AMERICAN EDUCATION / A HISTORICAL INVESTIGATION

The Lost Art of Critical Thinking



What schools taught. What changed. What the evidence can tell us about learning to think for ourselves.

42
CHAPTERS

94
SOURCE ENTRIES

1950s-2026
HISTORICAL FOCUS

JesseW22.com

Research through October 29, 2026

Second reviewed PDF edition / Third reviewed HTML text

BEGIN THE INVESTIGATION

EXPLORE CONTENTS

THE READING MAP

Explore the investigation

Read in order, or choose a chapter. Every title below is clickable.

I · The question

01 What exactly has been lost? 5

02 Critical thinking is more than disagreement 6

03 How to read the evidence 7

II · 1950s-1990s

04 The 1950s classroom: strengths and limits 8

05 Sputnik and the drive to think like a scientist 9

06 The named thinkers behind classroom methods 10

07 Debate, seminar, logic, and the uneven opportunity to use them 11

08 The 1970s-1990s: reform, backlash, and higher-order thinking 12

III · 2000-2014

09 Accountability changed the incentives 13

10 The science of learning versus the marketing of methods 14

11 Modern standards still ask students to reason 15

12 Why 2005-2014 deserves particular attention 16

IV · 2015-2026 in depth

13 A decade of overlapping changes 17

14 The downturn began before COVID 18

15 Teacher preparation: a concrete weakness to examine 19

16 Teacher preparation and ideology: what the documents say 20

17 Social-emotional learning: distinguish the versions 21

18 Have schools stopped debating ideas? 22

19 The boundary between teaching and prescribed agreement 23

V · Objectivity & racial ideology

20 The controversy is real: start with the original claims 24

21 The Smithsonian chart: institutional endorsement, then removal 25

CONTENTS / CONTINUED

V · Objectivity & racial ideology

22 Equitable Math: the disputed wording and the authors' answer 26

23 How influential were these ideas? 27

VI · Institutions & power

24 Teachers' unions: follow the actual mechanisms 29

25 The money is substantial, and categories matter 30

26 Pandemic reopening: a documented union connection 31

27 Does bargaining improve or damage learning? 32

28 Government responsibility: identify the level and decision 33

VII · Disruption & technology

29 2020-2021: instruction and measurement were both interrupted 34

30 Absence, recovery, and the missing practice 35

31 Phones and devices changed the environment of thought 36

32 Digital fluency did not guarantee source judgment 37

33 Grading and testing: when the signals stop agreeing 38

34 Generative AI: finished work is not necessarily learning 39

VIII · What the outcomes show

35 The 2024 results: serious academic warning signs 40

36 The 2026 update: recovery exists, and is uneven 42

37 What about twenty-somethings and everyday common sense? 43

IX · Rebuilding the practice

38 What the instructional evidence supports 44

39 A classroom that permits genuine disagreement 46

40 An evidence ledger: which explanations survived? 47

41 What families, school boards, and communities can do 49

42 The adult stakes, and the responsibility to improve 50

Sources & research notes / 51

A NOTE TO THE READER

Follow the evidence.
Keep your own judgment.

A historical investigation, a practical reading guide, and a record you can inspect.

Recent learning losses are real. Reasoning instruction has a longer, uneven history. This investigation follows the methods, the institutions, the controversies, and the results, then asks what communities can improve.

Using this PDF

The contents, chapter references, source citations, and footer navigation are clickable. The PDF bookmark panel provides another route through the chapters. Sources open the original research or document; the source index records context and retrieval limits.

Videos open online through the original provider. This edition preserves the research and reading sequence of the final HTML; interactive players, search, and text-size controls belong to the HTML version.

Designed first for a desktop browser or PDF reader. On a phone, portrait orientation and fit-to-width give the most comfortable starting point. A PDF has fixed pages; zoom remains available.

Research cutoff: October 29, 2026. This is a narrative investigation, not a new national assessment or a systematic causal review.

I · The question / CHAPTER 01

What exactly has been lost?

The title is a question worth investigating, not a conclusion we can assume.

A young person can be quick with a device yet struggle to judge whether an answer makes sense. A student can earn a strong grade yet have difficulty defending a claim. Those experiences deserve investigation. The central question is whether schools are developing the knowledge, habits, and independence needed to reason well, and whether the answer has changed over time.

This investigation finds a serious recent deterioration in several academic foundations of reasoning, documented weaknesses in evaluating online information, and substantial pressures on classroom discussion. It also finds explicit commitments to reasoning in modern standards, effective contemporary teaching approaches, and signs of recovery for some younger students. The historical record does not support a single national moment when schools abandoned thought.

The most useful explanation is layered: what students know; what teachers are prepared and given time to teach; what assessments reward; whether discussion permits genuine disagreement; and how technology changes the work students do themselves. Government, unions, universities, publishers, families, and technology companies all enter this story through particular decisions and incentives. None should receive either automatic blame or automatic exemption.

The standard used throughout

A documented policy establishes what an organization said or required. A survey establishes what respondents reported. An assessment establishes performance on particular tasks. An experiment may identify an effect in a particular setting. Moving from one kind of evidence to another requires an argument. This presentation makes those transitions visible.

The practical goal: help communities ask sharper questions, inspect actual materials, and improve what students learn to do independently. The investigation is national in scope, but its most actionable questions are often local.

Contents / Next chapter

Critical thinking is more than disagreement

Knowledge, reasoning, judgment, and the willingness to revise.

In this presentation, critical thinking means reaching and revising judgments through evidence and reasons: identifying the claim, checking relevant information, recognizing assumptions, comparing explanations, and deciding what follows. It includes knowing when a conclusion is uncertain and being willing to correct it.

John Dewey's *How We Think* (1910) treated reflective thought as deliberate inquiry rather than the stream of ideas that passes through the mind. A difficulty prompts investigation; possible explanations are considered; their implications are examined; and observations test them. His concern was already the relationship between schoolwork and thoughtful judgment. [S01]

Edward Glaser's 1941 formulation joined three elements: a disposition to consider problems thoughtfully, knowledge of logical inquiry, and skill in applying it. That combination helps explain why a slogan such as "question everything" is inadequate. Questioning without investigation can turn into indiscriminate suspicion; following a procedure without a willingness to reconsider can turn into rationalization. [S02]

A working example

Imagine a school announces that a new program raised the passing rate from 60% to 80%. A critical reader asks whether the test, passing threshold, students tested, and participation rate stayed comparable; whether another change occurred; and whether students can now solve unfamiliar problems. The improvement may be real. It still needs an explanation.

Independent thinking also does not require rejecting expertise. A sound decision may rely on specialists while checking their methods, incentives, uncertainty, and agreement with other evidence. Nor does reason require indifference to human experience. Testimony can reveal a problem; a fair investigation then asks how broadly it occurs and which explanations survive examination.

Previous chapter / Contents / Next chapter

I · The question / CHAPTER 03

How to read the evidence

What this investigation can establish, and where the record runs out.

The historical question is harder than comparing two test scores. There is no continuous, nationally representative examination of general critical thinking running from the 1950s through 2026. The National Assessment of Educational Progress (NAEP) began in 1969; its main subject assessments and separate long-term trend series have different designs. They supply indispensable evidence about relevant knowledge and skills, not a direct census of everyday judgment. [S79]

The strongest parts of the investigation combine original curriculum documents, federal assessments, original research papers, and organizations' own materials. Advocacy documents are valuable evidence of an organization's position. They do not, by themselves, establish that the position is effective or widely implemented. Book descriptions and accessible excerpts are identified as such rather than presented as full-book reviews.

The research cutoff is **October 29, 2026**. Assessment year and release year are kept separate: the 2024 main NAEP results appeared in 2025, while the 2025 long-term trend results appeared in June 2026. The presentation includes both. Where a direct page or full PDF could not be retrieved, the source notes disclose the limitation.

Four evidence labels

Documented: an attributable text, policy, financial disclosure, or measured result. **Supported explanation:** research connects a mechanism with an outcome, subject to its design. **Plausible risk:** a reasoned concern that has not been established as the national cause. **Not established:** the available evidence does not justify the claim.

These distinctions matter most when a vivid example is asked to explain an entire country. A training workbook may contain a troubling statement. Establishing its national effect would require adoption records, classroom implementation evidence, and credible comparisons of outcomes. This investigation found examples of circulation, but not a national prevalence estimate for teaching that objectivity is racist.

Previous chapter / Contents / Next chapter

II · 1950s-1990s / CHAPTER 04

The 1950s classroom: strengths and limits

A stable instructional tradition, not a universal school of debate.

The familiar mid-century classroom emphasized teacher explanation, textbooks, recitation, exercises, and tests. These routines could build fluent reading, arithmetic, memory, orderly work, and a body of shared knowledge. They could also leave students repeating a textbook's conclusion without examining how it was reached.

Larry Cuban's historical research describes substantial continuity in teacher-centered instruction across twentieth-century reforms, alongside hybrids and pockets of student-centered practice. His evidence challenges a simple story in which every earlier classroom cultivated independent thought and a later reform suddenly replaced it with conformity. The contrast between official reform language and daily practice is itself longstanding. [S03]

Analytical distinction: reciting a definition is different from applying it, just as memorizing multiplication is different from deciding which calculation a problem requires. Both can belong in a strong education. The weakness appears when either is treated as sufficient for the other.

Comparisons must also ask who was included. The National Center for Education Statistics (NCES) reports that 32.6% of adults aged 25 and older had completed high school or more in 1950, compared with 77.7% in 1990. Those are adult attainment figures, not graduating-class percentages, but they show how different the educational reach of the two eras was. An earlier generation's most educated students cannot stand in for all its children. [S04]

What is worth recovering?

Clear explanations, coherent subject knowledge, sustained reading, and personal responsibility for work remain useful. Recovering them does not require recreating every mid-century practice. The relevant test is whether students can remember, explain, apply, question, and revise what they learn.

Previous chapter / Contents / Next chapter

II · 1950s-1990s / CHAPTER 05

Sputnik and the drive to think like a scientist

1950s-1960s: inquiry was part of the reform, not something invented yesterday.

The Soviet launch of Sputnik in October 1957 intensified American concern about scientific education and national capacity. The institutional history of the National Science Foundation (NSF) describes expanded teacher institutes and new work in physics, biology, chemistry, and mathematics, while noting that concerns and projects preceded the satellite. The federal role in curriculum influence therefore did not begin with recent culture-war disputes. [S05]

The Biological Sciences Curriculum Study is a concrete example. BSCS's own history traces its beginnings to this reform period and NSF support for new biology materials. The point was to improve the substance and character of science learning, not merely increase the number of facts students could repeat. [S06]

Jerome Bruner's *The Process of Education* (1960) argued for introducing fundamental disciplinary ideas in accessible form and returning to them with greater sophistication. The spiral curriculum connects early intuition with later formal understanding. Its legacy helps explain recurring efforts to teach the structure of a subject and how knowledge is developed. [S08]

The enduring implementation problem

A laboratory task can cultivate reasoning when a student predicts, measures, compares an explanation with evidence, and considers error. The same apparatus can become a recipe: follow these steps, obtain the expected result, copy the conclusion. Calling an activity "inquiry" tells us less than examining what intellectual decisions students actually make.

That distinction also complicates nostalgia about traditional science. The period contained both demands for stronger knowledge and demands for more investigation. The present question is whether either aspiration receives adequate materials, preparation, and time.

Previous chapter / Contents / Next chapter

II · 1950s-1990s / CHAPTER 06

The named thinkers behind classroom methods

Different theories answer different educational problems.

Educational history contains overlapping traditions, not one method handed from thinker to thinker and then discarded nationwide. The following names explain recurring classroom choices; their influence is not evidence that every teacher used their work.

B. F. Skinner: in his 1958 *Teaching Machines* essay, active responses, immediate feedback, and individual pacing offered an alternative to passive reception. Small-step practice can build accuracy; it does not automatically teach evaluation of an unfamiliar argument. [S07]

Benjamin Bloom: the 1956 taxonomy distinguished cognitive objectives, including analysis, synthesis, and evaluation. The 2001 Anderson-Krathwohl revision uses the familiar sequence remember, understand, apply, analyze, evaluate, create. A taxonomy organizes objectives; it is not a requirement to postpone thought until all facts have been memorized. [S09]

Jean Piaget and Lev Vygotsky: the National Research Council's learning synthesis explains the importance of learners' existing concepts and active knowledge construction. Vygotsky's zone of proximal development highlights what a learner can accomplish with assistance beyond independent performance. Neither principle logically requires leaving novices to discover everything alone. [S15] [S16]

Mortimer Adler: the Paideia proposal of 1982 combined teaching information, coaching intellectual skills, and Socratic seminar. The combination matters: explanation, practice, and dialogue can support one another. Socratic questioning asks students to clarify and defend interpretations rather than merely announce a preference. [S11]

Our interpretation: these traditions are resources, not political team uniforms. A teacher may explicitly explain a concept, guide practice, invite competing interpretations, and require a revised written argument in the same unit. The strongest question is what each stage enables the student to do.

Previous chapter / Contents / Next chapter

II · 1950s-1990s / CHAPTER 07

Debate, seminar, logic, and the uneven opportunity to use them

An established tradition does not equal universal access.

Organized high-school debate has a long American institutional history. Bruno E. Jacob's National Forensic League began in 1925, and its first national tournament was held in 1931. Its history records expansion through the mid-century period and later developments such as Lincoln-Douglas debate. Students had established venues to research, construct arguments, answer objections, and speak publicly. That is evidence of a durable tradition, not proof that all students received it. [S12]

A debate team, a Socratic seminar, and a logic lesson serve different purposes. Competitive debate trains advocacy and rebuttal under rules. Seminar explores interpretation through shared texts. Logic examines whether a conclusion follows from premises. A classroom can benefit from each while also needing checks against winning through rhetorical speed, social status, or selective evidence.

What counts as a reasoning task?

A proof asks why a mathematical result follows. An experiment asks whether observations distinguish competing explanations. A history essay asks which account is best supported by documents. A practical project asks which design meets constraints. A formal logic course is valuable, but its absence does not establish the absence of reasoning instruction.

Nor did this investigation locate a national time series proving that all forms of debate disappeared. The stronger recent evidence concerns teachers reporting limits on political and social discussion, examined later. That should not be silently turned into a claim that mathematics explanations, literature discussions, or debate clubs ended everywhere.

A useful recovery target: give every student regular, supported opportunities to justify a claim, answer an objection, and revise. A few highly accomplished debaters cannot compensate for a wider curriculum in which most students rarely do those things.

Previous chapter / Contents / Next chapter

II · 1950s-1990s / CHAPTER 08

The 1970s-1990s: reform, backlash, and higher-order thinking

The concern about weak reasoning was already explicit.

The 1983 federal report *A Nation at Risk* called for stronger content, expectations, instructional time, teaching, and leadership. The report's existence is important historical evidence: national alarm about intellectual preparation is not a discovery of the social-media era. Its diagnosis and recommendations are a contemporary reform argument, not a controlled test of one teaching philosophy. [S13]

The 1989 mathematics standards movement also emphasized problem solving, mathematical communication, reasoning, connections, and estimation. A contemporary National Council of Teachers of Mathematics (NCTM) implementation monograph documents those priorities. Explaining a solution or relating representations was supposed to deepen mathematical understanding, not remove the need for valid answers. [S14]

Bloom's earlier mastery-learning proposal adds another strand. His 1968 paper treated time, instructional quality, persistence, and the ability to understand instruction as variables educators could change. Its essential distinction is between giving a student another opportunity to master demanding work and simply declaring the work mastered. A retake with feedback and a comparable assessment can be rigorous; an automatic passing grade cannot establish the same thing. [S10]

A recurring false choice

The political vocabulary changes, but schools repeatedly face a familiar dispute: knowledge or skills, discipline or freedom, direct instruction or inquiry. These oppositions can conceal the practical question of sequencing. Students may need clear explanation and practice before they can independently compare methods; they also need actual opportunities to apply what they know.

The historical conclusion is therefore qualified but clear: reasoning-oriented approaches were established well before 2000, and worries about their inadequate realization were established too. Evidence for a recent decline must come from recent outcomes and practice, rather than an imagined uniformly successful past.

Previous chapter / Contents / Next chapter

III · 2000-2014 / CHAPTER 09

Accountability changed the incentives

The difference between measuring learning and narrowing it.

No Child Left Behind (NCLB), enacted in 2002, made assessment-based accountability a major organizing force. The pressure had an understandable purpose: expose low performance that averages could conceal, including weaknesses affecting disadvantaged students. Once consequences depend strongly on particular measures, however, those measures can dominate decisions about time and instruction.

The evidence is not uniformly negative. Dee and Jacob compared achievement changes in states with different prior accountability histories. Their study estimated gains in mathematics, especially at grade 4, but did not find corresponding evidence of reading improvement in the working-paper analysis. This challenges both the claim that testing accomplished nothing and the claim that it solved the full learning problem. [S18]

A 2007 Center on Education Policy survey documented curriculum tradeoffs: 44% of surveyed districts reported reducing elementary time in one or more other subjects or activities to expand English language arts and mathematics. That is a reported allocation change, not direct proof that debate or critical thinking declined by a matching amount. It does show how a policy focused on a few outcomes can reshape the broader educational day. [S19]

The mechanism to inspect locally

A strong test may require inference and explanation. Preparation becomes educationally narrow when students mainly rehearse formats, hunt for answer cues, or repeat short passages without building connected subject knowledge. The proper criticism is directed at the tasks and incentives, rather than assuming every standardized question is rote.

A community should ask how much time goes to state tests, district benchmarks, vendor diagnostics, and preparation for each. It should also inspect the quality of the remaining work. Merely reducing test days does not guarantee that the recovered time will support better thinking.

Previous chapter / Contents / Next chapter

III · 2000-2014 / CHAPTER 10

The science of learning versus the marketing of methods

Student activity and effective instruction are not the same measurement.

The 2000 National Research Council synthesis *How People Learn* emphasized prior knowledge, organized understanding, and the active work of making sense of new information. It did not make factual knowledge obsolete. Prior conceptions affect what a learner notices and how new material is interpreted. [S15]

A different issue is the amount of instructional guidance. Kirschner, Sweller, and Clark's 2006 review argued that minimally guided methods often overload novices who lack the knowledge needed to direct their own search. The implication is to examine support, not to assume every project or classroom discussion is ineffective. [S22]

Lazonder and Harmsen's 2016 meta-analysis, covering 72 studies, found that guidance benefits inquiry learning, including an average benefit of half a standard deviation on learning outcomes. That is a standardized comparison across studies, not a 50% increase in scores. A guided investigation can therefore differ substantially from being told to figure it out. The relevant question is which prompts, examples, feedback, and prior teaching the learner receives. [S23]

An example that makes the distinction concrete

Giving students a pile of sources and asking them to discover whether a historical claim is true may reward students who already know the background. A guided version first explains context, models how to check a document's origin, and supplies questions that direct comparison. Students then make and defend their own judgment.

Pashler and colleagues' 2008 review found no adequate evidence base for assigning instruction according to assessed learning-style preferences, such as labeling a student visual or auditory. That finding does not erase disabilities, differences in prior knowledge, or the value of well-chosen diagrams. It challenges a particular matching claim. Schools should expect attractive training packages to meet the same evidentiary standards they ask students to use. [S26]

Previous chapter / Contents / Next chapter

III · 2000-2014 / CHAPTER 11

Modern standards still ask students to reason

The promise on paper is important, and insufficient.

Common Core English language arts standards require students to evaluate arguments and evidence, construct reasoned claims, use sources, and take part in collaborative discussion. This is direct evidence against a blanket claim that modern standards formally removed critical thinking. It is not evidence that every lesson meets those standards or that any particular implementation succeeded. [S20]

The 2013 College, Career, and Civic Life (C3) social studies framework organizes inquiry around developing questions, using disciplinary tools, evaluating sources and evidence, and communicating conclusions and taking informed action. The scientific practices in the National Academies' 2012 K-12 science framework similarly include investigation, interpreting data, and arguing from evidence. These documents explicitly join subject knowledge and intellectual practice. [S30] [S31]

There is also empirical evidence that disciplinary inquiry can work. Avishag Reisman's six-month *Reading Like a Historian* intervention involved 236 eleventh-graders in five San Francisco schools. Its comparison of classrooms, without randomly assigning students, found improvements in historical thinking, transfer to contemporary questions, factual knowledge, and reading comprehension. The local design limits generalization; the result still contradicts the assumption that source-based inquiry must displace knowledge. [S29]

The implementation gap

A standard may say "evaluate" while a worksheet asks students to locate one sentence. A lesson may say "discussion" while the teacher supplies the only acceptable conclusion. A writing task may require a citation without requiring the student to judge whether the source actually supports the claim.

The practical distinction is between verbs in curriculum documents and decisions in student work. A sound review should inspect completed assignments and feedback, including examples from average and struggling students. Standards are a starting point for accountability, not a substitute for looking at what happens.

Previous chapter / Contents / Next chapter

III · 2000-2014 / CHAPTER 12

Why 2005-2014 deserves particular attention

Testing incentives, curriculum redesign, and mobile internet overlapped before 2015.

The evidence does not identify one national turning-point year. It does identify a period when several influential changes converged. By the mid-2000s, NCLB accountability was shaping schedules and curriculum choices. Around 2009-2010, new federal incentives and standards initiatives accelerated reform. At the same time, the internet became increasingly portable and personally accessible.

GAO's 2011 Race to the Top review described almost $4 billion in awards to 12 states, aimed at standards and assessments, teacher and leader effectiveness, data systems, and low-performing schools. The competition could shape state choices even though Washington did not prescribe every lesson. Its significance is the use of incentives to promote a reform package; the oversight report does not establish that the package caused a national reasoning decline. [S88]

The Common Core developers date the final standards to June 2010, after draft releases in 2009 and early 2010. Their reasoning ambitions were discussed earlier; this chronology identifies when a major standards initiative entered the system. A release date is not an implementation date for every district, textbook, teacher, or assessment. [S89]

The information environment changed alongside school policy. Pew reported smartphone ownership among 12-17-year-olds rising from 23% in 2011 to 37% in 2013. Its 2015 survey found that 73% of 13-17-year-olds had or could access a smartphone. These are not directly identical measures: the later item includes access and uses a different age range. Even without treating them as a precise comparable growth rate, they document the rapid arrival of mobile connectivity. [S86] [S87]

When did outcomes signal trouble?

The contemporaneous 2015 NAEP release reported mathematics declines at grades 4 and 8 relative to 2013, a reading decline at grade 8, and no significant reading change at grade 4. It also reported performance above early-1990s baselines. The warning was a reversal or stall after earlier gains, not evidence of unbroken failure throughout the reform era. [S90]

Our chronology judgment: investigate the early and mid-2000s for policy incentives, 2009-2013 for overlapping reform and digital changes, and the following decade for amplification and disruption. There is no evidence here sufficient to select 2005, 2008, 2010, or 2015 as the single cause or onset of a national loss of critical thinking.

Previous chapter / Contents / Next chapter

IV · 2015-2026 in depth / CHAPTER 13

A decade of overlapping changes

There was no single switch, but several pressures arrived together.

The last decade brought changes to accountability, devices, public conflict, instructional priorities, and the conditions of attendance. Some began earlier. Their overlap matters because a student can simultaneously have weaker reading preparation, less sustained attention, fewer opportunities for open discussion, and easy access to a tool that produces finished answers.

The Every Student Succeeds Act (ESSA) was signed in December 2015, replacing NCLB. It retained statewide assessment while giving states greater discretion over accountability. That combination is important: claims that testing simply ended after 2015 confuse changes in consequences and administration with the existence of measurement itself. [S21]

The following chronology is a guide to the investigation, not a claim that every district adopted each practice at the same time.

Period	Documented change or development	Question it raises
2015-2019	ESSA implementation; continuing online life; new evidence of weak digital source evaluation	Were students learning to investigate information or mainly access it?
2020-2021	Closures, remote/hybrid instruction, assessment interruption, widespread disruption	Which students lost instruction and feedback, and for how long?
2021-2024	Intensified curriculum conflict; instructional restrictions; recovery efforts	Did fear or predetermined conclusions narrow discussion?
2022-2026	Generative AI, phone restrictions, literacy reforms, uneven recovery	Can students demonstrate learning without answer-producing assistance?

Analytical synthesis: these are interacting mechanisms, not an experimentally established formula. There is no credible percentage allocation here assigning the national decline to ideology, phones, unions, or one administration. The next chapters separate what is measured from what is proposed as an explanation.

Previous chapter / Contents / Next chapter

IV · 2015-2026 in depth / CHAPTER 14

The downturn began before COVID

The pandemic deepened a problem already visible.

The January 2025 release of 2024 main NAEP results describes reading declines that began before the pandemic and a decade-long widening between higher- and lower-performing students. Mathematics remained above early-1990s performance but below its earlier peak. This is a more informative historical comparison than claiming students have simply become worse every year since the 1950s. [S75]

The pre-pandemic timing rules out explanations that begin only with 2020 closures or the public arrival of ChatGPT in 2022. Those developments may worsen particular problems, but they cannot explain a trend that preceded them. The same discipline applies to a recent training document or state law: a cause must fit the timing and location of the outcome it supposedly caused.

Why weaker foundations matter for thinking

A student trying to judge an argument must understand its vocabulary, retain the relevant details, and know enough context to recognize what is missing. Weak reading or mathematics can make independent reasoning harder even when the student is curious and willing. Conversely, a student can read fluently and still reason selectively. Foundational achievement and critical judgment overlap without being identical.

There is also a distributional issue. An average can hide stronger students maintaining performance while struggling students fall further behind. A school that celebrates high achievers while ignoring its lower performers may appear intellectually healthy from a distance. Communities need subgroup and performance-distribution evidence, together with actual student work, to understand the breadth of the problem.

Supported conclusion: the academic foundation of critical thinking has been under pressure for longer than the pandemic. **Unresolved question:** which mix of earlier instructional, accountability, family, and technological changes accounts for that pre-pandemic deterioration? National score trends alone cannot isolate it.

Previous chapter / Contents / Next chapter

IV · 2015-2026 in depth / CHAPTER 15

Teacher preparation: a concrete weakness to examine

The issue is what candidates learn to teach, and how well they can teach it.

Teacher preparation affects students through subject knowledge, instructional methods, assessment, classroom management, and supervised practice. It is therefore a credible place to investigate. But blaming “education schools” as a single entity is less useful than identifying a demonstrable weakness and tracing it into instruction.

The National Council on Teacher Quality’s (NCTQ) 2023 Reading Foundations review provides one such warning. Using a scoring guide applied to course plans and materials, NCTQ found that only 28% of reviewed programs fully addressed all five core components of scientifically based reading instruction, while 22% did not adequately address any. The review had materials from 702 programs; 441 others did not supply materials or make them publicly available. NCTQ is an advocacy organization using its own evaluation method; it did not directly observe all teaching or establish what every graduate knew. Changes to its scoring guide also prevent straightforward comparison with earlier ratings. Even with those qualifications, the findings justify close scrutiny of preparation for elementary reading. [S32]

The underlying concern is not new. The National Reading Panel’s 2000 findings supported systematic phonics as an important part of reading instruction, integrated with other components. Decoding means using letter-sound relationships to read words; comprehension means understanding the text. A complete reading program also builds fluency, vocabulary, and the ability to hear and work with the individual sounds in words. A program that teaches children to guess words from context while neglecting reliable decoding deserves examination; so does a corrective program that reduces reading to phonics and neglects language and understanding. [S17]

What evidence would answer the local question?

Ask for required course content, the reading materials candidates use, supervised practice, and assessments of whether candidates can diagnose and address student errors. Ask how a preparation provider changes when graduates repeatedly struggle to teach foundational skills.

The defensible finding is that preparation can fail in identifiable ways. This evidence does not establish that every preparation program rejected effective methods, or that racial ideology caused the shortcomings. Those are separate claims requiring separate evidence.

Previous chapter / Contents / Next chapter

IV · 2015-2026 in depth / CHAPTER 16

Teacher preparation and ideology: what the documents say

Equity language is real; its educational meaning depends on the requirements attached.

The Council for the Accreditation of Educator Preparation's (CAEP) revised standards explicitly include reflection on personal biases and the practice of equity, diversity, and inclusion. They also require disciplinary concepts, instructional practice, assessment, clinical partnerships, and responsibility for learning. The presence of equity language is documented. The proposition that these standards instruct teachers to abandon evidence or logic is not established by that language. [S33]

There is a substantive intellectual history behind some contemporary approaches. Paulo Freire's critique of "banking" education opposed treating students as containers for information and emphasized dialogue, problem posing, praxis, and critical consciousness. His work links education with understanding and transforming social conditions. That is a political and philosophical orientation, not merely a technique for checking whether a conclusion follows. [S34]

Gloria Ladson-Billings and William Tate's 1995 article proposed a critical race theoretical approach to education, linking race, property, and inequality. This is an identifiable scholarly framework with claims that can be examined. It should not be confused with every discussion of racism, nor should its existence be used to infer that every classroom teaches the framework. [S35]

A necessary distinction

Critical thinking describes standards for evaluating claims. Critical pedagogy often includes a substantive analysis of power and an emancipatory purpose. They can overlap in examining assumptions; they can also conflict if students must accept the framework's conclusion in advance.

Our evaluation standard: candidates should be able to investigate disparities and teach difficult history while distinguishing evidence from interpretation. They should also be free to scrutinize the framework used in their training. Requiring a student or teacher to profess a contested political explanation as proof of intellectual competence would deserve criticism. Establishing that a specific program does so requires its actual assignments, grading criteria, and practice.

Previous chapter / Contents / Next chapter

IV · 2015-2026 in depth / CHAPTER 17

Social-emotional learning: distinguish the versions

A broad label can conceal different objectives.

Self-control, listening, cooperation, and the ability to tolerate frustration can help a student participate in demanding intellectual work. The research does not justify dismissing all social-emotional learning as ideological distraction. Durlak and colleagues' 2011 meta-analysis covered 213 school-based programs and found average benefits in several outcomes, including academic performance. Those findings belong to the programs and evidence reviewed, not every later product carrying the social-emotional learning (SEL) label. [S48]

The Collaborative for Academic, Social, and Emotional Learning (CASEL) describes "transformative SEL" as an approach that examines inequities, identity, agency, belonging, and community change. The organization distinguishes this from personally responsible and participatory forms, and says most current programs emphasize those other forms. This is documented variation within SEL, not proof that all SEL instruction follows one political program. [S47]

The questions that reveal the difference

Does a lesson help students regulate an emotional response so they can examine opposing evidence? Or does it ask them to classify disagreement as proof of harmful identity or intent? Does it develop a capacity that can be used across viewpoints, or require a particular political conclusion? Are families able to inspect the materials and objectives?

The same scrutiny belongs on claims made by vendors and critics. A broad historical meta-analysis cannot validate a specific new identity lesson. A provocative excerpt cannot establish what every SEL program does. Program titles are poor substitutes for reading the content and observing its effects.

Our recommendation: protect time for academic instruction and evaluate added programs against explicit purposes and outcomes. If a program supports attention, behavior, discussion, or learning, document that benefit. If it displaces valuable work or penalizes reasonable disagreement, document that cost. The relevant unit of judgment is the actual program and implementation.

Previous chapter / Contents / Next chapter

IV · 2015-2026 in depth / CHAPTER 18

Have schools stopped debating ideas?

The polarization hypothesis has evidence, with important boundaries.

There is direct evidence that the political climate constrains discussion. In RAND's 2023 teacher survey, 65% reported choosing to limit classroom discussion of political and social issues. Teachers described concern about upsetting parents and uncertainty about leadership support. Restrictions and community context were associated with those decisions, including spillovers where no formal state restriction applied. This is self-reported behavior, not a before-and-after national observation of all classroom debate. [S49]

Pew's 2024 report, based principally on an autumn 2023 survey, found that 41% of teachers said debates over race and LGBTQ topics harmed their ability to do their jobs. It also found substantial differences among teachers, families, and political groups. These data show real conflict; they do not establish that teachers hold one uniform position or that every discomfort is educationally damaging. [S50]

How polarization can reduce practice

If a teacher expects a recording, a complaint, administrative abandonment, or a reputational penalty, avoiding a difficult question may appear safer than guiding it. Students can learn a similar lesson: say the approved thing, stay silent, or perform certainty for the group. That mechanism is plausible and consistent with the survey reports; the size of its contribution to national academic decline has not been isolated.

The remedy is not to make every topic a free-for-all. Teachers need age-appropriate boundaries, knowledge of the subject, explicit discussion procedures, and backing for fair handling. A student should be able to criticize a policy or an interpretation without being allowed to harass classmates.

Finding: polarization is a supported explanation for some narrowing of discussion opportunities. **Not established:** that debate in every subject has disappeared or that polarization alone explains weak everyday judgment.

Previous chapter / Contents / Next chapter

IV · 2015-2026 in depth / CHAPTER 19

The boundary between teaching and prescribed agreement

Students should learn how to examine a framework, not merely repeat it.

An official NYC Public Schools staff resource explicitly endorses critical pedagogy, discussion of systemic racism and white supremacy, and students as agents of change. It also calls for learning and listening across perspectives. This is a documented example of a district publicly choosing a social and political educational orientation; it should not be dismissed as a claim invented by opponents. The page is evidence of guidance, not direct observation of every NYC classroom. [S45]

An opposing political example also belongs in the record. The Texas Republican Party's 2012 platform included opposition to "critical thinking skills" in a plank targeting outcome-based education, behavior modification, and the undermining of parental authority. The limiting context matters. It was a party platform, not evidence that Texas enacted a statewide ban on reasoning. Its wording nevertheless illustrates how a general educational ideal can become entangled with a faction's distrust of particular programs. [S51]

The same intellectual test applies

A curriculum can study systemic explanations of inequality, constitutional principles, religious history, economic institutions, or family values. The critical-thinking question is whether students can examine claims using evidence and offer a defensible alternative conclusion without a grading penalty for viewpoint alone.

There are also legitimate distinctions between established facts and open interpretations. A teacher need not pretend slavery did not occur, or make a false arithmetic result equally valid, to welcome discussion. The skill lies in identifying what is settled by the evidence, what remains contested, and which moral or policy judgments require additional premises.

Our recommendation: publish substantive objectives and grading rubrics, including what counts as evidence and how dissenting but well-supported work is evaluated. Transparency is more useful than assurances that a program is either entirely neutral or beyond criticism.

Previous chapter / Contents / Next chapter

The controversy is real: start with the original claims

A critique can be examined without assigning moral character by ethnicity.

Claims associating objectivity, logical thinking, or answer-getting with white supremacy are not all inventions or misunderstandings. Attributable materials use that language. A responsible investigation should identify who wrote it, what the surrounding argument says, how it circulated, and what is known about implementation.

Tema Okun's own account traces her characteristics framework to the late 1990s. Its list includes objectivity, perfectionism, individualism, and privileging the written word. She explicitly says she does not intend to deny anyone's ability to excel and warns that the framework has been misused to shame people. Her site even describes reports of the list becoming a weapon within organizations. Those cautions are part of the primary record. [S36]

An earlier version is credited to **Kenneth Jones and Tema Okun** in the 2001 workbook *Dismantling Racism*. Its objectivity section already questioned belief in being objective. That establishes an origin before the recent school controversies; it does not establish how widely schools used it in 2001. [S94]

The objectivity discussion goes beyond the uncontroversial observation that people can be biased. It challenges belief in neutrality and describes circumstances in which claims to rationality protect power or invalidate emotion. That argument contains two different propositions: people can misuse an appearance of objectivity; and objectivity itself is suspect. The first does not logically establish the second. [S37]

Our assessment

Experiences and emotions can identify harms, reveal neglected questions, and explain why an issue matters. The validity of a factual claim still depends on evidence and reasoning. A person's inability to be perfectly unbiased is a reason to use transparent checks, replication, competing explanations, and criticism. It does not establish that truth is determined by identity or power.

The ethnically fair standard is straightforward: no group owns reason, hard work, emotional insight, imagination, or error. Criticizing a training framework's claims does not denigrate the people it describes or the students it intends to help. Equally, the label of antiracism does not exempt a factual or philosophical claim from scrutiny.

Previous chapter / Contents / Next chapter

V · Objectivity & racial ideology / CHAPTER 21

The Smithsonian chart: institutional endorsement, then removal

A consequential 2020 episode with a limited evidentiary reach.

In 2020, the National Museum of African American History and Culture's online race-discussion resources included a chart associating "objective, rational linear thinking" with white culture, alongside other listed cultural assumptions. A subsequent government legal opinion preserves the wording and an archived-page reference. The original chart is no longer on the live page, so readers should distinguish the preserved record from a currently displayed museum statement. [S39]

The museum then removed the chart. Its own public statement, preserved in a transcript of its social-media thread, said the chart did not contribute to the productive discussion it intended and emphasized listening and reconsideration. The removal is part of the episode, not a detail to omit because it complicates an argument. [S38]

What the episode establishes

A prominent educational institution gave a cultural framework a platform. That matters because such institutions influence educators and public discussion. It is legitimate to question racial classifications that appear to attach ordinary intellectual or personal virtues to one group, even when the intended subject is historical cultural dominance.

What it does not establish

The episode does not demonstrate that the Smithsonian directed public-school curriculum, that teachers nationwide taught children to reject logic, or that the chart caused a measurable decline in student reasoning. A widely circulated screenshot cannot provide those missing steps.

Our judgment: the chart's formulation invited misleading ethnic generalization and confusion about the universality of intellectual capacities. Its removal was a correction worth recording. A productive classroom could examine the controversy itself: What was the claim? Was it descriptive, normative, or ambiguous? What evidence would establish it? Did the language communicate more than its authors intended?

Previous chapter / Contents / Next chapter

Equitable Math: the disputed wording and the authors' answer

A training resource whose real claims deserve a precise critique.

The 2020 *Dismantling Racism in Mathematics Instruction* workbook uses a white-supremacy-culture framework. In the indexed PDF passage retrieved for this review, the criticized condition is: "There is a greater focus on getting the 'right' answer than understanding concepts and reasoning." The surrounding text also challenges objectivity and argues that some mathematical problems can have multiple answers or no solution. The same passage acknowledges that most math problems have correct answers and explains how some word problems allow different interpretations. The full PDF could not be fetched for visual page verification; this discussion relies on the provider-hosted indexed passage and its surrounding retrieved text. [S40]

The authors' FAQ denies that math itself is racist or that Black and Latino students should receive lower standards. It frames the project as expanding access, engaging students, and developing deeper understanding. That is their stated position, and it deserves to be presented accurately. [S41]

What deserves criticism?

Distinguishing a correct answer from an explanation is educationally sound. It does not follow that correct answers or objective validity are racial properties. Multiple solutions to an equation, multiple methods of solving it, and multiple reasonable modeling assumptions are different mathematical situations. Each still permits tests of correctness under the stated conditions.

For example, 2 + 2 = 4 in ordinary integer arithmetic. A problem asking how to divide a budget can have many feasible allocations, evaluated against competing goals. Calling the latter open-ended does not make the former oppressive. Better instruction clarifies the kind of question and the criteria for a valid answer.

Our assessment: the workbook contains an identifiable pedagogical concern about shallow answer-getting, expressed through a contentious racial and philosophical framework. Communities can accept the need for conceptual reasoning while rejecting a racialized explanation of mathematical validity. This is a reason to inspect actual training and assignments, not to assert that all equity-oriented math teaching rejects correct answers.

Previous chapter / Contents / Next chapter

V · Objectivity & racial ideology / CHAPTER 23

How influential were these ideas?

Distribution is documented; national classroom prevalence is not.

The materials did circulate through educational channels. Student Achievement Partners offered a course series using Stride 1 as a foundational resource. Loyola Marymount University's Center for Equity for English Learners describes its reviewer role and links the toolkit and webinars. These are identifiable professional-development and university connections. They establish more than an isolated personal blog post. They do not establish the number of teachers trained, whether teachers followed the approach as intended, or effects on students. [S42] [S43]

California's adopted 2023 mathematics framework illustrates the need to separate related but nonidentical texts. The state's explanation explicitly includes rigorous standards, conceptual understanding, procedural fluency, argument, and routes to advanced mathematics. It also emphasizes equity and changes to instruction and grouping. The official account is evidence of intended guidance, not a guarantee that every recommendation is well-supported or that every district implements it well. [S44]

Three levels of influence

Availability: a provider posts a document. **Institutional use:** a university, district, or professional-development provider recommends or requires it. **Classroom implementation:** students actually encounter the associated practices. A fourth step, learning effects, needs its own evidence. A chain cannot be assumed complete because its first or second link is visible.

V · Objectivity & racial ideology / CHAPTER 23 / CONTINUED

The evidence reviewed here establishes documented circulation and particular institutional cases. It does **not** yield a defensible national percentage of schools teaching that objectivity is racist. Nor was there a credible national causal estimate linking these specific frameworks to declining NAEP results.

A further caution runs in both directions. Bonilla, Dee, and Penner studied a San Francisco ethnic-studies course by comparing students just above and below the cutoff used for course assignment. They found improvements in school completion and engagement. That local evidence does not validate every ideological framework; it does undermine the claim that any race-conscious curriculum must damage educational outcomes. [S46]

WATCH & EXAMINE

Dismantling Racism in Mathematics Instruction

Primary perspective: hear the proponents' framing, then compare it with the workbook and the critique in this presentation. Inclusion is an opportunity to examine the argument, not an endorsement. [S85]

Watch original video · Publisher / supporting page

Select a video link to watch online. Playback and captions are provided by the video host.

Previous chapter / Contents / Next chapter

Teachers’ unions: follow the actual mechanisms

Bargaining, elections, professional development, and advocacy.

Teachers’ unions can affect education through negotiated working conditions, pay, staffing rules, election activity, lobbying, litigation, member mobilization, and professional resources. Their educational effect is unlikely to be captured by asking whether an organization is simply good or bad. An agreement can protect planning time; a policy can also create barriers to changing assignments or responding to weak performance. The relevant question is the actual rule and its consequences.

The National Education Association’s (NEA) own 2021 assembly statement identifies racial-justice priorities and opposition to restrictions on teaching race and racism. Those are substantive public positions, not merely wage bargaining. They are evidence of the organization’s stated agenda, not proof that every member agrees or implements it. [S52]

The American Federation of Teachers (AFT) also publishes cognitive-science and instructional material. Daniel Willingham’s 2007 essay on knowledge and critical thinking and Barak Rosenshine’s 2012 principles of instruction appeared in its *American Educator*. That complicates any account in which union activity can only mean opposition to evidence-based teaching. The quality of a publication does not settle the effects of bargaining or politics, but it belongs in an honest institutional description. [S24] [S27]

What would establish responsibility?

Identify a specific union proposal, agreement, financial commitment, or intervention. Establish whether the decision-maker adopted it. Then examine implementation and outcomes. If a rule obstructs effective action, say so. If a training program improves practice, say so. If a political statement has no demonstrated classroom effect, do not invent one.

This investigation does not find evidence supporting the blanket conclusion that unions collectively eliminated critical thinking. It does find particular channels of influence substantial enough to examine, especially financial advocacy and pandemic reopening. Those deserve direct treatment.

Previous chapter / Contents / Next chapter

The money is substantial, and categories matter

Financial capacity is evidence of resources, not automatic proof of control.

The AFT national headquarters' LM-2 report for July 1, 2025 through June 30, 2026 lists approximately **$299.7 million in receipts**, **$297.0 million in disbursements**, and **$37.0 million in political activities and lobbying**. The filing's national-headquarters identity and reporting period matter. These are its reported federal disclosure categories, not a combined total for all affiliates or a measure of every school-related political dollar. The indexed filing was retrieved; direct-page access returned an error during review. [S53]

The NEA Fund for Children and Public Education's FEC summary shows approximately **$3.96 million in receipts** for the 2025-2026 cycle through October 14, 2026. This is one federal committee with one reporting window, not the NEA's entire budget. Its contribution disclosure says participation is voluntary. Union operating funds, lobbying, grants, federal candidate PACs, and state political activity must not be casually added together or treated as interchangeable. [S54] [S55]

Does this mean unions are more powerful than ever?

The figures establish substantial contemporary resources. They do not establish an all-time maximum of influence. That claim needs comparable long-term measures, adjusted for inflation, membership, legal authority, rival spending, and outcomes. This investigation did not establish that maximum.

The 2018 *Janus* decision, which barred compelled public-sector agency fees from nonconsenting nonmembers, is also inconsistent with a simple assumption that legal leverage has only increased. Different states and periods present different bargaining environments. [S60]

Our conclusion: unions have meaningful political capacity, and public scrutiny of spending and policy priorities is warranted. A dollar figure alone cannot prove a particular curriculum's adoption or explain a decline in students' reasoning. Follow the money to a specific activity, then follow that activity to a decision and a measurable result.

Previous chapter / Contents / Next chapter

Pandemic reopening: a documented union connection

A specific finding should neither be minimized nor stretched.

Marianno and colleagues studied 250 large school districts during fall 2020. Their measures of established institutional union power, particularly district size and collective-bargaining-agreement length, were associated with a lower likelihood of in-person reopening and fewer weeks of in-person instruction. Their social-media measures of more reactive union pressure were not similarly predictive. This is an observational finding, not an experiment, and it does not assign all closure decisions to unions. [S56]

That finding belongs next to evidence about instruction. Goldhaber and colleagues analyzed roughly 2.1 million students in about 10,000 schools and found remote instruction was a major contributor to widening achievement gaps, especially by school poverty. The study supplies a credible educational mechanism through which prolonged remote instruction could matter. [S66]

What follows, and what does not?

The two studies support serious scrutiny of union influence on reopening and of the costs of lost in-person instruction. They do not combine into a precise estimate of learning loss caused by unions. District politics, family preferences, health concerns, local conditions, and administrative decisions also affected reopening. Leaders made decisions under changing information; those circumstances do not exempt decisions from retrospective evaluation.

A fair review should ask who argued for which reopening conditions, what alternatives were available, what evidence was used at each point, and how learning and other outcomes were monitored. It should also distinguish the emergency in spring 2020 from later periods when conditions and knowledge differed.

Finding: a documented association connects established union power with slower reopening in large districts in fall 2020. **Our assessment:** this is stronger and more specific evidence than a generalized claim about agendas, and it should be included plainly. It still does not establish unions as the sole cause of pandemic damage or the earlier learning downturn.

Previous chapter / Contents / Next chapter

Does bargaining improve or damage learning?

The research does not produce a universal verdict.

Lovenheim and Willén's 2019 study used the historical timing of state duty-to-bargain laws to estimate long-run student outcomes. It found adverse effects on adult labor-market outcomes for men, with larger estimates among nonwhite men and no evident effects for women. The study is serious evidence of potential institutional costs. Its historical groups of students and labor-market outcomes do not directly measure current classroom reasoning. [S57]

Wisconsin's Act 10 illustrates how difficult a simple verdict can be. An earlier working paper by E. Jason Baron reported short-run test-score losses after the weakening of bargaining. A 2025 National Bureau of Economic Research (NBER) working paper by Barbara Biasi and Wayne Aaron Sandholtz instead reported test-score gains, lower union revenues, and political benefits for the reform's incumbent sponsor. Both exploit implementation differences, but they did not necessarily examine the same years, use the same comparisons, or estimate the same effects. [S59] [S58]

What should a reader do with conflicting findings?

Do not resolve them by selecting the result that fits a preferred view. Ask which students, years, outcomes, and comparison groups each study covers, whether later evidence revises an earlier interpretation, and which assumptions make the estimate credible. Working papers also require more caution than treating a finding as settled consensus.

Our synthesis: bargaining institutions can affect resources, management, incentives, and workforce behavior. Those effects can differ across reforms and contexts. There is evidence supporting criticism of particular arrangements, but this review cannot assign a single nationwide critical-thinking effect to unionization.

For local action, the practical unit is the contract clause, staffing practice, evaluation rule, or bargaining proposal. Review its purpose, costs, and evidence. A school board should be able to defend the educational consequences of an agreement, and a union should be able to explain how its proposals serve students as well as members.

Previous chapter / Contents / Next chapter

Government responsibility: identify the level and decision

Washington influences incentives; it does not write every local lesson.

The Education Department's May 2026 explanation states that federal law limits its direction or control over curriculum, instruction, personnel, and materials. Standards, graduation requirements, and curriculum decisions are primarily state and local responsibilities. This does not make federal policy irrelevant: funding rules, assessment requirements, civil-rights enforcement, grants, and research can strongly influence what other actors do. [S61]

Responsibility should therefore be traced through the decision chain. Congress can establish funding and accountability conditions. A state can set standards and adopt assessments. A district can purchase materials, require training, and set grading policies. A principal can determine schedules and instructional expectations. A teacher makes further choices within those constraints. A decision at one level should not be attributed to another without evidence.

Recent federal changes

The March 20, 2025 executive order directed steps toward closing the Department to the extent permitted by law while preserving delivery of services. By the research cutoff, the Department's own implementation page described interagency partnerships, including expanded administrative roles for other departments, while stating that statutory protections and authorities continue. This is an administrative transition, not evidence that all federal educational functions have disappeared. [S62] [S63]

Claims that these changes will improve reasoning are predictions until outcomes can be evaluated. The same applies to predictions of harm. A reorganization may reduce duplication or disrupt capacity; the answer depends on the implementation and evidence.

Our recommendation: judge governments by particular choices: the quality and comparability of assessment, evidence behind curricular guidance, transparency of expenditures, protection of inquiry, and support for effective instruction. Changing which office administers a program does not, by itself, improve what a student can explain or solve.

Previous chapter / Contents / Next chapter

2020-2021: instruction and measurement were both interrupted

Missing information should not be mistaken for missing damage.

The pandemic changed schooling abruptly. The Education Department invited waivers of 2019-2020 assessment and accountability requirements because closures made normal administration impractical. This was a documented interruption of measurement, not a permanent repeal of statewide testing. [S64]

For 2020-2021, federal data notes warn that assessment participation varied widely and results might not represent the full population or be comparable with other years. A higher passing rate among the students tested could therefore coexist with greater learning difficulty among those absent from the assessment. Before interpreting the number, the reader needs to know who was missing. [S65]

Instruction also changed. Remote classes could preserve contact and offer useful resources, but the ability to use them depended on circumstances and support. The large remote/hybrid study discussed in the reopening chapter found substantial differences associated with instructional mode and school poverty. The point is not that every remote lesson failed; it is that access to a screen was not an equivalent replacement for all the functions of school. [S66]

Why this could affect independence

A classroom ordinarily supplies repeated opportunities to attempt work, reveal misunderstanding, receive feedback, and try again. When those cycles weaken, gaps can accumulate unnoticed. A submitted assignment can conceal how much help produced it. A quiet student can disappear from the teacher’s view.

These are plausible mechanisms consistent with the broader learning-loss evidence. They do not establish a permanent incapacity in students who experienced the disruption. Recovery depends on subsequent instruction and opportunities. The response should be demanding support and reliable measurement, not writing off a cohort as incapable of thought.

Previous chapter / Contents / Next chapter

VII · Disruption & technology / CHAPTER 30

Absence, recovery, and the missing practice

A curriculum cannot work as intended when students repeatedly miss it.

Chronic absence usually means missing at least 10% of school days for any reason. Federal reporting placed the rate around 31% in 2021-2022 and 28% in 2022-2023. Illness, disengagement, caregiving, work, transportation, and other barriers can enter the problem; an absence rate alone cannot diagnose a student’s attitude. [S68]

The May 2026 Education Scorecard release, using linked state-test data through 2024-2025, reported a mixed recovery. It identified more than 100 districts improving faster than peers and larger gains among the highest- and lowest-income districts than among middle-income districts. It also reported chronic absence at 23% in 2024-2025, still above the 15% pre-pandemic level. These are findings from that research project and its linked assessments, rather than a replacement for NAEP’s distinct samples and measures. [S67]

A practical implication

Independent reasoning requires repeated encounters with problems and feedback. A student who misses the introduction, attends the application lesson, then misses the correction may encounter a confusing sequence even when each individual lesson is well designed. More ambitious assignments cannot by themselves repair the missing foundation.

Recovery therefore needs more than a promise to restore rigor. Schools must identify specific gaps, provide instruction at an appropriate level without closing off advancement, and check whether students can perform independently. Attendance improvement and academic improvement should be evaluated together.

A community also needs honest communication about the pace of recovery. Some improvement is not full recovery. A return to 2019 would not solve difficulties that existed in 2019. The recent record contains both encouraging examples and persistent deficits; local leaders should be expected to explain which description fits their own students.

Previous chapter / Contents / Next chapter

VII · Disruption & technology / CHAPTER 31

Phones and devices changed the environment of thought

The distinction is purpose, attention, and independent practice.

By the 2024-2025 school year, 88% of public schools reported a one-to-one computing program. In the same School Pulse Panel release, 53% of school leaders reported a negative effect of phones on academic performance and 73% on attention span. These are leaders' perceptions in an experimental survey product, not experimentally measured effects of removing a phone. They nevertheless document a substantial institutional concern and the scale of device availability. [S69]

The Programme for International Student Assessment (PISA), an international comparison of student skills, reported in its 2022 U.S. note that 30% of students said they were distracted by digital devices in most or all mathematics lessons. It also reported a U.S. mathematics decline from 2018 while reading and science were broadly stable. Technology cannot be used as a shortcut explanation that ignores the actual variation across measures. [S70]

What is educationally different?

A textbook does not normally notify a student about messages, social standing, entertainment, and unrelated information while a problem is being solved. A school device can provide extraordinary resources, accessibility, feedback, and simulations. The same device can also make task-switching and answer-finding routine. These are different uses of technology with different learning demands.

Our practical standard: identify the task the device improves, remove avoidable distractions, and test what students can do afterward without it. A spreadsheet can help students explore a model; they should still explain the assumptions and interpret the result. A search engine can locate a source; students should still judge its relevance and reliability.

Phone restrictions are a plausible tool for protecting attention, but their enforcement, exceptions, and effects need evaluation. A policy written in a handbook does not establish that interruptions have actually decreased. Communities should ask for classroom evidence, student experience, and independent learning outcomes.

Previous chapter / Contents / Next chapter

VII · Disruption & technology / CHAPTER 32

Digital fluency did not guarantee source judgment

A direct test of a modern reasoning problem.

Between 2018 and 2019, researchers assessed 3,446 high-school students using six open-internet tasks. The sample matched the national demographic profile. Two-thirds could not distinguish a news item from sponsored content on the task provided. On another task, 96% did not uncover a climate website's fossil-fuel-industry connections. More than half treated a misattributed video as strong evidence of U.S. voter fraud. These are striking failures on specific tasks, not a diagnosis of every student or proof that earlier generations would have succeeded. [S71]

The issue is deeper than knowing whether a website looks professional. Students may need to leave a site, investigate its publisher, find the original evidence, and compare independent accounts. A sponsor's identity does not automatically make a claim false; it provides context for scrutinizing what evidence has been selected or omitted.

There is evidence of teachability

A district-wide high-school government study examined six 50-minute lessons in lateral reading: investigating an unfamiliar source across the open web before relying on its self-description. The study reported gains in source evaluation relative to comparison classrooms. This is evidence for a concrete instructional approach, not a promise that one short course permanently inoculates students against deception. [S72]

A useful classroom task might ask students to identify the original date and location of a circulating video, then document the search path that established the answer. Assessment should reward the investigation, not whether the student already agrees with the teacher's favored news source.

Our conclusion: weak online source evaluation is directly documented, and explicit instruction can help. Digital literacy belongs within critical thinking, alongside sustained reading and subject knowledge. Familiarity with an interface is not evidence of sound judgment about what appears on it.

WATCH & EXAMINE

Check yourself with lateral reading

A practical demonstration of checking information across sources. Ask which underlying evidence supports a claim, including claims made by a trusted publisher. [S84]

Watch original video · Publisher / supporting page

Select a video link to watch online. Playback and captions are provided by the video host.

Previous chapter / Contents / Next chapter

Grading and testing: when the signals stop agreeing

A comforting grade can hide a genuine learning need.

ACT's subject-specific analysis reported that adjusted mathematics GPA increased from 3.02 to 3.32 between 2010 and 2022 without corresponding improvement in ACT achievement. It identified grade inflation across the core subjects. This is evidence from ACT-tested students and an assessment provider with an interest in the usefulness of external tests; it is not a census of every school or proof that every higher grade was undeserved. [S73]

The core question is whether a reported grade communicates what a student can do. Policies such as retakes, revisions, minimum grades, completion credit, and late-work flexibility have different purposes and designs. It would be inaccurate to treat all of them as equivalent or assume a single national policy. The same rule can have different effects depending on the work required and the information families receive.

An illustrative contrast

A student revises an argument after detailed feedback and then defends a new claim independently. Another student receives a passing mark for incomplete work without demonstrating the missing knowledge. Both may appear under a broad label such as flexibility. Only the first example supplies evidence of additional learning.

The remedy is not to worship a test score. Grades can reflect extended work and persistence that a short test misses. External assessments can reveal a discrepancy that classroom measures miss. Strong systems use multiple measures and explain the differences rather than selecting whichever number is most reassuring.

Communities should ask whether changes in passing rates reflect stronger work, easier thresholds, different participation, or changes in grading. They should also distinguish test-optional college admissions from K-12 assessment requirements. A college's application policy does not itself end a state's school-testing obligations.

Previous chapter / Contents / Next chapter

VII · Disruption & technology / CHAPTER 34

Generative AI: finished work is not necessarily learning

A new tool exposes an old distinction between assistance and independence.

The public spread of generative AI from late 2022 made it much easier to obtain a fluent explanation, completed essay, or worked solution. That timing means AI cannot explain earlier declines. It can, however, change how much practice a student performs and how much of a submitted product demonstrates the student's own understanding.

Bastani and colleagues' 2025 randomized field experiment involved nearly 1,000 high-school mathematics students in Turkey. Both a general AI interface and a guarded tutor improved assisted practice performance. When AI was removed, the general-interface group performed 17% worse than the control group; the guarded tutor largely avoided that harm but did not show a clear independent-test benefit. This is one setting and intervention, not a national U.S. estimate. The August 2025 correction fixed an author's university affiliation; it did not revise the reported findings. [S93] [S74]

The key distinction

A tool that supplies the answer may improve today's product while reducing the need to retrieve, attempt, fail, and correct. A tutor that gives a hint or asks for an explanation can preserve more of that work. Design, timing, and assessment matter more than the simple presence of AI.

Our recommendations: require an initial independent attempt, use assistance to address a specific obstacle, check factual claims against original sources, and include a brief unaided explanation or transfer problem. Students should be able to describe what the tool contributed and what they verified.

Teachers need analogous safeguards. AI-generated lesson plans and feedback should be checked for factual accuracy, curricular fit, and hidden assumptions. The goal is not a more polished stream of unexamined material. It is a better opportunity for the student to do the intellectual work that produces lasting understanding.

Previous chapter / Contents / Next chapter

VIII · What the outcomes show / CHAPTER 35

The 2024 results: serious academic warning signs

Reading, mathematics, history, and civics tell related but distinct stories.

The younger grades show both the scale of the reading setback and the limits of a blanket decline narrative. National main-NAEP reading averages, including public and private schools, followed this pattern. [S91]

Main NAEP reading	2019	2022	2024
Grade 4 average score	220	217	215
Grade 8 average score	263	260	258

These are scores on 0-500 scales, not percentages of questions correct or percentages of students passing. Both grades declined again after 2022; the source uses unrounded estimates for significance testing. Different students were sampled in each assessment, so this is not a record of the same children losing five points.

Mathematics differed: the grade-4 average improved by two points from 2022 to 2024, with gains concentrated among middle- and higher-performing students; lower performers did not show significant improvement. Eighth-grade mathematics showed no significant average recovery from the major 2022 drop. These differences help identify where the recovery is failing. [S92] [S75]

Students approaching graduation

The grade-12 results make the concern particularly concrete. In 2024, 32% of twelfth-graders performed below NAEP Basic in reading, compared with 30% in 2019. Thirty-five percent were at or above Proficient. NAEP explicitly warns that Proficient is not the same as a state or district’s grade-level benchmark. “Below Proficient” must not be translated into “unable to read.” [S76]

In grade-12 mathematics, 45% performed below Basic in 2024, five percentage points above 2019. This indicates a substantial weakness in measured mathematical preparation. The assessment uses a different scale from grades 4 and 8, so scores should not be combined across grades as if they were interchangeable. [S77]

Assessment	Comparison	Documented result
Grade 12 reading, below Basic	2019 to 2024	30% to 32%
Grade 12 mathematics, below Basic	2019 to 2024	40% to 45%
Grade 8 U.S. history, mean score	2018 to 2022	Down 5 scale points
Grade 8 civics, mean score	2018 to 2022	Down 2 scale points

The 2022 history and civics results add evidence about knowledge and reasoning relevant to citizenship. The mean declines were real; these are subject assessments, not a stand-alone national critical-thinking score. [S78]

What can be concluded: many students approach adulthood with serious gaps in skills and knowledge needed for independent judgment. **What cannot be concluded from these results alone:** which institution caused the gap, whether an individual lacks common sense, or that every educational method changed in the same way.

Previous chapter / Contents / Next chapter

The 2026 update: recovery exists, and is uneven

An honest account must include improvement as well as decline.

The 2025 NAEP long-term trend assessments, released in June 2026, showed four-point gains in both reading and mathematics for 9-year-olds compared with 2022. Their reading average was not significantly different from the pre-pandemic 2020 result; mathematics remained four points lower. For 13-year-olds, neither subject showed a significant improvement from 2023, and both remained below 2020. [S79]

These assessments are age-based and separate from the main grade-based NAEP program. The results should not be spliced into a grade-4 or grade-8 series. The proper conclusion is a differentiated one: improvement for younger students, continued difficulty for older students, and different historical comparisons by subject.

The accompanying federal release also reports that 14% of 13-year-olds said they read for fun almost every day in 2025, compared with 35% in 1984. That is a notable change in reported practice, not proof that a single device or curriculum caused it. The same release notes that age-13 mathematics remained above the earliest 1970s assessment, while reading was not significantly different from 1971. [S80]

Why the counterevidence matters

If the claim is that critical thinking is inevitably vanishing everywhere, improvement is inconvenient. If the goal is to improve education, improvement is valuable information. It shows why the investigation must examine which students, subjects, and settings are changing.

Our assessment: the record supports an urgent concern about adolescent preparation and persistent lower-performance gaps. It does not support an undifferentiated story of continuous decline across every measure. Effective local inquiry should look for improving peers as carefully as it investigates failures.

Previous chapter / Contents / Next chapter

What about twenty-somethings and everyday common sense?

The concern is reasonable; the evidence needs the right comparison.

Adults' difficulties with judging information, managing unfamiliar tasks, or making decisions cannot be explained solely by the latest school policy. Their schooling, work, family experiences, opportunities, and current information environment all matter. An anecdote can reveal a problem worth investigating, but a collection of memorable encounters cannot establish a population trend.

The Programme for the International Assessment of Adult Competencies (PIAAC) provides relevant U.S. adult evidence from 2023. Average literacy and numeracy declined from 2017, and the share at Level 1 or below rose from 19% to 28% in literacy and 29% to 34% in numeracy. NCES warns that administration and scoring changes require caution in cross-cycle comparisons. Its adaptive-problem-solving assessment was new in 2023, so it cannot establish a comparable historical decline in that measure. The overall adult results also do not isolate people in their twenties. [S81]

A useful cohort check

Someone aged about 29 in 2026 typically began school around 2002 and finished high school around 2015. Someone aged about 20 typically began around 2011 and finished around 2024. These are illustrative calculations, allowing for birthdays and different school paths, not observed outcomes. The range makes clear why 2015 cannot serve as the sole origin of a problem attributed to today's young adults.

An employer's observation that a new worker cannot troubleshoot may reflect missing subject knowledge, unfamiliarity with that job, reluctance to act without permission, weak reasoning, or some combination. A useful response tests the specific difficulty: Can the person explain the goal, identify constraints, propose alternatives, check the result, and learn from feedback?

The educational aim is broader than test preparation. It is the capacity to act responsibly when the answer is not supplied. That aim is best served by precise diagnosis and practice, not by dismissing an entire generation.

Previous chapter / Contents / Next chapter

IX · Rebuilding the practice / CHAPTER 38

What the instructional evidence supports

Knowledge, explicit modeling, dialogue, and independent application.

The evidence does not support a magical generic thinking skill detached from subject matter. Willingham’s account emphasizes that relevant knowledge and repeated practice make sound reasoning possible. A student can know the advice to consider alternatives without knowing enough to identify useful alternatives in a particular field. [S24]

Abrami and colleagues’ 2015 meta-analysis gives reason for optimism about instruction: across 341 estimates of instructional effects, the average benefit was 0.30 standard deviations, with substantial variation among studies. This standardized measure compares improvement across different tests; it does not mean a 30% gain. Dialogue, authentic problems, and mentoring were associated with stronger outcomes. This is not a guaranteed effect for every classroom, and it spans different levels and disciplines. It does show that explicit instructional effort can improve measured critical thinking. [S25]

Rosenshine’s principles supply practical building blocks: review earlier learning, teach in manageable steps, model processes, ask questions, check understanding, guide practice, and move toward successful independent work. These are compatible with demanding thought rather than limited to mechanical drill. [S27]

Nor should modern active learning be equated with intellectual softness. Freeman and colleagues’ meta-analysis found benefits across 225 undergraduate studies in science, technology, engineering, and mathematics (STEM). Its setting limits direct K-12 generalization, but it challenges the assumption that listening to a lecture is inherently more rigorous than supported problem solving. [S28]

IX · Rebuilding the practice / CHAPTER 38 / CONTINUED

Our proposed instructional sequence

Build knowledge to model reasoning to guide an attempt to compare alternatives to require independent application to revisit errors. This is a synthesis of the evidence and principles, not a separately tested branded program.

Within that sequence, students should encounter real objections, explanations of why an error is an error, and opportunities to revise. The endpoint is not a polished answer produced with unlimited help. It is increasingly competent judgment that the learner can demonstrate when the support is reduced.

WATCH & EXAMINE

Five tips to improve your critical thinking

A concise instructional introduction. Use it as a prompt for practicing the process, not as evidence about historical school performance. [S83]

Watch original video · Publisher / supporting page

Select a video link to watch online. Playback and captions are provided by the video host.

Previous chapter / Contents / Next chapter

A classroom that permits genuine disagreement

Clear standards can protect both dignity and inquiry.

A classroom can require civility without requiring agreement. The teacher’s task is to identify a worthwhile question, supply enough background for informed discussion, clarify which claims can be tested, and help students distinguish evidence, inference, and value judgments. Our proposed discussion structure follows.

First, prepare. Students read a common source or examine a shared problem. They identify the author, context, claim, and supporting evidence before being asked to take a position. Participation begins with something more substantial than an immediate reaction.

Second, compare. Students state an interpretation and then give a fair account of the strongest alternative. The discussion asks what evidence distinguishes the two, what assumptions each needs, and which new information would change the judgment.

Third, revise. After discussion, students write a short explanation of what they now think and why. They identify at least one objection and explain whether it changed their conclusion. Revision can mean stronger confidence as well as a changed mind, provided the reasons improve.

A proposed grading rubric

Dimension	What the teacher evaluates
Knowledge	Accurate understanding of the relevant content
Evidence	Relevant, credible support that actually bears on the claim
Reasoning	Clear connections between evidence, assumptions, and conclusion
Alternatives	Fair treatment of objections and competing explanations
Independence	Ability to explain and apply the judgment without answer-producing help

A disagreement should not become a personal attack. Equally, discomfort with an argument should not automatically settle whether it is valid. Students deserve protection from harassment and meaningful opportunities to examine ideas. Administrators should make that distinction clear enough that teachers need not invent it alone in the middle of a dispute.

Previous chapter / Contents / Next chapter

IX · Rebuilding the practice / CHAPTER 40

An evidence ledger: which explanations survived?

Several mechanisms have support; a single national culprit does not.

The table summarizes the investigation’s judgments. The linked chapters and sources contain the evidence and limits; the labels are editorial assessments, not numerical causal weights.

Proposed explanation	Assessment	Boundary
Students have recently lost academic foundations needed for reasoning	Strongly documented on several assessments	Does not directly measure every form of judgment
Schools formally removed all critical thinking	Contradicted as a blanket claim	Standards can promise what instruction fails to deliver
Test-based accountability narrowed some curriculum	Documented reports and credible incentive mechanism	Also associated with some mathematics gains
Teacher preparation has identifiable weaknesses	Supported, especially in the reviewed reading audit	Variation and audit limits prevent universal conclusions
Polarization discourages difficult discussion	Supported by teacher reports	National effect on reasoning proficiency not isolated

Proposed explanation	Assessment	Boundary
Some educational materials racialize objectivity	Documented	National classroom prevalence and learning effects unmeasured
Union power affected pandemic reopening	Documented association in large districts	Not a measure of all closures or all union effects
Phones and answer-producing technology can displace practice	Plausible, with direct evidence in some settings	Device use, purpose, and causal designs differ
The problem began in 2015	Not supported as a single starting point	Earlier reforms and weaknesses are substantial

The missing causal percentages

There is no defensible pie chart in this investigation assigning 30% of the problem to unions, 20% to ideology, or another fraction to parents or phones. Available studies examine different outcomes and populations. Their effects cannot be added as if they partitioned a common national total.

This is not a reason for inaction. A district can correct weak instruction, restore discussion, improve grading information, and reduce distraction without solving every national causal question first. It should measure whether the specific correction produces the intended result.

Previous chapter / Contents / Next chapter

IX · Rebuilding the practice / CHAPTER 41

What families, school boards, and communities can do

Examine student work and results, then choose a manageable improvement.

Begin with a small, concrete review. Ask to see a recent grade-level assignment requiring students to use evidence, a sample of completed work, and the rubric used to evaluate it. Include ordinary and struggling students' work, not only a showcase. Ask what students can now do independently that they could not do at the start.

Review the reading and knowledge foundations beneath the assignment. A student who cannot understand the source may need explicit instructional support before a debate task becomes useful. Review schedules and the opportunity cost of added tests, software, and training. Ask which decisions are required by the state and which were made locally.

Inspect professional-development content directly. If a training makes claims about race, objectivity, learning styles, or a program's effectiveness, request the original text and evidence. Identify whether it is optional or required, who approved it, its cost, and how its educational effect will be evaluated. Criticism should name the document and the decision rather than assign motives to every teacher.

A proposed 90-day local inquiry

Stage	Action	Evidence to retain
Days 1-30	Inspect assignments, assessment participation, attendance, training, and device rules	Baseline student work and a clear problem statement
Days 31-60	Pilot one change, such as weekly evidence discussions or protected independent practice	Comparable work, observation notes, and implementation record
Days 61-90	Compare outcomes and listen to students, families, and teachers	What improved, for whom, at what cost, and what remains unresolved

This schedule is a practical proposal, not a validated intervention. Avoid demanding a sweeping overhaul on the strength of one disappointing example. Equally, do not accept vague promises when repeated evidence shows a problem. Seek clear responsibility, a specific change, and an agreed way to judge whether it helped.

Previous chapter / Contents / Next chapter

The adult stakes, and the responsibility to improve

Independent judgment is a public good and a personal capability.

Reasoning matters when someone interprets a contract, compares financial offers, evaluates a medical claim, troubleshoots equipment, hires a worker, votes, or decides which information to share. These are illustrative applications of the capacities examined here, not measured effects of the reforms proposed here.

The danger is not simply that students know fewer facts. It is that fluent answers, pleasing grades, factional agreement, and confident presentation can become substitutes for understanding. A person may have access to abundant information and still lack the knowledge and habits needed to use it wisely.

Institutions should be held accountable for the parts they control. Preparation programs should teach methods with a clear basis in research. Districts should inspect the learning their purchases and policies produce. Unions should answer for their actual interventions. Governments should preserve useful measurement and justify their requirements. Families and students also have roles in attendance, practice, attention, and honest work. Responsibility is shared, but a shared responsibility must not become an excuse for nobody to answer.

Resources matter as well as incentives. Jackson, Johnson, and Persico’s finance-reform research linked sustained spending increases to improved later outcomes, particularly for children from low-income families. That does not prove every expenditure is effective. It does challenge the assumption that instructional conditions and resources are irrelevant. [S82]

The central conclusion: schools need to make a few demanding practices routine: teach enough knowledge for students to judge a claim, require evidence and clear explanations, allow thoughtful disagreement, and build the ability to work independently. Recent academic losses make that work urgent. Earlier policy changes, longstanding instructional weaknesses, and newer disruptions all deserve attention. Communities can act now by choosing a specific improvement and checking whether students actually learn more.

Previous chapter / Contents / Sources & research notes

THE RESEARCH RECORD

Sources & research notes

94 source entries. Original documents, research findings, and the boundaries of each.

The chapters contain direct, clickable source references. This index records what each source contributes and provides return links to the chapters that use it. Research cutoff: October 29, 2026.

How the material was used: original documents establish wording and stated policy; assessments describe measured performance; surveys describe reported practice; research designs support narrower explanations. Advocacy materials are attributed. A posted resource does not establish national classroom adoption. Sources with retrieval limits are identified below. This presentation is a narrative investigation, not a preregistered systematic review or a new causal study.

S01 · Original text

John Dewey · How We Think (1910)

Reflective inquiry, suspended judgment, and the relationship between information and thought. Public-domain full text.

Used in: Chapter 02

S02 · Author excerpt

Edward Glaser · definition and 1941 research

Reproduces Glaser’s definition from An Experiment in the Development of Critical Thinking. Not a national trend study.

Used in: Chapter 02

S03 · Historical research

Larry Cuban · How Teachers Taught, 1890-1990 (1993)

Publisher’s description and book contents document continuity and hybrid teaching practices. The full book was not read for this edition.

Used in: Chapter 04

S04 · Federal statistics

NCES · educational attainment, 1910-2022

Historical adult completion rates. These are attainment figures for adults 25 and older, not annual graduation rates.

Used in: Chapter 04

S05 · Institutional history

NSF · institutional history

Science curriculum reform and teacher institutes around Sputnik. Reform efforts were already underway before the launch.

Used in: Chapter 05

S06 · Institutional history

BSCS Science Learning · our history

Firsthand institutional account of the Biological Sciences Curriculum Study and NSF support.

Used in: Chapter 05

S07 · Original research essay

B. F. Skinner · Teaching Machines (1958)

Published in Science; self-paced responses and immediate feedback, with criticism of passive reception.

Used in: Chapter 06

S08 · Original book / publisher record

Jerome Bruner · The Process of Education

1960 work, revised edition 1977. The accessible description and excerpts explain the spiral curriculum and intuitive and analytic thought.

Used in: Chapter 05

S09 · University teaching guidance

Bloom’s taxonomy · Cornell teaching resource

Identifies the 1956 taxonomy and the Anderson-Krathwohl 2001 revision; used for the taxonomy’s structure, not evidence of national adoption.

Used in: Chapter 06

S10 · Original paper record

Benjamin Bloom · Learning for Mastery (1968)

Time, instructional quality, persistence, and feedback as variables in achieving mastery.

Used in: Chapter 08

S11 · Program description

Augsburg Paideia Program · training

Adler’s 1982 model combines didactic teaching, coaching, and Socratic seminar. Program descriptions are not independent impact evaluations.

Used in: Chapter 06

S12 · Institutional history

National Forensic League · year-by-year history

Speech and debate organization founded in 1925; history of participation and competition, not universal classroom coverage.

Used in: Chapter 07

S13 · Federal report record

National Commission · A Nation at Risk (1983)

Contemporary reform diagnosis and recommendations; useful evidence that concerns about rigor predate current disputes.

Used in: Chapter 08

S14 · Contemporary instructional record

NCTM standards implementation monograph (1989)

Problem solving, communication, reasoning, connections, and estimation in mathematics reform.

Used in: Chapter 08

S15 · Research synthesis

National Research Council · How People Learn (2000)

Prior knowledge, understanding, and active construction of knowledge; includes Piaget and Vygotsky.

Used in: Chapter 06 · Chapter 10

S16 · Research synthesis

National Research Council · children and learning

Vygotsky’s zone of proximal development and metacognitive development.

Used in: Chapter 06

S17 · Federal research synthesis

National Reading Panel · findings (2000)

Systematic phonics supports decoding; effective reading instruction also needs fluency, language, and comprehension.

Used in: Chapter 15

S18 · Quasi-experimental study

Dee & Jacob · NCLB achievement effects (2009 / 2011)

Comparative interrupted time series. Estimated mathematics benefits, with no corresponding evidence of reading gains in the working-paper abstract.

Used in: Chapter 09

S19 · District survey

Center on Education Policy · instructional time (2007)

Original CEP report preserved by CiteSeerX. Representative district survey of reported instructional-time changes, not direct observation of reasoning.

Used in: Chapter 09

S20 · Standards text

Common Core · English language arts standards

Arguments, source evaluation, discussion, and evidence. Standards specify expectations rather than prove successful implementation.

Used in: Chapter 11

S21 · Federal policy explanation

U.S. Education Department · ESSA

2015 replacement for NCLB; continued annual statewide assessments with greater state accountability discretion.

Used in: Chapter 13

S22 · Research review

Kirschner, Sweller & Clark (2006)

Critique of minimal guidance, especially for novices. Distinguish minimal guidance from guided inquiry.

Used in: Chapter 10

S23 · Meta-analysis

Lazonder & Harmsen (2016)

72 studies: guidance improves inquiry learning. Different outcomes and settings limit simple universal prescriptions.

Used in: Chapter 10

S24 · Cognitive-science essay

Daniel Willingham · critical thinking (2007)

Domain knowledge and repeated application are central to thinking well; published by the American Federation of Teachers.

Used in: Chapter 24 · Chapter 38

S25 · Meta-analysis

Abrami and colleagues · teaching critical thinking (2015)

341 effect sizes; average g = 0.30, with substantial variation. Dialogue, authentic tasks, and mentoring show promise.

Used in: Chapter 38

S26 · Evidence review

Pashler and colleagues · learning styles (2008)

Insufficient evidence for matching instruction to an assessed visual/auditory learning style. Preferences and disabilities are separate questions.

Used in: Chapter 10

S27 · Research-informed instructional guidance

Barak Rosenshine · principles of instruction (2012)

Review, small steps, models, questioning, guided practice, and independent practice. Published in American Educator.

Used in: Chapter 24 · Chapter 38

S28 · Meta-analysis

Freeman and colleagues · active learning (2014)

225 studies in undergraduate STEM. Strong results in that setting should not be treated as a K-12 national trend.

Used in: Chapter 38

S29 · Quasi-experimental study

Reisman · Reading Like a Historian (2012)

236 eleventh-graders in five San Francisco high schools; six-month document-based curriculum intervention.

Used in: Chapter 11

S30 · Professional framework

NCSS · C3 social studies framework

2013 inquiry arc: questions, disciplinary knowledge, evidence, conclusions, and informed action. Not a federal mandate.

Used in: Chapter 11

S31 · Consensus framework

National Academies · K-12 science framework (2012)

Science practices include models, investigations, data interpretation, and argument from evidence.

Used in: Chapter 11

S32 · Advocacy organization audit

NCTQ · Reading Foundations review (2023)

Syllabus/materials rubric, not direct observation of every preparation classroom. The 2023 rubric is not comparable to prior editions.

Used in: Chapter 15

S33 · Accreditation standards

CAEP · revised preparation standards

Explicit equity/diversity and bias reflection, alongside disciplinary knowledge, assessment, clinical practice, and technology.

Used in: Chapter 16

S34 · Intellectual tradition / advocate explanation

Freire Institute · key concepts

Banking education, problem posing, dialogue, praxis, and critical consciousness. Describes a philosophy, not a U.S. prevalence estimate.

Used in: Chapter 16

S35 · Original theoretical article

Ladson-Billings & Tate (1995)

Toward a Critical Race Theory of Education links race, property, and educational inequality. A framework, not an outcome experiment.

Used in: Chapter 16

S36 · Original advocacy framework

Tema Okun · characteristics and cautions

Author describes origins in the late 1990s and warns against shaming and weaponizing the list; website published 2021 and updated 2025.

Used in: Chapter 20

S37 · Original advocacy framework

Tema Okun · One Right Way / objectivity

Actual language about neutrality, logic, emotion, and power, allowing the reader to examine the argument directly.

Used in: Chapter 20

S38 · Institutional statement, preserved transcript

NMAAHC · July 2020 removal statement

Museum’s own posted statement, mirrored by Thread Reader. Official live chart is unavailable; this is a preservation source.

Used in: Chapter 21

S39 · Government legal opinion / archived reference

NMAAHC chart · preserved documentation

Footnote 10 records the removed chart’s wording and links its archival capture. Used to document the episode, not endorse the opinion’s legal argument.

Used in: Chapter 21

S40 · Original training workbook

Equitable Math · Stride 1 (2020)

Indexed PDF passage on answer-getting, concepts, and objectivity. Full PDF fetch failed during this review; exact-page visual verification was unavailable.

Used in: Chapter 22

S41 · Authors’ explanation

Equitable Math · authors’ FAQ

Explicitly denies that math itself is racist or that students should face lower standards. Read beside the workbook, not as a substitute for it.

Used in: Chapter 22

S42 · Professional development offering

Student Achievement Partners · course series

Documents distribution through an educator resource provider, without establishing how many classrooms implemented the approach.

Used in: Chapter 23

S43 · University participation / video directory

LMU · Math Equity Toolkit

LMU’s Center for Equity for English Learners describes its reviewer role and links original webinars.

Used in: Chapter 23

S44 · State guidance

California · adopted mathematics framework (2023)

Official description of standards, equity, rigor, acceleration, and grouping. Distinguish adopted guidance from draft claims.

Used in: Chapter 23

S45 · District guidance

NYC Public Schools · race and equity guidance

Official staff resource explicitly favors critical pedagogy and students as agents of change. A district case, not a national count.

Used in: Chapter 19

S46 · Regression-discontinuity study

Bonilla, Dee & Penner · ethnic studies (2021)

San Francisco course assignment near an eligibility threshold improved engagement and attainment; narrow local causal estimate.

Used in: Chapter 23

S47 · Program philosophy

CASEL · transformative SEL

Describes identity, agency, belonging, and inequity analysis; says most existing SEL programs emphasize other forms.

Used in: Chapter 17

S48 · Meta-analysis

Durlak and colleagues · SEL (2011)

213 school-based programs, 270,034 students; average improvements do not validate every later program or ideological adaptation.

Used in: Chapter 17

S49 · National teacher survey

RAND · restrictions and spillovers (2024)

2023 teacher reports of limiting political/social discussion; reports and associations do not directly measure discussion quality.

Used in: Chapter 18

S50 · National survey

Pew · race and LGBTQ issues in schools (2024)

Teacher, teen, and public attitudes; main teacher survey sampled 2,531 public K-12 teachers in autumn 2023.

Used in: Chapter 18

S51 · Historical political platform

Texas Republican Party · 2012 platform

Printed page 12, PDF page 20: critical-thinking language appears in opposition to specific outcome-based education programs. A platform, not enacted classroom law.

Used in: Chapter 19

S52 · Union position

NEA · 2021 assembly statement

The union describes racial-justice priorities and opposition to restrictions. Leadership statements do not establish every member’s beliefs.

Used in: Chapter 24

S53 · Federal financial disclosure

AFT · FY2026 LM-2 financial filing

National headquarters, file 000-012, July 2025-June 2026. Indexed Statement B retrieved; direct page fetch returned 502 during review.

Used in: Chapter 25

S54 · Federal campaign disclosure

NEA Fund · FEC committee summary

2025-2026 cycle; retrieved coverage through October 14, 2026. This committee is only one channel of union-related political activity.

Used in: Chapter 25

S55 · Union disclosure

NEA Fund · contribution disclosure

Federal candidate PAC contributions are voluntary. They must not be conflated with all dues-funded lobbying or advocacy.

Used in: Chapter 25

S56 · Observational district study

Marianno and colleagues · reopening (2022)

250 large districts, fall 2020. Institutional bargaining power was associated with less in-person instruction; not a randomized union treatment.

Used in: Chapter 26

S57 · Quasi-experimental historical study

Lovenheim & Willén · collective bargaining (2019)

State bargaining laws and adult outcomes; adverse estimates for men in historical cohorts, not a direct critical-thinking measure.

Used in: Chapter 27

S58 · Working paper

Biasi & Sandholtz · Wisconsin Act 10 (2025)

Reports test-score gains and union revenue reductions after reform; provisional evidence from one state.

Used in: Chapter 27

S59 · Working paper

Baron · Wisconsin Act 10 (2017, revised 2019)

Earlier short-run study reports test-score losses. Different estimates caution against a universal claim about bargaining and learning.

Used in: Chapter 27

S60 · Supreme Court decision

Janus v. AFSCME (2018)

Prohibits compelled public-sector agency fees from nonconsenting nonmembers. Used as historical context.

Used in: Chapter 25

S61 · Federal institutional explanation

Education Department · limits on federal role

May 2026 page explains statutory limits on federal curriculum control and separates NAEP from state tests.

Used in: Chapter 28

S62 · Executive policy text

Executive Order 14242 (March 20, 2025)

Directs steps toward closure to the extent permitted by law while continuing services. Does not itself establish learning effects.

Used in: Chapter 28

S63 · Current federal implementation account

Education Department · interagency partnerships

Describes redistribution of administrative work while retaining statutory protections and authorities; policy status is time-sensitive.

Used in: Chapter 28

S64 · Contemporaneous federal letter

Education Department · 2020 assessment waivers

Pandemic waiver invitation for 2019-2020 testing and accountability. An emergency interruption, not permanent elimination of testing.

Used in: Chapter 29

S65 · Federal data notes

Ed Data Express · assessment comparability notes

2020-2021 participation differences can make comparisons unrepresentative. Consult year/state data notes before trend claims.

Used in: Chapter 29

S66 · Large observational study

Goldhaber and colleagues · remote/hybrid instruction (2022)

2.1 million students, about 10,000 schools; remote instruction was a major contributor to widening achievement gaps.

Used in: Chapter 26 · Chapter 29

S67 · Research team release

Education Scorecard · May 2026 findings

Linked state assessments through 2024-2025; uneven recovery, literacy reforms, absenteeism, and improving districts.

Used in: Chapter 30

S68 · Federal data and research summary

Education Department · chronic absenteeism

Absence of at least 10% of school days; rates and contributing barriers, not a measure of student motivation alone.

Used in: Chapter 30

S69 · School Pulse Panel

NCES · phones and devices (February 2025)

Experimental survey data, December 2024. Leaders’ perceived effects are not experimental causal estimates.

Used in: Chapter 31

S70 · International assessment

OECD · PISA 2022 U.S. country note

Math declined; reading/science broadly stable compared with 2018. Sampling caveats and national standards differ from NAEP.

Used in: Chapter 31

S71 · Original study summary

Digital Inquiry Group · national online reasoning study

2018-2019 assessment of 3,446 high-schoolers matching the national demographic profile; no historic equivalent establishes a decades-long decline.

Used in: Chapter 32

S72 · Original study record

Wineburg and colleagues · lateral reading field study (2022)

Six lessons in high-school government; treatment/control comparison found improved source evaluation. Local implementation, not nationwide proof.

Used in: Chapter 32

S73 · Assessment-provider research

ACT · subject grades and inflation (2023)

Adjusted math GPA rose 2010-2022 without corresponding ACT gains. ACT-tested sample and provider interests should be considered.

Used in: Chapter 33

S74 · Randomized field experiment

Bastani and colleagues · generative AI and learning (2025)

Nearly 1,000 high-school math students in Turkey; answer-giving and guarded tutoring had different effects. The August 2025 correction concerns an author affiliation, not the results.

Used in: Chapter 34

S75 · Federal assessment release

NAGB · 2024 grades 4/8 results

Reading fell again; fourth-grade math improved; eighth-grade math showed no significant recovery. Released January 2025.

Used in: Chapter 14 · Chapter 35

S76 · Federal assessment

NAEP · grade 12 reading, 2024

35% at or above Proficient; 32% below Basic. Proficient does not mean a state’s grade-level passing threshold.

Used in: Chapter 35

S77 · Federal assessment

NAEP · grade 12 mathematics, 2024

45% below Basic, up from 2019. Grade 12 mathematics uses a different scale from grades 4/8.

Used in: Chapter 35

S78 · Federal assessment summary

NAGB · 2022 history and civics results

Eighth-grade history and civics mean scores declined. These are subject assessments, not a standalone national logic exam.

Used in: Chapter 35

S79 · Federal assessment

NAEP · 2025 long-term trend results

Released June 2026: younger students improved, older students did not significantly improve. Age-based LTT is separate from main NAEP.

Used in: Chapter 03 · Chapter 36

S80 · Federal assessment release

NAGB · June 2026 long-term trend release

Includes recreational-reading and algebra-course-taking survey trends. These are associations, not causal explanations.

Used in: Chapter 36

S81 · Adult skills assessment

NCES · PIAAC 2023 adult skills

Literacy/numeracy declines since 2017, with method-change caveats. Adaptive problem solving is a new domain with no comparable previous trend.

Used in: Chapter 37

S82 · Quasi-experimental historical study

Jackson, Johnson & Persico · school spending (2016)

Finance-reform-induced spending increases linked to improved adult outcomes, especially for low-income children.

Used in: Chapter 42

S83 · Educational video

TED-Ed · five critical-thinking tips

Samantha Agoos; instructional supplement, not evidence about historical U.S. school trends.

Used in: Chapter 38

S84 · Educational video

Crash Course · lateral reading

John Green, 2019; produced with MediaWise partners including the Stanford History Education Group.

Used in: Chapter 32

S85 · Original proponents' video

EdTrust-West · Stride 1 webinar

Lets readers hear the training's own framing. Linked by LMU; embedded playback may depend on provider permissions.

Used in: Chapter 23

S86 · National survey

Pew · Teens and Technology (2013)

Smartphone ownership among ages 12-17 increased from 23% in 2011 to 37% in 2013.

Used in: Chapter 12

S87 · National survey

Pew · Teens, Social Media & Technology (2015)

73% of ages 13-17 had or could access a smartphone. Access and ownership, and age ranges, differ from the 2013 item.

Used in: Chapter 12

S88 · Government oversight

GAO · Race to the Top (2011)

Almost $4 billion awarded to 12 states; teacher effectiveness, data systems, standards, and school turnaround. Documents incentives and implementation issues.

Used in: Chapter 12

S89 · Standards developer history

Common Core · development process

Drafts in 2009-2010; final standards released June 2010. Adoption and classroom implementation occurred on different schedules.

Used in: Chapter 12

S90 · Contemporaneous federal assessment

NCES · 2015 reading and mathematics results

Math declined at grades 4/8 from 2013, grade-8 reading declined, grade-4 reading showed no significant change. Long-run scores were still above early-1990s results.

Used in: Chapter 12

S91 · Federal assessment summary

NAEP · 2024 grades 4/8 summary tables

National reading scale-score trends, including public and private schools. Rounded scale scores are not percentages; significance uses unrounded estimates.

Used in: Chapter 35

S92 · Federal assessment

NAEP · grade 4 mathematics trends, 2024

Grade-4 mathematics recovery was concentrated among middle/higher performers; the average remained above 1990. Achievement-level and scale-score measures differ.

Used in: Chapter 35

S93 · Publisher correction

Bastani and colleagues · correction notice (2025)

Corrects Osbert Bastani’s institutional affiliation after a production error. It does not change the study results.

Used in: Chapter 34

S94 · Original advocacy text, preserved copy

Kenneth Jones & Tema Okun · 2001 framework

Attributes the earlier White Supremacy Culture text to the 2001 Dismantling Racism workbook. Includes the original objectivity section; not a school-adoption census.

Used in: Chapter 20