findings demonstrating the spacing effect, however, it does not have widespread application in the classroom. The spacing effect is 'a case study in the failure to apply the results of psychological research' (Dempster, 1988, p. 627).*
>
> *...*
>
> *When deciding on what instructional techniques to use (and when to use them), many teachers default to familiar methods (e.g., how they themselves were taught; Lortie, 1975) or rely on their intuitions, both less than ideal: Our intuitions about learning can sometimes be plain wrong, and it would be a waste to overlook the growing evidence base regarding the effectiveness of various teaching or learning strategies.*
>
> *...*
>
> *The second major hurdle is conventional instructional practice, which typically favors massed practice. Teaching materials and aids (e.g., textbooks, worksheets) are usually organized in a modular way, which makes massed practice convenient. After presenting a new topic in class, teachers commonly give students practice with the topic via a homework assignment. But aside from that block of practice shortly after the introduction of a topic, no further practice usually follows, until a review session prior to a major exam."*

Perhaps shockingly, Cepeda et al. (2009) observed that even many instructional design and educational psychology textbooks have little to no coverage of spaced repetition as a learning strategy:

> *"Failure to consider distributed practice research is evident in instructional design and educational psychology texts, many of which fail even to mention the distributed practice effect (e.g., Bransford, Brown, & Cocking, 2000; Bruning, Schraw, Norby, & Ronning, 2004; Craig, 1996; Gardner, 1991; Morrison, Ross, & Kemp, 2001; Piskurich, Beckschi, & Hall, 2000).*
>
> *Those texts that mention the distributed practice effect often devote a paragraph or less to the topic (e.g., Glaser, 2000; Jensen, 1998; Ormrod, 1998; Rothwell & Kazanas, 1998; Schunk, 2000; Smith & Ragan, 1999) and offer widely divergent suggestions – many incorrect – about how long the lag between study sessions ought to be (cf. Gagné, Briggs, & Wager, 1992; Glaser, 2000; Jensen, 1998; Morrison et al., 2001; Ormrod, 2003; Rothwell & Kazanas, 1998; Schunk, 2000; Smith & Ragan, 1999)."*

Typically, students learn a topic during class, practice it on the homework, and then forget about it until it's time to study for a test. After the test, students are rarely required to practice the topic again, unless it just happens that some new topic requires them to remember the old one. The end result is that students end up forgetting most of what they learn – and as Rohrer (2009) notes, the forgetting can be so severe that it can appear as though they never learned those things in the first place:

> *"The effects of forgetting are often neglected by learning theorists, but acquisition has little utility unless material is retained. Indeed, although poor performance on standard achievement tests is often attributed to the absence of acquisition, forgetting may often be the culprit.*
>
> *For example, in the 1996 National Assessment of Educational Progress, 50% of U.S. eighth graders were unable to correctly multiply -5 and -7, even though the question was presented in a multiple-choice format (Reese, Miller, Mazzeo, & Dosse, 1997). If any of these erring students knew the product previously, which seems likely, their error was likely due to forgetting."*

Often, students don't even realize how quickly they forget in the absence of spaced review. For instance, Emeny, Hartwig, & Rohrer (2021) found that students who engaged in spaced practice could predict their own future test scores fairly accurately, while students who engaged in massed practice were severely overconfident in their predictions:

> *"Following spaced practice, students predicted their future test scores very accurately, whereas massed practice yielded gross overconfidence. The overconfidence after massed practice might be due to the fluency or success with which students can solve a set of similar problems by merely repeating the same procedure over and over, giving the impression that students have mastered the content. In turn, overconfidence may lead students and their teachers to believe that further practice is unnecessary when, in fact, the gains will not be retained across time.*
>
> *Though massed practice produced overconfidence, we note that predictions following massed practice were only slightly greater than predictions following spaced practice. Thus, massed practice did not elevate predictions to an unrealistically high level but instead failed to help students recognize their low level of mastery."*

Another factor keeping spaced repetition out of STEM classrooms in particular is that, to our best knowledge, the current literature on mathematical models for determining optimal repetition spacing is limited to the setting of independent flashcard-like tasks. STEM subjects, in contrast, are highly connected bodies of knowledge. This introduces significant modeling complexities: for instance, repetitions on advanced topics should "trickle down" to update the repetition schedules of simpler topics that are implicitly practiced (while being discounted appropriately since these repetitions are often too early to count for full credit towards the next repetition).

To overcome this hurdle, Math Academy has developed a proprietary model of **Fractional Implicit Repetition** (**FIRe**) that not only accounts for implicit "trickle-down" repetitions but also

- minimizes the number of reviews by choosing reviews whose implicit repetitions "knock out" other due reviews (like dominos), and

- calibrates the spaced repetition process to each individual student on each individual topic (student ability and topic difficulty are competing factors – high student ability speeds up the overall student-topic learning speed, while high topic difficulty slows it down).

Our highly sophisticated spaced repetition model is the product of years of research, practice, and development since 2019, and it continues to be refined as we gain more data on student learning outcomes.

Spaced Repetition Improves Generalization

People usually think of spaced repetition as a process to remember isolated pieces of information. But in a highly connected body of complex skills, like math, spaced repetition can also promote generalized learning that can more easily transfer across different contexts.

To start off with some loose intuition, think about what happens when you reread a book or rewatch a movie that you haven't seen in a while. Often, you see things that you didn't notice before. You come in with a different mental state compared to the last time you watched, and you come out with some fresh perspectives and a more comprehensive understanding of the work.

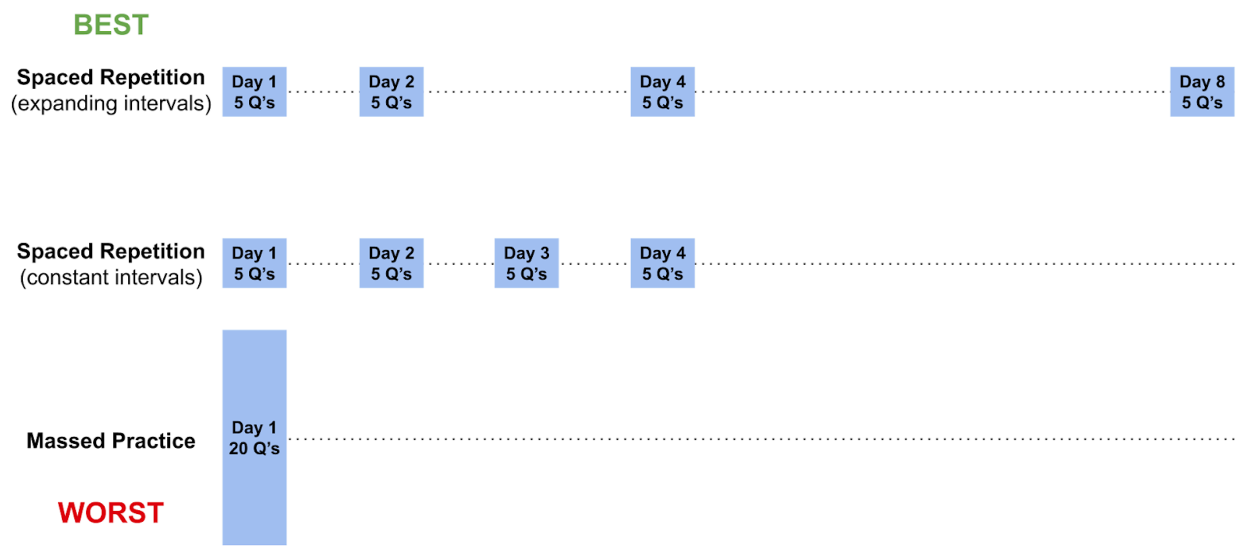
Indeed, a review by Smith & Scarf (2017) recounted multiple studies demonstrating that *"spacing not only benefits the learning and retention of specific items but improves the generalization of learning"*:

> *"Hagman (1980) had participants learn and practice electrical testing on the same equipment or different equipment, with practice massed all in 1 day or spaced on 3 consecutive days. On a transfer test after a 2-week delay, spaced practice on different equipment resulted in better transfer than spaced practice on the same equipment. Spaced practice on the same equipment resulted in better performance on the transfer test than massed practice on the same or different equipment. Moreover, massed practice on the same or different equipment resulted in equivalent performance on the transfer test, indicating that spacing was necessary for training variations to promote generalization.*
>
> *Similarly, Moulton et al. (2006) compared massed and spaced groups who practiced microsurgical skills on PVC-artery models and arteries from a turkey thigh, and tested to what extent their skills transferred to a live rat 1 month after training. Moulton et al. (2006) found that the spaced group performed significantly better on a variety of outcome measures than the massed group.*
>
> *Studies with children have investigated the impact of spacing on generalization using a greater range of spacing intervals relative to the adult literature. For example, Vlach and Sandhofer (2012) investigated the impact of spacing on the generalization of simple and complex science concepts in 5- to 7-year-olds. The children in their study completed 4 lessons on biomes, with*

> *each lesson involving a different context (desert, grasslands, artic, ocean or swamp), and a post-test 1 week after the last lesson. The massed group completed all four lessons in 1 day, the intermediate group completed 2 lessons per day for 2 days, and the Spaced Group completed 1 lesson per day for 4 days. For simple generalization, the spaced group showed significantly greater improvement from the pre- to post-test than the massed group, and the intermediate group's improvement was not significantly different when compared to the massed or spaced groups. In contrast, for complex generalization, the spaced group's improvement was significantly better than both the massed and intermediate groups. In fact, the data suggest that the spaced group is the only group to show an improvement in their gain scores as the questions moved from simple to complex, though unfortunately this trend was not tested for statistical significance. Spacing therefore may provide a greater benefit for more complex generalizations.*
>
> *Gluckman et al. (2014) replicated Vlach and Sandhofer's (2012) findings, but in the post-test they included questions on the children's memory for facts and concepts talked about during the lessons (e.g., what is a biome?), in addition to generalization questions. The spaced group showed significantly greater improvement than the massed group for simple and complex generalization questions and for memory questions. The reported means displayed the same pattern as above, with spacing benefiting complex generalizations more than simple generalizations."*

In a follow-up study, Vlach, Sandhofer, & Bjork (2014) also found that higher-fidelity spaced repetition with expanding intervals promoted even better generalization than spaced repetition with constant intervals, suggesting that optimizing the spaced repetition process can lead to significant gains in generalization:

> *"...[W]e examined whether an expanding learning schedule would promote generalization to a greater degree than would an equally spaced learning schedule ... [A]t the 24-hour delayed generalization test, we observed a significant difference between the two conditions: Children in the expanding learning condition significantly outperformed children in the equal spacing condition.*
> *...*
> *These findings suggest that the benefits of expanding schedules are not constrained to memory tasks, but that these learning schedules can promote multiple types of learning, such as the acquisition and generalization of information."*



Repetition Compression

A common criticism of spaced repetition is that it requires an overwhelming number of reviews. While this can be true if spaced repetition is used to learn unrelated flashcards, there is something special about the subject of mathematics that allows Math Academy to avoid this issue.

Unlike independent flashcards, mathematics is a hierarchical and highly connected body of knowledge. Whenever a student solves an advanced mathematical problem, there are many simpler mathematical skills that they practice implicitly. In other words, in mathematics, advanced skills tend to **encompass** many simpler skills.

As a result, whenever a student has due reviews, Math Academy is able to **compress** them into a much smaller set of learning tasks that implicitly covers (i.e. provides repetitions on) all of the due reviews. We call this process **repetition compression**.

To illustrate, consider the following multiplication problem, in which we multiply the two-digit number 39 by the one-digit number 6:

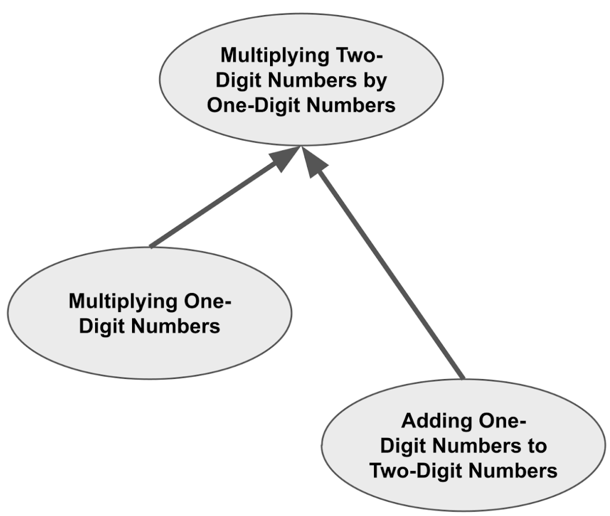
$$\begin{array}{r} {\scriptstyle 5} \\ 3\ 9 \\ \times \quad\ \ 6 \\ \hline 2\ 3\ 4 \end{array}$$

In order to perform the multiplication above, we have to multiply one-digit numbers and add a one-digit number to a two-digit number:

- First, we multiply 6 × 9 = 54. We carry the 5 and write the 4 at the bottom.
- Then, we multiply 6 × 3 = 18 and add 18 + 5 = 23. We write 23 at the bottom.

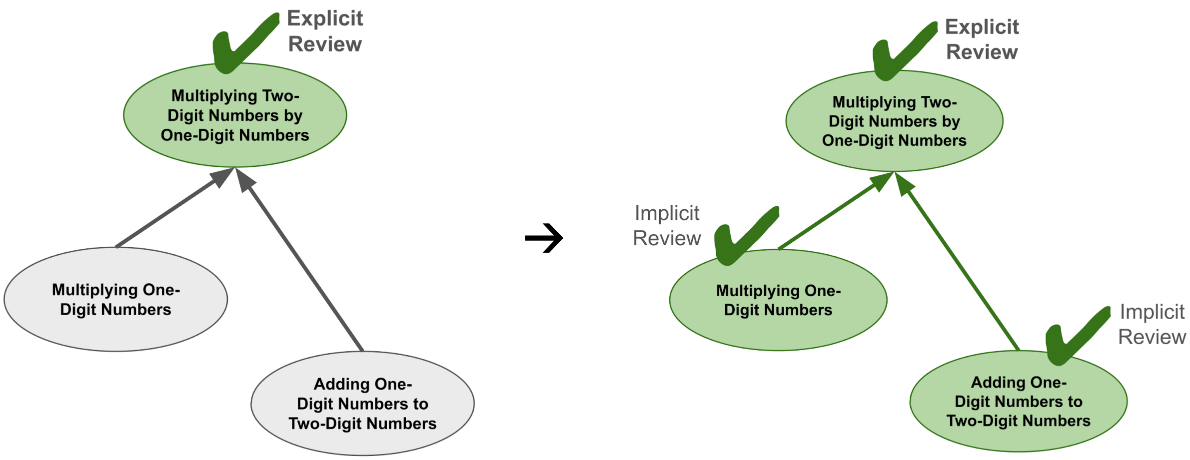
In other words, *Multiplying a Two-Digit Number by a One-Digit Number* **encompasses** *Multiplying One-Digit Numbers* and *Adding a One-Digit Number to a Two-Digit Number*.

We can visualize this using an **encompassing graph** as shown below. The encompassing graph is similar to a prerequisite graph, except the arrows indicate that a simpler topic is encompassed by a more advanced topic. (Encompassed topics are usually prerequisites, but prerequisites are often not fully encompassed.)



Now, suppose that a student is due for reviews on all three of these topics. Because of the encompassings, the only review that they will actually have to do is *Multiplying a Two-Digit*

Number by a One-Digit Number. When they complete this review, it will implicitly provide repetitions on the topics that it encompasses because the student has effectively practiced those skills as well.



In general, the more encompassings there are, the fewer reviews are actually required. And mathematics has *lots* of encompassings!

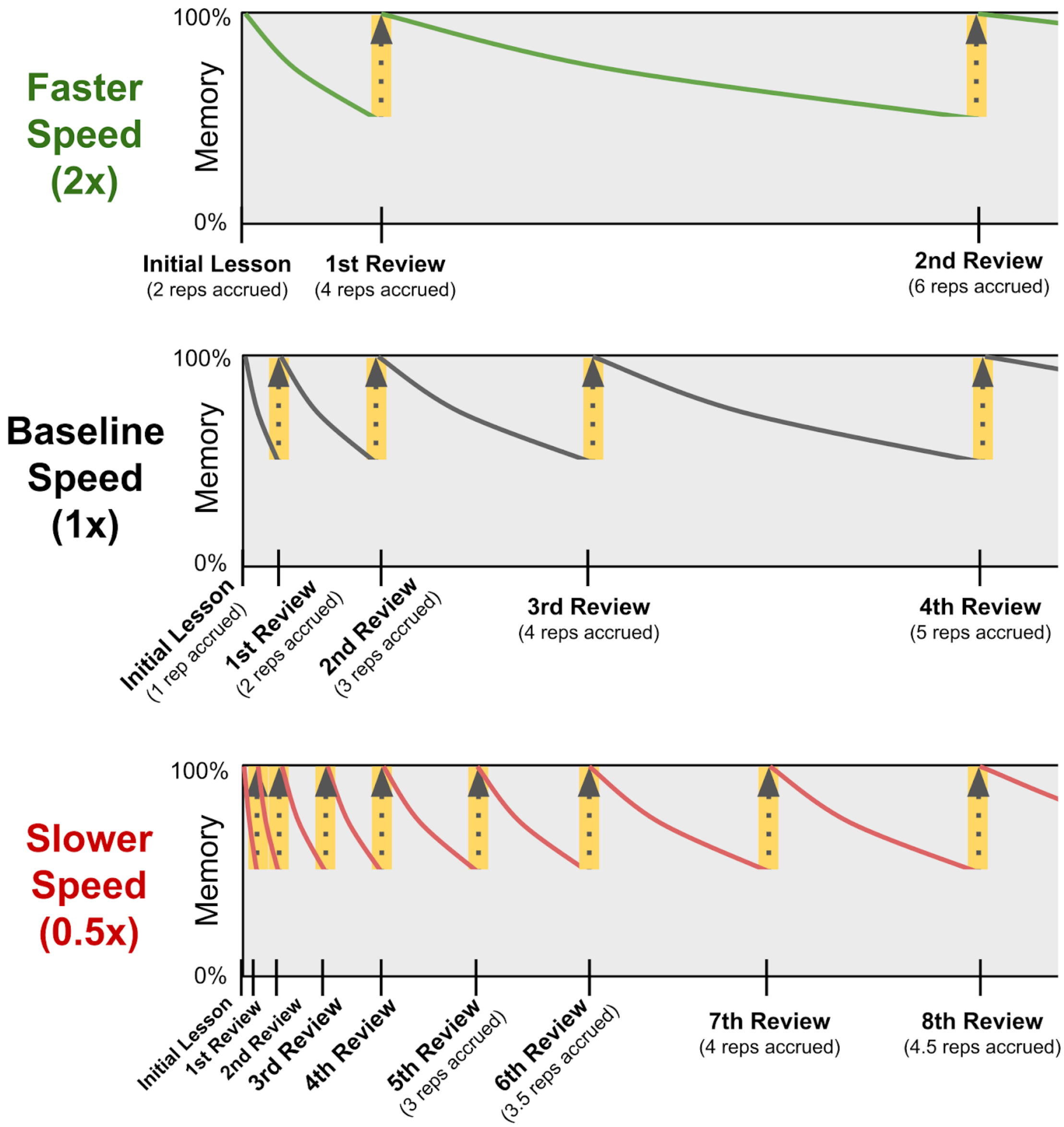
Calibrating to Individual Students and Topics

As discussed in chapter 7, the **speed** at which students learn (and remember what they've learned) varies from student to student. It has been shown that some students learn faster and remember longer, while other students learn slower and forget more quickly (e.g., Kyllonen & Tirre, 1988; Zerr et al., 2018; McDermott & Zerr, 2019). Similarly, learning speed also varies across topics: easier topics are learned faster and remembered longer, while harder topics take longer to learn and are forgotten more quickly.

So, for each student, each topic has a learning speed that depends on the student's ability and the topic's difficulty. Math Academy computes these **student-topic learning speeds** and uses them to adjust the speed of the spaced review process.

- For instance, if a student does a review on a topic for which their learning speed is 2x, then that review counts as being worth 2 spaced repetitions.

- Likewise, if a student does a review on a topic for which their learning speed is 0.5x, then that review counts as being worth 0.5 spaced repetitions.



Student-topic learning speeds are also considered within the fractional implicit repetition algorithm. Whenever a topic's spaced repetition process is being slowed down (i.e. whenever the student-topic learning speed is less than 1), we also shut down all incoming implicit repetition credit and instead force **explicit reviews**. The topic does not receive any implicit repetition credit that would normally "trickle down" from more advanced topics that encompass it.

We do this because, in practice, weaker students often have trouble absorbing implicit repetitions on difficult topics – they have a harder time generalizing that "what I learned earlier is a special case (or component) of what I'm learning now.'' The decision of whether or not to force explicit reviews is based on the student-topic learning speed because when a topic's spaced repetition process is being slowed down, it indicates that the topic is considered rather difficult relative to the student's ability.

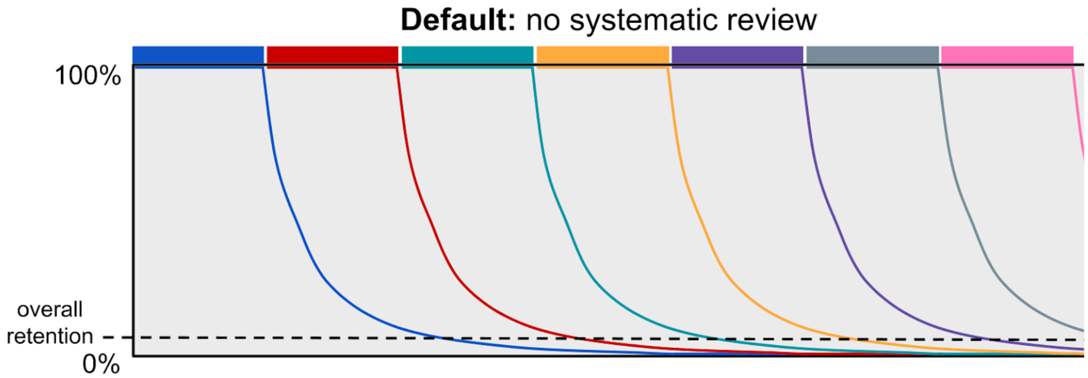
Lastly, it's important to realize that while there are a number of factors that could affect a student's learning speed, such as their aptitude, forgetting rate, level of interest/motivation, or how tired or distracted they typically are while working on learning tasks – these factors are ultimately only as relevant as their effects on the student's observed performance, which is what we use when adapting to each student's individual learning curve.

Spaced Repetition vs Spiraling

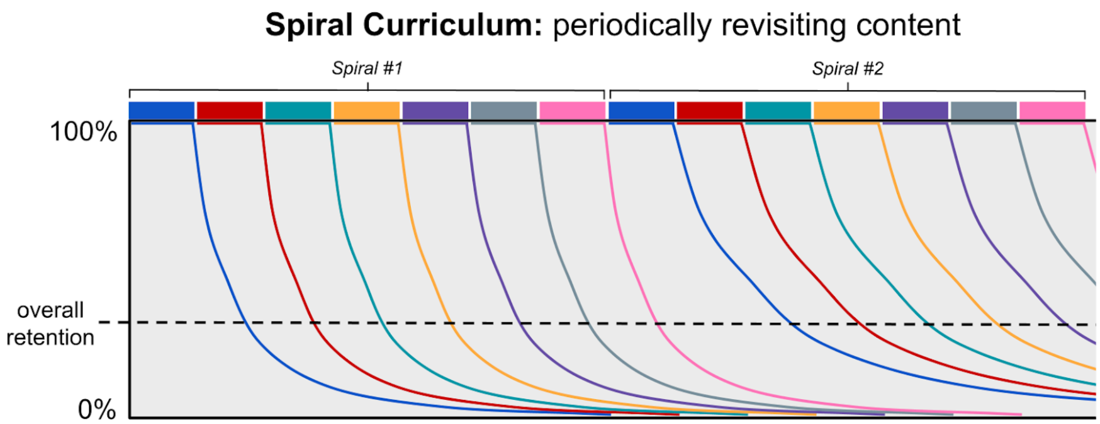
Some curricula adopt a **spiral approach** where material is naturally revisited and further built upon in later textbook chapters and/or grades. Spiraling is clearly an improvement over the default mode of instruction, which includes little to no systematic review – and it allows teachers to make use of the spacing effect to some extent while teaching manually at a group level without the assistance of technology. However, spiraling is still nowhere near the level of granularity, precision, and individualization that is required to capture the maximum benefit of true spaced repetition.

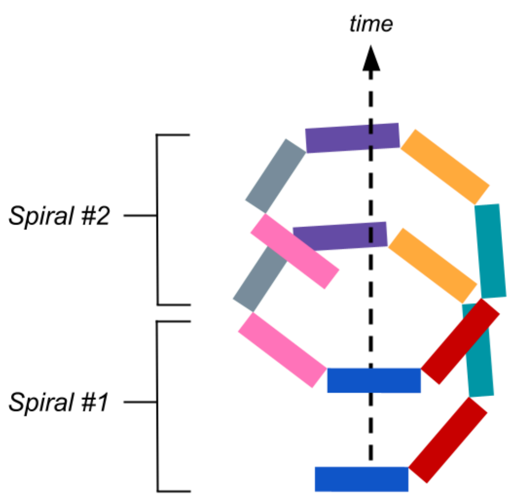
Note that while spiraling is sometimes conflated with discovery learning (both are widely attributed to psychologist Jerome Bruner in the 1960s), these are really two separate ideas, the latter of which we do not intend to endorse. There are plenty of spiral curricula (e.g., *Saxon Math*) that leverage direct instruction instead of discovery learning. The discussion here shall be concerned purely with the extent to which spiraling leverages the spacing effect, not the method by which instruction is delivered.

To understand the difference between spiraling and spaced repetition, it helps to visualize the corresponding forgetting curves. We start with the default mode of instruction, in which large groups of related material are covered all at once and are not systematically revisited in the future. Under this mode of instruction, students quickly forget what they learn, and their overall retention is extremely low.

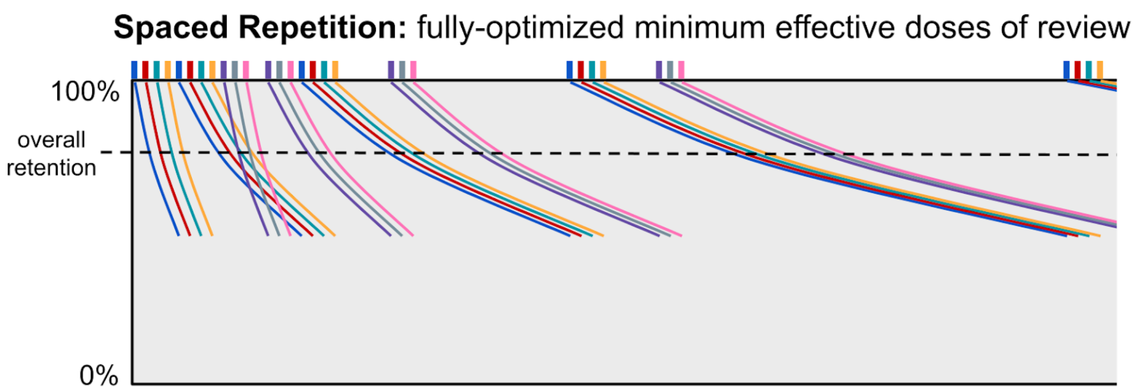


By periodically revisiting content, a spiral curriculum periodically restores forgotten knowledge and leverages the spacing effect to slow the decay of that knowledge. This raises students' overall retention of what they have learned. To illustrate, a forgetting curve for a spiral curriculum with two spirals is shown below:





Spaced repetition takes this line of thought to its fullest extent by fully optimizing the review process. It precisely calibrates the spacing of reviews so as to maintain a consistently high level of retention and slow down the decay (i.e. stretch out the decay curves) as quickly as possible. Review spacing continually adapts to student performance, expanding in response to good performance and shrinking in response to poor performance. By optimizing the review process to the fullest extent, spaced repetition further raises students' retention of what they have learned.



However, while spaced repetition is more optimal, it requires an inhuman amount of work from the instructor. Taken to its fullest extent, spaced repetition requires the instructor to keep track of a repetition schedule for every topic for every student and continually update that schedule based on the student's performance – and each time a student learns (or reviews) an advanced topic, they're implicitly reviewing many simpler topics, all of whose repetition schedules need to be adjusted as a result.

In this view, spiraling can be characterized as "the best an instructor can do" manually while teaching at a group level without the assistance of technology. Spaced repetition is the optimal solution to maximizing retention, but it is infeasible to perform spaced repetition manually, so spiraling is the next-best option that an instructor can actually perform without the assistance of technology.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Ebbinghaus, H. (1885). Memory: A contribution to experimental psychology, trans. *HA Ruger & CE Bussenius. Teachers College, Columbia University.*

  Importance: *When reviews are spaced out or distributed over multiple sessions (as opposed to being crammed or massed into a single session), memory is not only restored, but also further consolidated, which slows its decay. This is now known as the spacing effect.*

- Kang, S. H. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 12-19.

  Hartwig, M. K., Rohrer, D., & Dedrick, R. F. (2022). Scheduling math practice: Students' underappreciation of spacing and interleaving. *Journal of Experimental Psychology: Applied, 28*(1), 100.

  Rohrer, D. (2009). Research commentary: The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education, 40*(1), 4-17.

  Emeny, W. G., Hartwig, M. K., & Rohrer, D. (2021). Spaced mathematics practice improves test scores and reduces overconfidence. *Applied Cognitive Psychology, 35*(4), 1082-1089.

  Importance: *Hundreds of studies have demonstrated that spaced repetition produces superior long-term learning. However, spaced repetition deviates from traditional convention in education and consequently remains rarely used in classrooms. As a result, severe forgetting sets in quickly, and students and teachers are often overconfident about how well students will perform on tests.*

- Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4*(5), 316-321.

  Importance: *Using spaced repetition, memory can be retained up to at least (and likely longer than) 5 years, the*

longest delay period tested in the study.

- Smith, C. D., & Scarf, D. (2017). Spacing repetitions over long timescales: a review and a reconsolidation explanation. *Frontiers in Psychology, 8*, 962.

  Vlach, H. A., Sandhofer, C. M., & Bjork, R. A. (2014). Equal spacing and expanding schedules in children's categorization and generalization. *Journal of experimental child psychology, 123*, 129-137.

  Importance: *Multiple studies have demonstrated that spacing not only benefits the learning and retention of specific items but improves the generalization of learning. In a follow-up study, Vlach and colleagues also found that higher-fidelity spaced repetition with expanding intervals promoted even better generalization than spaced repetition with constant intervals, suggesting that optimizing the spaced repetition process can lead to significant gains in generalization.*

- Kyllonen, P. C., & Tirre, W. C. (1988). Individual differences in associative learning and forgetting. *Intelligence, 12*(4), 393-421.

  Zerr, C. L., Berg, J. J., Nelson, S. M., Fishell, A. K., Savalia, N. K., & McDermott, K. B. (2018). Learning efficiency: Identifying individual differences in learning rate and retention in healthy adults. *Psychological science, 29*(9), 1436-1450.

  McDermott, K. B., & Zerr, C. L. (2019). Individual differences in learning efficiency. *Current Directions in Psychological Science, 28*(6), 607-613.

  Importance: *Stronger students learn faster and remember longer, while weaker students learn slower and forget more quickly. (This implies that spaced repetition schedules should be calibrated to the strengths of individual students.)*

Additional Resources

- Carpenter, S. K., & Agarwal, P. K. (2019). How to use spaced retrieval practice to boost learning. *Iowa State University.*

- Rohrer, D., & Hartwig, M. K. (2023). Spaced and Interleaved Mathematics Practice. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 111-21). Society for the Teaching of Psychology.

- Pashler, H., Rohrer, D., & Cepeda, N. J. (2006). Temporal spacing and learning. *APS Observer, 19.*

Chapter 18. Interleaving (Mixed Practice)

Summary: *Interleaving (or mixed practice) involves spreading minimal effective doses of practice across various skills, in contrast to blocked practice, which involves extensive consecutive repetition of a single skill. Blocked practice can give a false sense of mastery and fluency because it allows students to settle into a robotic rhythm of mindlessly applying one type of solution to one type of problem. Interleaving, on the other hand, creates a "desirable difficulty" that promotes vastly superior retention and generalization, making it a more effective review strategy. But despite its proven efficacy, interleaving faces resistance in classrooms due to a preference for practice that feels easier and appears to produce immediate performance gains, even if those performance gains quickly vanish afterwards and do not carry over to test performance.*

Interleaving vs Blocking

In a traditional classroom, homework assignments usually focus on a single topic. For instance, if a student learns how to subtract multi-digit whole numbers during class, then their homework might contain 15 review problems to practice that skill. This is called **blocked practice** or **blocking**, in which a single skill is practiced many times consecutively.

While some initial amount of blocking is useful when first learning a skill, blocking is very *inefficient* for building long-term memory afterwards during the review stage. Instead of putting those 10 review problems on a single review assignment, it would be more effective to spread them out over multiple review assignments that each cover a broad mix of previously-learned topics.

For instance, one of those assignments might have the following breakdown of problems:

- (3 problems) *Subtracting Multi-Digit Whole Numbers*
- (3 problems) *Adding One-Digit Decimals*
- (3 problems) *Converting One-Digit Decimals Into Fractions*
- (3 problems) *Converting Improper Fractions Into Mixed Numbers*
- (3 problems) *Solving Word Problems Using Multi-Digit Addition*

This strategy is called **interleaving** (also known as **varied practice** or **mixed practice**).

Benefits of Interleaving

| Efficiency

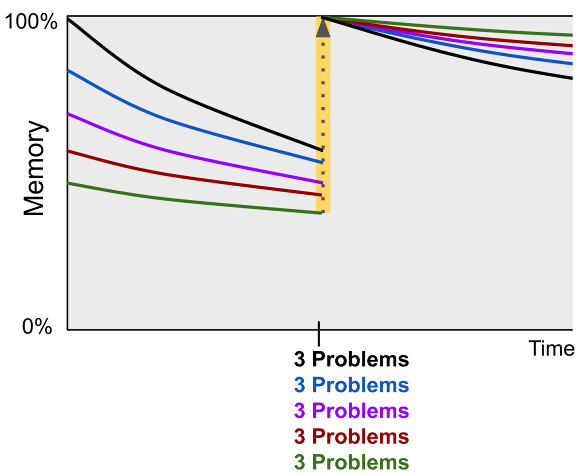
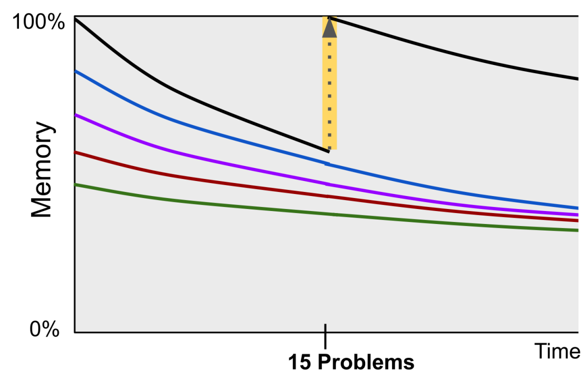
One benefit of interleaving is that it provides minimum effective doses of review for a handful of different topics, whereas blocked practice only hits a single topic and wastes most of the review effort in the realm of diminishing returns. As Rohrer & Pashler (2007) describe in a paper titled *Increasing Retention without Increasing Study Time*:

> *"Our results suggest that a single session devoted to the study of some material should continue long enough to ensure that mastery is achieved but that immediate further study of the same material is an inefficient use of time. ... The continuation of study immediately after the student has achieved error-free performance is known as overlearning. ... [W]hile overlearning often increases performance for a short while, the benefit diminishes sharply over time.*
> *...*
> *Because overlearning requires more study time than not overlearning, the critical question is how the benefits of overlearning compare to the benefits resulting from some alternative use of the same time period. ... [D]evoting this study time to the review of materials studied weeks, months, or even years earlier will typically pay far greater dividends than the continued study of material learned just a moment ago.*
>
> *In essence, overlearning simply provides very little bang for the buck, as each additional unit of uninterrupted study time provides an ever smaller return on the investment of study time."*

As quoted elsewhere:

> *"...[O]verlearning has the deficiencies of massed practice, and when the choice presents itself, our results suggest that overlearning will typically represent an inefficient use of study time."* – Pashler et al. (2007)
>
> *"...[A] typical mathematics assignment consists of many problems relating to the same skill or concept, yet evidence suggests that students receive little long-term benefit from working more than several problems of the same kind in immediate succession (e.g., Lyle, Bego, Hopkins, Hieb, & Ralston, 2020)."* – Rohrer & Hartwig (2020)

This can be visualized on forgetting curves (shown below), and it suggests an effective method that Math Academy uses to select topics for interleaved review: simply choose those topics whose spaced repetitions are due (or are closest to being due).

<table>
<tr><th>Interleaving</th><th>Blocking</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

| Discrimination and Category Induction Learning

Another benefit of interleaving is that, in addition to helping students practice carrying out solution techniques, it also enhances other types of learning that are necessary components of true mastery (see Rohrer, 2012 for a review):

- (*discrimination learning*) matching problems with the appropriate solution techniques – for instance, the equations $x^2 + 3x + 2 = 0$ and $x + 3x + 2 = 0$ look similar but require wildly different solution techniques.

- (*category induction learning*) recognizing general features that distinguish problems requiring different solution techniques

As Taylor & Rohrer (2010) elaborate:

> *"When practice problems are blocked, however, students can successfully solve a set of practice problems without learning how to pair a problem with the skill. Indeed, because all of the problems relate to the topic – typically the one presented in the immediately preceding lesson – students can choose the appropriate procedure for each practice problem before they read the problem. While this reduces the difficulty of the practice problems, students are effectively relying on a crutch. Unfortunately for students, this weakness is exposed when these same kinds of problems appear on a cumulative exam, standardized test, or during a subsequent research career.*

> *By contrast, interleaved practice gives students an opportunity to practice pairing each kind of problem with the appropriate procedure. Far from being limited to statistics courses, the difficulty of pairing a problem with the appropriate procedure or concept is ubiquitous in mathematics.*
>
> *For example, the notorious difficulty of word problems is due partly to the fact that few word problems explicitly indicate which procedure or concept is appropriate. For example, the word problem, 'If a bug crawls eastward for 8 m and then crawls northward for 15 m, how far is it from its starting point'? requires students to infer the need for the Pythagorean Theorem. However, no such inference is required if the word problem appears immediately after a block of problems that explicitly indicate the need for the Pythagorean theorem (e.g. if the legs of a right triangle have lengths 8 and 15 m, what is the length of its hypotenuse?). Thus, blocked practice can largely reduce the pedagogical value of a word problem.*
>
> *As a final example, it should be noted that blocked practice may facilitate students' failure to discriminate between different kinds of problems even when these kinds of problems are not superficially similar. In elementary school, for example, students are ordinarily taught to find both the greatest common factor of two integers and the least common multiple of two integers. Thus, the instructions for these two kinds of problems are easily distinguished from each other ('Find the greatest common factor ...' vs. 'Find the least common multiple ...'). However, if the practice problems of each kind are blocked, students can ignore the instruction and instead focus solely on the information that varies from problem to problem (i.e. the pair of integers). Students can then solve problems by merely repeatedly performing the same procedure without giving much thought to why it is appropriate."*

| Experimental Support

The benefits of interleaving are supported by numerous studies across a wide variety of domains including math, other academic subjects, raw cognitive tasks, motor skills, and even sports practice (see Rohrer, 2012 for a review). As summarized elsewhere by Rohrer (2009):

> *"Experiments have shown that test scores can be dramatically improved by the introduction of spaced practice or mixed practice, which are the two defining features of mixed review. Moreover, neither spacing nor mixing requires an increase in the number of practice problems, meaning that both features increase efficiency as well as effectiveness. ... Its effects on mathematics learning deserve greater consideration by teachers and researchers."*

While blocking leads to more rapid gains in performance (which makes it useful when first learning a skill), interleaving promotes vastly superior retention and generalization (which makes it a more effective review strategy). As Rohrer, Dedrick, & Stershic (2015) clarify elsewhere:

> *"...[A] small block of problems might be optimal, especially at the outset of an assignment given immediately after students are introduced to that kind of problem, perhaps because it gives students an opportunity to focus on the execution of a strategy (e.g., procedural steps and computation). Yet students who work more than a few problems of the same kind in immediate*

> *succession are likely to receive sharply diminishing returns on their additional effort (e.g., Rohrer & Taylor, 2006; Son & Sethi, 2006)."*

It's hard to overstate how beneficial interleaving is, especially in the context of mathematics. Taylor & Rohrer (2010) found that simply interleaving practice problems, as opposed to blocking them, doubled test scores. This phenomenon was observed again by Rohrer, Dedrick, & Stershic (2015) using different, older students and more advanced math problems. As summarized by *Scientific American* (Pan, 2015):

> *"The three-month study involved teaching 7th graders slope and graph problems. Weekly lessons, given by teachers, were largely unchanged from standard practice. Weekly homework worksheets, however, featured an interleaved or blocked design. When interleaved, both old and new problems of different types were mixed together. Of the nine participating classes, five used interleaving for slope problems and blocking for graph problems; the reverse occurred in the remaining four.*
>
> *Five days after the last lesson, each class held a review session for all students. A surprise final test occurred one day or one month later. The result? When the test was one day later, scores were 25 percent better for problems trained with interleaving; at one month later, the interleaving advantage grew to 76 percent."*

As Rohrer, Dedrick, & Stershic (2015) elaborate further, students whose practice was interleaved also demonstrated vastly superior retention of the tested material through a delay period:

> *"...[A]part from its superiority to blocked practice, interleaved practice provided near immunity against forgetting, as the 30-fold increase in test delay reduced test scores by less than a tenth (from 80% to 74%).*
> *...*
> *Another reason for the large effects of interleaving observed here and elsewhere is that interleaved mathematics practice inherently guarantees that students space their practice. That is, in addition to the juxtaposition of different kinds of problems within an assignment, problems of the same kind are spaced across assignments."*

Desirable Difficulty: Why Interleaving is Underused

It is natural to ask, then: why is interleaving so rarely leveraged in classrooms? The answer is all too familiar. In addition to deviating from traditional teaching convention, interleaving has been shown to suffer from the same misconception that plagues active learning: interleaving produces more learning by increasing cognitive activation, but students often *mistakenly* interpret extra cognitive effort as an indication that they are not learning as well, when in fact the opposite is true (Kornell & Bjork, 2008). Consider the following concrete example (Brown, Roediger, & McDaniel, 2014, pp.65):

> *"In interleaving, you don't move from a complete practice set of one topic to go to another. You switch before each practice is complete. A friend of ours describes his own experience with this:*
>
> *'I go to a hockey class and we're learning skating skills, puck handling, shooting, and I notice that I get frustrated because we do a little bit of skating and just when I think I'm getting it, we go to stick handling, and I go home frustrated, saying, 'Why doesn't this guy keep letting us do these things until we get it?''*
>
> *This is actually the rare coach who understands that it's more effective to distribute practice across these different skills than polish each one in turn. The athlete gets frustrated because the learning's not proceeding quickly, but the next week he will be better at all aspects, the skating, the stick handling, and so on, than if he'd dedicated each session to polishing one skill."*

Blocking, on the other hand, creates a more comfortable sense of fluent learning which artificially improves practice performance by reducing cognitive activation. When practicing a single skill many times consecutively, students settle into a robotic rhythm of mindlessly applying one type of solution to one type of problem. The mindlessness is quite literal: in a study that measured "mind-wandering" during practice, people were found to mind-wander much more while blocking than while interleaving (Metcalfe & Xu, 2016). But the artificially improved practice performance tricks students into thinking that they are learning better, even though the effect quickly vanishes afterwards and does not actually carry over to test performance.

As summarized by Rohrer (2009):

> *"A feature that decreases practice performance while increasing test performance has been described by Bjork and his colleagues as a **desirable difficulty**, and spacing and mixing are two of the most robust ones. As these researchers have noted, students and teachers sometimes avoid desirable difficulties such as spacing and mixing because they falsely believe that features yielding inferior practice performance must also yield inferior learning."*

In the literature, a practice condition that makes the task harder, slowing down the learning process yet improving recall and transfer, is known as a **desirable difficulty**. As Rohrer & Hartwig (2020) elaborate:

> *"Both spacing and interleaving are instances of a phenomenon known as a desirable difficulty (Bjork, 1994) – the focus of this forum. A desirable difficulty is a learning method that, when compared to an alternative, makes practice more difficult while nevertheless improving scores on a subsequent test (e.g., Bjork & Bjork, 2014; Bjork, 2018; Bjork & Bjork, 2019; Bjork & Kroll, 2015; Schmidt & Bjork, 1992)."*

Many types of cognitive learning strategies introduce desirable difficulties – for instance, Bjork & Bjork (2011) list a few more:

> *"Such desirable difficulties (Bjork, 1994; 2013) include varying the conditions of learning, rather than keeping them constant and predictable; interleaving instruction on separate topics, rather*

> *than grouping instruction by topic (called blocking); spacing, rather than massing, study sessions on a given topic; and using tests, rather than presentations, as study events."*

However, as Rohrer & Hartwig (2020) explain, the idea of desirable difficulties can be counterintuitive:

> *"That difficulties can be desirable is not intuitive. In fact, many people mistakenly assume that the degree of fluency achieved during practice is a good marker of a strategy's long-term efficacy (Bjork, Dunlosky, & Kornell, 2013). Indeed, many difficulties are undesirable in that they impede not only practice performance but also test scores, as might be true for students who do homework while watching television."*

Furthermore, as Robert Bjork (1994) explains, the typical teacher is incentivized to maximize the immediate performance and/or happiness of their students, which biases them against introducing desirable difficulties:

> *"Recent surveys of the relevant research literatures (see, e.g., Christina & Bjork, 1991; Farr, 1987; Reder & Klatzky, 1993; Schmidt & Bjork, 1992) leave no doubt that many of the most effective manipulations of training – in terms of post-training retention and transfer – share the property that they introduce difficulties for the learner.*
> *...*
> *If the research picture is so clear, why then are ... nonproductive manipulations such common features of real-world training programs? ... [T]he typical trainer is overexposed, so to speak, to the day-to-day performance and evaluative reactions of his or her trainees. A trainer, in effect, is vulnerable to a type of operant conditioning, where the reinforcing events are improvements in the [immediate] performance and/or happiness of trainees.*
>
> *Such a conditioning process, over time, can act to shift the trainer toward manipulations that increase the rate of correct responding – that make the trainee's life easier, so to speak. Doing that, of course, will move the trainer away from introducing the types of* ***desirable difficulties*** *summarized in the preceding section."*

What's more, most educational organizations operate in a way that exacerbates this issue:

> *"The tendency for instructors to be pushed toward training programs that maximize the performance or evaluative reaction of their trainees during is exacerbated by certain institutional characteristics that are common in real-world organizations.*
>
> *First, those responsible for training are often themselves evaluated in terms of the performance and satisfaction of their trainees during training, or at the end of training.*
>
> *Second, individuals with the day-to-day responsibility for training often do not get a chance to observe the post-training performance of the people they have trained; a trainee's later successes and failures tend to occur in settings that are far removed from the original training environment, and from the trainer himself or herself.*

> *It is also rarely the case that systematic measurements of post-training on-the-job performance are even collected, let alone provided to a trainer as a guide to what manipulations do and do not achieve the post-training goals of training.*
>
> *And, finally, where refresher or retraining programs exist, they are typically the concern of individuals other than those responsible for the original training."*

Micro- and Macro-Interleaving

| Macro-Interleaving

Math Academy practices interleaving within review and quiz tasks, where students interleave individual practice problems within the learning task. Lessons, on the other hand, involve minimal doses of blocked practice as this is more appropriate when a student is first learning new information.

However, by breaking up our curriculum into a massive number of bite-size, atomic lessons, we implement some degree of interleaving by doing a breadth-first (as opposed to depth-first) learning path through those lessons. We call this **macro-interleaving**, as opposed to **micro-interleaving** (which entails interleaving practice problems within a single learning task).

Most resources don't leverage macro-interleaving. For instance, when learning calculus in a typical school, a class might spend a month on limits, then a month on derivative rules, then a month on integration techniques, then a month on sequences and series – essentially, macro-blocking. The class spends all their time on one unit at a time before declaring it "done" and moving to the next one. To leverage macro-interleaving, it would be better to split up every hour-long class into 15 minutes learning one bite-size topic in each of the 4 categories.

| Micro-Interleaving

Math Academy uses macro-interleaving to its fullest extent. The same can be said about micro-interleaving, even though it may not appear that way on the surface.

On the surface, it may appear that micro-interleaving is not fully leveraged when lessons (blocked practice) provide implicit spaced repetition credit towards component skills in need of micro-interleaved review. Shouldn't every topic receive micro-interleaved review before appearing on a quiz?

However, this is actually the optimal solution to a crucial tradeoff.

- If you want to micro-interleave the problem types within every single topic before seeing them on quizzes, then you have to do an explicit review on every single topic before seeing it on the quiz.

- And if you have to do an explicit review on every single topic, then pretty soon you're going to have way too many reviews and your progress is going to grind to a halt because you're spending all your time reviewing instead of learning new material (this is a common complaint about spaced repetition systems).

So, you have to make a decision: should you

1. fully micro-interleave everything before quizzes, or

2. give up a little bit of micro-interleaving to enable spaced repetition optimizations leading to much faster progress through new material?

If you want to maximize your learning efficiency, the rate at which your learning effort gets transformed into educational progress, then option 2 is better.

Furthermore, in option 2, when engaging in repetition compression, very little micro-interleaving is actually being given up. Reviews micro-interleave not only the problem types in the original lesson, but also the component (prerequisite) skills -- and reviews are specifically chosen to cover as many component skills as possible that you need practice on, so you'll actually get an outsized dose of micro-interleaving compressed into each review.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*(4), 183-186.

  Importance: *Interleaving provides minimum effective doses of review for a handful of different topics, whereas blocked practice only hits a single topic and wastes most of the review effort in the realm of diminishing returns.*

- Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24*, 355-367.

  Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied cognitive psychology, 24*(6), 837-848.

  Importance: *Interleaving enhances discrimination learning and category induction learning. While blocking leads to more rapid gains in performance (which makes it useful when first learning a skill), interleaving promotes vastly superior retention and transfer (which makes it a more effective review strategy).*

- Rohrer, D. (2009). Research commentary: The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education, 40*(1), 4-17.

  Importance: *Test scores can be dramatically improved by the introduction of spaced practice or mixed practice, which are the two defining features of mixed review. Moreover, neither spacing nor mixing requires an increase in the number of practice problems, meaning that both features increase efficiency as well as effectiveness.*

- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society, 2*(59-68).

Rohrer, D., & Hartwig, M. K. (2020). Unanswered questions about spaced interleaved mathematics practice. *Journal of Applied Research in Memory and Cognition, 9*(4), 433.

Importance: *Desirable difficulties include varying the conditions of learning, rather than keeping them constant and predictable; interleaving instruction on separate topics, rather than grouping instruction by topic (called blocking); spacing, rather than massing, study sessions on a given topic; and using tests, rather than presentations, as study events. However, people may not capitalize on these advantageous forms of practice advantage because the fact that difficulties can be desirable is not intuitive*

- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205).

  Importance: *Many of the most effective manipulations of training – in terms of post-training retention and transfer – share the property that they introduce difficulties for the learner. The typical trainer is incentivized to maximize the immediate performance and/or happiness of trainees, which biases them against introducing these types of desirable difficulties. What's more, most training organizations are set up to exacerbate this issue.*

- Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied cognitive psychology, 24*(6), 837-848.

  Pan, S. C. (2015). The interleaving effect: mixing it up boosts learning. *Scientific American, 313*(2).

  Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology, 107*(3), 900.

  Importance: *In multiple studies, simply interleaving practice problems, as opposed to blocking them, doubled or nearly doubled test scores. Additionally, students whose practice was interleaved also demonstrated vastly superior retention of the tested material through a delay period.*

- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"?. *Psychological science, 19*(6), 585-592.

  Importance: *It turns out that interleaving is NOT the "enemy of induction," even though blocking apparently created a sense of fluent learning: participants rated blocking as more effective than interleaving, even after their own test performance had demonstrated the opposite.*

- Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 978.

  Importance: *Participants were found to exhibit more frequent "mind wandering" while blocking than while interleaving.*

Additional Resources

- Carpenter, S. K., & Agarwal, P. K. (2019). How to use spaced retrieval practice to boost learning. *Iowa State University.*

- Rohrer, D., & Hartwig, M. K. (2023). Spaced and Interleaved Mathematics Practice. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 111-21). Society for the Teaching of Psychology.

- Agarwal, P. K., & Agostinelli, A. (2020). Interleaving in Math: A Research-Based Strategy to Boost Learning. *American Educator, 44*(1), 24.

- Atchison, K. (2020). True Mastery: The Benefits of Mixed Practice for Learning. *EducationalRenaissance.com.*

- Are Spacing and Interleaving the Same Thing? InnerDrive.co.uk.

- Interleaving Do's and Don'ts. InnerDrive.co.uk.

Chapter 19. The Testing Effect (Retrieval Practice)

Summary: *The testing effect (or the retrieval practice effect) emphasizes that recalling information from memory, rather than repeated reading, enhances learning. It can be combined with spaced repetition to produce an even more potent learning technique known as spaced retrieval practice. Math Academy leverages the testing effect by continuously assessing students through quick, frequent, quizzes, and encouraging students to solve review problems without referring back to reference material. Measures are taken to reduce anxiety and promote a growth mindset during quizzes.*

Retrieval is the Most Effective Method of Review

To maximize the amount by which your memory is extended when solving review problems, it's necessary to avoid looking back at reference material unless you are totally stuck and cannot remember how to proceed. This is called the **testing effect** (also known as the **retrieval practice effect**): the best way to review material is to test yourself on it. As Yang et al. (2023b) summarize:

> *"...[P]ractice testing (i.e., practice retrieval) is one of the most effective strategies to consolidate long-term retention of studied information and facilitate subsequent learning of new information, a phenomenon labeled the testing effect, the retrieval practice effect, or test-enhanced learning (Carpenter et al., 2022; Pan & Rickard, 2018; Roediger & Butler, 2011; Shanks et al., 2023; Yang et al., 2021).*
>
> *It has been firmly established that retrieval practice is more beneficial by comparison with many other learning strategies, such as restudying (Roediger & Karpicke, 2006b), note-taking (Heitmann et al., 2018; Rummer et al., 2017), concept-mapping (Karpicke & Blunt, 2011) and other elaborative strategies (Larsen et al., 2013)."*

In other words, the testing effect exposes that "following along" is not the same as learning. Students often mistakenly believe that if they can follow along with a video, book, lecture, or any other resource, without feeling confused, then they're learning. However, if you define learning as a positive change in long-term memory, then you haven't learned unless you're able to consistently reproduce the information you consumed and use it to solve problems.

This doesn't happen when you just "follow along," even if you understand perfectly. It's the act of retrieving information from memory that transfers the information to long-term memory. If you don't practice retrieval, then the information quickly dissipates. It stays with you only briefly – just long enough to trick you into thinking it'll stick with you, when it's really on the way out the door.

Amusingly, the testing effect is one of the oldest cognitive learning strategies known to humankind – records date back as far as 1620, when Francis Bacon noted (pp. 76) the following:

> *"...[Y]ou won't learn a passage as well by reading it straight through·twenty times as you will by reading it only ten times and trying each time to recite it from memory and looking at the text only when your memory fails."*

Since the early 1900s, this observation has been experimentally supported by hundreds of studies across widely different memory tasks, content domains, and experimental methodologies, which have indicated that the benefits of retrieval practice are caused by increased cognitive effort (Rowland, 2014). In particular, the testing effect has been shown to carry over to classroom settings, where frequent quizzing (with feedback) promotes greater learning on both tested *and non-tested* material (McDaniel et al., 2007). Its reliability has even been explicitly demonstrated across individual cognitive differences like working memory capacity (Pastötter & Frings, 2019). As Yang et al. (2023b) summarize:

> *"The classroom testing effect generalizes to students across different educational levels (including elementary school, middle school, high school, and university/college), and across 18 subject categories (e.g., Education, Medicine, Psychology, etc.). More importantly, the results showed that classroom quizzes not only benefit retention of factual knowledge, but also promote concept comprehension and facilitate knowledge transfer in the service of solving applied problems. Test-enhanced knowledge transfer has also been observed in many other studies (for a review, see Carpenter, 2012)."*

Spaced Retrieval Practice

What's more, as Kang (2016) notes, the testing effect can be combined with spaced repetition to produce an even more potent learning technique known as **spaced retrieval practice**:

> *"Testing or spaced practice, each on its own, confers considerable advantages for learning. But, even better, the two strategies can be combined to amplify the benefits: Reviewing previously studied material can be accomplished through testing (often followed by corrective feedback) instead of rereading.*
>
> *In fact, many studies of the spacing effect compared spaced against massed retrieval practice, not just rereading (e.g., Bahrick, 1979; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). Spaced retrieval practice (with feedback) leads to better retention than spaced rereading.*
>
> *One study examined how type of review (reread vs. test with feedback), along with timing of review (massed vs. spaced), affected eighth-grade students' retention of history facts (Carpenter, Pashler,*

> *& Cepeda, 2009). On a final test 9 months later, spaced retrieval practice yielded the highest performance (higher than spaced rereading)."*

As Halpern & Hakel (2003) elaborate:

> *"The single most important variable in promoting long-term retention and transfer is 'practice at retrieval.' This principle means that learners need to generate responses, with minimal cues, repeatedly over time with varied applications so that recall becomes fluent and is more likely to occur across different contexts and content domains. Simply stated, information that is frequently retrieved becomes more retrievable.*
> *...*
> *The effects of practice at retrieval are necessarily tied to a second robust finding in the learning literature – spaced practice is preferable to massed practice. For example, Bjork and his colleagues recommend spacing the intervals between instances of retrieval so that the time between them becomes increasingly longer – but not so long that retrieval accuracy suffers."*

And as Yang et al. (2023a, pp.257) emphasize, frequent tests are ideal:

> *"Although it has been widely documented that a single test is sufficient to enhance memory compared to restudying, many laboratory studies have observed that repeated tests (i.e., with studied content tested repeatedly) produce a larger enhancing effect on knowledge retention and transfer than a single test (e.g., Butler, 2010; Dunlosky et al., this volume; Roediger & Karpicke, 2006b).*
>
> *The enhancing effect of repeated tests has been re-confirmed by many classroom studies. Moreover, Yang et al.'s (2021) metaanalysis coded the number of test repetitions (i.e., how many times the studied information was tested), and conducted analyses to quantify the relation between the magnitude of test-enhanced learning and the number of test (quiz) repetitions. The results showed a clear trend that the more occasions on which class content is quizzed, the more effectively quizzing aids exam performance."*

The Testing Effect is Underused

Unfortunately, the testing effect remains underused in traditional classrooms, where usually only a handful of tests are given throughout the entire duration of a course. As McDaniel et al. (2007) lament:

> *"...[D]espite this impressive body of evidence, the implications of the testing effect literature for educational practice have been virtually ignored by the educational community and educational research."*

Math Academy, however, leverages the testing effect to its fullest extent by testing frequently as a part of the learning process itself. We implement a form of **continuous assessment** with quick, frequent quizzes, and we also incorporate the testing effect into normal spaced reviews (i.e.

spaced retrieval practice) by encouraging students to solve problems without referring to worked examples (though they can go back to lesson and dig up a similar example for reference if they really get stuck on a review problem). Even multi-part problems – which pull together many earlier topics to explore a challenging, complex problem context one part at a time – leverage the testing effect by requiring the student to recall each previously-learned skill and apply it to a novel context.

Yet another inefficiency of traditional classrooms is that, in addition to occurring relatively infrequently, tests and quizzes are normally focused around a single set of closely-related topics. As we discussed in the context of interleaving, Math Academy achieves far greater efficiency by selecting a broad mixture of topics – not only on reviews and multi-part problems, but also on assessments. This produces much more forward progress in spaced repetition schedules and also helps students learn to match problems with the appropriate solution techniques.

Reducing Anxiety and Promoting a Growth Mindset

| Appropriate vs Inappropriate Usage of Timed Tests

Many people view tests, especially timed tests, as anxiety-inducing and consequently something to be avoided. However, it is important to realize that test anxiety can be mitigated, and often even reduced, by giving frequent, low-stakes quizzes on skills that a student is ready to be tested on.

Often, negative feelings toward timed tests are the result of *inappropriate usage* of the timed test, such as introducing it too early in the student's skill development process. A prerequisite for timed testing is that the student should be able to perform the tested skills successfully in an untimed setting. Timed testing demands a high level of proficiency, and anxiety can be produced if there is a mismatch between a student's level of proficiency and the performance expectations that are placed on them.

As Codding, Peltier, & Campbell (2023) summarize:

> *"Learners may benefit more or less from various instructional strategies or tactics, depending on the learners' stage of skill development (Burns et al., 2010). That is, are learners working on acquiring a math skill or concept, building skill fluency, generalizing or transferring a skill or concept, or using known skills and concepts to solve novel problems?*
>
> *Just because timed practice opportunities have been proven effective to build fluency, for example, does not mean that timed trials always benefit learners (Fuchs et al., 2021). Using timed trials*

> *with students who are working to acquire new knowledge or skills is an instructional mismatch; rather, students need to display accuracy with skills and concepts before building fluency. It is not the fault of the strategy; it is an issue with when to implement the strategy."*

| Desirable vs Undesirable Difficulties

More generally, while desirable difficulties are a necessary component of effective practice, they are only effective insofar as the learner is able to overcome them. Introducing an insurmountable difficulty is never desirable, even if that type of difficulty may be desirable later on in the learning process once the student has increased their proficiency. It is the act of *overcoming* a desirable difficulty that leads to greater learning. As echoed by Brown, Roediger, & McDaniel (2014, pp.98-99):

> *"Elizabeth and Robert Bjork, who coined the phrase 'desirable difficulties,' write that difficulties are desirable because 'they trigger encoding and retrieval processes that support learning, comprehension, and remembering. If, however, the learner does not have the background knowledge or skills to respond to them successfully, they become undesirable difficulties.'*
> *...*
> *Clearly, impediments that you cannot overcome are not desirable. ... To be desirable, a difficulty must be something learners can overcome through increased effort."*

As Bjork & Bjork (2023, pp.22) elaborate:

> *"...[I]t is necessary to consider what level of difficulty is appropriate in order for that level to enhance a given student's learning, and the appropriate level that is optimal may vary considerably based on a student's background and prior level of knowledge.*
>
> *To illustrate, while it is typically desirable to have learners generate a skill or some knowledge from memory, rather than simply showing them that skill or presenting that knowledge to them, a given learner needs to be equipped via prior learning to succeed at the generation task – or at least succeed in activating relevant aspects of the necessary skill or knowledge – for the act of generating to then potentiate their subsequent practice or study (e.g., Little & Bjork, 2016; Richland, Kornell, & Kao, 2009)."*

Indeed, Codding, VanDerHeyden, & Chehayeb (2023) found that when the type of instruction is mismatched against a student's level of proficiency, the instruction will not only be ineffective, but can also lead to anxiety:

> *"This study illustrated that when instructional strategies are misaligned with students' stage of skill development, even when the instructional target is appropriate, students' math performance will not improve. Furthermore, as suggested in this study, students may exhibit higher levels of anxiety and lower acceptability of misaligned instructional practices."*

| Appropriate Timed Testing Can Reduce Math Anxiety

However, when used *appropriately*, timed testing can be a valuable tool for overcoming math anxiety by building fluency and automaticity. According to VanDerHeyden & Codding (2020), who have extensive experience researching academic intervention in mathematics, the relationship between math anxiety and timed testing is unclear, but there is a clear relationship between math anxiety and math proficiency (lower proficiency promotes anxiety, which further hinders skill development), and timed tests are useful for building proficiency:

> *"Teachers and parents worry about math anxiety, and some math education experts caution against tactics used in math class, such as timed tasks and tests, that might theoretically stoke anxiety (Boaler, 2012). First, the evidence does not support that people are naturally anxious or not anxious in the context of math assessment and instruction (Hart & Ganley, 2019). Second, simply avoiding math or certain math tactics should not be expected to ameliorate anxiety in the long term. Third, preventing a student from full exposure to math assessment and intervention costs the student the opportunity to develop adaptive coping mechanisms to deal with possible anxiety in the face of challenging academic content.*
> *...*
> *Gunderson, Park, Maloney, Bellock, and Levine (2018) found a reciprocal relationship between skill proficiency and anxiety, such that weak skill reliably preceded anxiety and anxiety further contributed to weak skill development. They found that anxiety could be attenuated by two strategies: improving skill proficiency (this cannot be done by avoiding challenging math work and timed assessment) and promoting a growth mindset (as opposed to a fixed ability mindset) using specific language and instructional arrangements to promote the idea that I, as a student, can work hard and beat my score; I can grow today; my brain is like a muscle that gets stronger when I work it with challenging math content.*
> *...*
> *There is very little empirical evidence examining whether timed tests have a causal impact on anxiety, and the existing few studies that include school-age participants do not support the idea (Grays, Rhymer, & Swartzmiller, 2017; Tsui & Mazzocco, 2006). What is clear is there is a modest, negative bidirectional relationship between math anxiety and math performance (Namkung et al., 2019). These correlational data suggest that poor mathematics performance can lead to high math anxiety and that high math anxiety can lead to poor mathematics performance. The remedy that school psychologists can advocate for is to identify, through effective and efficient screening, the presence of high math anxiety and determine which students would benefit from supplemental and targeted mathematics supports. Intervention approaches should target math skill deficits, address high anxiety, and promote a growth mindset as well as monitor progress toward clearly defined objectives using tools that are brief (often timed), reliable, and valid."*

These sentiments are echoed by the U.S. Department of Education (Fuchs et al., 2021, pp.58), which recommends regularly using timed review activities to promote automatic retrieval of previously-learned material, since students will struggle to learn more advanced material unless they are able to automatically retrieve previously-learned material:

> *"Regularly include timed activities as one way to build students' fluency in mathematics. ... [However,] Do not use timed activities to introduce and teach mathematics concepts and operations.*

> *...*
> *Quickly retrieving basic arithmetic facts (addition, subtraction, multiplication, and division) is not easy for students who experience difficulties in mathematics. Without such retrieval, students will struggle to follow their teachers' explanations of new mathematical ideas. Automatic retrieval gives students more mental energy to understand relatively complex mathematical tasks and execute multistep mathematical procedures.*
>
> *Thus, building automatic fact retrieval in students is one (of many) important goals of intervention. In addition to basic facts, timed activities may address other mathematical subtasks important for solving complex problems."*

As summarized by Yang et al. (2023b), quizzes can increase students' skill proficiency and familiarity with the format of assessment, which can reduce their test anxiety:

> *"...[I]t is well-known that tests motivate students to study harder (Yang et al., 2017a), encourage them to read the assigned textbook materials to prepare for the lecture (Heiner et al., 2014), reduce mind wandering while learning (Szpunar et al., 2013), and increase class attendance (Schrank, 2016).*
>
> *These beneficial effects of practice tests [i.e., quizzes] may make students more prepared for tests and reduce their worry about poor test performance, therefore alleviating TA [test anxiety] (Brown & Tallon, 2015; Yusefzadeh et al., 2019). Furthermore, tests may inform students about the formats and contents of future assessments, hence reducing uncertainty (i.e., uncertainty about how and what content will later be assessed) and mitigating anxiety (Jerrell & Betty, 2005)."*

What's more, as Hattie & Yates (2013, pp.59) explain, performing well on a timed test has been shown to build confidence and promote positive feelings:

> *"...[S]tudies conducted under laboratory conditions show that, for both adults and children, speed of access in memory functions strongly predicts two other attributes: confidence and positive feelings. Whenever people are able to recall important information quickly there is an inherent sense in that the information is correct, together with a momentary flush of pleasure."*

Indeed, in a study of thousands of middle and high schoolers' reactions to frequent (at least weekly), low-stakes, immediate-feedback quizzes during class, Agarwal et al. (2014) found that most students felt it made them less nervous for higher-stakes tests, and students were more likely to report a decrease in overall test anxiety than an increase:

> *"We asked students whether clicker quizzes (i.e., retrieval practice) made them more or less nervous for unit tests and exams ... Remarkably, 72% of students reported that retrieval practice made them less nervous for tests and exams, 22% said they experienced about the same level of nervousness, and only 6% of students said clickers made them more nervous.*
>
> *Next, we asked students whether they experienced more, less, or about the same level of test anxiety for the class with retrieval practice compared to other classes in which they did not have retrieval practice ... [O]nly 19% of students reported experiencing more anxiety, while 81% of*

> *students said they experienced about the same level of test anxiety or less in the class with retrieval practice compared to their other classes (33% reported less nervousness).*
> *...*
> *[T]he use of clicker response systems reduced self-reported test anxiety. ... We hypothesize that students became familiar with taking quizzes, knew the course material better, and hence were less anxious when facing the unit test on which they would receive a grade."*

As echoed by Yang et al. (2023a):

> *"...[F]requent testing has little impact on or even reduces (rather than increases) test anxiety. For instance, in a large sample study (over 1,000 college participants), Yang et al. (2020) observed that interpolating tests across a study phase has minimal influence on participants' test anxiety. Szpunar et al. (2013) found that frequent tests significantly relieve test anxiety (for related findings, see Khanna, 2015). Furthermore, in a large-scale survey conducted by Agarwal et al. (2014), 72% of 1,306 middle and 102 high school students reported that frequent quizzes made them less anxious about exams, with only 8% reporting the opposite."*

In a separate meta-analysis, Yang et al. (2023b) summarized some other empirical studies observing that quizzes reduced test anxiety:

> *"...[I]n a quasi-experimental study conducted by Piroozmanesh and Imanipour (2018), two classes of nursing undergraduates took a coronary care course, with the experimental class taking class quizzes across the semester, whereas the control class did not take these quizzes. For both classes, students' TA was measured at the beginning of the semester (pretest) and one week before the final exam (posttest). The results showed that although there was minimal difference in TA during the pretest between the two classes, students in the experimental class were much less anxious before the final exam than those in the control class.*
> *...*
> *Szpunar et al. (2013) obtained consistent findings. Specifically, after both the test and the no-test group completed the interim test on Segment 4, both groups were told that they would take a cumulative test on all four segments and were instructed to report how anxious they were about the cumulative test. Consistent with the findings from Piroozmanesh and Imanipour (2018) and Brown and Tallon (2015), Szpunar et al. (2013) observed that participants in the test group were much less anxious than those in the no-test group."*

The meta-analysis, which included 24 studies across thousands of participants, ultimately concluded that quizzes reduce test anxiety about as much as they increase academic performance (in both cases, a medium effect size of about 0.5).

> *"The current review integrates results across 24 studies (i.e., 25 effects based on 3,374 participants) to determine the effect of practice tests (quizzes) on test anxiety (TA) and explore potential moderators of the effect. The results show strong Bayesian evidence (BF_{10}>25,000) that practice tests appreciably reduce TA to a medium extent (Hedges' g=-0.52), with minimal evidence of publication bias.*
> *...*
> *In a recent meta-analytic review, Yang et al. (2021) integrated data from over 48,000 students, extracted from 222 classroom studies, to determine whether class quizzes improve students'*

> *academic performance. The answer is affirmative: Class quizzes enhance students' academic attainment to a medium extent (Hedges' g=0.50)."*

| Implementing Appropriate Timed Testing

Granted, in a traditional classroom, it is difficult to keep instructional practices aligned to student proficiencies because each student develops their skills at a different rate. For any given skill, at any given time, some students may be ready for timed testing while others may need additional practice – but the teacher generally does not have enough bandwidth to manage different learning tasks for different students on different skills, and the best they can do is provide learning tasks that feel appropriate for the class "on average." Of course, those learning tasks will be inappropriate for some students and may lead to decreased learning and increased anxiety.

Math Academy, however, adapts the level of instruction to each individual student on each individual skill. Students initially learn skills during highly-scaffolded lessons, where they are given as much practice as they need to master the skills. Only after they demonstrate their ability to perform the skills do they begin seeing those skills on higher-intensity forms of practice like timed quizzes.

The quizzes are low-stakes and frequent, and are structured in a way that promotes an "I can do this" growth mindset. Whenever a student misses a question on a quiz, they receive a remedial review on the corresponding topic so that they can increase their proficiency in that area. If a student does less than "well" on a quiz, then they are also given the opportunity to retake the quiz to demonstrate their improved proficiency. The goal is not only to give students realistic feedback about their skill proficiency, but also to demonstrate to students that they can improve their proficiency by putting forth effort on their learning tasks.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological bulletin, 140*(6), 1432.

  Importance: *The testing effect has been experimentally supported by hundreds of studies across widely different memory tasks, content domains, and experimental methodologies, which have indicated that the benefits of retrieval practice are caused by increased cognitive effort.*

- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European journal of cognitive psychology, 19*(4-5), 494-513.

  Yang, C., Shanks, D. R., Zhao, W., Fan, T., & Luo, L. (2023). Frequent Quizzing Accelerates Classroom Learning. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 252-62). Society for the Teaching of Psychology.

  Importance: *The testing effect has been shown to carry over to classroom settings, where frequent quizzing (with feedback) promotes greater learning on both tested and non-tested material.*

- Pastötter, B., & Frings, C. (2019). The forward testing effect is reliable and independent of learners' working memory capacity. *Journal of cognition, 2*(1).

  Importance: *The reliability of the testing effect has even been explicitly demonstrated across individual cognitive differences like working memory capacity.*

- Codding, R. S., Peltier, C., & Campbell, J. (2023). Introducing the Science of Math. *TEACHING Exceptional Children*, 00400599221121721.

Codding, R. S., VanDerHeyden, A., & Chehayeb, R. (2023). Using Data to Intensify Math Instruction: An Evaluation of the Instructional Hierarchy. *Remedial and Special Education*, 07419325231194354.

Importance: *Often, negative feelings toward timed tests are the result of inappropriate usage of the timed test, such as introducing it too early in the student's skill development process. A prerequisite for timed testing is that the student should be able to perform the tested skills successfully in an untimed setting. Timed testing demands a high level of proficiency, and anxiety can be produced if there is a mismatch between a student's level of proficiency and the performance expectations that are placed on them.*

- VanDerHeyden, A. M., & Codding, R. S. (2020). Belief-Based versus Evidence-Based Math Assessment and Instruction. *Communique, 48*(5).

Fuchs, L. S., Bucka, N., Clarke, B., Dougherty, B., Jordan, N. C., Karp, K. S., ... & Morgan, S. (2021). Assisting Students Struggling with Mathematics: Intervention in the Elementary Grades. Educator's Practice Guide. WWC 2021006. *What Works Clearinghouse.*

Importance: *The relationship between math anxiety and timed testing is unclear, but there is a clear relationship between math anxiety and math proficiency (lower proficiency promotes anxiety, which further hinders skill development), and timed tests are useful for building proficiency. Timed review activities should be used to promote automatic retrieval of previously-learned material, since students will struggle to learn more advanced material unless they are able to automatically retrieve previously-learned material.*

- Agarwal, P. K., D'antonio, L., Roediger III, H. L., McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of applied research in memory and cognition, 3*(3), 131-139.

Yang, C., Shanks, D. R., Zhao, W., Fan, T., & Luo, L. (2023a). Frequent Quizzing Accelerates Classroom Learning. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 252-62). Society for the Teaching of Psychology.

Yang, C., Li, J., Zhao, W., Luo, L., & Shanks, D. R. (2023b). Do practice tests (quizzes) reduce or provoke test anxiety? A meta-analytic review. *Educational Psychology Review, 35*(3), 87.

Importance: *Frequent, low-stakes, immediate-feedback quizzes can make students less nervous for higher-stakes tests and reduce their overall test anxiety.*

Additional Resources

- Rohrer, D., Dedrick, R. F., & Agarwal, P. K. (2017). Interleaved mathematics practice: giving students a chance to learn what they need to know. *RetrievalPractice.org.*

- Agarwal, P. K., Roediger, H. L., McDaniel, M. A., & McDermott, K. B. (2013). How to use retrieval practice to improve learning. *Washington University in St. Louis.*

- Gonzalez, J. (2017). Retrieval Practice: The Most Powerful Learning Strategy You're Not Using. *CultOfPedagogy.com.*

Chapter 20. Targeted Remediation

Summary: *Math Academy provides automated, precise support to help students strengthen weaknesses on specific topics or component skills that are personal sources of struggle. The bar for success is never lowered; rather, students are given additional practice that helps them clear the bar fully and independently on their next attempt.*

High-Granularity, High-Integrity Remediation

In the academic literature, the term *targeted remediation* usually describes identifying individual students in need of broad remedial intervention such as tutoring, remedial courses, academic advisor meetings, etc.

But in the context of Math Academy, **targeted remediation** refers to fully-automated support mechanisms that are targeted to individual students *on individual topics* – and often even more precisely to the individual component skills that are causing a student to struggle on a topic.

Math Academy's targeted remediation is different from the concept of *adaptive feedback* in intelligent tutoring systems, which the *Handbook of Learning Analytics* describes as providing hints to the learner or recommendations for the instructional designer to better match a task to students' abilities (Pardo et al., 2017, pp.166):

> *"A large portion of the studies related to adaptive feedback have been developed through ... systems that provide a set of learning tasks to students in specific knowledge domains. ... the system commonly offers various types of task-level feedback, such as next-step hints (e.g., Peddycord, Hicks, & Barnes, 2014); correctness hints, also known as flag feedback (Barker-Plummer, Cox, & Dale, 2011); positive or encouraging hints (Stefanescu, Rus, & Graesser, 2014); recommendations on next steps or tasks [not for the students, but for the instructional designer to better match the task to students' abilities] (Ben-Naim, Bain, & Marcus, 2009); or various combinations of the above."*

Unlike the forms of adaptive feedback described above, which effectively lower the bar for success on the learning task, Math Academy's targeted remediation mechanisms keep the bar where it's at. Instead, we focus on actions that are most likely to strengthen a student's area of weakness and empower them to clear the bar fully and independently on their next attempt.

To our best knowledge, targeted remediation at Math Academy's level of **granularity** (individual students on individual topics) and **integrity** (maintaining the bar for success) has not been studied in academic literature.

As stated by the handbook (Pardo et al., 2017, pp.168):

> *"A dearth of research explores how students interact with and are transformed by algorithm-produced feedback. Furthermore, the relationship between the type of interventions that can be derived from data analysis and adequate forms of feedback remains inadequately explored."*

This may be for reasons of academic infeasibility: the expense (both time and money) to develop an automated learning system is large and increases proportionally with the granularity of the curriculum, making it a rather industrial endeavor.

Corrective Remediation

When students struggle, we follow up with **corrective remedial support** that is **targeted** to their specific point of struggle.

- If they struggle during a task, we give more questions – that is, more chances to learn and demonstrate their learning.
- If they fail a lesson, we give them a break and enable them to make progress learning unrelated topics before asking them to re-attempt the failed lesson. Usually, all it takes to rebound is a bit of rest and a fresh pair of eyes. On average, students pass their first attempt 95% of the time, and within two attempts 99% of the time, without any further intervention.
- However, if we detect that they get stuck again in the same place in a lesson, without making any additional forward progress, we give them **remedial reviews** to help them strengthen their foundations in the areas most relevant to their point of struggle.
- Whenever they miss a question on a quiz, we immediately follow up with a remedial review on the corresponding topic.

One challenge in properly targeting remedial reviews is that often, the key prerequisite concepts or skills required to solve a particular problem lie several steps back in the hierarchy of mathematical knowledge. However, when developing our content and building our knowledge graph, we explicitly keep track of the **key prerequisites** that are used in each part of each lesson. This allows us to pinpoint the exact topics that are necessary for successful remediation.

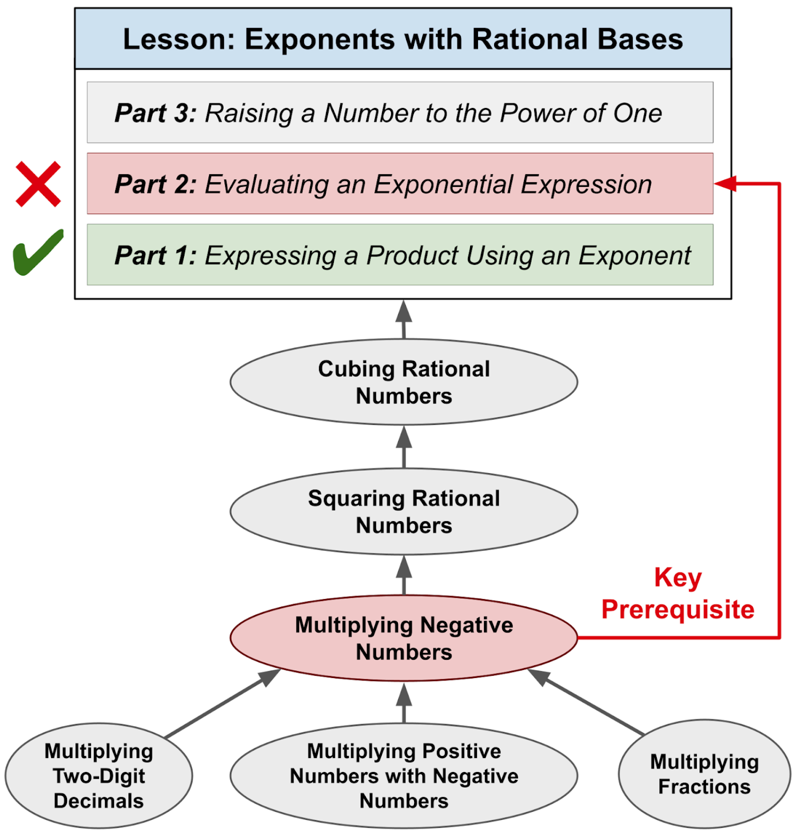
As a concrete example, suppose that while re-attempting the lesson *Exponents with Rational Bases*, a student

- manages to pass *Part 1: Expressing a Product Using an Exponent*, e.g. expressing $4 \times 4 \times 4$ as 4^3 , but

- gets stuck again at *Part 2: Evaluating an Exponential Expression*, e.g. computing $(-4)^3 = (-4) \times (-4) \times (-4)$.

In this situation, the student has demonstrated that they understand the concept of an exponent, but they are struggling to use multiplication to compute the result.

Although multiplication occurs several steps back in the sequence of prerequisites, we have linked *Part 2: Evaluating an Exponential Expression* to the key prerequisite topic *Multiplying Negative Numbers*, which allows us to automatically trigger a targeted remedial review on *Multiplying Negative Numbers*.



Preventative Remediation

We also attempt to predict struggle beforehand and leverage **preventative remediation** to avoid the struggle entirely. Conveniently, this happens naturally when we tailor the spaced repetition process to individual student-topic learning speeds.

The initial starting value of a student-topic learning speed is a prediction of how difficult that topic is going to be for the given student. The prediction is primarily based on learning speeds of other related topics – so if the predicted learning speed is low (i.e. we predict that the student is going to struggle on the topic), then it is low because one or more of the other related topics has a low learning speed.

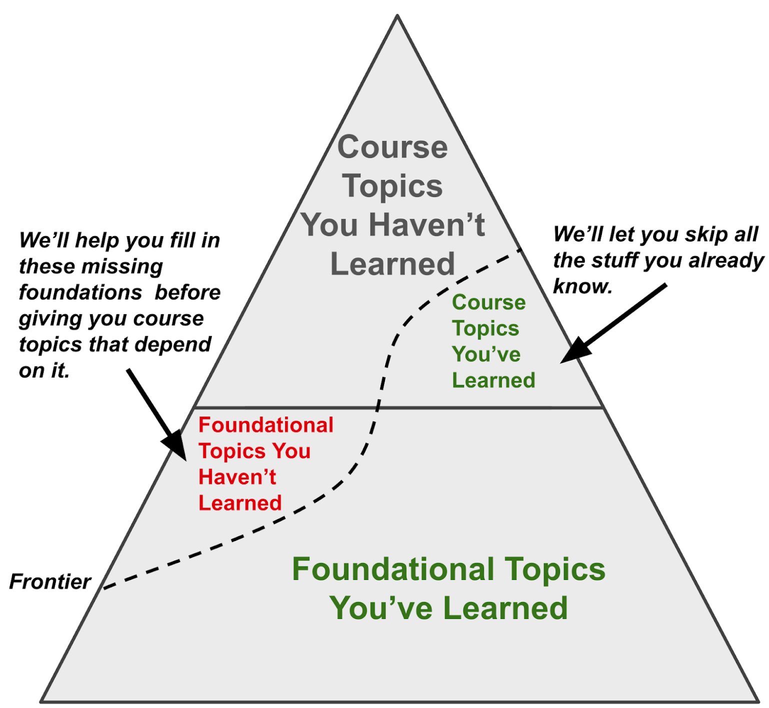
Those other related topics with low learning speed are the predicted points of failure in the student's predicted struggle, we are already performing preventative remediation on them by slowing down their spaced repetition processes and forcing explicit reviews. In other words, "post"-remediation of earlier topics naturally functions as "pre"-remediation for later topics.

Foundational Remediation

Math Academy's diagnostics are tailored to specific courses – but in addition to assessing knowledge of course content, they also assess knowledge of lower-grade **foundations** that students need to know in order to succeed in the course (i.e. they are prerequisites for the course).

For instance, students need to know plenty of arithmetic in order to solve problems in algebra. So, the foundations of algebra include those necessary topics from arithmetic. Likewise, the foundations of calculus include plenty of algebra and also some geometry, and the foundations of most university-level courses (such as multivariable calculus) include plenty of single-variable calculus and precalculus.

It is common for incoming students to lack some foundational knowledge that is necessary to succeed in their chosen course. While this could spell doom in a traditional classroom, Math Academy is able to estimate a student's knowledge frontier *even if it is below their course*, and help them fill in any missing foundational knowledge while simultaneously allowing them to learn course topics that don't rely on that missing foundational knowledge.



Math Academy also optimizes the timing of when to have students begin shoring up their missing foundations. Students are generally more excited to work on topics in the course that they are enrolled in, than they are to shore up missing foundations – and students tend to be more productive and consistent when they're excited about what they're doing. So, we allow students to start out completing the topics in their enrolled course that don't depend on their missing foundations. This helps students build up some momentum, make some progress towards their primary goal, and get into a habit of frequent learning. Once a student reaches the point where they need to shore up missing foundations in order to continue making progress in their enrolled course, they have built up plenty of momentum that will help carry them through the process of foundational remediation and make them far less likely to get discouraged and quit.

Content Remediation

As a mastery learning system, Math Academy holds its students accountable for learning – and in return, our students hold us accountable for providing material that is properly scaffolded and easy to learn from. If there is ever a topic that more than a tiny percentage of students struggle with, then we see it as an indication that we need to not only remediate the students, but also remediate our own content.

We take **content remediation** extremely seriously. Math Academy is like a tutor whose livelihood depends on the actual learning outcomes of its students – unlike many other learning platforms (and even many human teachers) that let students move on to more advanced content despite poor performance on prerequisite content. If a student can't succeed in mastering the material that we ask them to learn, then we are out of a job.

To help us remediate our content, we have developed learning analytics tools that allow us to analyze the performance of any piece of content, at any level of granularity: not just individual topics, but also each individual knowledge point within a topic, and each individual question within a knowledge point.

If the pass rate for any lesson is unacceptably low, we can pinpoint the exact knowledge point(s) within that topic where students are getting stuck, as well as any particular questions within that knowledge point that are causing issues.

By continually refining our content and algorithms over the course of many years, we have reached the point that **our students pass lessons 95% of the time on the first try and 99% of the time within two tries.** As we continue refining our content into the future, these pass rates will continue to increase.

It's worth emphasizing that when we refine and remediate our own content, we do **not** lower standards. The way we raise pass rates is by introducing more scaffolding into lessons to further reduce cognitive load. Sometimes this means improving the way a concept or worked example is explained; other times it means adding an intermediate knowledge point to a lesson, or occasionally even splitting an entire topic into two or more different topics that more specifically address different contexts of the original topic.

Chapter 21. Gamification

Summary: *Gamification, integrating game-like elements into learning environments, proves effective in increasing student learning, engagement, and enjoyment. Math Academy utilizes eXperience Points (XP) to gamify learning, incentivizing both quantity and quality of work. XP awards bonus points for stellar performance and introduces penalties for poor efforts, preventing exploitation by adversarial students. Math Academy also remedies loopholes that are typically found (and exploited) in traditional classrooms.*

Importance of Gamification

| Increasing Learning, Engagement, and Enjoyment

A common theme across many of the cognitive learning strategies described in this document has been that they produce more learning by increasing cognitive activation, which students find less enjoyable because it's more mentally taxing. Furthering the inconvenience, students often *mistakenly* interpret extra cognitive effort as an indication that they are not learning as well, when in fact the opposite is true.

Thankfully, the strategy of **gamification** behaves differently. Numerous studies have shown that when game-like elements (such as points and leaderboards) are integrated into student learning environments in ways that are

1. aligned with the goals of a course, the motivations of the students, and the context of the educational setting, and

2. robust to "hacking" or "gaming the system" (i.e. behaviors that attempt to bypass learning by exploiting loopholes in the rules of the game),

students typically not only learn more and engage more with the content, but also enjoy it more (Bai, Hew, & Huang, 2020; Looyestyn et al., 2017; Lei et al., 2022).

This applies not only to young students, but also to university-level students and even postgraduate students in technically-challenging courses. As the authors of a gamification study at Delft University of Technology describe (Iosup & Epema, 2014):

> *"Over the past three years, we have applied gamification to undergraduate and graduate courses in a leading technical university in the Netherlands and in Europe. Ours is one of the first long-running attempts to show that gamification can be used to teach technically challenging courses.*
>
> *The two gamification-based courses, the first-year B.Sc. course Computer Organization and an M.Sc.-level course on the emerging technology of Cloud Computing, have been cumulatively followed by over 450 students and passed by over 75% of them, at the first attempt.*
>
> *We find that gamification is correlated with an increase in the percentage of passing students, and in the participation in voluntary activities and challenging assignments. Gamification seems to also foster interaction in the classroom and trigger students to pay more attention to the design of the course. We also observe very positive student assessments and volunteered testimonials, and a Teacher of the Year award."*

| Increasing Learning Efficiency

Clearly, gamification is a potent strategy for maintaining student motivation and helping students feel good about hard work. (Any readers with experience in high-performance athletics will know the wonders that a bit of gamification can do for maintaining morale while working hard at practice – usually in the form of tracking personal progress or engaging in friendly competition with teammates.)

But even more importantly, gamification also functions as a lever by which to incentivize high-quality work. Because adaptive learning systems like Math Academy speed up or slow down based on student performance, a student's **learning efficiency** depends highly on the quality of their work:

- a student who performs well can make a lot of progress in a course by doing a relatively small amount of work, while

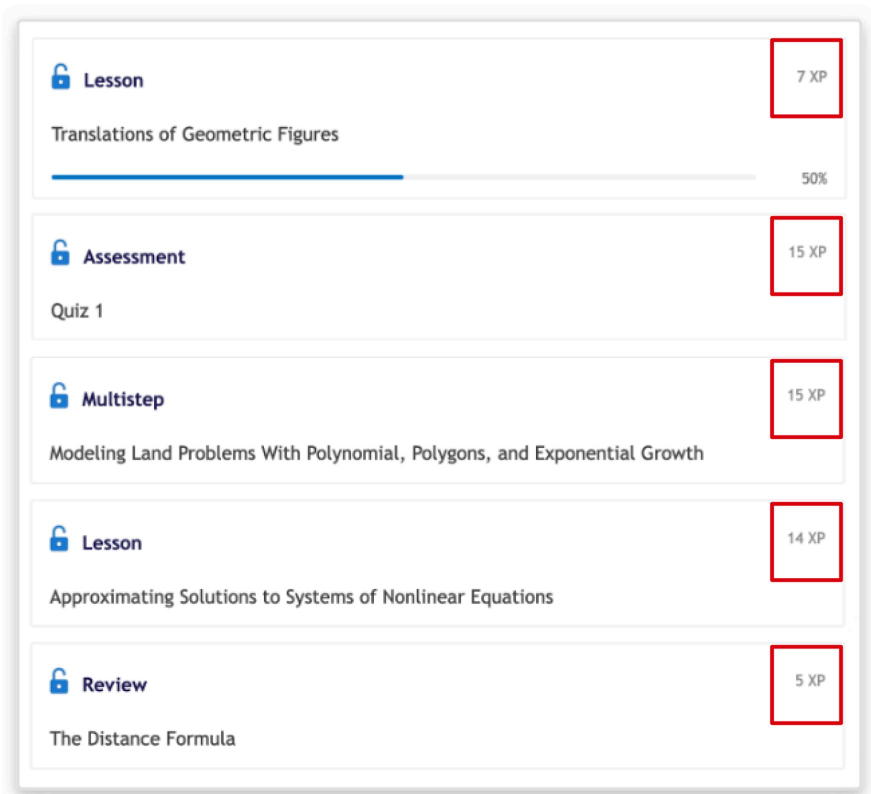
- a student who performs poorly will have to do significantly more work to make the same amount of progress.

In effect, for a student to make educational progress in an adaptive learning system like Math Academy, they have to put forth a sufficient amount of high-quality work.

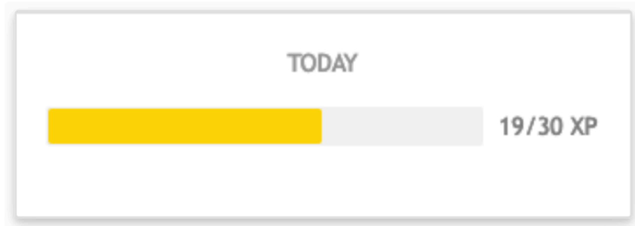
Incentivizing Quantity and Quality of Work

| XP-Time Equivalence

To incentivize both quantity and quality of work, Math Academy uses **eXperience Points** (**XP**) to implement a gamified reward system. Students earn XP upon successful completion of learning tasks, and XP is calibrated so that 1 XP represents 1 minute of fully-focused, fully-productive work for an average serious (but imperfect) student.

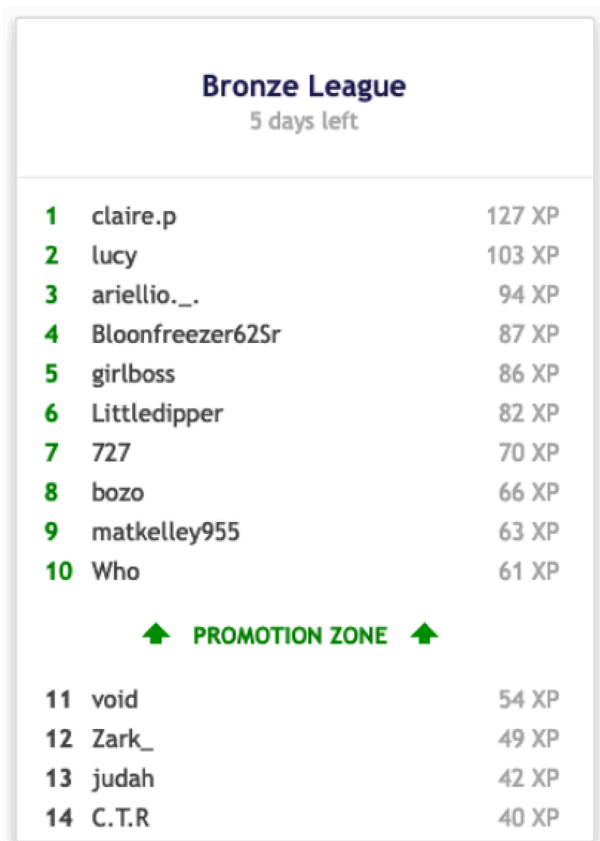


XP makes it easy for parents and teachers to set reasonable learning goals as a daily target number of XP, and it gives the system a lever by which to incentivize student behavior that is beneficial to learning.



| Incentivizing Quantity of Work

For instance, to incentivize students to put forth a sufficient quantity of work, we implemented (optional) **competitive weekly leaderboards** where students are grouped into smaller **leagues** with other students of similar competitive ability. If a student earns enough XP to end the week near the top of their league, they promote to a higher league. But if they end the week near the bottom of their league, they demote to a lower league.

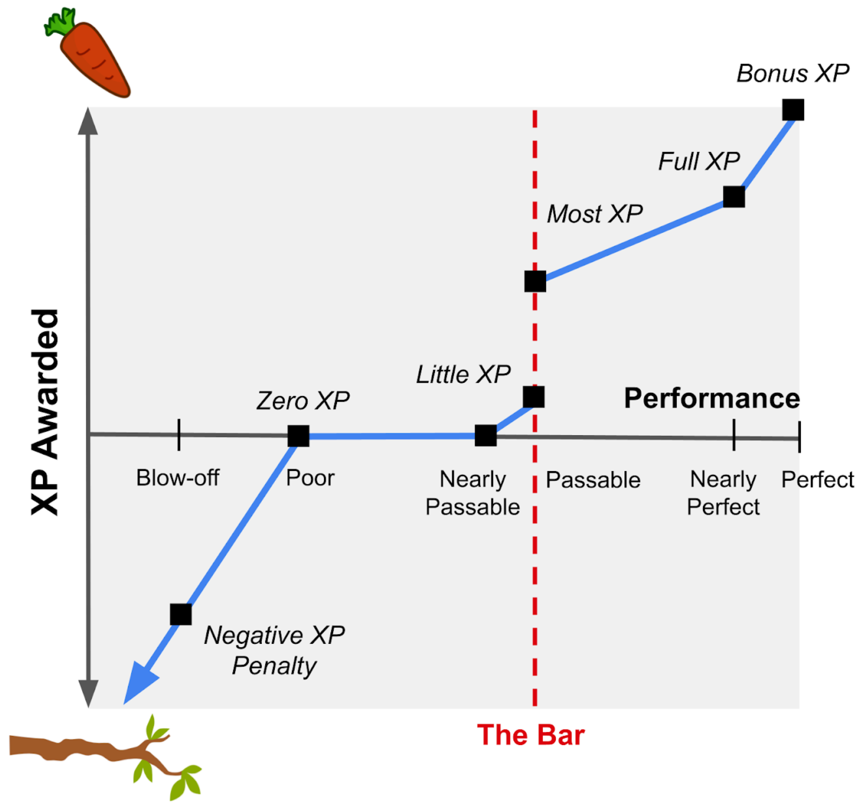


| Incentivizing Quality of Work

Likewise, to incentivize students to maintain high quality of work, we scale XP awards so that there is a large reward for doing a stellar job (as opposed to just "good enough"), and students must clear a bar in order to earn any XP. This stands in contrast to traditional aggregate-percentage grades, which provide minimal reward for going above and beyond while simultaneously often allowing students to "get by" with poor performance.

XP allows us to implement a "carrot and stick" approach to incentivizing student effort: we award

- bonus XP for perfect performance – awarding bonus points for high performance has been shown to increase performance (Egram, 1979),
- full XP for nearly perfect performance,
- most XP for otherwise passable performance,
- a little XP for nearly passable performance,
- zero XP for poor performance, and
- a negative XP penalty for blowing off a task.



Closing Loopholes

Most students use Math Academy properly and therefore rarely (if ever) see XP penalties. However, we have experienced on numerous occasions that in the absence of a penalty system, adversarial students will complete tasks that they feel are easy and then submit random guesses to intentionally fail out of tasks that require more effort. This tanks their performance, causing their adaptive learning schedule to slow down or even begin falling backwards, which drastically slows or even prevents progress towards their educational goals.

We call these students "XP hackers." They engage in this behavior because they are trying to minimize their effort per XP. Without XP penalties, the XP hacker strategy can be exploited indefinitely and students can rack up XP without making progress.

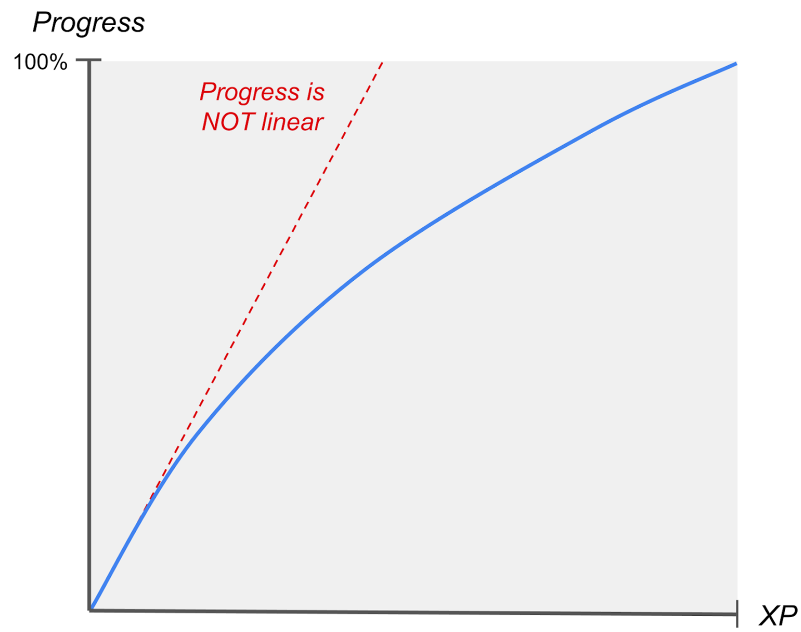
As Baker et al. (2006) noted, a way to prevent adversarial students from gaming the system is to tweak the rules in a way that *"change[s] the incentive to game – whereas gaming might previously have been seen as a way to avoid work, it now leads to extra work."* In our case, this means taking away some XP whenever a student blows off a task (and even more XP if they continue blowing off tasks). By introducing a penalty, we tweak the game so that the way to minimize effort per XP is to give a legitimate effort on every task.

In order to trigger and calibrate XP penalties appropriately, we interpret penalties as conveying how frustrated a teacher, tutor, or guardian sitting next to the student would become. After implementing XP penalties, we found that many adversarial students' rates of passing learning tasks jumped from under 50% to over 90%, while students who used the system properly and truly gave their best effort rarely (if ever) experienced penalties.

Math Academy also remedies loopholes that are typically found (and exploited) in traditional classrooms. For instance, the most obvious enabler of cheating in traditional classrooms is giving all students the same homework and assessments. But Math Academy customizes its learning path to each individual student, so it's unusual for classmates to have the opportunity to work on the same topic at the same time – and even if they do, then they are served different questions, since we have a large bank of questions for each topic. Our assessments are also fully individualized and even randomized, meaning that there is absolutely no edge that a student can gain from seeing a classmate's quiz. And if a student fails a task and has to re-attempt it, we change up the questions and even wait for a delay period before allowing the re-attempt (in the meantime, the student is able to continue making progress along other learning paths).

Progress vs XP

It's important to realize that a student's progress (percent of topics completed) in a course is highly correlated with, but fundamentally different from, the amount of XP that they have earned in the course. The only time a student's progress percent increases is when they complete a lesson. As a student gets further into their course (and math in general), more review is required to maintain their growing knowledge base. As a result, students make progress faster at the beginning of a course than they do at the end of a course.



Progress is nonlinear. Students make progress very quickly at the beginning of a course because they can focus primarily on learning new topics (i.e. lessons) as opposed to maintaining existing knowledge (i.e. reviews). But the more they learn, the more there is to review – so progress slows down. That said, we have a hard rule to ensure that on average, students have the opportunity to work on a lesson at least ~25% of the time or so at a minimum.

At surface level, it might seem like it would be more straightforward to measure progress in terms of XP completed relative to the total estimated XP in the course. However, this would create issues because the amount of XP in a course can change significantly in response to changes in student performance (because the spaced repetition process speeds up when students are doing well and slows down when students are struggling). If progress were measured in terms of XP, then a student could run into a situation where they are completing lessons but their progress is going down because their overall performance is decreasing, which would be far more counterintuitive and confusing.

It is also worth noting that progress naturally slow downs at the end of a course, when a student only has a handful of topics remaining. Often, when we give a student a new lesson, we are actually knocking out one or more due reviews with that lesson. The more lessons are on the student's "knowledge frontier," the more likely it is that we can find a new lesson to knock out some due reviews. The flipside is that when a student only has a handful of lessons left in a course, it severely restricts our ability to carry out this sort of optimization. To be clear, the system is not moving slowly in an absolute sense, just “less fast” relative to the normal turbo-boosted behavior, because it is unable to take advantage of a strategy that it normally uses to turbo-boost the rate at which students can make progress.

While this constraint can be circumvented by allowing the system to receive topics from the next course (that knock out some currently-due reviews) once they are in the last handful of topics of a course, that would lead to confusion, and in the big picture it would just be a micro-optimization that has negligible impact on total XP per course.

Key Papers

Note: *"Importance" blurbs may include pieces of direct quotes referenced earlier in this chapter. If citing this chapter, cite from the body (above).*

- Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review, 30*, 100322.

  Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. *PloS one, 12*(3), e0173403.

  Lei, H., Wang, C., Chiu, M. M., & Chen, S. (2022). Do educational games affect students' achievement emotions? Evidence from a meta-analysis. *Journal of Computer Assisted Learning, 38*(4), 946-959.

  Importance: *When game-like elements are properly integrated into student learning environments, students typically not only learn and engage more with the content, but also enjoy it more.*

- Iosup, A., & Epema, D. (2014, March). An experience report on using gamification in technical higher education. In *Proceedings of the 45th ACM technical symposium on Computer science education* (pp. 27-32).

  Importance: *The benefits of gamification apply not only to young students, but also to university-level students and even postgraduate students in technically-challenging courses. In addition to increasing learning and engagement, the authors note that gamification "seems to also foster interaction in the classroom and trigger students to pay more attention to the design of the course. We also observe very positive student assessments and volunteered testimonials, and a Teacher of the Year award."*

Chapter 22. Leveraging Cognitive Learning Strategies Requires Technology

Summary: *While there is plenty of room for teachers to make better use of cognitive learning strategies in the classroom, teachers are victims of circumstance in a profession lacking effective accountability and incentive structures, and the end result is that students continue to receive mediocre educational experiences. Given a sufficient degree of accountability and incentives, there is no law of physics preventing a teacher from putting forth the work needed to deliver an optimal learning experience to a single student. However, in the absence of technology, it is impossible for a single human teacher to deliver an optimal learning experience to a classroom of many students with heterogeneous knowledge profiles, each of whom needs to work on different types of problems and receive immediate feedback on each of their attempts. This is why technology is necessary. By automatically leveraging cognitive learning strategies to their fullest extent, Math Academy is able to deliver an optimized, adaptive, personalized learning experience to each individual student. Math Academy students are perpetually engaged in productive problem-solving, with immediate feedback (and remediation when necessary), on the specific types of problems (and in the specific types of settings) that will move the needle the most for their personal learning progress.*

High-Level Context

| The Problem: Cognitive Learning Strategies Remain Underused

It is common knowledge among researchers that the cognitive learning strategies discussed in previous chapters have the potential to drastically improve the depth, pace, and overall success of student learning. These strategies have been identified and researched extensively since the early to mid-1900s, with key findings being successfully reproduced over and over again since then. However, as discussed in chapter 2, the disappointing reality is that the practice of education has barely changed, and in many ways remains in direct opposition to these strategies.

| The Blame: Teachers are Victims of Circumstance

So, what happened? Why does the potential of these cognitive learning strategies remain unrealized, and who – or what – is to blame?

We do not wish to direct the blame at teachers. For instance, one cannot blame Sherman (1992) – who did everything in his power to leverage mastery-based learning within his own classroom and promote its widespread adoption – for having his efforts opposed and ultimately overpowered by various forces at play within the education system. Likewise, one cannot blame other teachers who have thought about ways to capitalize on these cognitive learning strategies to improve student learning, but, for one reason or another, found it too difficult to integrate them into their classroom in practice.

Teachers are victims of circumstance. The education system – like any other system whose intended function (promoting learning) is limited by the scarce resources (teachers and funding) available to achieve that function – has developed its own conventions while seeking the closest thing to a solution to an intractable problem. As the education system has evolved over hundreds of years, these conventions have accumulated and ossified into hard-baked constraints that outlive their usefulness. Many constraints are no longer helpful to the goal of promoting learning, yet remain deeply ingrained and act to resist change.

As summarized by Sherman (1992):

> *"...[T]he investment in keeping things as they are may be impossible to overcome. ... Improving instruction is the goal, but only in the context of not changing anything that is important to any vested interest. ... [When the role of the teacher] does not conform to what most people think of as teaching; this is a problem and an obstacle to implementation."*

| The Solution: Technology Changes Everything

In the past, scarcity of resources (teachers and funding) has made it impossible to fully leverage cognitive learning strategies in traditional classrooms. This scarcity persists today. However, a new variable has also entered the equation: technology.

Technology changes everything. Individualized digital learning environments are now technologically possible and commercially viable. Technology not only lets us circumvent the opposing inertia in the education system, but also helps us leverage cognitive learning strategies to a degree that would not be feasible for even the most agreeable and hard-working human teacher.

Resistance to Additional Effort

One force keeping cognitive learning strategies out of the classroom is that they require additional effort from teachers and students. Again, we do not say this with the intent to cast blame – it is simply a fact that as humans, we tend to resist additional effort, especially when we (like teachers) are already tired or (like young students) do not fully understand the long-term consequences of our decisions.

Teachers are already under a high level of baseline stress while facing strenuous – and often conflicting – demands from administrators, parents, and students. When it comes to promoting learning, teachers can keep all parties satisfied (or, perhaps, not too dissatisfied) by checking the boxes on long-standing conventions of the educational system: some lectures, some homework, several quizzes, and a couple tests. There's only so much you can deride a teacher for meeting the societal and institutional expectations that are placed on them, but not going above and beyond.

The same applies to students. Like a child who prefers to eat junk food and watch TV, but manages to complete their chores and eat the vegetables on their dinner plate, there's only so much you can deride a student for showing up to class, being undisruptive, and performing well enough on homework and tests to earn a passing grade, but not going above and beyond to maximize their learning and retention – especially when they are too young to fully grasp the long-term impact of their present habits on their future life.

Additionally, it is unreasonable to expect students to be highly motivated to maximize their learning in every subject when a reality of human nature is that most people are unmotivated to do most things. The tiny subset of things that a person is motivated to do in life are called their career and hobbies, and most people only have at most one career and a few hobbies. Everything else – i.e., the vast majority of things – are chores.

| Active Learning

Active learning requires teachers to spend more time and effort preparing and managing classroom activities. As we emphasized in chapter 10, true active learning requires every individual student to be actively engaged on every piece of material to be learned.

To implement true active learning in a math classroom, a teacher must continually supply problems, enforce that each student is attempting the problems, and check each student's solution to each problem, providing corrective feedback whenever a solution is incorrect. Enforcing that students are doing the problems can be particularly difficult and frustrating, since all but the most motivated students will typically avoid mentally taxing work when possible. (While it's true that more students may become motivated to put forth a high level of effort and maintain it in the absence of supervision if they enter a state of flow, there is typically an initial "activation energy" that must be overcome before reaching the flow experience, similar to how one might not look forward to working out but actually have a lot of fun and feel proud of their effort once they get going with it.)

Additionally, active learning requires the teacher to make lots of on-the-fly decisions, which can feel overwhelming to teachers who are more comfortable planning everything out beforehand. What the class does next should depend on whether students were able to do what the teacher originally asked them to do. These decisions can get especially tricky when the class becomes "split" with many students being able to do the original activity and being ready to move on to something more challenging, but many other students struggling and needing more practice (or even remedial support) with the original activity. No active learning lesson plan survives contact with a class full of students of varying abilities.

On the whole, it's way easier for a teacher to just talk and write on the board and "check the box" on active learning (without really leveraging it) by making sure that students appear to be listening, having some discussion with the smartest kids in the class, and maybe displaying a few problems and asking who wants to come up to the board to present a solution.

| Non-Interference, Interleaving, and Spaced Repetition

> Shuffling Instructional Material

As discussed in chapter 16, conceptually related pieces of knowledge can interfere with each other's recall, especially when taught simultaneously or in close succession. To minimize the impact of interference, new concepts should be taught alongside dissimilar material. However, it is easier for teachers to work in batch, creating week-by-week class lesson plans around groups of related material.

Additionally, the instructional material that's provided to teachers is typically structured around curricular units of related content. While it may make sense to structure a reference book this way (for ease of lookup), this organization does not reflect the optimal order to actually teach the material. As a result, a teacher wishing to leverage non-interference would have to invest additional time and effort into "shuffling" their instructional material while ensuring that every topic comes later than its prerequisites in the shuffled order.

Similar effort is required to leverage interleaving, which, as discussed in chapter 18, involves spreading out review problems over multiple review assignments that each cover a broad mix of previously-learned topics. Most textbooks are structured the opposite way – in blocks, where a single skill is practiced many times consecutively. As a result, a teacher typically cannot just grab an interleaved assignment "off the shelf" – rather, they will need to invest time and effort to manually allocate problems across interleaved assignments and keep track of how much practice they've given the class on each topic.

> Opening a Can of Worms on Forgetting

Interleaving can open a can of worms on who doesn't remember what: when students are doing a variety of different things and are not able to mindlessly apply one type of procedure to one type of problem, they may need reminders of how and when to apply various solution techniques, they may make a variety of different types of mistakes, and they may have scattered questions in class the next day about the previous day's homework. The same thing happens during spaced repetition, which, as discussed in chapter 17, involves spacing out reviews over time.

Opening this can of worms is actually a good thing because it provides an immense amount of information about what each student needs to work on – but it can feel overwhelming for teachers to have so many student needs at one time, especially when the teacher is under pressure for the class to cover a set amount of content by a fixed deadline, and the teacher feels like remediating student forgetting is "slowing down" their progress towards that goal.

Of course, good teachers understand the importance of continual review and periodically revisit previously-learned material to help their students retain it. However, as discussed in chapter 17, optimizing retention through true spaced repetition requires a massive, inhuman amount of bookkeeping and computation. Carrying out even a loose approximation of spaced repetition for the class as a whole requires an immense level of effort. Given the additional stress that continual review creates for teachers, it's easier to just stick to the status quo and cram a class or two of review before each major test.

| Testing Effect

Good, highly-engaged teachers understand the importance of quizzes and give regular weekly or biweekly quizzes (which is reasonably frequent, though a higher frequency would be ideal). However, unless these quizzes are built into some existing curriculum that they are working from, it takes a lot of work to create those quizzes, grade them, and go over mistakes with students. And that's not all:

- Ideally, students who don't do so well will be given the opportunity to demonstrate learning from their mistakes on a retake quiz with different (but similar) questions – which effectively doubles the teacher's workload relating to quizzes.

- In a class of more than a handful of students, there is always a good chance that one or more students will be absent due to sickness, medical appointments, or other things, in which case the teacher has to schedule make-up quizzes.

- Especially at the high school and university levels, a minority of students may cause further headache by routinely complaining that questions they missed were unfair (and should be thrown out) or begging for undeserved partial credit.

Considering that most full-time teachers teach about 5-6 different classes each day, it's an infeasible amount of work to quiz students every few days. Giving regular weekly or biweekly quizzes in all of one's classes is hard enough. Realistically, given the additional stress that teachers experience when they give additional quizzes, it's easier to just stick to the status quo and quiz students at the bare minimum frequency required to meet one's professional expectations.

| Gamification

Managing a gamified metric like eXperience Points (XP), and other gamified features like leaderboards, takes an immense amount of bookkeeping. It can be done, but it takes a really engaged teacher, and even then, it's typically too much work for a teacher to integrate every single learning task into the gamification structure. Gamification is typically not a part of a teacher's professional expectations, so it's easier for teachers to just forego it.

Tutoring the King's Kid: How Would You Teach if Your Life Depended On It?

The issues described above are not impossible to overcome manually. Each issue is solvable, but the solution requires a lot of work from the teacher. There is no law of physics preventing the teacher from putting forth that work, but the degree of accountability and incentives in place is not sufficient to motivate the teacher to do so.

(We again emphasize that this is not the fault of teachers, who are victims of circumstance in a profession lacking effective accountability and incentive structures. Who wants to work harder than necessary if they know they're not going to be rewarded for it, and there is no punishment for mediocre work? Nobody.)

To intuitively understand the importance of accountability and incentives, it may help to imagine yourself as an educator in a life-or-death situation, where the outcome of the situation depends on whether you can teach a student effectively enough that they are able to unequivocally demonstrate their learning to a third party. Below is a retelling of Jason Roberts's *Tutoring the King's Kid* anecdote:

> *Suppose that you are an educator back in medieval times, and you work within the kingdom of the wealthiest, but also the fiercest, king in all the world. The king's child has participated in a school within the kingdom, but the king has been unhappy with the results: the child has gone to school for over a year, and has learned how to count, but remains unable to solve any problem requiring simple application of arithmetic.*
>
> *One day, the king sends for you to appear immediately at his throne. When you show up, he commands you to teach simple arithmetic to his child as your sole duty for the next month. The child shall spend the entirety of each school day with you, and in exactly one month, the king shall ask his child five questions, each one requiring the addition, subtraction, multiplication, or division of two numbers, each number being one or two digits long. The child will have two minutes to complete each question, and their performance on this test will determine your fate.*

> *Being the wealthiest king in all the world, he has decided that if the child answers at least four of the five questions correctly, then he will grant you a fortune so extravagant that you can live out the rest of your life to the same level of luxury as a lesser king. However, if the child answers three or fewer questions correctly, then you shall be executed the following day.*

In this situation, you're motivated to put in the work to overcome each of the issues described earlier. The instructional experience becomes entirely student-centered, leveraging cognitive learning strategies as much as humanly possible.

> *Suddenly, you realize that you don't care at all about how much time, effort, and stress you have to endure to make this child learn. Your own feelings are completely out of the picture. All that matters is whether the child comes to know their arithmetic facts by heart, understand the meaning of the operations deeply enough to know when to apply each one in problem-solving contexts, and quickly and reliably calculate the result of any arithmetic operation with numbers up to two digits long.*
>
> *To accomplish this, every moment you have with the child will be devoted to getting the child to the point where they are able to do all of these things independently.*
>
> *• You will of course introduce each skill along with a quick demonstration, but you won't ramble about anything that's irrelevant because your goal will be to have them start attempting to solve problems on each skill as soon as possible.*
>
> *• You will provide corrective feedback on every single problem that they solve, talking them through the correct solution whenever they make a mistake. If they do well, you will quickly move them onwards to more difficult problems, but if they struggle, you will give them however much practice they need to master the skill before moving forwards.*
>
> *• You will cover a mix of different topics every day and continually feed them review problems on previously-learned skills (but not too much review – just a "minimal effective dose" to restore their memory on any topics that they might be in danger of forgetting).*
>
> *• You will also provide frequent timed quizzes on a mixture of different problem types, go over their quizzes with them, give them more practice on anything they missed on the quizzes, and give them a retake to make sure they learned from their mistakes.*
>
> *• Lastly, you will gamify the experience in a way that incentivizes the child to put their best effort forward all the time.*

Heterogeneity of Student Knowledge Profiles

| Tutoring the King's Kid vs Teaching Many Kings' Kids

The "tutoring the king's kid" anecdote illustrates that there is no law of physics preventing a teacher from putting forth the work needed to deliver an optimal learning experience to a single student: rather, it is a matter of accountability and incentives.

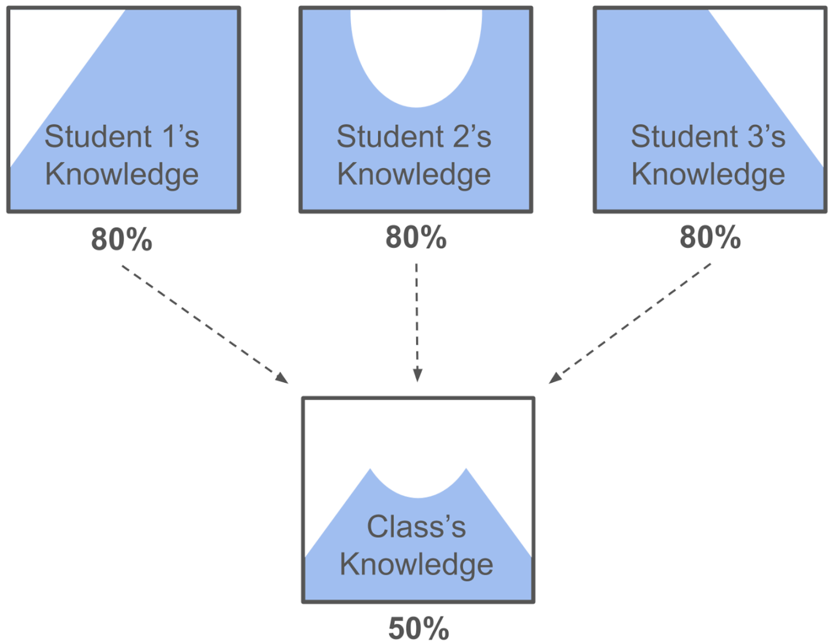
However, a key assumption in the anecdote is that the teacher is working with a single student. If the same story were told with a class full of 30 children, each from a king who will execute the teacher if their own child fails the test, then it would emphasize a different perspective: the teacher may be doomed because no matter how hard the they try, they will not be able to deliver that same optimal learning experience to every student in the class. Regardless of the level of accountability and incentives, the amount of work required would be inhuman.

Loosely speaking, it boils down to the physics of learning. The reason why it's so much harder to teach 30 students than to teach a single student is that the 30 students all have unique, heterogeneous knowledge profiles.

| Differences in Background Knowledge

Students who earned different grades in a prerequisite math course typically come into the next course with vastly different knowledge profiles. For instance, students who received a C in the prerequisite course typically have far more foundational knowledge gaps than students who received an A (though even students who received an A usually have *some* foundational knowledge gaps, even if they tend to be fewer and/or less severe).

Moreover, and more subtly, even students who earned the exact same grade in a prerequisite math course typically have vastly different knowledge profiles from each other. Any two students who mastered the same amount of material in the prerequisite course may *completely differ* in the material that they were unable to master. One student may have struggled with fractions, while another may have struggled with decimals. One student may have struggled with solving equations, while another may have struggled with graphing functions.



| Student Knowledge Profiles Naturally Tend Towards Heterogeneity

Even in an unrealistic hypothetical scenario where all the students in a class were academic "clones" of one another with exactly the same knowledge profiles, learning speeds, and levels of motivation, their knowledge profiles would naturally diverge over time as the class went on. Despite having the same academic profile, each student would be missing class or spacing out at different times, and as a result, some students would struggle with some topics more than others. (Missing class and spacing out are effectively the same thing, just on different time scales: they differ only in frequency and duration.)

Everyone spaces out sometimes – even adults. It happens constantly, even to people who are consciously trying to pay attention. People have a hard time focusing when they have other things on their minds: what they're going to eat for lunch, their plans for the weekend, anxiety about a personal relationship, etc. The author of this book spaced out at least twice while writing the four paragraphs in this subsection.

This is especially true for students, who also face an endless list of mini-distractions in a classroom. For instance, a student might need to spend 30 seconds ruffling through their backpack for another pencil/pen or piece of paper (or their friend might ask them for one of those things). Or, a student may need to miss several minutes of class to use the bathroom.

Regardless of whether it is their fault or not, students are momentarily distracted at different times and they miss things. These differences compound over time unless the teacher immediately detects and fully remediates them at the instant that they arise – but this requires an inhuman amount of work, so teachers aren't doing it unless they have technology that does it.

| Every Student in the Class Effectively Needs a Private Tutor

The heterogeneity of student knowledge profiles means that different students need different amounts of practice, on different skills, at different times. Consequently, to deliver an optimal learning experience to all students in the class, the teacher must effectively function as a private tutor for every individual student. Needless to say, no matter how a teacher attempts this, it's an intractable problem if their class consists of more than a few students. Even if a teacher tries their hardest, they will not be able to deliver an optimal learning experience to every student in the class.

To fully leverage the cognitive learning strategies discussed in this book, and deliver an optimal learning experience to every student in the class, every individual student needs to be fully engaged in productive problem-solving, with immediate feedback (including remedial support when necessary), on the specific types of problems, and in the specific types of settings (e.g., with vs without reference material, blocked vs interleaved, timed vs untimed), that will move the needle the most for their personal learning progress at that specific moment in time. This needs to be happening throughout the entirety of class time, the only exceptions being those brief moments when a student is introduced to a new topic and observes a worked example before jumping into active problem-solving.

However, when students have heterogeneous knowledge profiles, it's at best extremely difficult, and at worst (and most commonly) impossible, to find a type of problem that is productive for all students in the class. Even if a teacher chooses a type of problem that is appropriate for what they perceive to be the "class average" knowledge profile, it will typically be too hard for many students and too easy for many others (an unproductive use of time for those students either way).

To even know the specific problem types that each student needs to work on, the teacher has to separately track each student's progress on each problem type, manage a spaced repetition schedule of when each student needs to review each topic, and continually update each schedule based on the student's performance (which can be incredibly complicated given that each time a student learns or reviews an advanced topic, they're implicitly reviewing many simpler topics, all of whose repetition schedules need to be adjusted as a result, depending on how the student performed). This is an inhuman amount of bookkeeping and computation.

Furthermore, even on the rare occasion that a teacher manages to find a type of problem that is productive for all students in the class, different students will require different amounts of practice to master the solution technique. Some students will catch on quickly and be ready to move on to more difficult problems after solving just a couple problems of the given type, while other students will require many more attempts before they are able to solve problems of the given type successfully on their own. Additionally, some students will solve problems quickly while others will require more time.

In the absence of technology, it is impossible for a single human teacher to deliver an optimal learning experience to a classroom of many students with heterogeneous knowledge profiles, who all need to work on different types of problems and receive immediate feedback on each attempt. However, technology changes everything. By automatically leveraging cognitive learning strategies to their fullest extent, Math Academy is able to deliver an optimized, adaptive, personalized learning experience to each individual student. Math Academy students are perpetually engaged in productive problem-solving, with immediate feedback (and remediation when necessary), on the specific types of problems (and in the specific types of settings) that will move the needle the most for their personal learning progress.

IV. COACHING

[In Progress] Chapter 23. Parental Support

Summary: *(in progress)*

Why Parental Support is (Usually) Necessary

As discussed in Chapter 11, developing expertise requires accumulating large amounts of deliberate practice – but deliberate practice tends to be more effortful and less enjoyable, so people tend to avoid it.

A practical consequence of this is that, for a child to develop a high level of talent in mathematics or any other skill domain, they will typically require a high level of parental support.

In particular, while some children take to Math Academy naturally, most children require parental support in maintaining proper usage habits.

What, then, must parents do to support and motivate their children to engage in deliberate practice?

The Bare Minimum: Incentives and Accountability

To develop their talents, most children need a responsible adult – such as a parent or teacher – to incentivize them and hold them accountable for their behavior.

For instance, here is what often happens when a parent signs their child up for Math Academy but does not set up an incentive and accountability structure with their child:

- Child puts forth little to no effort: they skim instead of reading carefully, rush and guess on practice questions instead of referring back to the worked example, and ultimately fail their learning tasks. Even if they appear to work for the expected amount of time, they are actually spending most of that time distracted by other activities (browsing other websites, socializing with friends, doodling or daydreaming, etc.).

- Adult checks in, realizes the student is not accomplishing anything, and asks the student what's going on.

- Child says that the system is too hard or otherwise doesn't work.

- Adult might take the student's word at face value. Or, if the adult notices that the child hasn't actually attempted any work and calls them out on it, the scenario repeats with the child putting forth as little effort as possible – enough to convince the adult that they're trying, but not enough to really make progress.

In these situations, here is what needs to happen:

- The adult needs to sit down next to the student and force them to actually put forth the effort required to use the system properly.

- Once it's established that the student is able to make progress by putting forth sufficient effort, the adult needs to continue holding the student accountable for their daily progress. If the student ever stops making progress, the adult needs to sit down next to the student again and get them back on the rails.

- To keep the student on the rails without having to sit down next to them all the time, the adult needs to set up an incentive structure. Even little things go a long way, like *"if you complete all your work this week then we'll go get ice cream on the weekend,"* or *"no video games tonight until you complete your work."* The incentive has to be centered around something that the student actually cares about, whether that be dessert, gaming, movies, books, etc.

Even if an adult puts a child on a system that leverages every single cognitive learning strategy to its fullest extent, if the adult clocks out and stops holding the student accountable for completing their work every day, then the overall learning outcome is going to be poor.

The Platonic Ideal

To paint a picture of the Platonic ideal of parental support, consider Bloom's description of the level of parental support that was required to produce the world-class concert pianists that were studied in *Developing Talent in Young People* (1985, pp. 453-58):

> *"Most of the pianists' parents monitored the amount of daily practice in the home. They listened or watched to ensure the quality of time spent. The children were not allowed to "play around," skip drills, or quit before the designated time. Practice had priority and was to be done every day, despite the inconvenience of schedules.*
> *...*
> *In addition to monitoring the amount of practice time, the parents did whatever they could to make the practice productive and enjoyable. ... The parents also applauded and encouraged the child's efforts and tried to convey to the child their interest and involvement.*
> *...*
> *'I would always sit down with him [to practice]. ... And I think that helped, especially when they're young. Because it's pretty hard to just sit down and practice without someone there beside you (M of P-4).'*
> *...*
> *[The parents] knew what the instructional goals were from their involvement in daily practice. They were learning more and more about the field – the rates of progress that were reasonable to expect and what the child's next goals would be.*
> *...*
> *The child's efforts in the field became a central part of the family's life. Discussions at the dinner table often focused on practice, the child's progress, future competitions, or the performance of other talents in the field. ... Close bonds were also developed with other families who had similar interests.*
> *...*
> *In addition to providing an opportunity for the family to pursue activities together, the talent field also became a means of translating the value of achievement into specific behaviors. The importance of goals and self-discipline were evident in the rules and expectations surrounding lessons and practice. The parents saw to it that the child worked consistently toward the goals set by the teachers or coaches.*
>
> *Progress was monitored by the parents at practice and at public performances. In some families, goals and progress were recorded on charts or in notebooks. When progress faltered, the parents discussed possible causes with the child and/or the teacher and sought solutions to the problem immediately.*
>
> *Doing one's best was stressed continuously, with respect not only to public performances but also to daily practice. "Slacking off" during practice or repeating mistakes were cause for reprimand. As might be expected, the parents had different methods of handling this situation. Some appealed to the child's professed love of the field or reprimanded the child of goals and accomplishments that lay ahead. Others emphasized the time, energy, and resources already committed. Still others threatened to discontinue their support and provision of resources if the child was not dedicated to working hard.*
>
> *Along with self-discipline and doing one's best were rewards and praise for a job well done. Ribbons and trophies decorated the family room; scrapbooks were filled with newspaper clippings. The joys and pride in winning were stressed, as was the satisfaction of doing your best even if you*

> *weren't first – this time. The parents were there with applause and verbal praise when goals were attained, with solace and encouragement when goals were not quite reached."*

Similar descriptions are echoed throughout the literature. For instance, according to Sloboda (1996), who studied factors underlying variability in musical achievement:

> *"Parents of Group 1 students [the higher-achieving group] were more likely to attend instrumental lessons with their children, obtain detailed feedback and instructions from teachers, and actively supervise daily practice on a moment-to-moment basis, often at some considerable cost to their own schedule.*
> *...*
> *Parents of low-achieving children were less likely to have meaningful contact with the teacher, and were likely to confine their domestic interventions to telling children to "go and do your practice," without any direct involvement in it.*
>
> *In sum, therefore, it seems that abnormally high levels of early practice are sustained by abnormal levels of social and cognitive support, mainly from parents."*

[In Progress] Chapter 24. In-Task Coaching

Summary: *(in progress)*

The Vicious Cycle of Forgetting

In an attempt to mitigate the effects of forgetting, students sometimes solve problems alongside reference material. However, this actually has the opposite effect: when a student continually looks back at a reference, the information doesn't stay in their brain. They hold the information in short-term memory, but only temporarily – it dissipates after use.

This leads to a vicious cycle of forgetting. The reference material becomes a crutch, and the student is lost without it. They miss out on making connections and understanding things deeply. They learn slower, and forget faster, until eventually they grind to a halt. They might think it's because they're not doing enough review problems, when it's really because they're not doing those review problems properly (by pulling information from memory).

Even people who are serious about their learning sometimes fall into this vicious cycle of forgetting. They might take great notes and then refer back to those notes all the time instead of trying to pull the information from memory.

To break – or better, avoid – the vicious cycle of forgetting, students need to practice retrieving information without assistance so that it actually transfers into long-term memory. Every time a student successfully recalls a fuzzy memory, it stays intact longer before getting fuzzy again.

If a student can't recall something after trying, it's okay to check reference material, but only as a last resort. They should peek once, then solve the problem without looking again. Challenging oneself to remember may be tough initially, but it pays off in the long run.

V. TECHNICAL DEEP DIVES

Chapter 25. Technical Deep Dive on Spaced Repetition

Note: *this chapter elaborates on concepts introduced in chapter 17.*

Summary: *Math Academy employs Fractional Implicit Repetition (FIRe), a novel spaced repetition algorithm, to calculate student learning profiles. FIRe generalizes spaced repetition to hierarchical knowledge, allowing repetitions on advanced topics to implicitly trickle down to simpler topics. The algorithm handles partial encompassings and extends repetition flows through fractional encompassings, optimizing credit distribution. The speed of the spaced repetition process is calibrated to each individual student on each individual topic, where student ability and topic difficulty are competing factors.*

Fractional Implicit Repetition (FIRe)

To calculate student spaced repetition profiles, Math Academy uses a novel spaced repetition algorithm called **Fractional Implicit Repetition** (**FIRe**). FIRe generalizes spaced repetition to hierarchical bodies of knowledge where

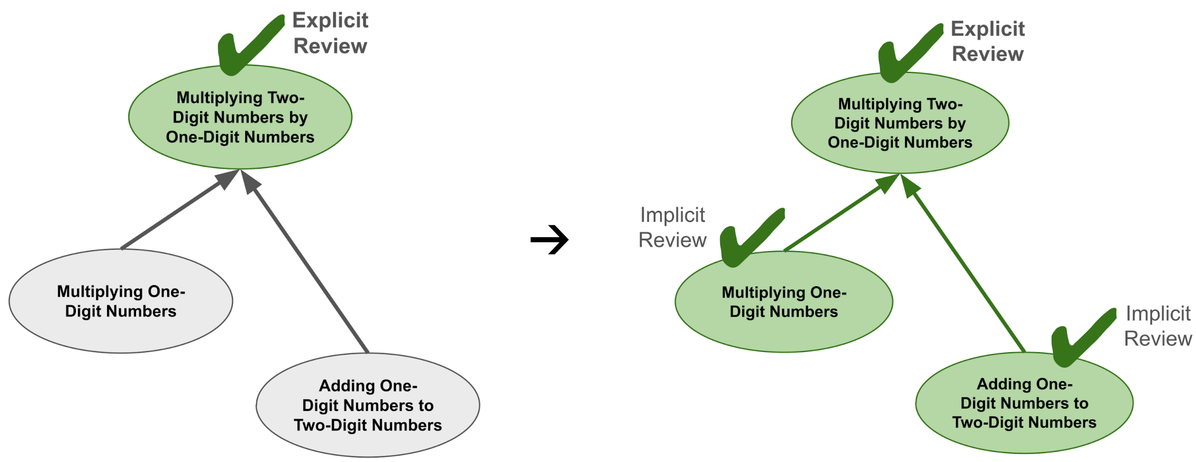
1. repetitions on advanced topics "trickle down" implicitly to simpler topics through encompassing relationships, and

2. simpler topics receiving lots of implicit repetitions discount the repetitions appropriately (since they are often too early to count for full credit towards the next repetition).

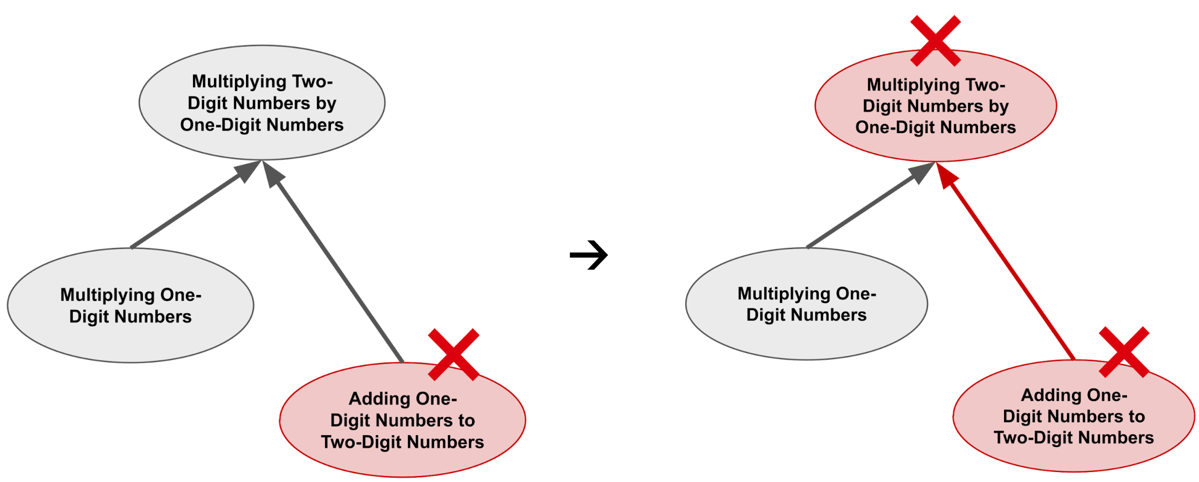
| Concrete Example

As a concrete example, recall that *Multiplying a Two-Digit Number by a One-Digit Number* encompasses *Multiplying One-Digit Numbers* and *Adding a One-Digit Number to a Two-Digit Number*.

If you pass a review on *Multiplying a Two-Digit Number by a One-Digit Number*, then the repetition will also flow backward to reward *Multiplying One-Digit Numbers* and *Adding a One-Digit Number to a Two-Digit Number* because you've just shown evidence that you still know how to perform these skills.



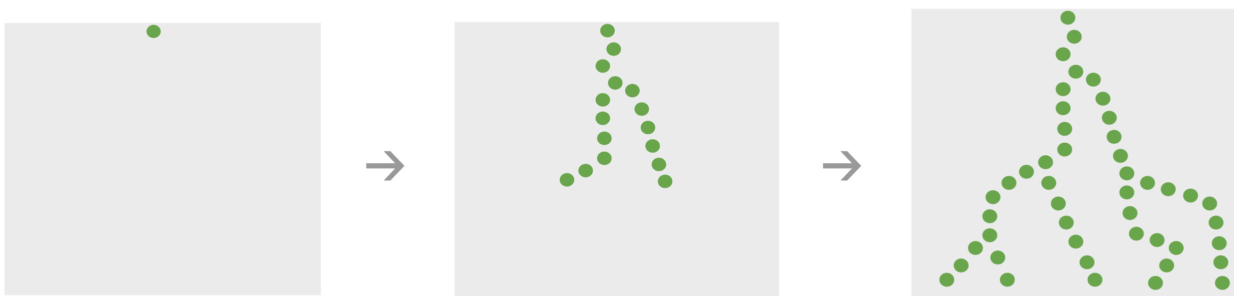
On the other hand, if you fail a repetition on *Adding a One-Digit Number to a Two-Digit Number*, then the failed repetition will also flow forward to penalize *Multiplying a Two-Digit Number by a One-Digit Number*. If you can't add a one-digit number to a two-digit number, then there's no way you're able to multiply a two-digit number by a one-digit number. The same thing happens if you fail a repetition on *Multiplying One-Digit Numbers*.



Visualizing Repetition Flow

Note that repetition flows can extend many layers deep – not just to directly encompassed topics, but also to "second-order" topics that are encompassed *by the encompassed topics*, and then to third-order topics that are encompassed by second-order topics, and so on.

Visually, credit travels downwards through the knowledge graph like lightning bolts.



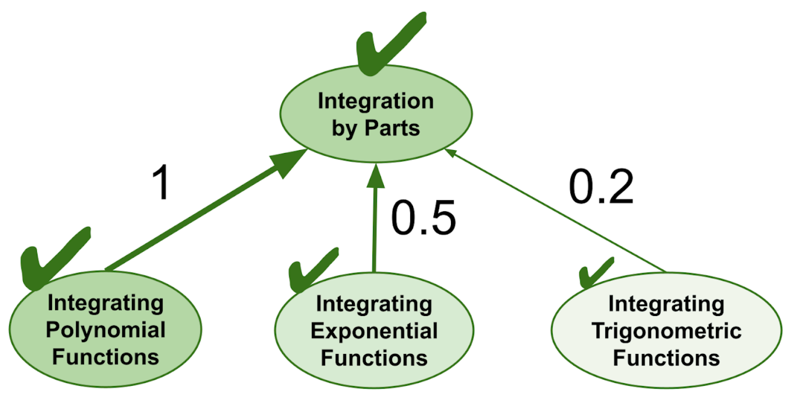
Penalties travel upwards through the knowledge graph like growing trees.

Partial Encompassings

FIRe also naturally handles cases of **partial encompassings**, in which only some part of a simpler topic is practiced implicitly in an advanced topic. This occurs more frequently in higher-level math.

For instance, in calculus, advanced integration techniques like integration by parts require you to calculate integrals of a variety of mathematical functions such as polynomials, exponentials, and trigonometric functions. But some of those functions might only appear in a portion of the integration by parts problems. So, if you complete a repetition on integration by parts, you should only receive a fraction of a repetition towards each partially-encompassed topic.

In the diagram below, we label encompassings with numerical **weights** that represent what fraction of each simpler topic is practiced during the more advanced topic. You can loosely interpret each weight as representing the probability that a random problem from the advanced topic encompasses a random problem from the simpler topic.



FIRe extends repetition flows many layers deep through fractional encompassings as well. The end result is that repetitions

1. travel unhindered along a "trunk" of full encompassings, and
2. fade off along partial encompassings branching outwards from the trunk.

Setting Encompassing Weights Manually

| Direct and Key Prerequisites are Sufficient

Because encompassing weights are set manually, it is not feasible to set an explicit weight between every pair of topics in the graph. We have thousands of topics, so the full pairwise weight matrix would contain tens of millions of entries. How do we set all those weights?

It turns out that it is not actually necessary to explicitly set every weight in the matrix. It suffices to set only the weights for topic pairs where

1. the weight has a nontrivial value,
2. the weight cannot otherwise be inferred using repetition flow, and
3. the distance between the topics in the prerequisite graph is low,

and assume that all other weights not computed implicitly during repetition flow are zero. The reasoning behind these conditions is as follows:

1. The magnitude of the weight represents the magnitude of the implicit repetition credit. In order for an implicit repetition to make an impact on staving off explicit reviews, it has to be associated with a nontrivial amount of credit.
2. If repetition flow can infer a weight, then nothing will change if the weight is set manually (unless the manually-set weight is being used to correct a value that would otherwise be inferred by repetition flow).
3. If two topics are far apart in the prerequisite graph, then their weight will not make much of an impact on staving off reviews, even if it is a full encompassing. In that case, by the time the student reaches the more advanced topic, they will already have done most of their explicit reviews on the simpler topic.

Conveniently, the weights that satisfy the above conditions tend to be those along direct and key prerequisite edges, the number of which scales linearly with the number of topics. This makes it feasible to set encompassing weights manually: one weight for each direct or key prerequisite

Note that it is not unusual to find direct and key prerequisite edges with a weight as low as zero. This can happen when a topic requires some amount of conceptual familiarity with the prerequisite, but does not require the student to actually have mastered the prerequisite to the point of being able to solve problems in the prerequisite topic.

| Non-Ancestor Encompassings and Mastery Floors

To comply with course standards, it is sometimes necessary to have **equivalent topics** spread out across multiple courses, with the equivalent topics in higher courses covering a more advanced treatment of the same skills taught in lower courses. However, the simpler equivalent topics are usually not required as prerequisites for the more advanced equivalent topics.

For instance, courses on algebra-based statistics and calculus-based statistics would have many equivalent topics that cover the same skills. Although the calculus-based statistics course would provide more advanced treatments of these skills, the corresponding equivalent topics in the algebra-based statistics course would not be prerequisites.

Even though simple equivalent topics would not be ancestors of advanced equivalent topics via direct or key prerequisite paths, we can still set full-encompassing edge weights between them so that a student who completes an advanced topic will implicitly receive credit for any simpler equivalent topics as well. These are called **non-ancestor encompassings**.

Non-ancestor encompassings, along with course-based **mastery floors** (lower-course topics that are automatically considered mastered by any student taking the course), can also be useful for assigning credit to leaf topics in lower-level courses. The mastery floor of a course consists of the lower-course topics that

- are "far back" enough that it is safe to automatically consider them mastered, or
- lie "below" the simplest topics that could be assessed on the course's diagnostic.

Intuitively, the top of the mastery floor marks the dividing line regarding whether it is at all feasible for a student to take a course.

Student-Topic Learning Speeds

| Ratio of Student Ability and Topic Difficulty

Student ability and topic difficulty are competing factors – high student ability speeds up the overall student-topic learning speed, while high topic difficulty slows it down. So, to compute a student-topic learning speed, we compute

1. the speedup due to student ability,
2. the slowdown due to topic difficulty, and then
3. their ratio.

$$\text{student-topic learning speed} = \frac{\text{speedup due to student ability}}{\text{slowdown due to topic difficulty}}$$

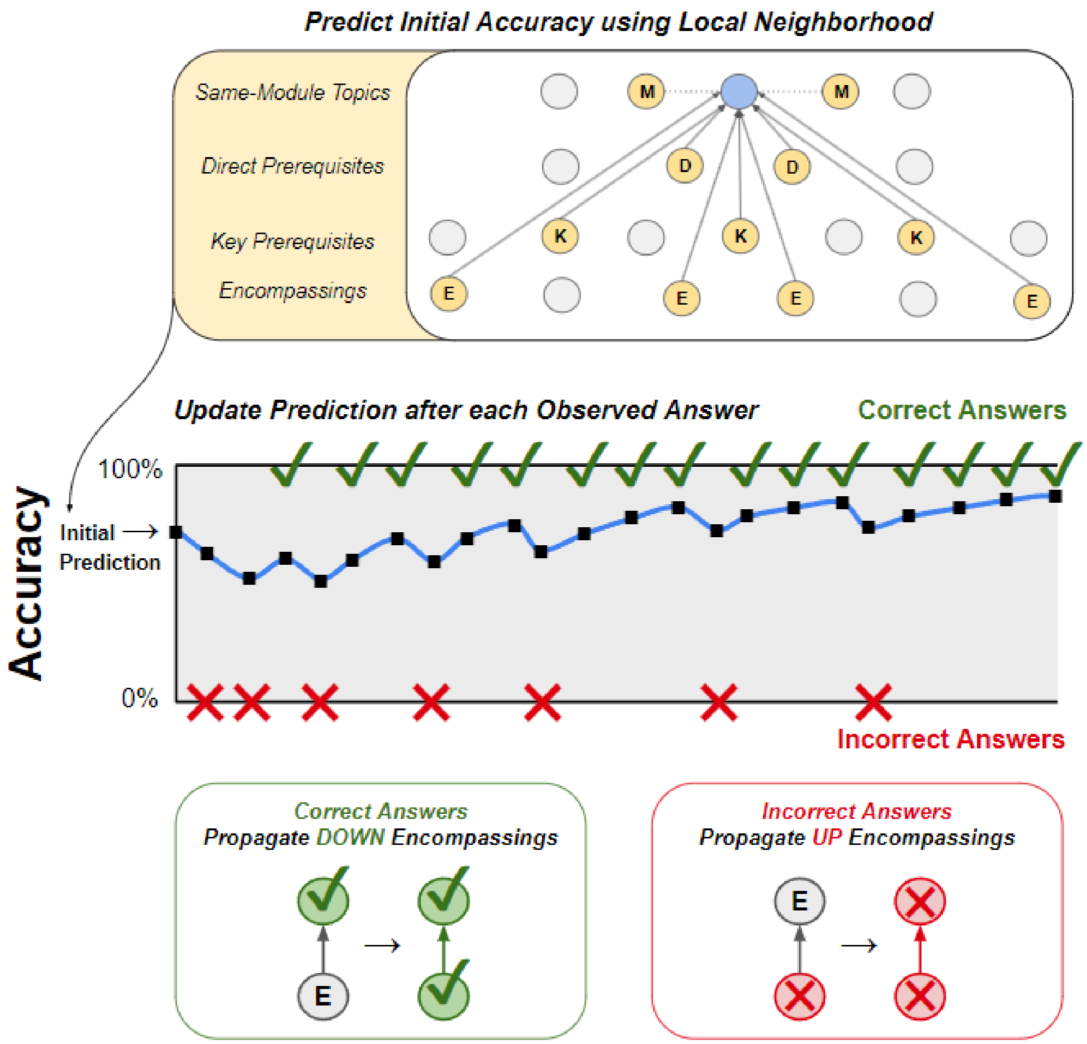
Student-Topic Learning Speed vs Student Ability and Topic Difficulty		*Student Ability*		
		Strong	*Moderate*	*Weak*
Topic Difficulty	*Easy*	fastest	faster	baseline
	Moderate	faster	baseline	slower
	Hard	baseline	slower	slowest

| Measuring Student Ability at the Level of Individual Topics

Student ability is measured at the granular level of individual topics – we keep track of accuracy across answers, giving more weight to recent answers, and also propagating

- correct answers down to simpler encompassed topics and
- incorrect answers up to more advanced topics that encompass the answered topic.

To choose the initial starting value for a topic's accuracy, we make a prediction based on the accuracy values of the topic's **local neighborhood** consisting of its direct prerequisites, key prerequisites, encompassings, and same-module topics.



| Measuring Topic Difficulty

Topic difficulty is measured by computing the topic's accuracy across all instances when one of its questions was answered by a serious student on an assessment.

In theory, if student abilities could be measured on each topic with perfect fidelity, then topic difficulties would no longer be needed and student-topic learning speeds could be based entirely on student abilities. But in practice, there are two reasons why it is helpful to rely on topic difficulty as well:

1. *It improves the initial prediction.* Although we already have information about the particular student's learning speed on other topics, the topic difficulty provides information about the particular topic's learning speed for other students. This is an independent, information-rich signal.

2. *It naturally acts as a correction factor.* When topic difficulty is high, it decreases the learning speed – which is desirable given that high topic difficulty is caused by low assessment performance, which is in turn (largely) caused by students not getting enough review. Similarly, when topic difficulty is low, it increases the learning speed – which is desirable given that low topic difficulty is caused by extremely high assessment performance, which indicates that students might not need as much review as they are receiving.

High-Level Structure

At a high level, the structure of Math Academy's spaced repetition model can be summarized as follows:

$$\begin{aligned} repNum &\to \max\left(0,\ repNum + speed \cdot decay^{failed} \cdot netWork\right) \\ memory &\to \max\left(0,\ memory + netWork\right)(0.5)^{days\,/\,interval} \end{aligned}$$

- *repNum* = how many successful rounds of spaced repetition the student has accumulated on a particular topic.

  In following definitions, a "repetition" is a successful review at the appropriate time.

- *days* = how many days it's been since the previous repetition.

- *interval* = the ideal number of days to be spaced between repetition *repNum* and repetition *repNum+1*.

- *memory* = how well the student is expected to remember now that it's been some time since the previous repetition. Memory decays over time and the next repetition becomes due when the memory becomes sufficiently low.

- *speed* = the learning speed for the student on this particular topic, based on how well the student is performing. Governs how quickly the student moves forwards or backwards through the spaced repetition process.

- *failed* = 1 if repetition was failed and 0 if it was passed.

- *netWork* = how much net work the student completed during the repetition. *netWork* is positive if the repetition was passed and negative if failed.

  The higher the quality of work in a passed repetition, or the worse the quality of work in a failed repetition, the larger the magnitude of *netWork*.

  The magnitude of *netWork* is also discounted if the repetition was completed early relative to the desired *interval*, i.e., if *memory* is sufficiently high.

  Note that successful work on an advanced topic is also counted towards any simpler topics that are implicitly practiced as component skills, and unsuccessful work on a simpler topic is also counted towards any more advanced topics of which that simpler topic is a component skill.

- *decay* = the speed at which the student moves backwards in the spaced repetition process, relative to their forwards speed, if they fail a repetition.

  decay is a positive quantity that starts at 1 and grows larger as the repetition becomes more overdue relative to its ideal *interval*, i.e., as *memory* becomes severely low.

  decay was introduced to model severe knowledge decay like the notorious "summer slide," where topics learned shortly before the end of the previous school year may be forgotten so severely over summer vacation that they need to be reviewed more frequently or even completely re-taught at the start of the following year.

High-Level Structure of Math Academy's Spaced Repetition Model

`repNum` = how many successful rounds of spaced repetition the student has accumulated on a particular topic.

A "repetition" is a successful review at the appropriate time.

`speed` = the learning speed for the student on this particular topic, based on how well the student is performing.

Governs how quickly the student moves forwards or backwards through the spaced repetition process.

`failed` = 1 if repetition was failed and 0 if it was passed.

`decay` = the speed at which the student moves backwards in the spaced repetition process, relative to their forwards speed, if they fail a repetition.

`decay` is a positive quantity that starts at 1 and grows larger as the repetition becomes more overdue relative to its ideal `interval`, i.e., as `memory` becomes severely low.

`decay` was introduced to model severe knowledge decay like the notorious "summer slide," where topics learned shortly before the end of the previous school year may be forgotten so severely over summer vacation that they need to be reviewed more frequently or even completely re-taught at the start of the following year.

$$\boxed{repNum} \to \max\left(0, \boxed{repNum} + \boxed{speed} \cdot \boxed{decay^{failed}} \cdot \boxed{netWork}\right)$$

$$\boxed{memory} \to \max\left(0, \boxed{memory} + \boxed{netWork}\right)(0.5)^{\boxed{days\,/\,interval}}$$

`memory` = how well the student is expected to remember now that it's been some time since the previous repetition.

Memory decays over time and the next repetition becomes due when the memory becomes sufficiently low.

`netWork` = how much net work the student completed during the repetition. `netWork` is positive if the repetition was passed and negative if failed.

The higher the quality of work in a passed repetition, or the worse the quality of work in a failed repetition, the larger the magnitude of `netWork`.

The magnitude of `netWork` is also discounted if the repetition was completed early relative to the desired `interval`, i.e., if `memory` is sufficiently high.

Note that successful work on an advanced topic is also counted towards any simpler topics that are implicitly practiced as component skills, and unsuccessful work on a simpler topic is also counted towards any more advanced topics of which that simpler topic is a component skill.

`days` = how many days it's been since the previous repetition.

`interval` = the ideal number of days to be spaced between repetition `repNum` and repetition `repNum+1`.

Chapter 26. Technical Deep Dive on Diagnostic Exams

Note: *this chapter elaborates on concepts introduced in chapter 4 and chapter 20.*

Summary: *Math Academy uses adaptive diagnostics to infer each incoming student's knowledge profile. The novel diagnostic assessment algorithm leverages causal relationships and correlation-based inference in the knowledge graph to efficiently gauge a student's knowledge frontier to a sufficient level of detail while minimizing the number of questions on the diagnostic. It measures knowledge confidence and handles conflicting information, adapting its diagnosis to nuanced scenarios like prerequisite-postrequisite and accuracy-time conflicts. Conditional completion is employed for areas where knowledge was inferred with low confidence, allowing the system to continue fine-tuning a student's placement as it collects more data about their performance.*

Minimizing the Number of Questions

Without any clever algorithms, it would take a massive number of diagnostic questions to infer a student's knowledge frontier. Courses often contain up to several hundred topics, plus twice as many foundational topics – which means that if we started at the bottom and asked you a diagnostic question for every topic up until the point that you could no longer answer them correctly, we'd end up asking you 500+ questions in total.

However, Math Academy is able to cut down this number of diagnostic questions by an order of magnitude using a novel diagnostic question selection algorithm. Our diagnostics generally take only 20-40 questions for lower-grade courses (like Prealgebra) and 40-60 questions for higher-grade courses (like Calculus).

We're able to achieve this level of diagnostic efficiency for two reasons:

1. In addition to leveraging "causal" relationships, i.e. encompassings, we also leverage looser forms of correlation-based inference.

2. We compress the knowledge graph beforehand into the smallest number of topics that fully "covers" a course and its foundations at a desired level of granularity.

As a result, our diagnostic assessment algorithm is highly leveraged on both accuracy and precision. While a perfect-accuracy, perfect-precision diagnostic could require up to a thousand questions, we are able to reduce this by an order of magnitude by giving up a negligible amount of accuracy and precision.

(Of course, highly leveraged algorithms have a higher risk of being thrown off by false input – but we mitigate this risk by detecting and re-assessing questions that we suspect may have been incorrectly assessed.)

Knowledge Confidence and Conditional Completion

| Theory

During diagnostics, Math Academy also measures **knowledge confidence**, i.e., our confidence in our classification of whether the student knows or does not know a topic. Most diagnostics complete with fairly high confidence across the student's entire knowledge profile, but occasionally, there can be areas of low confidence. These do not arise from a lack of diagnostic coverage, but rather, from conflicting evidence in a student's responses.

There are two main types of conflicts:

1. *Prerequisite-Postrequisite Conflict:* a student answers a more advanced topic correctly but a simpler question incorrectly, which may indicate a gap in the student's knowledge.

2. *Accuracy-Time Conflict:* a student submits a correct answer but takes an excessively long time to solve the problem, which may indicate that they have not yet mastered the corresponding topic.

To handle these conflicts, we carefully weight positive and negative evidence against each other to form a highly nuanced diagnosis of student knowledge that adapts appropriately to future observations, just like a tutor would.

In particular, if the evidence balances out to "just barely" place a student out of some topics, the system will consider those topics **conditionally completed**: the student will initially be given tasks under the assumption that they know those topics, but if the student struggles, then the system will immediately begin "falling backwards" along the appropriate learning paths.

| Example

As an example, suppose

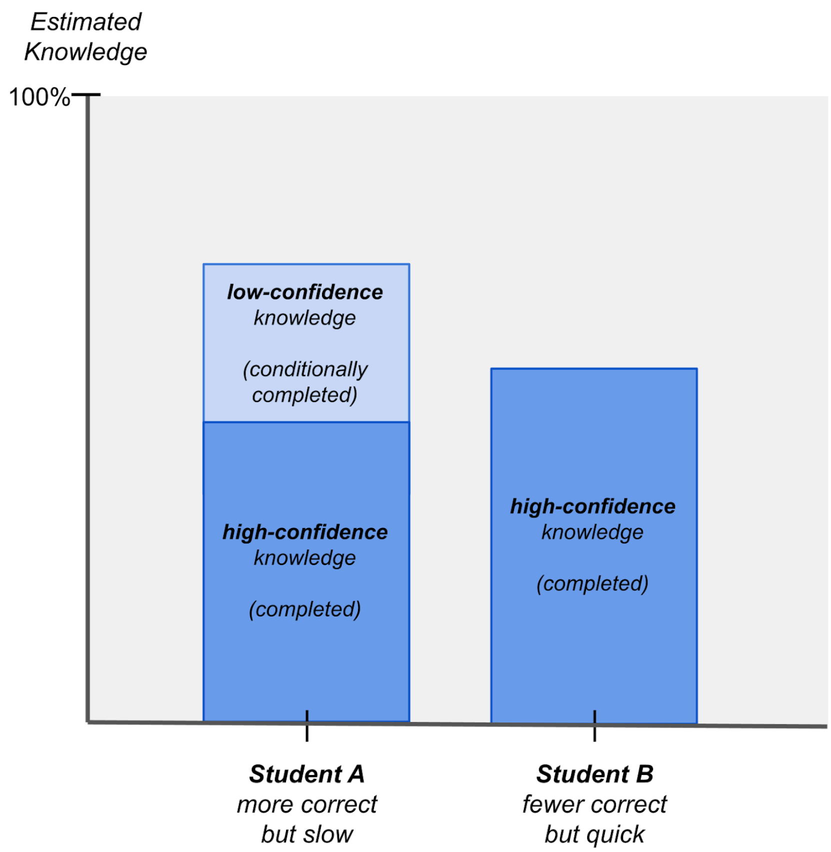
- student A answers many questions correctly on their diagnostic but takes excessively long on most questions, while

- student B answers fewer questions correctly but supplies correct answers quickly and confidently.

Then

- student A will have a higher amount of overall knowledge, a significant portion of which is low-confidence (and may be quickly pruned back if they struggle), while

- student B will have less overall knowledge but may have more high-confidence knowledge.



| Implementation

To achieve this behavior, we weight diagnostic evidence using a **plus-minus balance** at each topic.

- The **sign** (positive or non-positive) represents the prediction of whether the student knows the topic.

- The **magnitude** of the balance represents the degree of confidence in the prediction.

Each answer is associated with a **weight** that represents its contribution to the plus-minus balance of the corresponding topic. By default, an answer's weight is equal to one. However, if a student submits a correct answer but takes an excessively long time relative to the expected time for a student who has mastered the topic, the answer weight is diminished. The answer still gives the student positive credit, but the slower the student is to solve the problem (beyond a reasonable time threshold), the smaller the amount of credit.

Once the weight of an answer is determined, it propagates throughout the knowledge graph updating plus-minus balances of topics as appropriate:

- A *correct* answer *increases* the plus-minus balances of the answered topic and its *prerequisites* (and their prerequisites, and so on), while

- an *incorrect* answer *decreases* the plus-minus balances of the answered topic and its *post-requisites* (and their post-requisites, and so on).

After all diagnostic questions have been processed, any topics with positive plus-minus weights are credited with a number of repetitions equal to their plus-minus balance.

Conservative vs. Aggressive Edge of Mastery

There have been times when a student completed a course, took a placement diagnostic for the same course, and was surprised that the placement told them their knowledge frontier was lower in the course. But this is actually an expected result, especially for students who are weaker.

Just because a student successfully completes all the homework for a course, doesn't mean that they're going to ace a comprehensive final exam over the course. This is especially true for the placement exam, which is even harder than a normal exam: unlike a normal exam,

- a placement exam has to cover every topic (including all of the hardest topics), and

- it has to cover the most advanced question type from each of those topics (we can't place a student out of a topic if they only know how to do the simpler cases).

Additionally, although we often talk about a student's "edge of mastery" as though it were a single line across the student's knowledge profile, a student really has a "zone of mastery" that is bounded below by a conservative edge of mastery and above by an aggressive edge of mastery. A placement exam measures the conservative edge of mastery, while mastery-based learning with layering operates on the aggressive edge of mastery.

When a student successfully completes a lesson, they've mastered the topic well enough to continue layering on top of what they've learned, but they probably haven't reached the point of automaticity with the topic yet. By way of analogy, whenever a gymnast learns a new skill at practice, it takes some more time and practice before they are ready to showcase that skill in competition, but that doesn't stop them from continuing to work on more advanced skills during practice in the meantime.

Supplemental Diagnostics

As topics are added and connectivity is revised in the knowledge graph, the knowledge profile inferred from a student's initial placement diagnostic can get a little out of date. When this happens, we assign tiny diagnostics called **supplemental diagnostics** to bring the student's knowledge profile back up to date.

The topics on a supplemental diagnostic are those that were not directly assessed on the original diagnostic and have plus-minus balances of zero when considering all the assessment answers since (and including) the original diagnostic. The same form of highly efficient inference is used on supplemental diagnostics, which means supplemental diagnostics are generally quite small, consisting of at most a handful of questions.

Selecting Good Diagnostic Questions

There's a lot of nuance that goes into selecting a good diagnostic question for a given topic.

- On one hand, diagnostic questions can't be too easy. Each diagnostic question should be difficult enough that if a student were to answer it correctly, then an expert tutor or teacher would infer that they have fully mastered the corresponding topic.

- Additionally, a diagnostic question should exercise all of the prerequisites of the corresponding topic. Otherwise, if there's a prerequisite that the question doesn't exercise, then a student who doesn't know the prerequisite could still answer the question correctly and erroneously receive credit for the prerequisite.

- That said, a diagnostic question should generally not be the most difficult question in its corresponding topic. The more complicated a question, the higher the likelihood that a student might make a silly mistake.

So, each diagnostic question should be chosen as the simplest question that

1. would convince an expert tutor or teacher that the student has mastered the corresponding topic, and

2. exercises all of the topic's prerequisites.

In practice, due to the level of nuance required, we select diagnostic questions manually.

Chapter 27. Technical Deep Dive on Learning Efficiency

Note: *this chapter elaborates on concepts introduced in chapter 17 and chapter 21.*

Summary: *We introduce the concept of theoretical maximum learning efficiency, analogous to the physical phenomenon that nothing can travel faster than the speed of light. The maximum learning efficiency for a given knowledge graph depends on its encompassing density – however, a knowledge graph does not have to be fully encompassed, or even nearly fully encompassed, for its maximum learning efficiency to approach the theoretical limit. Math Academy's mathematical knowledge graph contains enough encompassings that its maximum learning efficiency is close to the theoretical limit. In practice, the actual learning efficiency attained by an individual student depends primarily on the student's quality of performance, and to a lesser extent on their pace (the average amount of work completed each day).*

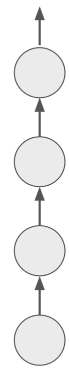
What is Learning Efficiency?

| Theoretical Maximum Learning Efficiency

In physics, nothing can travel faster than the speed of light. It is the theoretical maximum speed that any physical object can attain. A universal constant.

In the context of spaced repetition, there is an analogous concept: **theoretical maximum learning efficiency.** In theory, given a sufficiently encompassed body of knowledge, it is possible to complete all your spaced repetitions without ever having to explicitly review previously-learned material.

As a simple demonstration, consider a sequence of topics where the first topic is fully encompassed by the second, which is fully encompassed by the third, which is fully encompassed by the fourth, and so on.



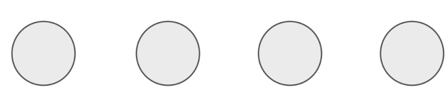
Each time you learn the next topic, all the topics below receive full implicit repetitions. Assuming that you never run out of new topics to learn, the only reason you would ever need to do an explicit repetition is if you get stuck repeatedly attempting and failing to learn the next topic.

It's important to realize that *a graph does not have to be fully encompassed, or even nearly fully encompassed, for its maximum learning efficiency to approach the theoretical limit.* Even if most relationships between topics are non-encompassing, a considerable minority of encompassings goes a long way.

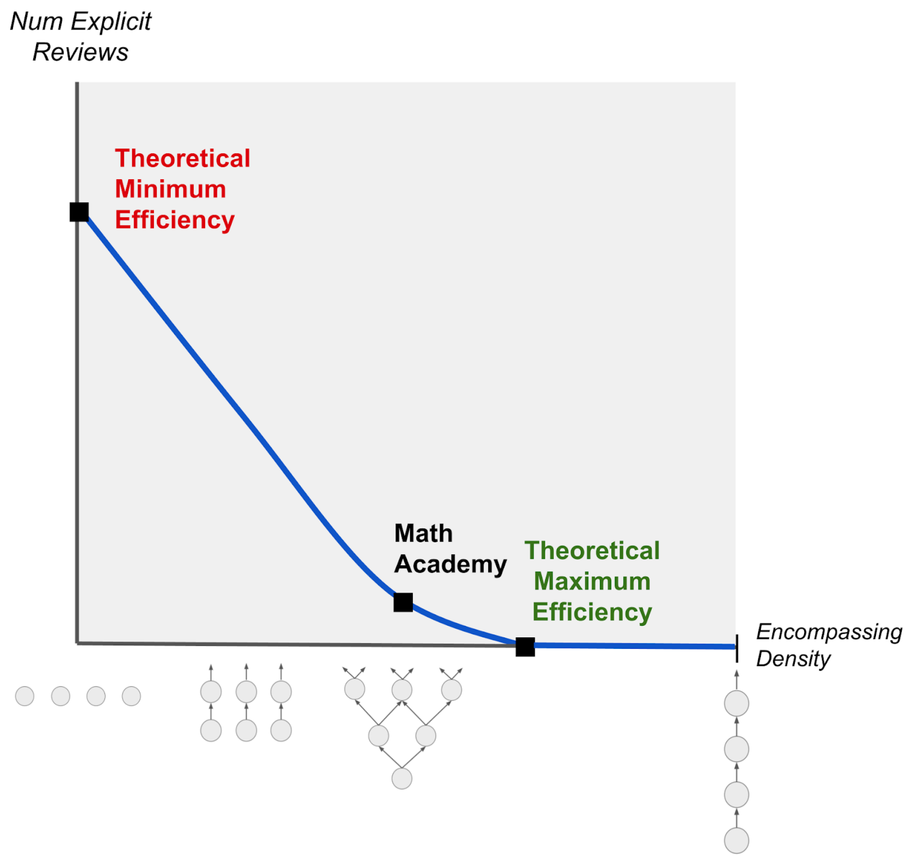
For instance, Math Academy's mathematical knowledge graph contains enough encompassings that its maximum learning efficiency is close to the theoretical limit. We have empirically observed that, in practice, most mathematical courses can be learned with roughly only one explicit review per topic on average. In theory, a perfect student who aced every single learning task would need even fewer explicit reviews.

| Theoretical Minimum Learning Efficiency

By contrast, there is also a concept of **theoretical minimum learning efficiency.** This is precisely the setting of independent flashcards – or equivalently, a set of topics without any encompassings.



In this setting, no topic can receive implicit repetitions from any other topic. Every single review must be done explicitly.



It's worth emphasizing that, unlike Math Academy, other spaced repetition systems do *not* leverage the power of encompassings and therefore implement theoretical *minimum* learning efficiency.

Factors that Impact Learning Efficiency

Remember that to achieve maximum learning efficiency, Math Academy uses a process that we call **repetition compression**. We gather all topics that have due repetitions, and then compress this set into a much smaller set of tasks that

1. covers all of the due repetitions, and

2. will lead to the greatest overall gain in spaced repetitions across your entire knowledge profile.

You can think of Math Academy as a turbo-boosted educational engine, where repetition compression is our combustion mechanism.

But remember that an engine can't actually move a car unless it is supplied with gas and oil. The gas is needed to produce the energy that moves the car, and the oil is needed to prevent friction from locking up the engine.

The same applies to Math Academy. In order to experience the turbo-boosting,

1. you have to put in a sufficient **amount** of work that can be converted into educational progress, and

2. the **quality** of your work has to be high enough to avoid excessive friction during the learning process.

| Performance

By looking at your **performance** (pass rate and accuracy) across various types of learning tasks, we are able to calculate a **learning efficiency percentage** that estimates how close you are to the maximum possible efficiency for your course.

If you maintain a high learning efficiency, then you can make a lot of progress in your course by doing a relatively small amount of work. But if you have a low learning efficiency, then you will have to do significantly more work to make the same amount of progress.

Learning Efficiency	How Much Work to Complete a Course
1.00	1x
0.80	1.25x
0.67	1.5x
0.50	2x
0.25	4x

| Pace

In Math Academy, work is measured in **eXperience Points** (**XP**). One XP represents one minute of fully-focused, fully-productive work for an average serious (but imperfect) student. The amount of XP that you complete per weekday (on average) is called your **pace.**

> Learning Efficiency vs Pace

Although the quality of your work is the single greatest factor that affects your learning efficiency, your pace can affect your learning efficiency as well. The faster you push your knowledge frontier forward, the further your knowledge frontier is ahead of your due reviews, and the more likely it is that we can find good topics to "knock out" a large number of your due reviews.

We empirically determined the following relationship:

$$\text{learning efficiency} \propto \text{pace}^{0.1}$$

This means that if you double your pace, your learning efficiency increases by about $2^{0.1}$ = 7%. Likewise, if you cut your pace in half, your learning efficiency decreases by about 7%.

> Pace vs Time to Completion

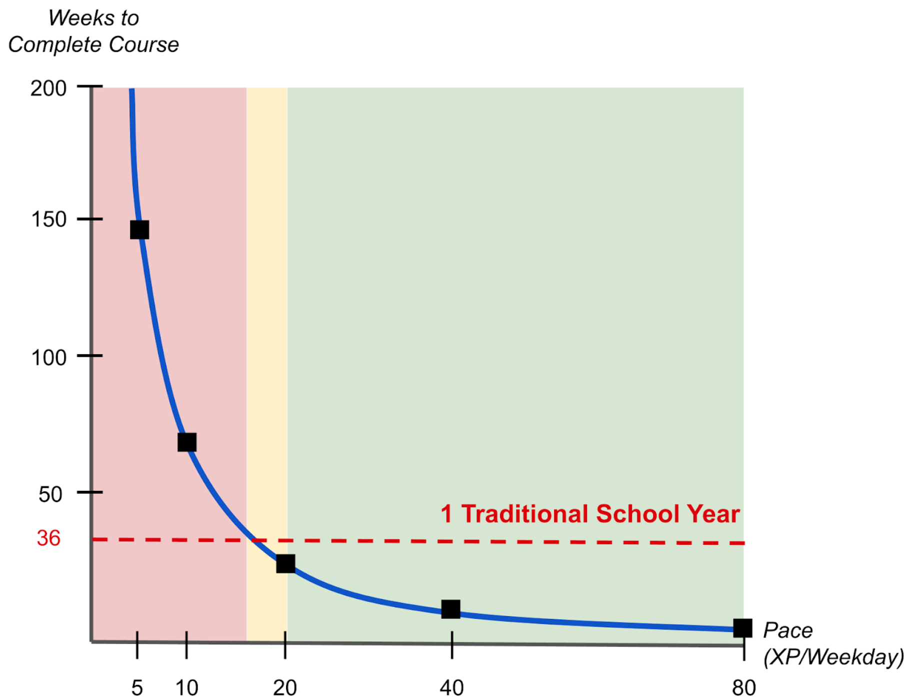
In a normal class period during a school day, it's reasonable to expect at least 40 minutes of fully-focused, fully-productive work. This corresponds to a baseline pace of 40 XP per weekday.

When we benchmark the amount of XP in our courses, we simulate an average student who is serious but imperfect and works at a pace of 40 XP per weekday. On average, courses contain about 3000 XP assuming that a student knows all the necessary prerequisites (though this can vary a lot depending on the amount of material that must be covered, e.g. prealgebra is ~2000 XP but precalculus is ~4000 XP).

Below is a table that shows how long it would take you to complete a 3000 XP course, depending on your pace. Note that learning efficiencies are computed relative to the baseline pace of 40 XP/weekday (so a pace of 40 XP/weekday corresponds to an efficiency of 1, and higher paces correspond to efficiencies greater than 1).

Pace *(XP/weekday)*	**Efficiency Multiplier** $(pace/40)^{0.1}$	**How Long To Complete a Course Benchmarked at 3000 XP** *weekdays = 3000/(pace*multiplier)*
160	1.15	3 weeks
80	1.07	7 weeks
40	1.00	15 weeks
20	0.93	32 weeks
10	0.87	69 weeks (~1.3 years)
5	0.81	148 weeks (~3 years)

To put this in perspective: in a traditional classroom, each weekday involves 50 minutes of class plus the same duration of homework after school. On this schedule, it takes students a full school year (36 weeks) to complete a course.



But if you spent the same amount of time working on Math Academy (100 XP/weekday), you could finish in just 5-6 weeks! That's more than a 6x speedup – and you don't have to be a genius to achieve it. Remember, we're talking about an average student who is serious but imperfect.

Even for a massive course like AP Calculus BC (which is benchmarked at about 6000 XP, twice as big as an average course), the speedup is still over 3x. And if you factor in all the extra time you'd spend studying for quizzes, midterms, finals, and the AP test itself in a traditional class, which is already included in Math Academy's 6000 XP benchmark, it's a 4x speedup.

On the flipside, if you tried to use Math Academy like a phone game and only did a couple of minutes per day, it could take you nearly a decade to learn a traditional school year's worth of math.

For this reason, we highly recommend that you maintain a pace of *at least* 15 XP per weekday if you want to experience the benefits of Math Academy. But really, the higher your pace, the better.

Chapter 28. Technical Deep Dive on Prioritizing Core Topics

Note: *this chapter elaborates on concepts introduced in chapter 4.*

Summary: *Math Academy prioritizes core topics that are most relevant in the "big picture" of mathematics. No topics are skipped, but because all topics are continually reviewed, covering core topics first allows students to get more practice and therefore develop a greater degree of automaticity on the core topics by the end of the course. This is advantageous because the core topics are the ones that appear more frequently as prerequisites of other topics in mathematics. We also observed that roughly a third of 4th grade through AP Calculus BC topics, while necessary to meet standards, were not actually prerequisites for university math, so we created a streamlined Mathematical Foundations (MF) course sequence that cuts out those topics and consists of mostly core topics. The MF sequence is geared towards adult learners who want to pursue advanced university courses as soon as possible but lack the necessary foundational knowledge.*

Core and Supplemental Topics

When a student progresses through a course, Math Academy prioritizes **core topics** first, that is, the topics that are most relevant in the "big picture" of mathematics. For instance, in calculus, the product rule would be a core topic, while the intermediate value theorem would be a **supplemental topic**.

Of course, a student taking the calculus course will of course cover both core and supplemental topics. No topics are skipped; it's just a matter of the order in which they are covered. Because students cover core topics first and continue practicing them throughout the course, they get more practice and therefore develop a greater degree of automaticity on the core topics by the end of the course. This is advantageous because the core topics are the ones that appear more frequently as prerequisites of other topics in mathematics.

Math Academy employs a proprietary intelligent algorithm to automatically identify core topics in its knowledge graph. At a high level, the idea is to satisfy two competing conditions:

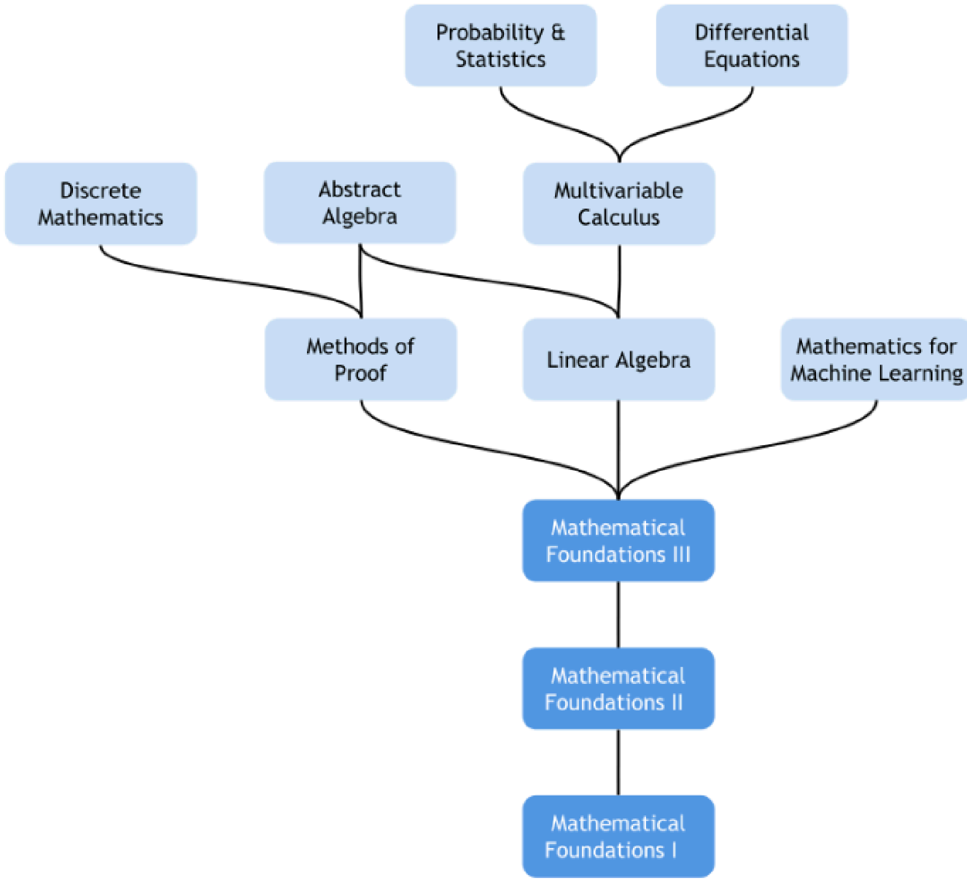
1. If a topic is core, then all of its ancestor topics (i.e. its prerequisites, their prerequisites, and so on) must also be core.

2. However, in each standard course in the knowledge graph, there must be some balance between core and supplemental topics – for instance, we should not label all topics as core, or all topics as supplemental, even though both of those cases would technically satisfy the preceding condition. The specific balance may vary depending on the connectivity of topics in the course and their relationship to topics in other courses.

The Mathematical Foundations Sequence

After developing a comprehensive curriculum that covers all the standards for 4th grade through AP Calculus BC, as well as plenty of advanced university courses, we found that roughly a third of 4th grade through AP Calculus BC topics were not actually prerequisites for university math. So, we created a streamlined **Mathematical Foundations** (**MF**) course sequence that cuts out those topics and consists of mostly core topics.

The MF sequence is geared towards adult learners who want to pursue advanced university courses as soon as possible but lack the necessary foundational knowledge. Whether an adult is starting off again with the basics or just needs to brush up on calculus, our Mathematical Foundations sequence is the fastest and most efficient way to get up to speed with the mathematical concepts and tools that are necessary to excel in university-level mathematics.


Probability &
Statistics
Differential
Equations
Discrete
Mathematics
Abstract
Algebra
Multivariable
Calculus
Methods of
Proof
Linear Algebra
Mathematics for
Machine Learning
Mathematical
Foundations III
Mathematical
Foundations II
Mathematical
Foundations I

Frequently Asked Questions

The Practice Experience

| Active Learning, Scaffolding, and Automaticity

> How does a lesson work?

At the most fundamental level, a lesson is a sequence of slides with instructional text, interspersed with active problem-solving.

Each lesson starts out with an introduction, and then moves to a worked example, followed by 2-5 practice questions on the same type of problem as the worked example. The worked example and active problem-solving we collectively refer to as a "knowledge point" or KP.

The number of practice questions in a KP adapts to the student's performance: practice questions continue until the student demonstrates sufficient mastery of the KP to continue building more advanced learning on top of that understanding.

After completing a KP, the student moves on to the next KP. A typical lesson has about 3 or 4 KPs increasing in difficulty. The first KP covers the simplest case to introduce a new concept/skill, and the following KPs build on this concept/skill, extending it to progressively more advanced cases.

(If a student is unable to demonstrate mastery of a KP within 2-5 questions, then they "fail" the lesson and spend some time working on other lessons before coming back to re-attempt the originally failed lesson. On average, students pass lessons on the first attempt 95% of the time and within two tries 99% of the time.)

> Solving problems breaks my flow of learning. Is it really necessary?

Active problem-solving is where the learning happens. It may *feel* like learning when you're following along while reading/skimming a book, but that comfortable fluency is completely artificial. It arises from the fact that the surrounding context is already on your mind, and you're not actually being made to pull it from memory.

If you define learning as a positive change in long-term memory, then you haven't learned unless you're able to consistently reproduce the information you consumed and use it to solve problems. This doesn't happen when you just "follow along," even if you understand perfectly. It's the act of retrieving information from memory that transfers the information to long-term memory. If you don't practice retrieval, then the information quickly dissipates. It stays with you only briefly – just long enough to trick you into thinking it'll stick with you, when it's really on the way out the door.

The most effective way to avoid this problem and maximize your learning – not just the perception of learning – is to switch over to active problem-solving immediately after consuming a minimum effective dose of information. While this may initially feel a bit jarring, it isn't slowing down your learning – it's only exposing the fact that your perception of learning does not accurately reflect actual learning. In reality, it's speeding up your actual learning, and the only thing it's slowing down is your perception of learning.

You might say "but I had learned so much, and I had it down pat, and then I forgot it all when I focused my effort on solving a problem." But the thing is, if you can't retrieve that information from memory at the snap of a finger, after thinking about other things or zooming in to focus on a specific problem, it means you didn't really have it down pat. You just felt like you did because you weren't being made to attempt to regenerate the information from scratch, from memory. What you're really saying is "I was juggling a lot of information in my working memory (WM), and I thought it was in my long-term memory (LTM), and then I cleared a lot of it out from my WM when I focused my effort on transferring some of that WM into LTM."

> Don't students need to struggle for long periods of time, without too much guidance, to train their general problem-solving ability?

For students (not experts), empirical results point in the opposite direction. One key empirical result is the expertise reversal effect, a well-replicated phenomenon that instructional techniques that promote the most learning in experts, promote the least learning in beginners, and vice versa. It's true that many highly skilled professionals spend a lot of time solving open-ended problems, and in the process, discovering new knowledge as opposed to obtaining it through direct instruction. But that doesn't mean beginners should do the same. The expertise reversal effect suggests the opposite – that beginners (i.e., students) learn most effectively through direct instruction.

Additionally (and relatedly), as discussed in chapter 8: there's a mountain of empirical evidence that you can increase the number of examples & problem-solving experiences in a student's knowledge base – but a lack of evidence that you can increase the student's ability to generalize from those examples (by doing things other than equipping them with progressively more advanced examples & problem-solving experiences). In other words, research indicates that the most effective way to improve a student's problem-solving ability in any domain is simply to equip them with more foundational skills in that domain. The way to increase a student's ability to make mental leaps is not by having them jump further, but by having them build bridges from which to jump.

There does not seem to be any tangible, empirically-supported reason for a student to struggle with a problem for a long period of time as opposed to using that time to learn more content. For instance, in an hour-long training session, a student will make a lot more progress by solving numerous "deliberate practice" problems that each take a small amount of time given their current level of knowledge, than by attempting a single problem that they struggle with for a long period of time. (To be clear: the deliberate practice problems must be grouped into minimal effective doses, well-scaffolded & increasing in difficulty, across a variety of topics at the edge of the student's knowledge.)

As Sweller, Clark, and Kirschner sum it up in their 2010 article *Teaching General Problem-Solving Skills Is Not a Substitute for, or a Viable Addition to, Teaching Mathematics*:

> *"Although some mathematicians, in the absence of adequate instruction, may have learned to solve mathematics problems by discovering solutions without explicit guidance, this approach was never the most effective or efficient way to learn mathematics.*
> *...*
> *In short, the research suggests that we can teach aspiring mathematicians to be effective problem solvers only by providing them with a large store of domain-specific schemas. Mathematical problem-solving skill is acquired through a large number of specific mathematical problem-solving strategies relevant to particular problems. There are no separate, general problem-solving strategies that can be learned."*

Another good reference is *Putting Students on the Path to Learning: The Case for Fully Guided Instruction* by the same authors (Clark, Kirschner, & Sweller, 2012). It's an expanded version of the 2010 article.

> Is automaticity really required to move up to the next level? Doesn't it just come with time?

While we emphasize the importance of building automaticity over time (see Chapter 14), we do not mean to suggest that students have to learn skills to the point of automaticity before moving forward.

Before moving forward, students must reach a "baseline mastery" performance threshold indicating that they have learned the material well enough to solve problems successfully (and do this consistently). This level of baseline mastery is necessary for students to continue layering on additional knowledge where baseline-mastered skills are exercised as component sub-skills within more complex skills. However, the performance threshold for baseline mastery is not as high as the performance level for automaticity, which is realized over a longer period of time.

As discussed in Chapter 15, layering more advanced skills is one of the most efficient ways to achieve automaticity: as students learn progressively more advanced material, they reinforce and deepen their foundational knowledge. However, the efficiency of layering is conditional on students being able to successfully execute the foundational skills, which requires a baseline level of mastery.

Furthermore, while one may hope to naturally develop automaticity on lower-level skills by layering on more advanced skills, it is still necessary to check that this is happening and take swift action if it's not. This is one reason why Math Academy leverages frequent timed assessments and immediately follows up with remedial support on any questions a student misses.

Checking for automaticity will continue to grow as a centerpiece of the Math Academy system, especially in the context of teaching "math facts" like addition and multiplication tables, since it is easier for a lack of automaticity to fly under the radar in those areas (due to how simple and quick the problems are). To provide a concrete example: sometimes a student will default to recalculating (or even finger-counting) every single fact instead of first trying to retrieve it from memory. At first their speed and accuracy will increase because they're getting better at recalculating, but these gains will asymptote off before the student reaches anywhere near the range of automaticity. This kind of student will never develop the necessary automaticity unless somebody intervenes to break them out of their habit and support them with flashcard-style practice. As Math Academy develops a "math facts" curriculum, these automaticity interventions will be built directly into the system.

> If worked examples are necessary to maximize learning efficiency, then why am I able to solve problems just fine without them?

When students start out learning math, it sometimes feels easy to the point that they can solve problems reasonably quickly without having to see worked examples. But this phase is temporary: as the level of math rises, solving problems without worked examples quickly becomes overwhelming and inefficient. Without worked examples, learners reach a point where unguided problem-solving overwhelms their working memory and puts them in a state of cognitive overload where they feel frustrated, confused, and are unable to solve the problem. They flat-out stop making progress, and no more learning happens.

Even before a lack of worked examples becomes a complete roadblock to successful problem-solving, it will inflate the amount of time needed for a student to successfully solve problems, thereby throttling the volume of deliberate practice cycles that can be achieved in any given amount of time. This is problematic because (as discussed in chapter 11) the accumulated volume of action-feedback-improvement cycles is the single biggest factor responsible for individual differences in performance among elite performers across a wide variety of talent domains.

In summary: math gets hard for different students at different levels – it can be as early as high school algebra or as late as graduate-level Algebraic Topology – but everyone eventually reaches a level where things no longer feel obvious and they can't figure things out as quickly on the fly. That's where worked examples and instructional scaffolding come in to keep students making fast progress. If you don't have worked examples and instructional scaffolding to help carry you through once math becomes hard for you, then every problem basically blows up into a "research project" for you. That's okay if you're a research mathematician at the edge of your field, but if you're a student who still has a ways to go before reaching the edge of human mathematical knowledge, then it's far less efficient (even if you have fun with it).

Of course, if you really want to solve problems without referring to worked examples, nobody is stopping you from skipping worked examples and trying your hand at solving the corresponding problems without guidance. It can be a fun challenge! You just need to make sure that you're still solving the problems quickly and accurately – if you start slowing down and/or becoming less accurate (or, more subtly, if you start to doubt yourself and lose interest), then that's an indication you need to start leveraging those worked examples.

> Why aren't Math Academy's university courses structured like typical higher math textbooks with minimal scaffolding?

Higher math textbooks and classes are typically not aligned with (and are often in direct opposition to) decades of research into the cognitive science of learning. Higher math is heavily g-loaded, which creates a cognitive barrier for many students. The goal of guided and scaffolded instruction is to help boost students over that barrier. (To be clear, we do not mean to imply that higher math would be "easy" if taught properly – just that many more people would be able to learn it, than are currently able to learn it.)

Why do higher math textbooks lack such scaffolding? For one, the amount of work it takes to create a textbook explodes with the level of guidance and scaffolding, so in practice there is a limit to the amount of boosting that is feasible, especially if the textbook is written entirely by a single author.

That said, most higher math textbooks don't even come close to the theoretical limit for a single author, much less the theoretical limit for a team of content writers. Why is that? First, consider the following problem that has affected anyone who has ever tried to learn math from a textbook: a worked example demonstrates a special case, but a practice problem requires a logical leap that wasn't explicitly covered. A number of textbooks seemingly attempt to solve this problem by side-stepping the need for a large amount of scaffolding (worked examples and practice problems increasing in difficulty), and instead focus the effort on trying to teach general problem-solving skills with challenging problems that require large mental leaps.

However, as discussed in Chapter 8, there is a mountain of evidence in the cognitive science literature that you can increase the number of examples and problem-solving experiences in a student's knowledge base, but a lack of evidence that you can increase the student's ability to generalize from those examples. In other words, research indicates that the best way to improve one's problem-solving ability in any domain is simply to acquire more foundational skills in that domain. The way you increase your ability to make mental leaps is not actually by jumping farther, but rather, by building bridges that reduce the distance you need to jump.

Higher math textbooks and courses often focus on trying to train jumping distance instead of bridge-building – especially once a student gets into serious math-major courses like Real Analysis and Abstract Algebra. However, what actually works in practice is simply creating more worked examples, organizing them well, and giving students practice with problems like each worked example before moving them onto the next worked example covering a slightly more challenging case. Students can successfully climb to higher-than-expected levels of math with this approach, but many educational resources shy away from it because it takes so much work to create all the necessary content.

| Spaced Repetition and Interleaving

> Why can't I just learn one unit at a time? Interleaving feels disorienting.

We realize it may *feel* easier to learn one unit at a time without interleaving. However, that feeling is completely artificial: students measurably learn better when they vary up their practice after a minimum effective dose of initial learning.

As discussed in Chapter 18, a common finding in the research is that when students do not interleave, learning tasks are made artificially easy because the surrounding context is already on one's mind – one does not have to pull the context from memory again. This produces a comfortable sense of fluency, but that feeling is completely artificial.

Think about it this way: maybe you go through a lesson on limits and you're feeling really good, and then we change things up on you and give you a lesson on derivatives, and then integrals, and then sequences & series. And after that we go back to limits. You might say *"You messed up my learning! I had limits down pat and now you made me forget it by making me think about other stuff."* But the thing is, if you can't retrieve that information from memory instantaneously, after thinking about other things, it means you didn't really have it down pat. You just felt like you did because you weren't being made to attempt to regenerate the information from scratch, unassisted, from memory. And the only way to get better at regenerating the information from scratch, unassisted, from memory, is by having to practice doing that – which only happens when you interleave.

> Is the spaced repetition happening? I started recently and all I have is lessons. Where are the reviews?

A lot of the review happens implicitly by having you learn new topics that encompass previously learned topics as subskills. At the beginning, when you have a small body of knowledge to review, we're able to pick new lessons that knock out all your reviews without you having to explicitly do any review tasks. However, as you build up a larger body of knowledge to review, you'll start to see explicit review tasks on topics that we are not able to knock out explicitly.

So, basically: the spaced repetition has already started kicking in, but we do a lot of optimization to make that happen simultaneously while having you learn new material. The only time you'll get an explicit review is when we're not able to knock it out implicitly while having you learn something new. (Though, after a quiz, you'll also get explicit reviews immediately on any questions you miss.)

> Math Academy reviews feel challenging. Aren't reviews supposed to be easy if I learned the material properly?

If you're actually trying to maximize learning efficiency, then reviews should feel tough. Why? Because recalling tricky information improves memory, while recalling easy information doesn't.

That's the whole idea behind spaced repetition: your memory has to get a bit fuzzy before the next repetition, otherwise the desired effect – slowing the rate of forgetting and remembering longer next time – doesn't happen (or at least not nearly as much). It's the act of successfully retrieving fuzzy memory, not clear memory, that extends the memory duration.

If review problems are easy, not actually extending your memory duration, then what's the point? It's better to learn something new. A maximum-efficiency teacher will intentionally let your memory fade a bit before review so that the act of refreshing your memory actually deepens your long-term encoding, and they'll use the extra time to cover more new material.

Reviews should feel as mentally taxing as initial lessons. You're getting better, but the bar for success also is getting higher. Your brain has to hold the memory for a longer period of time – just like a muscle holding a weight.

The analogy to weightlifting runs deep. In the context of spaced repetition, the way you increase the weight is by waiting longer before retrieving the knowledge again. But you also don't want to wait too long to retrieve the knowledge, because then you won't be able to successfully retrieve it. This is just like how in weightlifting, you need to increase the weight to the point where you struggle to lift it, but you are able to overcome the struggle. That's how you build muscle, and that's also how you build long-term memory. Spaced repetition = "wait"lifting.

> Does Math Academy's spaced repetition system provide enough practice? I learned some new information on Math Academy but I am not confident in my ability to retrieve it from memory unassisted.

Math Academy's spaced repetition system should be sufficient for remembering everything: concepts, procedures, definitions, theorems, formulas, etc. However, it can take several weeks of consistent practice before you really feel confident in retrieving recently learned information. During their first several weeks on Math Academy, it is not uncommon for students feel unsure whether the information they are learning is going to stick in their brain long-term – but after a month or so of consistent usage, they notice something has changed and they're able to pull a lot of this information from memory without much effort (it feels kind of like magic).

This phenomenon can be explained by the following dynamics within the spaced repetition process:

- Early on, forgetting happens so rapidly that the spaced repetition process is unforgiving to imprecision: if you are slightly late to the next repetition, your memory may have decayed enough since the ideal repetition time that you need a retrieval cue or reminder during the next repetition (and on the flipside, if you are slightly early, then the repetition may lose a lot of its effectiveness in slowing your rate of forgetting).

- However, by the time you get through a handful of repetitions, your rate of forgetting has slowed enough to make the spaced repetition process more robust to imprecision: even if you are slightly late to the repetition, this lateness is small relative to the repetition interval which is now large, so your memory hasn't decayed much more than desired and you're still able to recall successfully without reference material. (Likewise, if you are slightly early, the repetition still retains most of its effectiveness.)

Note, however, that even with consistent practice, this "magical transition" depends on properly engaging in retrieval practice, trying your best to recall from memory instead of automatically going back to reference material whenever you feel your memory is a bit fuzzy. Successfully retrieving a fuzzy memory is the very thing that slows future forgetting and extends the memory's duration. If you just load up the information into your brain by looking at a reference, then you may refresh the information, but you don't actually slow the rate of forgetting, so you end up stuck in a vicious cycle of constant forgetting and reliance on reference material.

> After I complete a course and move to the next course, won't I forget what I've learned in the first course unless I keep on reviewing it?

Every topic a student learns on Math Academy will automatically be reviewed into the future, even if they switch to a different course. Note that these reviews will be implicitly "knocked out" by material in the new course when possible, so students may not see very many reviews that are explicitly on topics from past courses, even though those lower-course topics are indeed continually being reviewed into the future in accordance with the usual spaced repetition procedure.

> Given that lessons can "knock out" reviews, should students always give preference to lessons over reviews if both activity types are available?

It doesn't really matter. If a review is on a student's dashboard it means we weren't able to knock it out by having the student do a lesson instead. Whenever it is possible for a due review to be knocked out by a lesson, we will only offer the student the lesson. We will not offer them a review that is made redundant by a lesson already on their dashboard.

> Math Academy maximizes learning efficiency if a student is willing to engage in forms of training that are highly effortful. What about for students who don't have as much energy and motivation?

Math Academy teaches math as though we were training a professional athlete or musician, or anyone looking to acquire a skill to the highest degree possible. When a student signs up for Math Academy, it's like going to a gym where one of the personal trainers was a former Olympic sprinter, and telling them "I'm going to show up 40 minutes per day, 5 days per week, and I want you to use whatever methods of training are going to make the most improvements on my 100-meter dash time. I don't care how exhausting they are; I am willing to work hard."

Many students do not want to devote this much effort to their learning, just like many people who sign up for the gym do not want to do an Olympic-intensity workout most days of the week. While it's true that willingness to work hard is a bottleneck for many students, such students are not part of our target market. If a student is not willing to put forth a high degree of effort engaging in the most effective training techniques, which are taxing, then the Math Academy system is not a good fit for them.

| Remediation

> If a student fails a task, why does the system ask them to re-attempt the task later? Why doesn't it just peel back their knowledge profile immediately?

How quickly the system peels back a student's knowledge profile in response to a failed task depends on how much evidence the student has demonstrated for knowing the prerequisite topics. If a student has demonstrated strong evidence for knowing the prerequisites, then the system will be very slow in peeling back the student's knowledge profile. If they fail a lesson twice in a row without making additional progress, the system will provide support in the form of remedial reviews on the key prerequisite material implicated in the student's point of struggle, and if the student fail those remedial reviews, then they will be given the corresponding lessons, and so on.

This is a slow process because it has to be resistant to adversarial students "gaming the system." If we peeled back a student's knowledge profile quickly in response to failing a task, even when there is strong evidence that a student knows the prerequisite content, then it would create an exploit: whenever tasks begin to feel challenging, an adversarial student could intentionally fail a number of tasks to peel back their knowledge profile until they reach the point where they have days of super easy work ahead of them that they already know how to do.

That said, if a student supplied evidence for knowing material, but the amount of evidence is very low, then the system will be faster to adapt. This process is explained in more detail in the answer to the FAQ entry *"After the diagnostic, why did failing a single task bring my progress down multiple percent?"*.

Student Behavior

| Usage of Paper and Pencil

> Should Math Academy students take notes during lessons?

Note-taking should not be necessary. Our spaced repetition system takes care of review, and on the rare occasion that you do happen to forget something, you can always look back at prerequisite lessons in "reference mode" to brush up on whatever you may have forgotten.

Even further, we would actively recommend *against* taking notes. To transfer information into long-term memory, you need to practice retrieving it without assistance – but when you take great notes, you're tempted to refer back to those notes all the time instead of trying to pull information from memory. As a result, notes can turn into a crutch that spirals you into a vicious cycle of forgetting. The same reasoning applies to any sort of reference material.

Of course, if you can't recall something after trying your hardest, it's okay to check reference material, but only as a last resort. Even then, do not solve the problem alongside the reference material – peek once, and then try to solve the problem without looking again.

That said, we wish to make a distinction between "note-taking" and "listening on paper." While we do not recommend transcribing information for later use, we see no issue with jotting down key bits of information to maintain focus and draw connections while being presented with new material. "Listening on paper" is not necessary or even helpful for all students, but some find that it helps them engage in "active listening" and deepen their processing of the material being learned.

Again, however, a student who practices "listening on paper" must always take care to avoid the following pitfalls:

- **Pitfall 1:** Transcribing, or, more generally, slowing down the rate of ingesting new information without actually deepening the processing of that information.

- **Pitfall 2:** Solving problems alongside notes, or, more generally, using a reference as a crutch to avoid or reduce effort towards proper retrieval practice.

> When should a student solve a problem in their head vs writing it out on paper?

We would suggest that the most technically correct general rule is this: Do a problem (or sub-component of that problem) in your head when

1. you're able to do it in your head reliably, with very little effort, and

2. it feels like you're only holding one thing in your head – like, you're working with a solid, cohesive "chunk" of information as opposed to having to "juggle" multiple components of that information to keep it in your brain.

Here's the science behind that.

It's well established in cognitive science, specifically under the umbrella of "cognitive load theory," that your working memory has a limited capacity to hold new information – and when you push your working memory close to that limit, you become more likely to make mistakes and less likely to complete the training task, which impedes your learning.

Reducing cognitive load is the goal – not just "a" goal, but in fact "the" goal, the whole point – of structured education. The more instructional scaffolding is provided, the lower the student's cognitive load, and the less cognitive "friction" there is to slow the student's acquisition of the knowledge covered in the curriculum.

It's the same way with solving problems on paper: the goal is to lower your cognitive load. By writing down intermediate steps on paper, you can temporarily remove information from your working memory to make room for new information, and then quickly load up the original information by looking back at the paper when needed.

(In a sense, the paper functions as an artificial long-term memory bank where you can store new information that you don't already have encoded in your brain's long-term memory. It takes a lot of time and effort for your brain to encode information to its own long term memory, but writing information on paper enables you to sidestep these biological limits.)

The Asymmetric Tradeoff

To be clear, there *is* a hidden tradeoff. If you overdo the scaffolding, or you write down more than you need to on paper, then it's going to inflate the amount of time that it takes you to complete the curriculum or solve the problem, respectively. If your cognitive load is already low, there is no benefit to lowering it further – all that does is create more mechanical work for you that burns your time.

However, the tradeoff is asymmetric:

- If you undershoot the scaffolding or don't write down enough work on paper, and you blow your working memory capacity, then you hit a brick wall. You're simply unable to complete the task. Your learning progress grinds to a halt, full-stop. Even if you just "come close" to full capacity, your error rate skyrockets, impeding your learning.

- On the other hand, if you overshoot the scaffolding or write down more than you needed to on paper, then sure, it will technically be suboptimal, but typically not by much. You wrote down an extra line or two on paper than you really needed to? Big whoop, it took you an extra couple seconds to solve the problem. You could have saved a couple seconds by not writing those steps down, but that would also have put you dangerously close to holding too much in your head and making a mistake. Just like in your bank account, having a little buffer is not a bad thing.

Because the tradeoff is so asymmetric, it's best to err on the side of caution, writing down potentially a bit more than you need to. When in doubt, write it out.

Building Good Habits

In addition to guarding yourself against being on the wrong side of the asymmetric tradeoff, another reason why you should err on the side of caution (i.e., writing down too much as opposed to too little) is that you need to build good habits for the future.

As you climb up the levels of mathematics, the level of technical sophistication increases, and consequently, so does the level of cognitive effort. Even if you are able to do problems entirely in your head at lower levels of math, you will not be able to do so indefinitely into the future. You will eventually get to a point where you are unable to do problems in your head without blowing your working memory capacity, and at that point, the only way to continue making progress will be to use paper and pencil as a tool to reduce your cognitive load.

However, the longer you go without using paper and pencil, the more solidified that habit will be, and the harder it will be to get yourself to change it. So, even if it's not strictly necessary, it's a good idea to get in the habit of writing at least some work out. If you don't, then you may cling to the habit of doing all the work in your head for too long, well past the point when you really need to start writing work down on paper – which will gradually eat away at your performance and progress, eventually bringing you face to face with a "day of reckoning" where your entire mathematical future is on the line.

We have seen many bright students breeze through basic math refusing to write down any work, only to struggle in intermediate or advanced math simply because they stubbornly continue refusing to write down their work.

| Reliance on Reference Material

> During a quiz, if I can't remember a "fact" (e.g., a definition or theorem) but I remember the process for using it to solve problems, should I look it up?

We would recommend *not* to rely on any external material during the quiz, as it will cause the system to make decisions as if you got the question right without relying on any reference material. In the case described, you would benefit from getting a refresher on those topics via follow-up reviews after the quiz – but you won't get that support if the system thinks you were able to answer the question without looking at a reference.

That said, we *would* recommend taking your best guess even if you are not confident about your answer. If you're fuzzy on information, then it's possible to retrieve the information successfully despite having a low degree of confidence. (That said: even if you manage to get it right, definitely look at the solution afterwards to further refresh your memory.)

More generally, we (people in general) are often not good at judging how well we know something, especially if it's something that we learned recently and are not super confident about. The Math Academy system has been designed around the idea that the student should just take their best attempt at whatever they're doing, and the system will take corrective action after any attempt that turns out to be unsuccessful.

> What should a student do if they are unsure how to solve a problem despite trying their best to refer back to the supporting instructional content?

This situation should almost never happen if a student reads the supporting instructional content carefully and attempts to solve the problem step-by-step, writing every step down on paper, and referring back to the worked example at each step when stuck.

However, if this situation does still arise, and the student spends 5 minutes stuck at a particular step without making further headway on it (despite reviewing the key prerequisites and checking earlier parts of the lesson for information they may have missed), then the most productive use of time is to submit a "best guess" and then study the solution carefully after the question is graded (regardless of whether the best guess was correct or incorrect). The student should not move forward to the next question until they have worked out the original question themselves, on paper, following along with the solution and ensuring that they understand the rationale behind each step.

If a student ever finds something that could have been explained better in the supporting instructional content, they are encouraged to submit a flag explaining the improvements that they think should be made. We continually refine our content based on student feedback, and while the content is very solid by now after years of refinement, we are always on the lookout for ways to continue improving it.

XP and Practice Schedules

| XP System

> If I pass a lesson but don't get full XP, does that mean I only understood part of the material? If so, how does Math Academy fill in the rest of my understanding?

Math Academy only allows a student to pass a lesson if they evidence sufficient mastery to continue building on the knowledge covered during the lesson. So, if a student passes a lesson, then no remedial support is needed.

Of course, mastery is not synonymous with maximum understanding (i.e., "completely intuits everything covered in the lesson to the point of full automaticity"), and students will be at varying degrees between mastery and maximum understanding after completing a lesson. However, as students continue reviewing and layering more advanced knowledge on top of that topic, their understanding will become further solidified and they will move closer and closer to the point of maximum understanding.

> I can do way more than 1 XP per minute!

The XP is calibrated to 1 XP = 1 minute for an average Math Academy student *at that level of math*. A student may be able to move significantly faster if they are particularly mathematically inclined relative to other students at that level, especially if they are already halfway familiar with some of the material.

> As I progress through Math Academy's curriculum, it is gradually taking me longer to earn XP. Why does this happen, and can anything be done about it?

As a student progresses into more advanced content, the amount of time it takes to earn XP will increase gradually, even though the average is staying at 1 XP = 1 minute for students at that level of math.

This happens because, as the math gets more advanced, students who take longer to earn XP are more likely to drop out, which effectively raises the bar for being an "average" student at the next level of math. In general, the further you go in any skill domain, the higher levels you reach, the more talented the other people at that level, and the harder you have to work to get to the next level.

That said, if a student is taking a long time to earn XP, then it is always worth checking for an unnecessary time sink that forms a bottleneck in their learning process.

- For instance, a student might skip over the lesson and then spend a long time getting through each problem because they didn't read carefully and are trying to solve the problem without relying on any scaffolding.

- Alternatively, a student may spend an excessively long time studying a worked example, hung up on a minor detail or phrase that they feel they are not fully confident in understanding, when it would be more productive to move on to active problem-solving. (Often, actively working through a problem in a slightly different context can clear up minor confusions that may arise when passively viewing a worked example.)

Another possibility:

- A student might make a lot of silly mistakes due to rushing and doing all the work in their head, causing them to have to do many more problems than if they just worked problems out more carefully and accurately on paper.

- On the flipside, a student might spend too much time unnecessarily double and triple-checking against the worked example to make sure they solved the problem correctly. (While it's good to be diligent and work problems out carefully, it's also possible for a student to go overboard and move too slowly because they spend too much time minimizing the risk of getting something wrong.)

Yet another possible cause of taking a long time to earn XP is when a student over-relies on reference material, solving problems alongside worked examples. After a student has read the worked example and moved on to solving problems, the student should refer back to the worked example only as a last resort when they have tried their hardest and failed to remember the next step in the problem-solving process. And even then, whenever a student refers back to the worked example, they should only peek at the part they're stuck on before trying to solve the rest of the problem unassisted.

There are many possible bottlenecks that may occur – too many to list exhaustively. However, in general, it is often possible to increase one's XP per unit time by tracking where the big time sinks are and trying out strategies to speed up those parts of the learning pipeline.

> I don't think XP is a perfect measurement of effort.

The "1 XP = 1 minute of focused effort" metric is an average: when we take a large number of serious students and a large number of tasks and compute the XP per time, it comes out to about 1 XP per minute. While we do our best to assign each learning task an XP that accurately represents the amount of work needed to complete it, the observed XP-to-time ratio may vary for any particular student doing any particular task.

Additionally, we recognize that some components of effort are not currently taken into account when assigning XP, and it is on our roadmap to make XP grading even more granular to incentivize those components of effort. For instance, one thing that we'd like to factor into the XP grading is how often a student refers back to the worked example. Getting questions right with minimal reliance on the worked example is more effortful and will move the needle on a student's learning more than if they refer back to the worked example all the time, by default, instead of trying their best to pull information from memory. This would be part of a larger push on "in-task coaching," that is, encouraging students to engage in micro-behaviors that enhance learning but aren't yet incentivized through our XP system.

| Practice Schedules

> If I have a limited amount of time to devote to Math Academy each week, should I allocate that time into longer, less-frequent sessions or shorter, more-frequent sessions?

When learning math, it's best for study sessions to be short and frequent (as opposed to long and sparse). For instance, suppose you're budgeting 3 hours per week to learn math. It would be better to study 30 minutes six days per week, as opposed to 90 minutes twice a week. There are a handful of reasons why.

1. You want to form a habit. The more consistently you study math, the more it will become a habit that you naturally do each day without thinking, just like (hopefully!) taking a shower and brushing your teeth.

2. You want to operate at peak productivity during your session. During a short 30-minute session, it's easy to maintain a high level of focus and intensity – whereas, during the second half of a long 90-minute session, fatigue will set in and make you significantly less productive.

3. You want to minimize the amount you forget between sessions. When you have multi-day gaps between study sessions, you'll have to spend more time revisiting previously covered material. (Just ask any teacher how much their students forget over weekends, and how much valuable class time they have to spend on Monday re-teaching the things that they covered on Thursday and Friday.)

However, there are some caveats to consider.

- Whenever you switch to a different activity, it takes a few minutes for your brain to catch up and enter a state of flow in the new context. This is called "context switching cost," and if you make your sessions too short (less than 20 minutes or so), then the proportion of study time that is wasted on context switching will outweigh the other benefits of daily practice. Consequently, it's best to spread out your practice as much as possible *subject to the constraint that each session is sufficiently long for the context-switching cost to be proportionally negligible.*

- Additionally, if you have a hectic schedule and "six days per week" in theory ends up being just "three days per week" in practice, then you'll need longer sessions just to achieve the same volume of practice.

> If a student completes a Math Academy course very rapidly, working several hours per day for several weeks, will they still learn the material properly?

Yes, the student will still properly learn and master the material. The only catch is that if they were to flat-out stop doing math afterwards, they would forget the material sooner than if they spread it out over a longer period of time (this is simply a consequence of the spacing effect and the mechanics of spaced repetition, discussed in chapter 17). However, if the student continues working on Math Academy afterwards, continuing to learn more math, then they will layer on top of their knowledge and receive any additional reviews that are necessary to maintain their knowledge, so they won't forget what they've learned.

> Can Math Academy be used for very casual learning, an hour or two per month?

Math Academy focuses on students who are trying to acquire math skills to the highest degree possible. We teach math as if we were training a professional athlete or musician. We maximize learning efficiency in the sense that we minimize the amount of work required to learn math to the fullest extent. Learning math to the fullest extent requires a dedicated effort of at least a couple hours per week.

We realize that there are many learners who only want to devote an hour or two per month, but, at least right now, such learners would be better served elsewhere. It's a totally different optimization problem – maximize surface-level coverage subject to some fixed, miniscule amount of work – and as a result it would require a different curriculum and possibly different training techniques (or at least, differently calibrated techniques).

Diagnostics and Curriculum

| Diagnostics

> There were way more questions than I expected on the diagnostic!

As discussed in Chapter 26, diagnostics require a larger-than-expected number of questions because

1. our curriculum is hyper-scaffolded, and

2. we assess students not only on the course content, but also on any lower-level foundations they might be missing.

For higher-level math courses, diagnostics may need to assess your knowledge of over 1,000 math topics! Even if it feels like there are many questions on the diagnostic, each individual question provides decisive information about your knowledge of about 10 different topics on average.

Think of it like going to a serious gym where you start with a body composition analysis and strength/flexibility tests at every muscle/joint in your body. In order to get you progressing towards your mathematical goals as efficiently as possible, we need to figure out exactly where your strengths and weaknesses are so that we can perfectly calibrate your workouts to your personal needs.

Of course, if you just want to go to the gym on weekends and walk around the track a few times, then this is probably not a great fit! But if you want to come for a serious workout most days each week and reach a serious level of fitness as quickly as possible, then this is the way to do it.

> Can't you improve the diagnostic algorithm to cut down on the number of questions?

We have spent a lot of time optimizing our diagnostic algorithm to be as quick and efficient as possible. There are some hard limits in the physics of how small we can reduce the number of diagnostic questions subject to additional constraints on the precision and robustness of the overall conclusions drawn from that information. Our diagnostics are optimized to the point that if we forcibly cut the number of questions in half, your placement wouldn't match up well enough with your true knowledge frontier, and you'd get frustrated doing tasks that are too easy (or worse, too hard).

Yes, people quit if the diagnostic is too long, but they also quit if they're not placed accurately. Math Academy's approach to that tradeoff caters to students who are serious about putting in a large amount of work to learn an even larger amount of math. For such users, an hour or two on the diagnostic is proportionally negligible compared to the amount of work that they plan to commit to learning math. Such users typically find it worthwhile to spend a proportionally tiny amount of extra time at the beginning to ensure a smooth mathematical journey indefinitely into the future.

> There is a topic that I know how to do, but the diagnostic didn't ask me about it and I didn't get credit for it.

The diagnostic is fully comprehensive; it continues asking questions until it has evidence of knowledge (or lack of knowledge) for every single topic in the student's course and foundations. Whatever topics the student is not given credit for, it's because the student submitted incorrect answers on those topics or their prerequisites. While it is *sometimes possible* to solve questions from a topic despite not fully grasping a prerequisite, this indicates the presence of "holes" in the student's mathematical knowledge, and the diagnostic intentionally places students at the bottom of their lowest knowledge holes so that these holes can be filled in.

Placing students at the bottom of their lowest knowledge holes is absolutely critical to ensure student success. If the diagnostic did the opposite, placing students at the *top* of their *highest* knowledge holes, then students might initially feel like they are closer to their goals as a result of receiving more credit, but these knowledge holes would sooner or later (and likely sooner) derail the student by causing them to become "stuck" while learning new topics that make deeper use of the prerequisite knowledge.

That said, it is not uncommon for adult students to be extremely rusty on their math while taking the initial diagnostic, and then have an outsized portion of their memory come rushing back afterwards as they complete learning tasks. When this happens, it is sometimes possible for a student to place significantly further by retaking the diagnostic.

Additionally, we are working on a button where students can say "I already know this" on any lesson that they receive and evidence their knowledge by answering a couple advanced questions on the topic. That way, it will be fast and easy for a student to continue fine-tuning their knowledge profile after the diagnostic.

> After the diagnostic, what if I am asked to complete a lesson for which I have not learned a prerequisite?

Math Academy's diagnostic exam is highly accurate, but not necessarily 100% perfect, as guaranteeing a 100% perfect placement would require an infeasibly large number of diagnostic questions to be answered. To drastically cut down on the number of questions, our diagnostic exam leverages some loose forms of inference that – rarely, but occasionally – may place a student slightly behind or ahead (but more likely behind) their true knowledge frontier along some learning path.

(For instance, if a student answers a question correctly on a "leaf topic" in some module, then we treat that question as a "representative" for the module and award some credit to other leaf nodes in the same module. Otherwise the diagnostic would have to explicitly assess every single leaf topic, which would make the number of questions blow up. The idea is that if a student knows a maximally advanced technique within some cohesive group of topics, then they probably know any other advanced techniques within that group, or they should have enough prior knowledge to brush up on the fly.)

In practice, on the rare occasion that a student's level of knowledge is overestimated and they receive a lesson for which they have not learned a prerequisite, the degree of overestimation is small enough that the student is able to learn the prerequisite by clicking on the prerequisite and viewing its lesson in reference mode. To the best of our knowledge, there has never been an instance where a student was unable to do so, provided that they took the diagnostic properly and submitted answers reflective of their true knowledge. (In the small number of instances where a student was placed too far beyond their true knowledge to make progress on the system, it has always turned out that the student used an external resource for help during the diagnostic.)

Additionally, even on the rare occasion that a student may have to learn a prerequisite by viewing its lesson in reference mode, this issue will quickly disappear as the student completes more work on the system: the student will quickly reach a point where, for all their available lessons, they have explicitly completed lessons on all the prerequisites.

> After the diagnostic, why did failing a single task bring my progress down multiple percent?

After the diagnostic, there may be pockets of math where the system infers student knowledge but has low confidence in that inference. As the student completes learning tasks, the system will adapt more quickly to the student's performance in these areas of low confidence. As discussed in Chapter 26, these topics are called "conditionally completed" because while the student (just barely) received credit for them based on the diagnostic, retaining this credit is conditional on the student passing tasks that assume knowledge of these topics.

If a student fails a task on a conditionally completed topic, then the system reasons as follows:

> *Wait, we thought they knew that topic and some more advanced topics that build on it – but they just barely placed out of those topics on the diagnostic, so the fact that they're struggling indicates*

> *that they probably don't actually know it. Time to revise our earlier decision and remove the credit we originally awarded.*

This mechanism may be subtle in tasks like multisteps and quizzes, where every question is linked to a different topic, because a student may lose credit for a particular topic (from incorrectly answering a question linked to that topic) while gaining credit for other topics (from correctly answering questions linked to other topics) and passing the task overall.

| Curriculum

> With short lessons, is the curriculum really comprehensive?

Yes, our curriculum is fully comprehensive – in fact, we perform curriculum comparisons against all the major textbooks to ensure that we're covering a superset of the material.

How do we cover all the necessary material if the lessons are so limited in scope? By breaking up each course into many, many lessons. A typical course ranges from 150-300 lessons, each lesson containing about 3-4 "knowledge points" of increasing difficulty, each knowledge point consisting of a worked example followed by 2-5 questions of active problem-solving where the number of questions adapts to the student's performance.

Basically, the "learning staircase" is being chopped up into a massive number of tiny stairs. It still reaches all the way to the top, but the individual stairs are small enough that students don't get stuck unable to climb a stair that's too big for them.

> Will taking a Math Academy course for a standardized exam (e.g., AP Calculus BC) fully prepare a student for the actual exam?

Math Academy courses cover the content knowledge that a student needs to be successful on a standardized test. While this constitutes the vast majority of the work necessary to prepare for a standardized test, it is not fully sufficient. After finishing their Math Academy course, a student must also take a number of practice exams for the specific exam that they are planning to take.

The reason why it's so important to take practice exams is that – in addition to being timed – standardized exams will "package" the content knowledge within various question framings, phrasings, and general contexts that may initially be unfamiliar to the student. While nobody can predict the exact problems that will appear on the exam, students *can* train themselves on the same statistical distribution that the exam problems are going to come from. This is accomplished by working through as many practice exams as possible – ideally real exams from the past, or at least practice exams that come directly from the organization who creates the exam. (If such resources are unavailable, it is critical to acquire practice exams from another organization that has a good reputation for matching up its problem types up accurately against the real exam.)

Whenever a student misses a question on a practice exam (or answers the question with a low degree of confidence), they should refer to the solution, identify their mistake, and immediately work it out again correctly. The next day, they should try working out the same problem unassisted. If they solve it correctly, they should wait another several days before attempting the problem again; otherwise, if their attempt is unsuccessful, they should go back to the beginning of this process (refer to the solution, identify their mistake, immediately work it out again correctly, and re-attempt the next day). This is essentially performing spaced repetition on the student's areas of weakness. This process should continue for multiple rounds through all the practice exams, continuing all the way up until the day before the actual exam.

At the same time, the student should also enable "Test Prep Mode" in their Math Academy course so that they continue receiving reviews on course content instead of being promoted to the next course. It is necessary to continue completing these reviews on Math Academy so that the student does not get rusty on any of the course content. At the time the student takes the exam, they need to be 100% solid and 100% fresh on all their Math Academy course content as well as all the problem types covered in the practice exams.

As a rule of thumb, we recommend that at a minimum, students finish their Math Academy course at least 6 weeks prior to their exam, and go through at least 6 practice tests leading up to the exam. During this time, we recommend about 1.5 hours of prep per day: 30-45 XP of review on Math Academy, and 45-60 minutes taking and re-attempting problems from practice tests (the time it takes to grade the practice test should not be counted towards this time). In the week before the exam, we would recommend increasing the test prep sessions to 2-2.5 hours per day, spending the extra time on practice tests.

The practice tests need to be timed, but they can be broken up into smaller increments. For instance, if an entire test is 2 hours long, then a 30-minute segment can be constructed by doing every problem number that is a multiple of 4. It is important to slice the exam longitudinally like this, as opposed to just taking the first fourth of the exam, because exam problems are often arranged from easy to hard, with hard questions expected to take more time.

Lastly, note that when a student is solid on their course content knowledge and starts taking actual practice exams, they will probably be surprised at how low their score is initially. This is normal and it just takes a bit of exposure to get used to the time limit, question types, and phrasing of the exam questions. Math Academy has extensive hands-on experience preparing students for the AP Calculus BC exam, which is graded on a 1-5 scale (5 being the best), and in our experience, even students who end up getting a 5 often start out getting a 2 or maybe a 3 on their first practice exam. By their second exam, they may get a 3 or 4, and then a solid 4 or maybe just barely a 5 on their third exam, and then a more solid 5 on their fourth exam, and then deeper and deeper into 5 territory on their fifth and sixth practice exams.

> Does Math Academy explain the "why" behind procedures? Are all the concepts taught first?

Math Academy explains the "why" behind procedures all throughout the curriculum. Just to name a particular instance:

- In algebra, when we teach how to solve equations, the very first thing we do is introduce the idea of a solution of an equation: it's just a number that can be substituted for the variable to make the equation come out true. If you have the equation 2x=6, that's just saying "2 times something makes 6," and you don't even need algebra to know that the solution is x=3 (since 2 times 3 makes 6).

- We first give students some practice solving simple equations like that, without algebra, and only afterwards do we start talking about algebraic manipulations like "start with 2x=6, divide both sides of the equation by 2, get x=3." That way, students see algebraic manipulations as an extension of the intuitive reasoning that they were using to begin with.

However, concepts and procedures are intermingled throughout our curriculum. We do not teach "all the concepts first" and then "all the procedures after" because it's not possible to do one in proper depth without the other. Concepts and procedures have a bidirectional, mutually reinforcing relationship. In other words, they build on each other:

- low-level concepts support low-level procedures,
- low-level procedures support higher-level concepts, and
- higher-level concepts support higher-level procedures.

To provide a concrete example:

- a student must be able to count to understand the concept of a number,
- a student must understand the concept of a number to carry out arithmetic procedures, and
- a student must be able to carry out arithmetic procedures to understand the concept of a variable or an algebraic equation.

> Do your university courses have exercises with proofs or is it just computation?

Our Methods of Proof course is proof-based. As of the time of writing (November 2024), our other university courses are computation-based, but that's just because they represent the first course in each subject, whereas a proof-based course would come second. We will eventually be building out those additional proof-based courses, but our current computation-based courses will be prerequisites.

There is sometimes confusion about people thinking Math Academy is stopping at its current level of depth/difficulty, when in reality, we are still building out the curriculum. It is nowhere near finished.

- Does our Methods of Proof course cover epsilon-delta proofs? Yes. Does it cover Papa Rudin? No, because that's well out of scope. Does that mean we're stopping short of Papa Rudin? No, it just means we haven't built up to that yet.

- Another example: we have a computation-based Linear Algebra course that will be a prerequisite for a proof-based Abstract Linear Algebra course later down the road. That second Linear Algebra course will go deeper into the theory and proofs that one might encounter while taking a linear algebra course at an elite university that uses, say, Axler's book.

  Unfortunately, people sometimes move the goalposts and say "your [first] Linear Algebra course is not as intense as Axler," when this isn't even an apples-to-apples comparison. That's like pointing at a high school calculus class and complaining that it's not as intense as a real analysis course – of course it's not! It's a completely different course; in fact, a prerequisite course; and it's not meant to cover the same material.

  Axler is really a second course in linear algebra, even if some universities throw students into it as their first course (which ends up causing a lot of unnecessary struggle). We

> often joke that Axler's book *Linear Algebra Done Right* should really be called *Linear Algebra Done a Second Time.*

In general, we are building our curriculum from the ground up, we are scaffolding everything to the max, and it's mastery-based (students are only asked to learn things after having mastered the prerequisites) – so, naturally, we are going to be doing computation-based versions of courses before proof-based versions.

But that doesn't mean we're not going to be getting the proof-based versions eventually! The proof-based courses are ultimately just different courses, and we are getting the prerequisite courses in place to build up to them. The proof-based courses are well on our roadmap.

> It's hard to believe that 5 hours a week for a year starting from basic multiplication tables will have me completely prepared for university courses. I'd prefer an explanation for people who are not familiar with the XP system.

We realize that this can be a bit shocking and sound too good to be true! For this to feel more realistic, it's important to first understand that not all the math that children cover in school is necessary for our university math courses. (To be clear: our university-level courses are comprehensive and cover all the material that you would expect to see at a top university. They contain a superset of the topics covered in the most widely utilized textbooks.)

As described in Chapter 28, we developed a Mathematical Foundations course sequence specifically for the purpose of getting adults up to speed as quickly and efficiently as possible with all the prerequisites that they would need to know (fractions through calculus) for university-level math courses. The difference between our Foundations sequence and our Traditional sequence is that our Traditional sequence contains a number of topics that are required by school standards but do not actually come up as prerequisite material in university-level math. Roughly a third of topics in the Traditional sequence fall into this category and have been stripped out of the Foundations sequence.

Now, knowing that the Foundations sequence covers about two-thirds the content in the Traditional sequence, the rest of the argument follows from the success of our original in-school program in Pasadena where 6th graders started at various places in Prealgebra, did about 40-50 minutes of fully-focused work per school day for the next 3 years, and covered all of Prealgebra, Algebra 1, Geometry, Algebra 2, Precalculus, and AP Calculus BC, passing the AP exam by the end of 8th grade. While this may also seem shocking, Math Academy has the AP scores to prove it, and there has been plenty of news coverage over the past decade.

Now, look at the numbers: 40-50 fully focused minutes per school day × 180 school days year × 3 years comes out to about 24000 minutes or 400 hours, and our Foundations series is about two-thirds the size of that (since roughly a third of topics are not actually prerequisites for university math), which comes out to about 267 hours. Divide by 52 weeks in a year, and you're at about 5 hours per week.

> Why is "holistic mode" (in which students also fill in any missing knowledge in lower-grade topics that are not prerequisites of their enrolled course) disabled for university courses?

Students are most likely to succeed when they break up long-term goals into short-term goals, maintain momentum, and experience plenty of small wins along the way. Suppose an expert tutor works with a Linear Algebra student who makes the following request:

In addition to helping me out with Linear Algebra and any missing prerequisites, can you fill in all my knowledge gaps in all the math I would be expected to know by now?

In this situation, the expert tutor should try to dissuade the student:

Are you sure you want to do this? It's possible, but I wouldn't recommend it. I will have to assess you on 4 years' worth of math and then teach you whatever you're missing, which will probably be the equivalent of 1 or more full years of math. This is not just an extra 15 minutes on top of each tutoring session. It's going to at least double your workload. And that extra work is not going to get you through Linear Algebra any faster.

Instead of trying to eat the whole elephant in one bite, why don't we just focus on Linear Algebra right now (and whatever math you're missing that's necessary for Linear Algebra) and then fill in the rest of your math background as you move on to other university-level courses that require it.

For instance, I know you're shaky on your calculus, but most of that isn't necessary for Linear Algebra, so instead of trying to build that up now let's wait until you get to Multivariable Calculus and build it up then. It will feel more relevant that way. We can do the same thing with your probability/stats knowledge when you take Probability & Statistics after Multivariable Calculus.

It will be more motivating this way. You'll learn linear algebra in 8 months, multivariable calculus in another 8 months, and probability & statistics in another 8 months, and you'll be filling in your math background all throughout that time. But if we front-load it and fill in all of your math background right now, it will take 14 months just to get through linear algebra. You'll get through multivariable calculus in 5 months after that, and probability & statistics in another 5 months after that, assuming that you don't quit in those initial 14 months.

> *Either way, you'll be at the same point 2 years from now. But it's going to be way more motivating if your wins are spaced 8 months apart, compared to if your first win doesn't happen for an entire 14 months.*

If a student gets as far as university level math but has missing background knowledge, it would be a mistake to front-load that missing knowledge and fill it all in while they complete their first university course. It would reduce friction and increase motivation to instead spread out the work, which is what will happen naturally as the student takes more university courses in non-holistic mode.

Miscellaneous

| Math Academy Itself

> Why use Math Academy instead of self-studying a textbook or free online resource?

Math Academy's key value proposition is that it maximizes every student's learning efficiency. While it is possible for sufficiently motivated students to learn math by self-studying a textbook or free online resource, there are many sources of inefficiency:

- *Not hyper-scaffolded.* Students will periodically run into situations where they are confused about a logical leap that has taken place. It often takes a long time to resolve the confusion and figure out the logical rationale (if the student figures it out at all).

- *Doesn't track student knowledge and implement mastery learning (i.e., does not ensure that the student has mastered the prerequisites before moving on to new material).* Students will feel a large gap between their level of knowledge and the new material, which leads to more confusion and time wasted trying to figure out what prerequisite knowledge they are missing and how to learn it. Often, students will be unable to pinpoint all their missing prerequisite knowledge and will consequently be unable to fully grasp new material, even if they grasp it partially.

- *No spaced review.* Students will quickly become rusty on the material that they learn. Not only will students come out of their course of study having forgotten much of the content, but even during the course, they will constantly be forgetting the prerequisites for new material that they attempt to learn.

- *Doesn't adapt to the student's level of performance*. Students waste a lot of time doing the wrong amount of work. Sometimes a student will grasp a topic quickly and do far more practice than is necessary; other times they will struggle with a topic and not get enough practice to reach mastery.

- *Leaves the definition of mastery open to interpretation by the learner.* It is difficult for a student to know when they have mastered a topic well enough to continue moving forward. Even in good faith, students often think that they have learned a topic well enough when they actually haven't – and they will not realize this unless their mastery is being evaluated by an expert. On the flipside, students can also take things too far in the way of perfectionism, spinning their wheels on the same topic for days or when there is a minor point that doesn't make perfect intuitive sense, when it would be more productive to keep moving forward and solidify their understanding by building on top of it.

This list could be continued endlessly with other items discussed earlier in the body of this book, but the point is that all of these sources of inefficiency introduce unproductive friction into the learning process, lowering a student's educational progress per unit time and effort that they put towards learning.

Math Academy removes as much of this learning friction as possible, maximizing student learning efficiency. That is our main proposition: sure, it's possible to learn math elsewhere, but it's way more efficient with Math Academy.

It's worth noting that efficiency is important not only because students make faster progress, but also because they are less likely to quit. Typically, people get off the train and stop learning math once it begins to feel too inefficient relative to other opportunities in life. In anything one does, once the progress-to-work ratio becomes too low, one will lose interest and focus on other endeavors where their progress-to-work ratio is higher. Efficiency keeps that progress-to-work ratio as high as possible, keeping students on the math learning train as long as possible.

> Why isn't Math Academy free?

Math Academy requires payment because it takes so much time & effort to build. Additionally, it must be priced in a way that the company's solvency is not dependent on a massive user base.

Math Academy is intent on using the most effective training techniques, but most people are not *that* serious about their learning. Maximum-efficiency learning feels like a sweaty, exhausting workout with a personal trainer, for at least several hours spread across several sessions each week.

When an education company depends on a massive base of learners, most of whom are not serious enough to engage in that level of intensity and frequency in their training, it requires the company to employ ineffective learning strategies that do not repel unserious students. The company must convince their students that they've managed to learn things despite putting in little to no work. (This can be accomplished, for instance, by cherry-picking the simplest cases of each topic and letting students move on despite poor performance on prerequisite material.) Unlike such companies, Math Academy is in the business of optimizing real learning – not just the perception of it – for students who are willing to put in serious work.

At the same time, of course, we do want to make mathematical talent development accessible to more and more people. As discussed in Chapter 1: before Math Academy, if a student wanted to replace their traditional schooling with the equivalent duration of 1-on-1 coaching from a personal trainer who develops their mathematical talent using a personalized training program that is tailored and constantly adapting to their individual needs, they would have to obtain it from a private tutor for a typical price of at least \$50/hour. Year-round talent development, with a daily work time that is in line with the amount of time that students would be working anyway during the school year (conservatively, 1 hour per weekday) would cost \$50 × 5 days/week × 52 weeks/year = \$13,000/year.

Bringing that figure down to \$499/year (26x cheaper) via Math Academy makes mathematical talent development accessible to many, many more people. That's not everyone, and there are still people who are priced out, but providing a 26x cheaper option is a good starting point towards a goal of making mathematical talent development accessible to more and more people.

> Where do the exercises and content on Math Academy come from? Are they all made in-house or pulled from other materials?

All of our content and exercises are created in-house, carefully crafted over many years by a team of math experts. We perform curriculum comparisons to ensure that our content is comprehensive, but everything on Math Academy has been carefully crafted by an expert human.

Why generate questions manually instead of algorithmically? Beyond simple arithmetic, there is so much dimensionality and nuance in math questions that generating questions algorithmically would require a custom algorithm for each question type, which would take far longer to build than just biting the bullet and manually generating a sufficiently large pool of questions. Furthermore, a sufficiently large pool of questions does not have to be *that* large: because Math Academy students work hyper-efficiently, engaging in minimum effective doses of practice across a highly segmented curriculum, we don't actually need that many questions within each question type to ensure that students are not seeing the same questions over again. About 20 questions per knowledge point is more than sufficient. (A typical course consists of about 500-1,000 knowledge points, requiring a pool of about 10,000-20,000 questions.)

Additionally, all of the information in our knowledge graph – tens of thousands of prerequisite links, estimates for the time it takes to work out questions, and encompassing relations (i.e., how much "credit" should a simpler topic get for doing an advanced topic where the simpler topic is a component skill), have also been carefully crafted by expert humans.

If that sounds like a ridiculous amount of work, then, well... that's about right and it gives you an idea of why this kind of system is so difficult to build. Even without all the fancy tech, the amount of content that's needed is enough to form the basis of a full-fledged publishing company. You could say the same about the software: even without the core base of content, the amount of software that's needed is enough to form the basis of a full-fledged tech company. Even individual components of the software – the content management system, the student interface, the expert system (i.e. AI system) that makes all the complicated behind-the-scenes decisions regarding what the student needs to work on – could each on their own form the basis of a full-fledged company.

| Features that Do Not Exist for a Good Reason

> I don't really want to do any of the learning tasks that Math Academy presents to me. There are other topics I would rather learn. Why can't I choose my own tasks?

Math Academy's main value proposition is maximizing student learning efficiency. That is our top priority. When a student signs up for Math Academy, we are making a promise to them that their learning experience is going to be as efficient as possible. The student is going to learn the most math possible in the time that they're devoting to study.

In order to keep good on that promise, we have to use a lot of sophisticated algorithms to analyze the student's knowledge profile and select their tasks. The whole system has been built around that concept.

We do have some ideas for features that will give students more agency over what they're learning, but it's going to take some work because we have to be careful not to allow students to make decisions that throttle their learning efficiency. The approach that we've been thinking about is less like "select whatever topic you want at any time" and more like "tell us what your specific goal is and we'll put you on the most efficient path to that goal." Of course, none of that is fully-baked yet, but it's something that's on our mind and that we're working on.

> Why can't I edit my knowledge profile?

Learners have a tendency to massively overestimate self-reported knowledge, and then fault the resulting instruction for moving too quickly, not explaining enough, or otherwise being too challenging, when the issue is really that they lack sufficient mastery of prerequisites. To construct an accurate knowledge profile, the system must infer it from a student's demonstrated ability to solve problems.

> Why isn't there an "I don't know" button on questions during tasks other than the diagnostic?

If a student ever gets stuck during a lesson, then they always can and should go back to the preceding worked example and follow along carefully to identify what they missed when they read it the first time. If a student ever gets stuck during a review or multistep, then they always can and should go back to the corresponding lesson in reference mode. The information that a student needs to solve the problem should always be there.

For instance, review questions are pulled from the same exact pool of questions that a student might see during the lesson. So, any review question that a student receives will line up with one of the question types covered in a worked example in the lesson. (To be clear: the review question will be chosen as a different question within that pool – the student will not be given a review question that is exactly the same as one they already answered during the lesson.)

A student should never be in a position where they are being asked to perform a skill that's not covered in the supporting lesson, so we do not wish to communicate otherwise (which is what an "I don't know" button would do).

That said, it is true that students should not refer to reference material during quizzes, so why isn't there an "I don't know" button on quizzes? This is something we considered when first implementing quizzes – but, working with a large number of students, we've experienced that many students will abuse an "I don't know" button if it's provided.

- This can be intentional, e.g., adversarial students (especially kids who are using the system for school and have a mentality that is not fully aligned with the learning process) will click "I don't know" simply to avoid doing work. When we first deployed the automated system in school classes, there was a period of time where the system was getting attacked left and right by adversarial students trying to game the system (or otherwise create chaos that they could leverage to confuse their parents and get out of doing work). It took a lot of effort to patch up exploits, and whenever we make adjustments to the system, we're always on the lookout for any ways that it can be exploited (because if it can, then it will, and the behavior will spread).

- Or it can be unintentional, e.g., underconfident learners may underestimate their ability and give up too early. When a tutor is working with a student on a problem, it is not uncommon that a student will claim not to know how to do the problem, but when the tutor asks the student to make their best guess, the "guess" is correct – and when the tutor asks the student about their thought process afterwards, it turns out that the student knew how to solve the problem, but they weren't confident about it and they didn't want to risk getting it wrong.

While it may be subtle, removing the "I don't know" button is a crucial safeguard to protect many students against self-destructive behavior. In theory, no such safeguards should be necessary, but in practice, a vital component of a functioning learning system is that it must be robust to all sorts of unexpected behavior arising from the various human emotional experiences associated with learning and intense training. Often, these emotional experiences can be intense and (if the option is provided) lead people to make short-sighted decisions that ultimately hinder their educational progress.

> Why doesn't Math Academy use large language models (LLMs) to engage students in conversational dialogue?

Many people who have (unsuccessfully) attempted to apply AI to education have focused too much on the "explanation" part and not enough on scaffolding, navigating, and managing the entire learning process.

It's easy to go on a wild goose chase building an explanation AI. You fall in love with the idea of AI having conversational dialogue with students, and then you get lost in the weeds of complexity. You solve just enough of the problem to produce a cool demo, yet you're still hopelessly far away from self-service learning in real life.

Dialogue isn't even necessary. We simply hardcode explanations into bite-size pieces, served at just the right moment. And we close the feedback loop by having students solve problems, which they need to do anyway. (Their "response" is whether they got it correct.)

Sure, hard-coding explanations feels tedious, takes a lot of work, and doesn't have produce the same "wow" factor as an AI that generates responses from scratch – but it's a practical solution that lets us move on to other components of the AI that are just as important (i.e., the entirety of this book). Just to name a few such components:

- After a minimum effective dose of explanation, the AI needs to switch over to active problem-solving. Students should begin with simple cases and then climb up the ladder of difficulty, covering all cases that they could reasonably be expected to solve on a future assessment.

- Assessments should be frequent and broad in coverage, and students should be assigned personalized remedial reviews based on what they answered incorrectly.

- Students should progress through the curriculum in a personalized mastery-based manner, only being presented with new topics when they have (as individuals, not just as a group) demonstrated mastery of the prerequisite material.

- Students should progress through the curriculum in a personalized mastery-based manner, only being presented with new topics when they have (as individuals, not just as a group) demonstrated mastery of the prerequisite material.

- After a student has learned a topic, they should periodically review it using spaced repetition, a systematic way of reviewing previously-learned material to retain it indefinitely into the future.

- If a student ever struggles, the system should not lower the bar for success on the learning task (e.g., by giving away hints). Rather, it should strengthen a student's area of weakness so that they can clear the bar fully and independently on their next attempt.

> If I get stuck, is there somewhere that I can ask for help or a further explanation?

We don't offer human tutoring services. However, we've been quantitatively analyzing and refining our content for years, smoothing out sections where anyone has struggled. Thousands of learners have successfully made it through our courses, and on average, students pass lessons 95% of the time on the first try and 99% of the time on the second try.

Furthermore, on the rare occasion that a learner does happen to get stuck and fail a lesson twice in the same place, the system will automatically have them review the prerequisite knowledge that is most relevant to their area of struggle before having the student re-attempt the lesson.

Glossary

abstraction – the ability to see "the forest for the trees" by learning underlying rules as opposed to memorizing example-specific details. Abstracting underlying rules improves one's ability to extrapolate knowledge to new contexts. Abstraction ability is known to vary among individuals and depends on **working memory capacity**.

abstraction ceiling – a practical limitation on the amount of math that one can learn, resulting from the phenomena that people have limited abstraction abilities and that higher levels of math become increasingly abstract and technical. The abstraction ceiling is not a "hard" threshold, a level at which one is suddenly incapable of learning math, but rather a "soft" threshold, a level at which the amount of time and effort required to learn math begins to skyrocket until learning more advanced math is effectively no longer a productive use of one's time. That level is different for everyone.

academic acceleration – the practice of allowing students to learn academic material at a younger age and/or faster rate than is typical.

active learning – learning in which students are actively performing learning exercises as opposed to passively consuming educational content. The most effective active learning technique is **deliberate practice**.

automaticity – the ability to execute low-level skills without having to devote conscious effort towards them. Automaticity is necessary because it frees up limited **working memory** to execute multiple lower-level skills in parallel and perform higher-level reasoning about the lower-level skills. When you develop automaticity on a skill or piece of information, it is stored in your **long-term memory**, where indefinitely many things can be held for indefinitely long without requiring cognitive effort.

blocked practice (**blocking** or **massed practice**) – a type of practice in which a single skill is practiced many times consecutively. While some initial amount of blocking is useful when first learning a skill, blocking is very *inefficient* for building long-term memory afterwards during the review stage. See also: **interleaving**.

category induction learning – recognizing general features that distinguish problems requiring different solution techniques. See also: **discrimination learning**.

chunk – a coherent, meaningful group of related pieces of information.

cognitive control – see **executive function**.

cognitive load – the amount of **working memory** required to complete a task.

cognitive overload – when the **cognitive load** of a task exceeds one's **working memory**, they are not able to complete the task.

conditional completion – if the evidence balances out to "just barely" place a student out of some topics, the system will consider those topics **conditionally completed**: the student will initially be given tasks under the assumption that they know those topics, but if the student struggles, then the system will immediately begin "falling backwards" along the appropriate learning paths.

consolidation – the process of storing new information in **long-term memory**.

core topics – topics that are most relevant in the "big picture" of mathematics. For instance, in calculus, the product rule would be a **core topic**, while the intermediate value theorem would be a **supplemental topic**. Core topics are the ones that appear more frequently as prerequisites of other topics in mathematics.

course graph – a highly-compressed version of a **knowledge graph** where a single entity represents hundreds of topics. It is important to realize that each course is ultimately just a set of topics in the knowledge graph. The knowledge graph is the ultimate source of truth; a course graph simply summarizes and communicates information about the high-level structure of a knowledge graph so that humans can understand it.

deliberate practice – individualized training activities that are specially chosen to improve specific aspects of one's performance through repetition and successive refinement. Deliberate practice is the opposite of mindless repetition, and it has been shown to be one of the most prominent underlying factors responsible for individual differences in performance, even among highly talented elite performers.

desirable difficulty – a practice condition that makes the task harder, slowing down the learning process yet improving recall and transfer. Desirable difficulties make practice more representative of true assessment conditions.

diagnostic – an adaptive exam that leverages the **knowledge graph** to quickly identify a student's knowledge frontier.

direct instruction – instruction that teaches knowledge to students explicitly as opposed to attempting to have students "construct their own knowledge" through unguided activities.

discrimination learning – matching problems with the appropriate solution techniques. For instance, the equations $x^2 + 3x + 2 = 0$ and $x + 3x + 2 = 0$ look similar but require wildly different solution techniques. See also: **category induction learning**.

distributed practice – see **spaced repetition**.

dual-coding theory – a theory of cognition in which the mind processes information along two different channels: verbal and visual. Instructional materials can help students avoid **cognitive overload** by distributing cognitive load more evenly between these two channels.

Ebbinghaus – known for discovering the **spacing effect**.

edge of mastery – see **knowledge frontier**.

effect size – when a group of students undergoes an intervention that is intended to improve learning, the effect size measures the degree of improvement relative to a *control group*, a group of students who did not receive the intervention. Specifically, effect size is calculated as the number of standard deviations (also called *sigmas*) by which the mean performance increases. For instance, if an intervention increases the average exam score by 20%, and the standard deviation of exam scores is 10%, then the effect size is 20% / 10% = 2 sigmas. Effect sizes can also be reported in percentiles: an effect size of 2 sigmas indicates that the average student who experienced the intervention learned more than 98% of students in the control group. (For 1 sigma, the corresponding percentile is 84%.)

encoding – the interpretation of the brain's information-processing pipeline that emphasizes that the pipeline converts or "encodes" information from the outside world into a representation that can be stored in **long-term memory** and later recalled.

encompassing – advanced mathematical problems implicitly practice or "encompass" many simpler skills. Using sophisticated algorithms that capitalize on these encompassings, we enable students to spend most of their time learning new material while simultaneously making sure they keep getting practice on things they've previously learned. This results in turbo-boosted learning speed.

equivalent topic – to comply with course standards, it is sometimes necessary to have **equivalent topics** spread out across multiple courses, with the equivalent topics in higher courses covering a more advanced treatment of the same skills taught in lower courses.

executive function (**cognitive control**) – the interpretation of the brain's information-processing pipeline that emphasizes that the pipeline is centered around **working memory**, which pulls relevant information from **sensory** and **long-term memory** into an area where it can be combined, transformed, and used to guide behavior to achieve goals.

eXperience Points – see **XP**.

expertise reversal effect – the instructional techniques that promote the most learning in beginners, promote the least learning in experts, and vice versa.

facilitation – when a new task exercises knowledge learned in a prior task, learning can be facilitated in two ways:

- (*Retroactive Facilitation*) The new task can restore memory of prior knowledge to the same extent as identical repetition of the prior task, leading to long-lasting retention (Ausubel, Robbins, & Blake, 1957; Arzi, Ben-Zvi, & Ganiel, 1985).
- (*Proactive Facilitation*) Knowledge acquired during the prior task can improve the acquisition of knowledge that is specific to the new task (Arzi, Ben-Zvi, & Ganiel, 1985).

forgetting curve – a graph of memory versus time that shows memory decaying over time.

foundations (**foundational knowledge**) – lower-grade topics that students need to know in order to succeed in their enrolled course (i.e. foundations are prerequisites for the course).

Fractional Implicit Repetition (**FIRe**) – Math Academy's novel spaced repetition model that generalizes spaced repetition from independent flashcard-like tasks to highly connected bodies of knowledge where repetitions on advanced topics should "trickle down" to update the repetition schedules of simpler topics that are implicitly practiced.

frontier – see **knowledge frontier**.

gamification – when game-like elements (such as points and leaderboards) are integrated into student learning environments in ways that are aligned with the goals of the course, the motivations of the students, and the context of the educational setting, students typically not only learn more and engage more with the content, but also enjoy it more. However, these gamified elements need to be resistant to "hacking" behaviors that attempt to bypass learning by exploiting loopholes in the rules of the game.

growth mindset – the belief that one's current level of knowledge is not "fixed" or set in stone, but rather, can be increased through practice. To maximize student growth, it is necessary to give each student enough practice to achieve mastery and allow them to move on to more advanced skills immediately after mastering the prerequisites. The necessary amount of practice to achieve mastery will vary depending on the particular student and the particular skill.

illusion of competence / comprehension – it is easy for students (and their teachers) to vastly overestimate their knowledge if they do not leverage **desirable difficulties** during practice.

induction learning – see **category induction learning**.

interference – see **non-interference**.

interleaving (**varied practice, mixed practice**) – the effectiveness of practice is diminished when a single skill is practiced many times consecutively beyond a minimum effective dose. Review problems should be spread out or *interleaved* over multiple review assignments that each cover a broad mix of previously-learned topics. In addition to being more efficient, this also helps students match problems with the appropriate solution techniques (**discrimination learning**) and recognize general features that distinguish problems requiring different solution techniques (**category induction learning**).

key prerequisite – each **knowledge point** is linked to one or more **key prerequisite** topics that represent the prerequisite knowledge that is most directly being used in that knowledge point. If a student ever fails a lesson twice at the same knowledge point, we automatically provide remedial reviews on the key prerequisites. This helps the student strengthen their foundations in the areas where they are most in need of additional practice, so that they are better prepared to pass the lesson the next time around.

knowledge frontier (**edge of mastery**) – the boundary between what a student knows and does not know, which indicates what topics they are ready to learn. Following a student's initial **diagnostic**, whenever a student is served new lessons, those lessons always cover topics that are on the student's knowledge frontier. See also: **zone of proximal development**.

knowledge graph – organizes our curriculum in a way that enables algorithmic decision-making. Contains multiple thousands of interlinked topics, with each linkage between topics indicating a relationship between them (such as one topic being a prerequisite for another topic). Knowledge graphs can encode a lot of complicated information that would otherwise be hard to describe and reason about.

knowledge point – each topic involves a lesson that is broken down into several key pieces of learning called **knowledge points**. Each knowledge point contains a worked example and asks questions similar to the worked example. Knowledge points build on each other to help scaffold students through the lesson: the first knowledge point covers the most basic idea or skill of the lesson, and later knowledge points gently introduce more advanced cases. To demonstrate mastery of a topic, a student must answer sufficiently many questions correctly (with sufficiently few mistakes) in each successive knowledge point in the lesson. Once this is accomplished, more advanced topics become available for the student to work on.

knowledge profile – measures a student's knowledge at every topic in a **knowledge graph**. Loosely speaking, a student's knowledge profile represents how "developed" their mathematical brain is. Every time they learn a new math topic, it's as if they grow a new brain cell and connect it to existing brain cells. Initially, this new brain cell is weak and requires frequent nurturing, but over time it becomes strong and requires less frequent care. See also: **knowledge frontier**.

layering – learning is about making connections: the more connections there are to a piece of knowledge, the more ingrained, organized, and deeply understood it is, and the easier it is to recall. The most efficient way to increase the number of connections to existing knowledge is to continue layering on top of it – that is, continually acquiring new knowledge that exercises prerequisite or component knowledge.

league – to incentivize students to put forth a sufficient quantity of work, we implemented (optional) **competitive weekly leaderboards** where students are grouped into smaller **leagues** with other students of similar competitive ability. If a student earns enough XP to end the week near the top of their league, they promote to a higher league. But if they end the week near the bottom of their league, they demote to a lower league.

learning – a positive change in **long-term memory**. At a physiological level, learning involves the creation of strategic electrical wiring between **neurons** ("brain cells") that improves the brain's ability to perform a task.

learning efficiency – the amount of progress that a student makes in their course, relative to the amount of time that they spend working.

learning rate – rate at which one's ability to perform a task improves over the course of exposure, instruction, and practice on the task. Learning rate is known to vary among individuals and depend on **working memory capacity** (**WMC**).

long-term memory – effortlessly holds indefinitely many facts, experiences, concepts, and procedures, for indefinitely long, in the form of strategic electrical wiring between **neurons**. Wiring induces a "domino effect" by which entire patterns of neurons are automatically activated as a result of initially activating a much smaller number of neurons in the pattern. See also: **consolidation**.

massed practice – see **blocked practice**.

mastery – to demonstrate **mastery** of a topic, a student must answer sufficiently many questions correctly (with sufficiently few mistakes) in each successive **knowledge point** in the lesson. Once this is accomplished, more advanced topics become available for the student to work on.

mastery floor – lower-course topics that are automatically considered mastered by any student taking the course.

mastery learning – each individual student needs to demonstrate proficiency on prerequisite topics before moving on to more advanced topics. True mastery learning at a fully granular level requires fully individualized instruction, which is only attainable through one-on-one tutoring.

Mathematical Foundations (**MF**) **sequence** – a streamlined sequence of courses that covers elementary mathematics through calculus but cuts out roughly a third of topics that are not actually prerequisites for university math.

mixed practice – see **interleaving**.

narrow limits of change principle – The severe limitation of the **working memory** when processing novel information. Most people can only hold about 7 digits (or more generally 4

chunks of coherently grouped items) simultaneously and only for about 20 seconds. And that assumes they aren't needing to perform any mental manipulation of those items – if they do, then fewer items can be held due to competition for limited processing resources.

neuromyth – a common yet scientifically inaccurate misunderstanding about the brain. Neuromyths can often be characterized as the oversimplification, misinterpretation, and/or misapplication of a nuanced complex scientific finding. One of the most widespread – and most widely debunked – neuromyths is that people learn better when they receive information in their preferred "learning style."

neuron – a cell that transmits information through electrical activity. The brain is a gigantic network of roughly 100 billion neurons that are "wired up" with over 100 trillion connections between them.

non-ancestor encompassing – even though simple **equivalent topics** would not be ancestors of advanced equivalent topics via **direct prerequisite** or **key prerequisite** paths, we can still set full-encompassing edge weights between them so that a student who completes an advanced topic will implicitly receive credit for any simpler equivalent topics as well. These are called **non-ancestor encompassings**. See also: **encompassing**.

non-interference – conceptually related pieces of knowledge should be spaced out over time so that they are less likely to interfere with each other's recall. New concepts should be taught alongside dissimilar material.

pace – the amount of XP that a student completes per weekday (on average).

partial encompassing – an **encompassing** where only some part of the simpler topic is practiced implicitly in the more advanced topic. Partial encompassings occurs more frequently in higher-level math.

placement – see **diagnostic**.

proactive facilitation – see **facilitation**.

radical constructivism – a philosophy in which knowledge does not reflect an "objective" ontological reality, but exclusively an ordering and organization of a world constituted by our experience.

reconsolidation – the process of updating information that has already been **consolidated** into **long-term memory**.

rehearsal – the process of activating neural patterns and persistently maintaining their simultaneous activation, by which the brain holds information in **working memory**.

repetition – see **spaced repetition**.

repetition compression – whenever a student has due reviews, Math Academy is able to compress them into a much smaller set of learning tasks that implicitly covers (i.e. provides repetitions on) all of the due reviews. This is accomplished by choosing reviews whose implicit repetitions "knock out" other due reviews (like dominos).

retrieval practice – see **the testing effect**.

retroactive facilitation – see **facilitation**.

scaffolding – support given to a student to reduce the cognitive load of a learning task.

schema – the underlying structure or framework of one's knowledge.

sensory memory – temporarily holds a large amount of raw data observed through the senses (sight, hearing, taste, smell, and touch), only for several seconds at most, while relevant data is transferred to **short-term memory** for more sophisticated processing.

short-term memory – has a much lower capacity than **sensory memory**, but can store information about ten times longer.

sigma – see **effect size**.

standard deviation – see **effect size**.

spacing effect – when reviews are spaced out or *distributed* over multiple sessions (as opposed to being crammed or *massed* into a single session), memory is not only restored, but also further **consolidated** into long-term storage, which slows its decay. A profound consequence of the spacing effect is that the more reviews are completed (with appropriate spacing), the longer the memory will be retained, and the longer one can wait until the next review is needed. This observation gives rise to a systematic method for reviewing previously-learned material called **spaced repetition** (or **distributed practice**).

spaced repetition (**distributed practice**) – reviews should be spaced out or *distributed* over multiple sessions (as opposed to being crammed or *massed* into a single session) so that memory is not only restored, but also further consolidated into long-term storage, which slows its decay. A **repetition** is a successful review at the appropriate time.

spaced retrieval practice – an especially potent learning strategy that combines spaced repetition with retrieval practice by testing (instead of simply re-studying a reference) during reviews.

speed of learning – see **learning rate**.

spreading activation – a method by which connections between information can be used to recall information in response to a stimulus. The stimulus activates some piece(s) of information, and the activity flows through connections to other pieces of information.

spiral curriculum – a curriculum in which material is naturally revisited and further built upon in later textbook chapters and/or grades.

subgoal labeling – the act of grouping steps into meaningful units with labels. Subgoal labeling can help students grasp the structure of the problem, thereby enabling the learning to transfer to other problems in the same category, and minimize the number of **chunks** of information that they need to store in their working memory, thereby reducing cognitive load.

supplemental diagnostic – as topics are added and connectivity is revised in the knowledge graph, the knowledge profile inferred from a student's initial placement **diagnostic** can get a little out of date. When this happens, we assign tiny diagnostics called **supplemental diagnostics** to bring the student's knowledge profile back up to date. See also: **diagnostic**.

supplemental topic – see **core topic**.

targeted remediation – in the academic literature, the term *targeted remediation* usually describes identifying individual students in need of broad remedial intervention such as tutoring, remedial courses, academic advisor meetings, etc. But in the context of Math Academy, **targeted remediation** refers to fully-automated support mechanisms that are targeted to individual students *on individual topics* – and often even more precisely to the individual component skills that are causing a student to struggle on a topic.

testing effect (**retrieval practice**) – to maximize the amount by which your memory is extended when solving review problems, it's necessary to avoid looking back at reference material unless you are totally stuck and cannot remember how to proceed. For this reason, it's necessary to test frequently as a part of the learning process itself.

topic – see **knowledge graph** and **knowledge point**.

tragedy of the commons – in the absence of accountability and incentives that promote collective interests, people will focus on behaviors that benefit themselves as individuals, and pay less attention to how their actions affect the group as a whole. As a result, when a group is given responsibility for the maintenance and improvement of a shared resource, the resource will typically deteriorate. While some individuals may care for the resource properly, they are typically unable or unwilling to pick up the slack of those who do not.

two-sigma problem – In 1984, educational psychologist Benjamin Bloom published a landmark study comparing the effectiveness of one-on-one tutoring and traditional classroom teaching. The difference was monumental: the average tutored student performed better than 98% of the students in a traditional class. This finding led to a challenge widely known as *Bloom's two-sigma problem:* can we develop methods of group instruction that are as effective as one-on-one tutoring? (The terminology "two-sigma" comes from statistics, where the effects of interventions are often measured in standard deviations or *sigmas*. An **effect size** of 98% is slightly more than two sigmas.)

varied practice – see **interleaving**.

worked example – a problem along with a step-by-step-demonstration of how to solve it. See also: **knowledge point**.

working memory – consists of **short-term memory** along with capabilities for organizing, manipulating, and generally "working" with the information stored in short-term memory. See also: **cognitive load**, **cognitive overload**, **narrow limits of change principle**, and **rehearsal**.

working memory capacity (**WMC**) – the maximum amount of information that one can hold and manipulate in working memory. Working memory capacity is known to vary between individuals and is known to influence perceived effort, cognitive control, mind-wandering,

abstraction ability, learning outcomes, and **learning rate**. See also: **narrow limits of change principle**.

XP (**eXperience Points**) – the currency of Math Academy's gamified reward system. Students earn XP upon successful completion of learning tasks, and XP is calibrated so that 1 XP represents 1 minute of fully-focused, fully-productive work for an average serious (but imperfect) student.

zone of proximal development – the range of tasks that a student is able to perform while supported, but cannot do on their own. Students maximize their learning when they are completing tasks within this range.

References

Agarwal, P. K., & Agostinelli, A. (2020). Interleaving in Math: A Research-Based Strategy to Boost Learning. *American Educator, 44*(1), 24.

Agarwal, P. K., D'antonio, L., Roediger III, H. L., McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of applied research in memory and cognition, 3*(3), 131-139.

Agarwal, P. K., Roediger, H. L., McDaniel, M. A., & McDermott, K. B. (2013). How to use retrieval practice to improve learning. *Washington University in St. Louis.*

Allen-Lyall, B. (2018). Helping students to automatize multiplication facts: A pilot study. *International Electronic Journal of Elementary Education, 10*(4), 391-396.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of experimental child psychology, 106*(1), 20-29.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological review, 94*(2), 192.

Anderson, J. R., Reder, L. M., Simon, H. A., Ericsson, K. A., & Glaser, R. (1998). Radical constructivism and cognitive psychology. *Brookings papers on education policy,* (1), 227-278.

Arzi, H. J., Ben-Zvi, R., & Ganiel, U. (1985). Proactive and retroactive facilitation of long-term retention by curriculum continuity. *American educational research journal, 22*(3), 369-388.

Atchison, K. (2020). True Mastery: The Benefits of Mixed Practice for Learning. *EducationalRenaissance.com.*

Ausubel, D. P., Robbins, L. C., & Blake Jr, E. (1957). Retroactive inhibition and facilitation in the learning of school materials. *Journal of Educational Psychology, 48*(6), 334.

Ayres, P. L. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology, 26*(2), 227-248.

Bacon, F. (1620). The new organon: Or true directions concerning the interpretation of nature.

Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 302*(1110), 311-324.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4*(5), 316-321.

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review, 30*, 100322.

Baker, R. S. D., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., ... & Beck, J. E. (2006). Adapting to when students game an intelligent tutoring system. In *Intelligent Tutoring Systems: 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26-30, 2006. Proceedings 8* (pp. 392-401). Springer Berlin Heidelberg.

Bernstein, B. O., Lubinski, D., & Benbow, C. P. (2021). Academic acceleration in gifted youth and fruitless concerns regarding psychological well-being: A 35-year longitudinal study. *Journal of Educational Psychology, 113*(4), 830.

Berzsenyi, G. (2019). Talent search versus talent development. *Notices of the American Mathematical Society, 66*(9).

Betts, K., Miller, M., Tokuhama-Espinosa, T., Shewokis, P. A., Anderson, A., Borja, C., ... & Dekker, S. (2019). International Report: Neuromyths and Evidence-Based Practices in Higher Education. *Online Learning Consortium.*

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society, 2*(59-68).

Bjork, E. L., & Bjork, R. A. (2023). Introducing Desirable Difficulties Into Practice and Instruction: Obstacles and Opportunities. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 111-21). Society for the Teaching of Psychology.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205).

Bloom, B. S. (1968). Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation comment, 1*(2), n2.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher, 13*(6), 4-16.

Bloom, B. S., ed. (1985). *Developing Talent in Young People.* New York: Ballantine Books.

Bloom, B. S. (1986). Automaticity: "The Hands and Feet of Genius." *Educational leadership, 43*(5), 70-77.

Bloom, B. S., & Sosniak, L. A. (1981). Talent development vs. schooling. *Educational Leadership, 39*(2), 86-94.

Bloom, G. (2002). Role of the elite coach in the development of talent. *Psychological foundations of sport*, 466-483.

Borland, J. H. (1989). *Planning and Implementing Programs for the Gifted.* New York: Teachers College Press.

Brandt, R. S. (1985). On Talent Development: A Conversation with Benjamin Bloom. *Educational Leadership, 43*(1), 33-35.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly journal of experimental psychology, 10*(1), 12-21.

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Harvard University Press.

Buskist, W., Cush, D., & DeGrandpre, R. J. (1991). The life and times of PSI. *Journal of Behavioral Education, 1*, 215-234.

Cacciola, S. (2020). How Kobe Bryant Created His Own Olympic Dream Team. *New York Times.*

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral cortex, 15*(8), 1243-1249.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current biology, 16*(19), 1905-1910.

Campbell, J. I. (1987). The role of associative interference in learning and retrieving arithmetic facts. *Cognitive processes in mathematics*, 107-122.

Carpenter, S. K., & Agarwal, P. K. (2019). How to use spaced retrieval practice to boost learning. *Iowa State University.*

Catrambone, R. (1995). Aiding subgoal learning: Effects on transfer. *Journal of educational psychology, 87*(1), 5.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental psychology, 56*(4), 236-246.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In *Psychology of learning and motivation* (Vol. 16, pp. 1-58). Academic Press.

Clark, R., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator, 36*(1), 5-11.

Clear, J. The Beginner's Guide to Deliberate Practice. *JamesClear.com.*

Codding, R. S., Peltier, C., & Campbell, J. (2023). Introducing the Science of Math. *TEACHING Exceptional Children*, 00400599221121721.

Codding, R. S., VanDerHeyden, A., & Chehayeb, R. (2023). Using Data to Intensify Math Instruction: An Evaluation of the Instructional Hierarchy. *Remedial and Special Education*, 07419325231194354.

Conway, A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. (2007). Variation in Working Memory: An Introduction. In Conway, A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. (Eds.), *Variation in working memory* (pp.3-17). Oxford University Press.

Coughlan, E. K., Williams, A. M., McRobert, A. P., & Ford, P. R. (2014). How experts practice: A novel test of deliberate practice theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 449.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences, 24*(1), 87-114.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research, 169*, 323-338.

Cross, K. P. (1999). Learning is about making connections. *The Cross Papers; 3*.

Dallimore, E. J., Hertenstein, J. H., & Platt, M. B. (2013). Impact of cold-calling on student voluntary participation. *Journal of Management Education, 37*(3), 305-341.

Deakin, J. M., & Cobley, S. (2003). A search for deliberate practice. *Expert performance in sports*, 115-36.

Debatin, T., Hopp, M. D., Vialle, W., & Ziegler, A. (2023). The meta-analyses of deliberate practice underestimate the effect size because they neglect the core characteristic of individualization – An analysis and empirical evidence. *Current Psychology, 42*(13), 10815-10825.

del Rosario Zavala, Maria, Ma Bernadette Andres-Salgarino, Zandra de Araujo, Amber G. Candela, Gladys Krause, and Erin Sylves (2020). The Mo(ve)ment to Prioritize Antiracist Mathematics: Planning for This and Every School Year. *TODOS: Mathematics for All.*

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*(39), 19251-19257.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 761-772.

Dori, Y. J., & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts?. *The journal of the learning sciences, 14*(2), 243-279.

Doz, D. (2021). Students′ mathematics achievements: A comparison between pre-and post-COVID-19 pandemic. *Education and Self Development, 16*(4), 36-47.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students′ learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest, 14*(1), 4-58.

Ebbinghaus, H. (1885). Memory: A contribution to experimental psychology, trans. *HA Ruger & CE Bussenius. Teachers College, Columbia University.*

Emeny, W. G., Hartwig, M. K., & Rohrer, D. (2021). Spaced mathematics practice improves test scores and reduces overconfidence. *Applied Cognitive Psychology, 35*(4), 1082-1089.

Engram, D. (1979). *Bonus points as tokens: their effect on academic performance* (Masters thesis).

Engström, M., Landtblom, A. M., & Karlsson, T. (2013). Brain and effort: brain activation and effort-related working memory in healthy participants and patients with working memory deficits. *Frontiers in human neuroscience, 7*, 140.

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. *The psychology of problem solving*, 31-83.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. *The Cambridge handbook of expertise and expert performance, 38*(685-705), 2-2.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Academic emergency medicine, 15*(11), 988-994.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review, 100*(3), 363.

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard business review, 85*(7/8), 114.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415.

Fuchs, L. S., Bucka, N., Clarke, B., Dougherty, B., Jordan, N. C., Karp, K. S., ... & Morgan, S. (2021). Assisting Students Struggling with Mathematics: Intervention in the Elementary Grades. Educator's Practice Guide. WWC 2021006. *What Works Clearinghouse.*

Gershenson, S. (2018). Grade Inflation in High Schools (2005-2016). *Thomas B. Fordham Institute.*

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6*(3), 225-255.

Goldhaber, D., & Young, M. G. (2023). Course Grades as a Signal of Student Achievement: Evidence on Grade Inflation Before and After COVID-19. CALDER Research Brief No. 35.

Gonzalez, J. (2017). Retrieval Practice: The Most Powerful Learning Strategy You're Not Using. *CultOfPedagogy.com.*

Grospietsch, F., & Lins, I. (2021). Review on the prevalence and persistence of neuromyths in education–Where we stand and what Is still needed. In *Frontiers in Education* (Vol. 6, p. 665752). Frontiers Media SA.

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M., Dodge, K., ... & Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review, 113*(3), 738-765.

Halpern, D. F., & Hakel, M. D. (2003). Applying the Science of Learning. *Change*, 37.

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in cognitive sciences, 17*(3), 111-120.

Hartwig, M. K., Rohrer, D., & Dedrick, R. F. (2022). Scheduling math practice: Students' underappreciation of spacing and interleaving. *Journal of Experimental Psychology: Applied, 28*(1), 100.

Hattie, J., & Yates, G. C. (2013). *Visible learning and the science of how we learn.* Routledge.

Hofstadter, D., & Carter, K. (2012). Some Reflections on Mathematics from a Mathematical Non-mathematician. *Mathematics in School, 41*(5), 2-4.

Iosup, A., & Epema, D. (2014). An experience report on using gamification in technical higher education. In *Proceedings of the 45th ACM technical symposium on Computer science education* (pp. 27-32).

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol., 59*, 193-224.

Joy Cumming, J., & Elkins, J. (1999). Lack of automaticity in the basic addition facts as a characteristic of arithmetic learning problems and instructional needs. *Mathematical Cognition, 5*(2), 149-180.

Kang, S. H. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 12-19.

Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own?. *Memory, 17*(4), 471-479.

Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250-266.

Kersey, A. J., & James, K. H. (2013). Brain activation patterns resulting from learning letter forms through active self-production and passive observation in young children. *Frontiers in psychology, 4*, 567.

Kirschner, P., & Hendrick, C. (2024). *How learning happens: Seminal works in educational psychology and what they mean in practice.* Routledge.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist, 41*(2), 75-86.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"?. *Psychological science, 19*(6), 585-592.

Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher, 51*(7), 500-506.

Kulasegaram, K. M., Grierson, L. E., & Norman, G. R. (2013). The roles of deliberate practice and innate ability in developing expertise: evidence and implications. *Medical education, 47*(10), 979-989.

Kulik, C. L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of educational research, 60*(2), 265-299.

Kulik, J. A., & Kulik, C. L. C. (1984). Effects of accelerated instruction on students. *Review of educational research, 54*(3), 409-425.

Kyllonen, P. C., & Tirre, W. C. (1988). Individual differences in associative learning and forgetting. *Intelligence, 12*(4), 393-421.

Lehtinen, E., Hannula-Sormunen, M., McMullen, J., & Gruber, H. (2017). Cultivating mathematical skills: From drill-and-practice to deliberate practice. *ZDM, 49*, 625-636.

Lei, H., Wang, C., Chiu, M. M., & Chen, S. (2022). Do educational games affect students' achievement emotions? Evidence from a meta-analysis. *Journal of Computer Assisted Learning, 38*(4), 946-959.

Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. *PloS one, 12*(3), e0173403.

Luo, L., & Kiewra, K. A. (2021). Parents′ roles in talent development. *Gifted Education International, 37*(1), 30-40.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning?. *American psychologist, 59*(1), 14.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European journal of cognitive psychology, 19*(4-5), 494-513.

McDaniel, M. A., Cahill, M. J., Frey, R. F., Rauch, M., Doele, J., Ruvolo, D., & Daschbach, M. M. (2018). Individual differences in learning exemplars versus abstracting rules: Associations with exam performance in college science. *Journal of Applied Research in Memory and Cognition, 7*(2), 241-251.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General, 143*(2), 668.

McDermott, K. B., & Zerr, C. L. (2019). Individual differences in learning efficiency. *Current Directions in Psychological Science, 28*(6), 607-613.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological science, 21*(7), 914-919.

Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 978.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63*(2), 81.

Nairne, J. S., & Neath, I. (2012). Sensory and working memory. *Handbook of Psychology, Second Edition, 4.*

National Council of Supervisors of Mathematics and TODOS: Mathematics for ALL. (2016). Mathematics education through the lens of social justice: Acknowledgment, actions, and accountability. *Joint Position Paper.*

Pan, S. C. (2015). The interleaving effect: mixing it up boosts learning. *Scientific American, 313*(2).

Pardo, A., Poquet, O., Martínez-Maldonado, R., & Dawson, S. (2017). Provision of data-driven student feedback in LA & EDM. *Handbook of learning analytics*, 163-174.

Park, G., Lubinski, D., & Benbow, C. P. (2013). When less is more: Effects of grade skipping on adult STEM productivity among mathematically precocious adolescents. *Journal of Educational Psychology, 105*(1), 176.

Parrish, S. The Ultimate Deliberate Practice Guide: How to Be the Best. *FS.blog.*

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic bulletin & review, 14*(2), 187-193.

Pasquinelli, E. (2012). Neuromyths: Why do they exist and persist?. *Mind, Brain, and Education, 6*(2), 89-96.

Pastötter, B., & Frings, C. (2019). The forward testing effect is reliable and independent of learners' working memory capacity. *Journal of cognition, 2*(1).

Pedersen, B., Makel, M. C., Rambo-Hernandez, K. E., Peters, S. J., & Plucker, J. (2023). Most mathematics classrooms contain wide-ranging achievement levels. *Gifted Child Quarterly, 67*(3), 220-234.

Plant, E. A., Ericsson, K. A., Hill, L., & Asberg, K. (2005). Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary educational psychology, 30*(1), 96-116.

Qadir, J., & Imran, M. A. (2018). Learning 101: The untaught basics. *IEEE Potentials, 37*(3), 33-38.

Reber, P. J., & Kotovsky, K. (1997). Implicit learning in problem solving: The role of working memory capacity. *Journal of Experimental Psychology: General, 126*(2), 178.

Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review, 27*(4), 617-633.

Reeves, W. R. (2014). *Creativity as a learned skill: The role of deliberate practice in the development of creativity* (Doctoral dissertation).

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology, 69*(10), 1969-1995.

Rohrer, D. (2009). Research commentary: The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education, 40*(1), 4-17.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24*, 355-367.

Rohrer, D., Dedrick, R. F., & Agarwal, P. K. (2017). Interleaved mathematics practice: giving students a chance to learn what they need to know. *RetrievalPractice.org.*

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology, 107*(3), 900.

Rohrer, D., & Hartwig, M. K. (2020). Unanswered questions about spaced interleaved mathematics practice. *Journal of Applied Research in Memory and Cognition, 9*(4), 433.

Rohrer, D., & Hartwig, M. K. (2023). Spaced and Interleaved Mathematics Practice. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 111-21). Society for the Teaching of Psychology.

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*(4), 183-186.

Ross, P. E. (2006). The expert mind. *Scientific American, 295*(2), 64-71.

Roth, J. K., & Courtney, S. M. (2007). Neural system for updating object working memory from different sources: sensory stimuli or long-term memory. *Neuroimage, 38*(3), 617-630.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological bulletin, 140*(6), 1432.

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology, 23*(08), 577-589.

Pashler, H., Rohrer, D., & Cepeda, N. J. (2006). Temporal spacing and learning. *APS Observer, 19.*

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International journal of Educational research, 13*(1), 89-99.

Sanchez, E. I., & Moore, R. (2022). Grade Inflation Continues to Grow in the Past Decade. Research Report. ACT, Inc.

Schmidt, S., & Eisend, M. (2015). Advertising repetition: A meta-analysis on effective frequency in advertising. *Journal of Advertising, 44*(4), 415-428.

Shamloo, F., & Helie, S. (2016). Changes in default mode network as automaticity develops in a categorization task. *Behavioural Brain Research, 313*, 324-333.

Sherman, J. G. (1992). Reflections on PSI: Good news and bad. *Journal of Applied Behavior Analysis, 25*(1), 59.

Simonton, D. K. (2007). Talent and expertise: The empirical evidence for genetic endowment. *High Ability Studies, 18*(1), 83-84.

Simonton, D. K. (2013). If innate talent doesn't exist, where do the data disappear?. *The complexity of greatness: Beyond talent or practice*, 17-26.

Sloboda, J. A. (1996). The acquisition of musical performance expertise: Deconstructing the "talent" account of individual differences in musical expressivity. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 112-114). Psychology Press.

Smith, C. D., & Scarf, D. (2017). Spacing repetitions over long timescales: a review and a reconsolidation explanation. *Frontiers in Psychology, 8*, 962.

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research, 86*(4), 849-899.

Sternberg, R. J. (2014). Costs of expertise. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 347-354). Psychology Press.

Subirana, B., Bagiati, A., & Sarma, S. (2017). *On the Forgetting of College Academics: at "Ebbinghaus Speed"?.* Center for Brains, Minds and Machines (CBMM).

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of educational psychology, 96*(3), 471.

Swanson, H. L., & Siegel, L. (2011). Learning disabilities as a working memory deficit. *Experimental Psychology, 49*(1), 5-28.

Sweller, J., Ayres, P. L., Kalyuga, S., & Chandler, P. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and instruction.*

Sweller, J., Clark, R., & Kirschner, P. (2010). Teaching general problem-solving skills is not a substitute for, or a viable addition to, teaching mathematics. *Notices of the American Mathematical Society, 57*(10), 1303-1304.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory.* Springer Science+Business Media.

Talcott, J. B. (2021). The neurodevelopmental underpinnings of children's learning: Connectivity is key.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied cognitive psychology, 24*(6), 837-848.

Tharayil, S., Borrego, M., Prince, M., Nguyen, K. A., Shekhar, P., Finelli, C. J., & Waters, C. (2018). Strategies to mitigate student resistance to active learning. *International Journal of STEM Education, 5*, 1-16.

TODOS: Mathematics for All (2020). *Where Is Manuel? A Rejection of 'Learning Loss'*.

Underwood, B. J., & Ekstrand, B. R. (1967). Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology, 74*(4p1), 574.

Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & cognition, 33*(2), 213-220.

VanDerHeyden, A. M., & Codding, R. S. (2020). Belief-Based versus Evidence-Based Math Assessment and Instruction. *Communique, 48*(5).

Vlach, H. A., Sandhofer, C. M., & Bjork, R. A. (2014). Equal spacing and expanding schedules in children's categorization and generalization. *Journal of experimental child psychology, 123*, 129-137.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748-751.

Von Glasersfeld, E. (1984). An introduction to radical constructivism. *The invented reality, 1740*, 28.

Wai, J. (2015). Long-term effects of educational acceleration. *A nation empowered: Evidence trumps the excuses holding back America's brightest students, 2*, 73-83.

Weinstein, Y., Madan, C. R., & Sumeracki, M. A. (2018). Teaching the science of learning. *Cognitive research: principles and implications, 3*(1), 1-17.

Westervelt, Eric (2016). A Nobel Laureate's education plea: revolutionize teaching. *NPR.*

Wright, R. E. (1981). Aging, divided attention, and processing capacity. *Journal of Gerontology, 36*(5), 605-614.

Yang, C., Shanks, D. R., Zhao, W., Fan, T., & Luo, L. (2023a). Frequent Quizzing Accelerates Classroom Learning. In C. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In Their Own Words: What Scholars and Teachers Want You to Know About Why and How to Apply the Science of Learning in Your Academic Setting* (pp. 252-62). Society for the Teaching of Psychology.

Yang, C., Li, J., Zhao, W., Luo, L., & Shanks, D. R. (2023b). Do practice tests (quizzes) reduce or provoke test anxiety? A meta-analytic review. *Educational Psychology Review, 35*(3), 87.

Y Combinator (2018). *Peter Reinhardt on Finding Product Market Fit at Segment.*

Zerr, C. L., Berg, J. J., Nelson, S. M., Fishell, A. K., Savalia, N. K., & McDermott, K. B. (2018). Learning efficiency: Identifying individual differences in learning rate and retention in healthy adults. *Psychological science, 29*(9), 1436-1450.

Zou, X., Ma, W., Ma, Z., & Baker, R. S. (2019). Towards helping teachers select optimal content for students. In *Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part II 20* (pp. 413-417). Springer International Publishing.

Notes for Future Additions

While this working draft contains a lot of information, it's not even halfway done. I am still building out this book and there is still lots of key information that remains unaddressed. Below are my working notes on items that I need to incorporate into the main body of the book.

Parental Support / Coaching

- Why Does Parents' Involvement Enhance Children's Achievement? The Role of Parent-Oriented Motivation https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8a7a096cf1d1ad9e783b3d26f32110a4ddf049f9
  - Parental involvement is not just about keeping kids on the rails with deliberate practice, it's also about being someone that the kid respects and trusts and wants to impress or at least meet their standards. Not just drill sergeant forcing them to meet standards, but also mentor / role model of whom the kid wants to gain the approval.
  - "Children's motivation in school is parent oriented when it is driven by a concern with meeting parents' expectations in the academic arena so as to gain their approval."
  - "Over time, the more involved parents were in children's learning, the more motivated children were to do well in school for parent-oriented reasons, which contributed to children's enhanced self-regulated learning and thereby grades. Although children's parent-oriented motivation was associated with their controlled and autonomous motivation in school, it uniquely explained the positive effect of parents' involvement on children's grades."
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=coach+athlete+relationship+communication+incentives&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=coach+athlete+relationship+communication&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=coach+athlete+relationship+incentives&btnG=

- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=coach+athlete+relationship&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=coach+athlete+relationship+elite&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=mathematician+parent+support&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=math+parent+support&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=math+parent+coach&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=parent+athlete+motivation&btnG=
- https://scholar.google.com/scholar?start=10&q=successful+mathematician+coaching&hl=en&as_sdt=0,22
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=successful+mathematicians&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=aspiring+mathematician&btnG=
- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=parent+oriented+motivation&btnG=&oq=parent-oriented
- Science of coaching
- Emotional side of learning – in particular, how to deal with feedback. Not easy for a student to be told that their attempt at a given problem is wrong.

Direct Instruction

After elaborating more on the decades of research behind direct instruction, need to break direct instruction into a front-and-center chapter where we go really in-depth -- probably right after active learning. Direct instruction is a cognitive learning strategy in its own right.

In that section, I can also talk about the need for specific direct instruction, not "general" domain-independent problem solving:

- Project Follow Through
- Zig Engelmann
  - https://education-consumers.org/pdf/CT_111811.pdf (CLEAR TEACHING: With Direct Instruction, Siegfried Engelmann Discovered a Better Way of Teaching. By Shepard Barbash)

    - https://www.zigsite.com/
- Teaching General Problem Solving Skills Is Not a Substitute for, or a Viable Addition to, Teaching Mathematics https://www.ams.org/notices/201010/rtx101001303p.pdf
- Mathematical Ability Relies on Knowledge, Too https://files.eric.ed.gov/fulltext/EJ909939.pdf
- Consequences of History-Cued and Means-End Strategies in Problem Solving https://www.jstor.org/stable/1422136
- Response to De Jong et al.'s (2023) paper "Let's talk evidence – The case for combining inquiry-based and direct instruction" https://www.sciencedirect.com/science/article/pii/S1747938X23000775
- Should also mention that in recent years, unguided instruction is still around and is still leading to subpar learning outcomes: The Efficacy of Inquiry-Based Instruction in Science: a Comparative Analysis of Six Countries Using PISA 2015 https://link.springer.com/article/10.1007/s11165-019-09901-0

Knowledge Spaces

Explain similarities & differences between ALEKS knowledge space theory and MA's approach.

Knowledge spaces are another way to describe mastery learning on a knowledge graph data structure. The researchers behind ALEKS and knowledge space theory did a great job formalizing, setting up definitions, proving rigorous theorems about mastery learning and diagnostic exams in a knowledge graph, but in order for our system to do what we want it to do, we had to introduce a bunch of other cognitive learning strategies and capabilities resulting in needing to wrangle a lot more complexity.

In particular, the combinatorial perspective taken in the knowledge space papers & textbooks becomes intractable, which we've worked around by taking the approach of having quantities physically flowing through the graph – "constructing" the particular knowledge state in question as opposed to filtering it out of an exhaustive list of possible knowledge states. You kind of have to take the constructive, flowing-quantities approach when you're incorporating spaced repetition into the system.

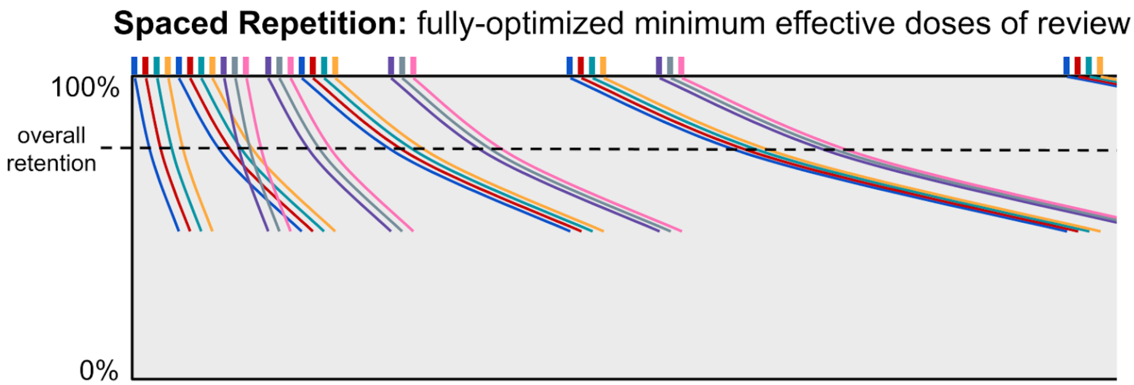
There are also problems of scale. Several years ago I reimplemented the ALEKS diagnostic assessment algorithm and found that it became prohibitively computational expensive when the knowledge graph contained more than a hundred or so nodes – whereas MA's diagnostic algorithm can and needs to handle many hundreds, even sometimes over a thousand nodes since our content is so much more scaffolded, we're assessing prerequisite knowledge, and we're going deep into high-level math that has tons of prerequisites.

Here are some of my old notes from that reimplementation project:

- The adaptive assessment method in the ALEKS paper seems intractable.
- Running it on a straight graph line, the number of knowledge states is roughly equal to the number of nodes.
- Running it on a binary tree graph, we find that when we double the number of splits, the number of knowledge states (KS) gets squared:
    - ~2 splits --> ~4 KS
    - ~4 splits --> ~16 KS
    - ~8 splits --> ~256 KS
    - ~16 splits --> ~65,536 KS
    - ~32 splits --> ~4,294,967,296 KS
    - ~64 splits --> ~1.8 x 10^19 KS
    - ~128 splits --> ~3.2 x 10^38 KS
- For recombining binary tree, it's a comparable (but less precise) trend (the numbers fit into the range suggested by the non-recombining binary tree).
- This places constraints on the graph -- beyond about 20 splits, it becomes intractable. This matches up with what's in their picture: about 20 splits (about 100 nodes) for Beginning Algebra, corresponding to 60,000,KS
- We have about 150 splits (about 300 nodes) in our MVC/LA course, so this isn't going to work for us.

Spaced Repetition Visualization

I created the following visualization of spaced repetition. It's already worked into the main body of the document.



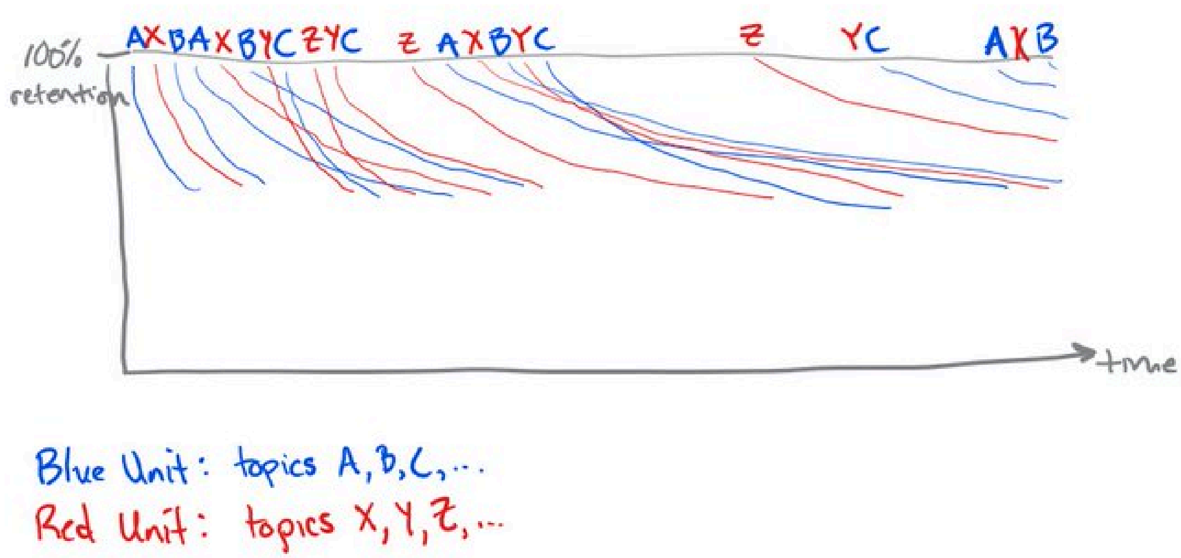
However, it needs some additional clarification. In the spaced repetition graphic, the first blue rectangle corresponds to initial learning of the first topic within the blue unit, learning to the level of mastery.

The next blue rectangle with a slower forgetting curve would correspond to the first spaced review of that topic. And so on for future spaced reviews.

Same for the red rectangles: the first one corresponds to initial learning of the first topic within the red unit learning to the level of mastery. The next red rectangle with a slower forgetting curve would correspond to the second spaced review of that topic. And so on.

This same process would be happening for the second topic in each unit, the third topic, and so on, all interleaved together. For simplicity, this is not shown in the graphic.

Just considering Blue and Red units, it would look something like this:



Chunking

I haven't yet done any explicit literature searches on chunking. I should make sure I'm referencing "chunking" vocabulary whenever I'm essentially describing it in the body of this book, and check if there's any other relevant info surrounding chunking that I need to address in this book.

Biological Basis of Neuroplasticity

Every concept is ultimately represented as a pattern of neural activity. Your brain can only devote so much effort to maintaining neural activity, but as you practice activating those neural patterns, biological changes occur that make the patterns easier to activate with less effort. As I recall, these biological changes are thought to primarily take place within synapses (https://en.wikipedia.org/wiki/Synaptic_plasticity), though there is also research into changes that occur elsewhere, e.g., in dendrites.

Automaticity and Intuition

Need to lay out an explicit connection between repetition, automaticity, intuition, and creativity. I have a section on the relationship between automaticity and creativity, and the relationship between automaticity and intuition should fit right in around there.

Also need to really emphasize how important repetition is in mathematics to get to the point where you can even attempt to think creatively.

Doing the grunt work yourself really drills into you what those operations are. Like, you don't just think about the definition from afar, you really "feel" what it is, close up and in your bones, almost in a physical sense.

Talk about the importance of developing computational (procedural) & conceptual knowledge in tandem.

More Case Studies

Look into other cases where a university changed their instructional methods and got some kind of feedback on the outcome.

For instance: https://people.math.harvard.edu/~knill/pedagogy/harvardcalculus/

Spaced Repetition in Math vs Language

Mark (on Discord) mentioned that spaced repetition systems (SRS) have been shown to have very positive effects when used for discrete unrelated pieces of information (e.g. state capitals or the pronunciation of letters in a syllabary) but for language learners a focus on SRS decks of words typically underperforms extensive reading/conversation. It makes perfect sense that our fractional implicit repetition (FIRe) solution works well for math, but it is covered later on and it is only tangentially explained how FIRe solves those shortcomings affecting SRS decks with language learning. We need to address this more directly and early on in the spaced repetition section.

Disambiguating Interleaving, Non-Interference, and Spaced Repetition

May also want to have a section on Interleaving, Non-Interference, and Spaced Repetition that clarifies the difference between all these things and how we balance all of them. Related resources:

- https://blog.innerdrive.co.uk/interleaving-dos-and-donts
- https://blog.innerdrive.co.uk/are-spacing-and-interleaving-the-same-thing

Visualization of Macro-Interleaving

Maybe show some diagrams of interleaved paths through the knowledge graph vs non-interleaved paths.

Elaborative Interrogation

Should also talk about how we scaffold elaborative interrogation once we have more of that in the system (and should also have a chapter on elaborative interrogation as its own learning strategy). Some resources:

- https://www.learningscientists.org/blog/2016/7/7-1
- https://blog.innerdrive.co.uk/retrieval-practice-generative-learning

Optimal Manual Teaching

Another chapter about how teachers can leverage these cognitive learning strategies as much as possible if they do not have access to technology. (How would you get the most bang for your buck teaching manually, without totally blowing up your workload to an inhuman degree? Obviously not going to capitalize on the full effects of what our tech can do, but you can still do a lot better than what the status quo is.)

Elaborative Interrogation

Why/how questions like this: "Why does the parabola y-k=a(x-h)^2 have its vertex at (h,k)?" "Starting with the formula for slope, how do you get to the point-slope formula for a line?"

We don't do much of that at the moment, but now that we have select questions, we can probably use those for elaborative interrogation.

Coaching Science

It's really interesting how many parallels there are to coaching. I've got it on my reading list to check out some literature in coaching science. I anticipate it's also going to be a gold mine for motivational techniques, way moreso than the standard "science of learning" literature, since coaching literature seems to focus more on high-performance settings whereas the science of learning literature tends to focus on fairly low-performance settings. (Which is unsurprising since there are way better incentive structures and accountability mechanisms to support high performance in the field of athletics, as compared to the field of education.)

Miscellaneous Resources to Check Out

- Go through my own recent blog posts and check for anything to be worked in
- Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students' (mis)judgments of learning and teaching effectiveness. Journal of Applied Research in Memory and Cognition, 9(2), 137–151. https://doi.org/10.1016/j.jarmac.2019.12.009
- Book: Uncommon Sense Learning
- https://www.thescienceofmath.com/timed-tests-cause-math-anxiety
- https://x.com/seventhmeal/status/1838296108369612918
  https://dominiccummings.com/the-odyssean-project-2/
- https://x.com/justinskycak/status/1841508577443496260
- https://gwern.net/doc/psychology/chess/2014-hambrick.pdf
- https://notes.andymatuschak.org/zBmSSpM1WfFDehxNCBcqSZp?stackedNotes=zMX9Lfuz8sGfDUivWZcyWT
- Tons of great references here:
  https://scienceoflearning.substack.com/p/should-we-teach-children-to-memorize?triedRedirect=true

- https://scottbarrykaufman.com/wp-content/uploads/2011/06/Protzko-Kaufman-2010.pdf
- https://www.colorado.edu/ics/sites/default/files/attached-files/91-06.pdf
- https://www.researchgate.net/publication/51129143_Spaced_Retrieval_Absolute_Spacing_Enhances_Learning_Regardless_of_Relative_Spacing
- https://arxiv.org/abs/2006.01581
- https://gwern.net/doc/psychology/chess/2014-hambrick.pdf Deliberate practice: Is that all it takes to become an expert?
- https://www.rocketmath.com/about-rocket-math/research_studies-and-results/
- https://www.rocketmath.com/wp-content/uploads/2016/03/Math-Facts-research.1.pdf
- https://www.rocketmath.com/wp-content/uploads/2016/03/Third-stage-of-Learning-Math-Facts.pdf
- https://www.rocketmath.com/wp-content/uploads/2016/03/How-fast-is-fast-enough-to-be-automatic.pdf
- Individual Differences in Arithmetic: Implications for Psychology, Neuroscience and Education by Ann Dowker
- Working Memory and Learning: A Practical Guide for Teachers by Susan Gathercole and Tracy Packiam Alloway
- Children's Mathematical Development: Research and Practical Applications by David C Geary
- Visible Learning: Feedback by John Hattie & Shirley Clarke
- Visible Learning: The Sequel: A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement by John Hattie
- Acquisition of Complex Arithmetic Skills and Higher-Order Mathematics Concepts, Volume 3 edited by David C. Geary, Daniel B. Berch, Robert Ochsendorf, & Kathleen Mann Koepke
- Cognitive Foundations for Improving Mathematical Learning, Volume 5 edited by David C. Geary, Daniel B. Berch, & Kathleen Mann Koepke
- Summing up hours of any type of practice versus identifying optimal practice activities: Commentary on Macnamara, Moreau, & Hambrick (2016). (link)
- Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research. (link)
- Given that the detailed original criteria for deliberate practice have not changed, could the understanding of this complex concept have improved over time? A response to Macnamara and Hambrick (2020). (link)
- Self-Construction, Self-Protection, and Self-Enhancement: A Homeostatic Model of Identity Protection. (link)
- Self-enhancement and self-protection: What they are and what they do (link)

- Illusions of comprehension, competence, and remembering. (link)
- Assessing our own competence: Heuristics and illusions. (link)
- Can research inform classroom practice?: The particular case of buggy algorithms and subtraction errors (link)
- https://www.johndcook.com/blog/2013/02/04/four-hours-of-concentration/

Books to check out

- Teach Like a Champion
- A Coach's Guide to Teaching
- Teach to Learn (Catherine Scott)
- Accelerated Expertise - Robert R. Hoffman
- Ultralearning – Scott Young
- The Science of Rapid Skill Acquisition – Peter Hollins
- Hidden Potential by Adam Grant