WHY GOOD IDEAS FAIL

Institutions, Change, and the Science-Practice Gap in Education

Applied Pedagogy Research Lab

Guido Bartolucci, Principal Investigator

guido@appliedpedagogy.com

LAB.APPLIEDPEDAGOGY.COM

L1-007 · March 2026

CONTENTS

Part I

THE PROBLEM

THE IMPLEMENTATION PROBLEM

The Applied Pedagogy Research Lab now knows a great deal about how humans learn. L1-002 established that self-determination theory provides a robust framework for sustaining motivation — autonomy, competence, and relatedness are non-negotiable needs, and environments that thwart them reliably destroy engagement. L1-003 showed that assessment is simultaneously the most powerful lever for learning and the most common mechanism for undermining it. L1-004 demonstrated that instruction should adapt to learner expertise — novices need explicit guidance, intermediates benefit from productive failure, and advanced learners need independence. L1-009 found that the institutional environment is a first-order determinant of whether upper-layer competence (judgment, metacognition, character) develops or degrades. L1-010 showed that training contexts — military, aviation, medicine — produce better learning outcomes than schools in part because their institutional structures create tighter feedback loops and higher psychological safety.

None of this matters if institutions cannot or will not implement it.

This is the implementation problem, and it is the central challenge this investigation addresses. The lab's research has identified clear, evidence-based principles for effective education. The question is no longer primarily *what* should be done but *why it isn't being done* — and what conditions would make implementation possible. The answer lies not in cognitive science or instructional design but in institutional theory, organizational sociology, political science, and the history of education. It requires understanding why schools look the way they do, how the "grammar of schooling" operates, what makes reform possible or impossible, and how the gap between learning science and educational practice can be closed.

This investigation is about where research meets reality. It is, therefore, about power, path dependence, organizational survival, and the deeply conservative nature of institutions that are charged with society's most important task.

The investigation draws on several intellectual traditions that do not usually appear together in educational research. Institutional theory (Scott, DiMaggio & Powell, Meyer & Rowan) explains why organizations persist in their current forms. The sociology of education (Tyack & Cuban, Hafferty, Jackson) explains the specific structures and hidden curriculum of schools. Political science (Kingdon, Blyth) explains when and how institutional change becomes possible. Organizational learning (Argyris & Schön, Senge, Edmondson) explains how institutions can — or cannot — learn and adapt. And the philosophy of science (Kuhn, as translated in the lab's archives) provides a framework for understanding paradigmatic change that is strikingly applicable to institutional change in education.

The synthesis of these traditions produces a picture that is sobering but not hopeless. The grammar of schooling is extraordinarily resilient, and most reform efforts fail. But the conditions for successful reform are identifiable: reforms must target the instructional core, build capacity rather than mandate change, maintain coherence across levels, and be sustained over years. Crisis can open policy windows, but only if the alternative is prepared in advance. And the evidence base for what good institutional design looks like — synthesized across the lab's L1 investigations — is now substantial enough to serve as the foundation for a serious alternative.

THE GRAMMAR OF SCHOOLING: WHY SCHOOLS LOOK THE WAY THEY DO

2.1 THE CONCEPT

David Tyack and Larry Cuban coined the term "grammar of schooling" in their landmark 1995 work *Tinkering Toward Utopia: A Century of Public School Reform* — a book whose field-weighted citation impact of 193.66 makes it one of the most influential works in the history of education (Tyack & Cuban, 1995). The concept was elaborated in a 1994 article with William Tobin (Tyack & Tobin, 1994), which itself carries an FWCI of 77.76.

The grammar of schooling refers to the organizational regularities that structure the daily life of schools: age-graded classrooms, subject-based curriculum divisions, Carnegie units measuring seat time, one teacher per classroom, 50-minute class periods, letter grades, and standardized testing as the primary accountability mechanism. These structures are so deeply embedded in how society understands "school" that they feel natural, inevitable — as grammar feels to a native speaker. A native English speaker does not think about subject-verb-object order; it simply feels like the way language works. Similarly, educators and the public do not think about age-graded classrooms or Carnegie units; they simply feel like the way school works.

This naturalization is the grammar's greatest power. Reforms that work *within* the grammar — new textbooks, revised standards, additional subjects, improved test instruments — are adopted relatively easily because they do not challenge the organizational logic. Reforms that challenge the grammar — multi-age grouping, team teaching, competency-based progression, project-based curriculum, elimination of grade levels — face enormous structural resistance. The resistance is not primarily ideological. It is organizational. Changing the grammar requires simultaneously altering scheduling, staffing, facilities, evaluation, reporting, teacher training, community expectations, and the legal-regulatory framework that encodes these structures. The coordination costs are prohibitive for most institutions.

2.2 HOW THE GRAMMAR EMERGED

The grammar of schooling is not ancient. Most of its features crystallized in the late nineteenth and early twentieth centuries as American education industrialized. The age-graded classroom was an innovation of the 1840s–1860s, replacing the one-room schoolhouse. The Carnegie unit — defining credit hours by time spent in class — was established in 1906 by the Carnegie Foundation, originally as a mechanism for standardizing college admissions and teacher pensions. Subject-based departmental divisions in secondary schools emerged alongside the Committee of Ten report in 1893. The 50-minute class period was a scheduling convenience that became a structural given. Letter grades spread unevenly through the early twentieth century, gradually displacing narrative evaluations and descriptive reports.

Each of these innovations solved a real organizational problem at the time. Age-grading allowed mass education to scale. Carnegie units created legibility across institutions. Subject departments allowed teacher specialization. But solutions designed for the organizational challenges of 1900 became the unquestioned assumptions of 2025. The structures persisted not because they con-

tinued to serve learning but because they became embedded in every other structure — teacher training programs prepare teachers for subject-based, age-graded classrooms; textbook publishers produce materials for grade-level, subject-specific courses; state accountability systems measure performance within grade levels and subject areas; parent expectations are calibrated to report cards with letter grades.

2.3   PATH DEPENDENCE

The concept of path dependence — borrowed from economics and historical institutionalism — helps explain why the grammar persists. Early organizational decisions create self-reinforcing feedback loops that make alternatives progressively more costly to adopt, even when the alternatives would be superior. The QWERTY keyboard layout is the standard example, but the grammar of schooling is a far more consequential one.

Consider what it would take to replace age-graded classrooms with competency-based progression — an approach that L1-004's expertise-adaptive model, L1-003's mastery learning evidence, and L1-010's training science all support. You would need to: redesign the physical school building (currently optimized for age-graded homerooms); retrain every teacher (currently trained to teach a specific grade level or subject); redesign every assessment instrument (currently normed by grade level); change state accountability systems (currently structured around grade-level testing); change parent communication systems (currently built around grade-level report cards); change college admissions criteria (currently based on Carnegie-unit transcripts); and change the legal framework governing compulsory education (currently defined by age, not competency). Each of these changes is individually feasible but collectively prohibitive. The path-dependent lock-in is not any single barrier but the reinforcing structure of all of them together.

This is why Tyack and Cuban found that over a century of reform efforts — from progressive education in the 1920s to open classrooms in the 1960s to standards-based reform in the 1990s — have produced remarkably little change in the basic organizational structure of schooling. The grammar absorbs, dilutes, or rejects reforms that challenge it. Progressive education's project-based curriculum was domesticated into "enrichment activities" that supplemented rather than replaced the traditional curriculum. Open classrooms were physically constructed but then operated as traditional classrooms with the walls removed. Standards-based reform produced more testing but did not change the fundamental instructional model.

2.4   INSTITUTIONAL ISOMORPHISM: WHY SCHOOLS CONVERGE

If path dependence explains why individual schools resist change, institutional isomorphism explains why schools across an entire system converge on similar structures even when those structures don't serve learning well. DiMaggio and Powell (1983), in one of the most influential papers in organizational sociology, identified three mechanisms through which organizations in a field come to resemble each other.

Coercive isomorphism results from formal and informal pressures exerted by organizations upon which other organizations depend. In education, this includes state and federal mandates (No Child Left Behind, Every Student Succeeds Act), accreditation requirements, and legal regulations. Schools that deviate too far from the expected model face loss of funding, loss of accreditation, and legal liability. A school that eliminates letter grades may find its students unable to obtain college admission. A school that abandons age-grading may run afoul of compulsory education laws that define requirements by age.

Mimetic isomorphism results from uncertainty. When organizations face ambiguous goals or unclear technologies for achieving them — and education faces both — they model themselves on other organizations perceived as successful. Schools copy other schools, not because the copied structures have been shown to work, but because copying reduces uncertainty and confers legitimacy. This is why educational innovations spread through emulation rather than evidence — a school district adopts a program because a prestigious neighboring district adopted it, not because the evidence base supports it.

Normative isomorphism results from professionalization. Teachers are trained in similar programs, absorb similar norms, and bring those norms into their practice. The professionalization of teaching creates a shared cognitive framework about what school should look like, how classes should be structured, and what counts as good teaching. This normative pressure operates below the level of conscious decision-making — teachers do not decide to replicate the grammar of schooling; they have been trained within it and cannot easily imagine alternatives.

The combined effect of coercive, mimetic, and normative isomorphism is that even when individual schools attempt innovation, the pressures of the organizational field push them back toward conventional structures. This is the "drift problem" that L1-006 identified in alternative education — Montessori schools gradually compromise their distinctive features, charter schools converge toward traditional models, and progressive schools slowly reintroduce the structures they were founded to escape. Isomorphic pressure is the gravitational field of the educational system; escaping it requires escape velocity, not just a good idea.

2.5 DECOUPLING: THE GAP BETWEEN STRUCTURE AND PRACTICE

Meyer and Rowan (1977) identified a phenomenon that is central to understanding why education reform so often fails to change what actually happens in classrooms: decoupling. Their insight is that organizations adopt formal structures — mission statements, curricula, policies, organizational charts — to maintain legitimacy with their external environment, but these formal structures often decouple from actual organizational practice.

In education, decoupling is pervasive. A school district may adopt a formal curriculum aligned with the latest standards, but individual teachers close their classroom doors and teach as they always have. A school may declare a commitment to formative assessment, but teachers continue to rely primarily on summative grades. A university may espouse student-centered learning in its mission statement while its reward structure incentivizes research over teaching. The formal structure satisfies the demands of the external environment (regulators, accreditors, parents, politicians), while the actual practice satisfies the demands of the internal environment (teacher preferences, student expectations, organizational routines).

Decoupling is not hypocrisy, though it can look like it. It is an organizational survival strategy. Schools face legitimacy demands from multiple audiences with conflicting expectations. Parents want their children to be happy and well-educated. Politicians want measurable accountability. Teachers want professional autonomy. Students want grades that will get them into college. Administrators want smooth operations. No single structural arrangement can satisfy all of these demands simultaneously, so schools adopt the formal structures that satisfy the most powerful external demands while preserving internal flexibility through loose coupling between formal structure and actual practice.

The implication for reform is devastating. A policy mandate may change the formal structure without changing the practice. A new curriculum may be adopted on paper without being implemented in classrooms. A new assessment system may be installed without changing how teachers

actually give feedback. The reformers declare victory because the formal structure has changed; the practitioners shrug because their daily work has not. This is why Cuban (1993) found, in his study of American classroom practice from 1890 to 1990, that despite waves of reform, the basic mode of instruction — teacher-centered, lecture-and-recitation — remained remarkably stable. The grammar persists not because reformers fail to change the policy but because policy changes decouple from practice changes.

THE HIDDEN CURRICULUM: WHAT INSTITUTIONS TEACH BY THEIR STRUCTURE

3.1   BEYOND THE FORMAL CURRICULUM

Philip Jackson (1968) introduced the concept of the hidden curriculum in *Life in Classrooms*, arguing that schools teach powerful lessons through their organizational structures, social norms, and reward systems that are distinct from — and sometimes contradictory to — the formal curriculum. Frederic Hafferty (1998), in a paper that has now been cited 1,783 times, extended the concept to medical education, demonstrating that the hidden curriculum operates through institutional policies, resource allocation, and institutional "slang" — the informal language and attitudes that communicate what an institution actually values as opposed to what it claims to value.

The hidden curriculum is not the same as bad teaching or institutional failure. It is a structural feature of all institutions. Every organizational design teaches something beyond its intended purpose. The question is not whether a hidden curriculum exists but whether it is recognized, and whether it can be deliberately shaped.

3.2   WHAT SCHOOLS TEACH IMPLICITLY

The standard grammar of schooling — age-graded classrooms, competitive grading, 50-minute periods, subject silos — teaches the following implicit lessons:

Compliance over inquiry. The dominant mode of classroom interaction — teacher asks question, student provides expected answer, teacher evaluates — trains students in compliance, not inquiry. Students learn that the purpose of intellectual activity is to produce the answer that the authority figure expects, not to genuinely investigate a question. This is reinforced by assessment systems that reward correct recall and penalize deviation from expected responses.

Performance over learning. When grades are the primary currency of the educational system — determining class rank, honor roll, college admission, parental approval — students learn that the purpose of school is to produce performances that earn high marks, not to develop genuine understanding. The distinction between learning orientation and performance orientation, extensively documented in the motivation literature (L1-002), is not merely a student characteristic; it is an institutional product. Schools that organize their reward systems around grades are *teaching* performance orientation through their structure, regardless of what they say about valuing learning.

Time-based rather than mastery-based progress. The Carnegie unit structure teaches students that learning is measured by time spent, not by competence achieved. A student who masters algebra in two weeks and a student who has not mastered it after a semester both receive the same credit if they complete the same seat-time requirement. This teaches, implicitly, that time is more important than mastery — a lesson that directly contradicts the evidence from mastery learning (Bloom, 1968), training science (L1-010), and competency-based education.

Knowledge as compartmentalized and inert. Subject-based departmental divisions teach students that knowledge comes in discrete categories — "math," "English," "science," "history" — that have little to do with each other. A student may learn about the Enlightenment in history class

and about the scientific method in science class without anyone connecting these two topics, which are historically inseparable. The compartmentalization of knowledge produces what Whitehead (1929) called "inert knowledge" — knowledge that can be retrieved on a subject-specific test but is never activated in real-world situations where problems do not come pre-sorted by academic discipline.

Social hierarchy as natural. Competitive grading and tracking systems teach students that some people are inherently "smarter" than others and that the purpose of evaluation is to sort people into hierarchies. This contradicts the evidence on growth and development (ability is not fixed), reinforces the undermining effect on motivation (L1-002), and creates exactly the performance-oriented, psychologically unsafe environments that L1-009 identified as competence-destroying.

3.3   THE HIDDEN CURRICULUM AND THE COMPETENCE STACK

The connection between the hidden curriculum and the competence stack (COMPETENCE-TARGET.md) is direct and consequential.

The environmental multiplier — question 5 in the diagnostic framework: "Are they allowed to tell the truth in that system?" — is a hidden curriculum question. An institution can espouse intellectual honesty in its mission statement while its grading system penalizes honest self-assessment ("I don't understand this") and rewards confident performance ("I got an A"). Over time, students in such systems learn to perform confidence rather than exercise it. They learn to avoid revealing uncertainty because uncertainty is punished — not formally, but structurally, through the grading system that treats uncertainty as ignorance rather than as a metacognitive achievement.

L1-009 found that environments that penalize honesty do not merely suppress truth-telling; over time they degrade the capacity to *perceive* truth. The hidden curriculum of most schools operates exactly this way at layers 3–5 of the competence stack. It does not merely fail to develop judgment, metacognition, and epistemic character; it actively degrades them by creating conditions in which honest self-assessment is costly, intellectual risk-taking is punished, and the performance of knowledge is rewarded over the genuine admission of ignorance.

This is why the "environment first" finding that converges across L1-002, L1-006, L1-009, and L1-010 is fundamentally an institutional analysis finding. The environment is not separate from the institution; the environment *is* the institution, experienced from the inside.

3.4   THE HIDDEN CURRICULUM AND MOTIVATIONAL DECLINE

L1-002 documented one of the most important and depressing findings in the motivation literature: intrinsic motivation for learning declines systematically across the years of schooling. Students enter school curious and eager to learn; by middle school, many have become disengaged; by high school, many are going through the motions. L1-002 found that the causal mechanisms are underresearched, but the institutional analysis suggests a powerful explanation: the hidden curriculum of schooling systematically teaches students that learning is a performance to be evaluated rather than an activity to be enjoyed.

Consider the trajectory from the student's perspective. In early elementary school, the hidden curriculum is relatively benign — many classrooms emphasize exploration, play, and intrinsic engagement. As students advance, the hidden curriculum shifts. Grades become more prominent. Standardized testing begins. Competition for class rank intensifies. The Carnegie unit structure forces learning into fixed time blocks regardless of interest or mastery. By middle school, the

predominant institutional message is: learning is work to be endured for the reward of grades, and the purpose of intellectual activity is to demonstrate performance, not to satisfy curiosity.

This trajectory is not an accident; it is a structural feature of the grammar of schooling. The grammar was designed for efficiency and accountability, not for motivation. Age-grading produces social comparison (I'm in the "smart" group or the "slow" group). Letter grades produce performance orientation (what matters is the A, not the learning). Subject compartmentalization prevents the kind of sustained, interest-driven inquiry that sustains motivation over time. The cumulative effect of these structural features is a hidden curriculum that teaches students, year by year, that learning is extrinsically motivated, externally evaluated, and disconnected from personal meaning.

The stage-environment fit theory, referenced in L1-002, suggests that motivational decline occurs when the institutional environment becomes increasingly misaligned with students' developmental needs. As adolescents develop greater needs for autonomy, competence, and relatedness, the institutional environment becomes more controlling (more rules, more testing, more evaluation), less competence-supportive (larger class sizes, less individual attention, more social comparison), and less relational (departmental organization, rotating teachers, larger schools). The grammar of schooling creates environments that are increasingly hostile to the basic psychological needs identified by SDT — not out of malice but out of structural logic.

This institutional analysis of motivational decline has a hopeful implication: if motivational decline is caused by institutional structures rather than by developmental inevitability, it can be prevented by different institutional structures. An institution that maintains autonomy support, competence feedback, and relational warmth across the years of education — that does not shift from exploration to evaluation as students age — might sustain the curiosity and engagement that the current grammar systematically extinguishes.

3.5 CAN THE HIDDEN CURRICULUM BE DELIBERATELY DESIGNED?

Hafferty's work in medical education provides the strongest evidence that the hidden curriculum can be made visible and deliberately shaped. Medical schools that recognized the hidden curriculum — particularly its tendency to erode idealism and empathy during clinical training — were able to intervene by changing institutional structures: modifying the evaluation system, changing the clinical rotation structure, creating protected time for reflection, and training clinical faculty to model the values the institution espoused (Lempp & Seale, 2004; Brown et al., 2020).

The key insight from Hafferty's framework is that the hidden curriculum operates through three channels: (1) institutional policies, including resource allocation and evaluation criteria; (2) the attitudes and behaviors modeled by people in positions of authority; and (3) the physical and organizational environment. Changing the hidden curriculum requires changing all three channels simultaneously. A school that changes its grading policy (channel 1) but whose teachers continue to use evaluative language (channel 2) in a competitive classroom environment (channel 3) will not change the hidden curriculum. Alignment across all three channels is necessary.

For Applied Pedagogy, this means that institutional design is not a separate consideration from curriculum design — it *is* curriculum design at the most fundamental level. The institution teaches through its structure at least as powerfully as it teaches through its content. Every design decision — the assessment system, the physical space, the authority structure, the language used by instructors, the response to error and uncertainty — is a hidden curriculum decision.

Part II

REFORM AND THE GAP

4.1 A CENTURY OF FAILED REFORMS

Tyack and Cuban's most sobering contribution is their documentation of a century of reforms that were absorbed, diluted, or rejected by the grammar of schooling. The pattern is remarkably consistent:

1. A reform movement emerges, often driven by external crisis (economic competition, war, social upheaval) or by new evidence about learning.
2. The reform is adopted at the policy level — new legislation, new standards, new mandates.
3. Schools adopt the formal structures of the reform while maintaining their existing practices (decoupling).
4. Over time, the reform is either domesticated (its most challenging features are dropped, leaving only the features compatible with the existing grammar) or abandoned.
5. A new reform cycle begins.

This pattern played out with progressive education in the 1920s–1930s, life adjustment education in the 1940s–1950s, the new math and open classrooms in the 1960s–1970s, standards-based reform in the 1990s, and No Child Left Behind in the 2000s. Each wave left some residue — progressive education contributed the idea that children's interests matter; standards-based reform contributed clearer learning objectives — but none fundamentally changed the grammar.

Mehta (2013) argues in *The Allure of Order* that American education reform has been trapped in a recurring cycle of rationalization: the belief that if the right standards, the right tests, and the right accountability systems could be designed, the system would improve. Each iteration of this belief fails because it addresses the formal structure (what schools are supposed to do) without addressing the instructional core (what teachers and students actually do in classrooms). The allure of order — the idea that complex human processes can be managed through top-down standardization — is itself a manifestation of the grammar's logic. It assumes that the problem is insufficient control rather than insufficient capacity.

4.2 WHAT MAKES REFORM SUCCEED: CONDITIONS FOR CHANGE

Not all reform efforts fail. The evidence points to several conditions that distinguish successful reforms from unsuccessful ones.

Reforms that succeed target the instructional core. Elmore (2004) argues in *School Reform from the Inside Out* that the only reforms that produce lasting improvement are those that change the relationship between teacher, student, and content — what he calls the "instructional core." Policy reforms that do not reach the instructional core — new standards, new tests, new organizational charts — produce changes in form without changes in substance. Reforms that directly change classroom practice — through new instructional methods, new materials, intensive coaching, and protected time for professional learning — have a much better track record.

Reforms succeed when they build capacity, not just mandate change. Fullan (2007) argues in *The New Meaning of Educational Change* that sustainable reform requires building the capacity of

teachers and schools to implement change, not merely mandating what they should do differently. Mandates without capacity-building produce compliance without commitment — the decoupling phenomenon. Capacity-building includes professional development, coaching, collaborative planning time, adequate resources, and institutional structures that support experimentation and learning from failure.

Reforms succeed when they are coherent across levels. Honig and Hatch (2004), in a study with an FWCI of 10.99, found that schools that successfully managed multiple external reform demands did so by "crafting coherence" — strategically selecting, integrating, and adapting external demands to align with their own instructional goals. Schools that treated each new mandate as an independent requirement became fragmented and exhausted. Schools that filtered mandates through a coherent instructional vision were able to make reform productive.

Reforms succeed when they have sustained support over time. Educational change is slow. Fullan estimates that meaningful instructional change takes 3–5 years of sustained, focused effort. Most reform initiatives lose political support, funding, or leadership attention before they have time to mature. The reform cycle — adoption, partial implementation, abandonment, new reform — is driven partly by the political cycle, which demands visible results faster than genuine change can produce them.

4.3   THE FINNISH EXCEPTION

Finland's education reforms, beginning in the 1970s and culminating in the system that achieved top PISA results from 2000 onward, represent the most-cited example of successful systemic education reform. The Finnish approach differed from the American reform pattern in several ways that illuminate the conditions for success.

Finland invested heavily in teacher quality — requiring a master's degree for all teachers, making teaching a competitive and prestigious profession, and giving teachers substantial professional autonomy. Rather than imposing external accountability through testing, Finland trusted teachers' professional judgment and provided them with the training and support to exercise that judgment well. The Finnish system eliminated tracking, grade retention, and competitive grading in the early years, creating an equitable foundation that reduced the need for later interventions.

Critically, Finland's reforms were sustained across decades and across changes in government. They were not driven by crisis or political posturing but by a deliberate, long-term strategy to build educational capacity. Sahlberg (2011) describes this as the "Finnish Way" — trust in teachers, equity before excellence, collaborative rather than competitive school cultures, and a rejection of the "Global Education Reform Movement" (GERM) that emphasizes standardized testing, competition, and accountability.

The Finnish case illuminates a paradox: the most successful education reform in the modern era succeeded by doing the *opposite* of what most reform efforts attempt. Rather than increasing external control (testing, accountability, competition), Finland increased internal capacity (teacher quality, professional trust, collaboration). Rather than imposing uniformity (standardized curriculum, standardized tests), Finland created the conditions for professional judgment and local adaptation.

However, the Finnish case also has important limitations as a model. Finland is a small, relatively homogeneous country with high social trust and low inequality. Whether its approach can be replicated in larger, more diverse, more unequal societies is genuinely unknown. Finland's more recent PISA declines have also prompted debate about whether the Finnish model is sustainable or was a product of specific historical conditions.

4.4 JAPAN'S LESSON STUDY: REFORM FROM THE INSIDE

While Finland and Singapore represent systemic, government-led reform, Japan's lesson study tradition represents a different model — reform driven by teachers' own professional practice, embedded in the institutional structure of schools. Lesson study (*jugyou kenkyuu*) is a structured process in which teachers collaboratively plan a lesson, one teacher delivers it while others observe, and the group then debriefs — analyzing student learning, identifying what worked and what didn't, and revising the lesson. The cycle is repeated across the school year.

Lesson study is institutionally significant because it addresses several of the barriers to reform that this investigation has identified. It targets the instructional core directly (Elmore's criterion). It builds teacher capacity rather than mandating change (Fullan's criterion). It creates time and structure for evidence-based practice (addressing the time barrier to evidence use). It embeds professional learning in the daily work of teaching rather than isolating it in separate professional development sessions. And it creates a culture of collaborative inquiry that aligns with the psychological safety and error management norms that L1-009 and L1-010 identify as prerequisites for organizational learning.

The Japanese experience also illustrates the importance of institutional infrastructure for sustained practice. Lesson study works in Japan because the school schedule allocates time for collaborative planning and observation, because the profession values this activity, and because the institutional culture supports honest critique of practice. Attempts to import lesson study to American schools have had mixed results — not because the practice itself doesn't work, but because the institutional infrastructure to support it (time, norms, expectations) doesn't exist in most American schools. This is the contextual variation problem that Joyce and Cartwright (2019) identify: a practice that works in one institutional context may fail in another if the supporting conditions are absent.

4.5 THE SINGAPORE MODEL

Singapore's education transformation provides a different model of successful reform — one that is more centrally directed than Finland's but equally focused on building capacity. Singapore moved from a low-performing system in the 1960s to one of the world's highest-performing systems through a deliberate, multi-decade strategy that included: investing heavily in teacher training and professional development; aligning curriculum, assessment, and instruction; creating career pathways that allowed excellent teachers to remain in classrooms; and building institutional infrastructure for continuous improvement.

Unlike many reform efforts, Singapore's approach was explicitly evidence-informed — drawing on cognitive science and instructional design research to inform curriculum development and teacher training. The Singapore math curriculum, for example, incorporates worked examples, the concrete-pictorial-abstract progression, and systematic variation — all features supported by cognitive load theory and the worked example effect.

Singapore's success suggests that centrally directed reform can work when it targets the instructional core, builds teacher capacity, and maintains focus over decades. But Singapore, like Finland, is a small city-state with a strong government and distinctive cultural conditions. The scalability question remains.

4.6 CRISIS AS CATALYST: BLYTH, KLEIN, AND KINGDON

The PI's prior reading in institutional change theory — Scott, Blyth, Klein — identifies crisis as a central mechanism of institutional change. The connection to education reform is direct and illuminating.

Mark Blyth (2002), in *Great Transformations*, argues that institutional change requires what he calls "ideational change" — a shift in the ideas that legitimate existing institutional arrangements. Existing institutions are sustained not just by material interests but by shared ideas about how things should work. Crisis destabilizes these ideas by making the existing institutional framework visibly inadequate. When the existing framework is perceived as failing, previously marginal ideas become viable alternatives.

Naomi Klein (2007), in *The Shock Doctrine*, documents how crises are exploited to implement institutional changes that would be impossible under normal conditions. Her account is more critical than Blyth's — emphasizing how powerful actors use crises to impose changes that serve their interests rather than the public good. But the structural insight is the same: crisis opens policy windows that are normally closed.

John Kingdon (1984), in *Agendas, Alternatives, and Public Policies*, provides the most detailed framework for understanding when policy change becomes possible. Kingdon argues that three "streams" must converge for policy change to occur: a problem stream (a recognized problem demanding attention), a policy stream (a viable solution that has been developed and is available), and a politics stream (political conditions favorable to change). When all three streams converge — what Kingdon calls a "policy window" — change becomes possible. Policy windows are often opened by crisis, but they close quickly if the available solution is not ready for implementation.

The Kuhn translation available in the lab (Chapter 7: Crisis and the Emergence of Scientific Theories) provides a striking parallel. Kuhn argues that paradigm change in science is preceded by crisis — a prolonged period in which the existing paradigm fails at its own problems, generating "pronounced professional insecurity." The parallel to education is obvious: the grammar of schooling is a paradigm, and successful reform requires not just a better alternative but a crisis that makes the existing paradigm's failures undeniable.

4.7 COVID-19: THE NATURAL EXPERIMENT

The COVID-19 pandemic was the most significant disruption to global education in a century. Schools worldwide closed, millions of students shifted to remote learning, and the grammar of schooling was suspended — not by policy choice but by public health necessity. Zhao (2020), writing in the early months of the pandemic, argued that COVID could serve as a catalyst for educational change, opening a Kingdon-style policy window.

What actually happened is more instructive than what was predicted. The pandemic revealed the deep dependence of the educational system on the grammar of schooling. Remote learning was widely experienced as inferior — not because technology failed but because the grammar's structures (scheduled class periods, grade-level progression, teacher-led instruction) were crudely replicated online rather than redesigned for a different medium. Schools attempted to reproduce the grammar of schooling through Zoom, and the result was widely experienced as a degraded version of the original.

When schools reopened, the dominant response was reversion to pre-pandemic structures, not reform. The grammar reasserted itself with remarkable speed. Some innovations persisted — hybrid learning options, increased comfort with technology, some flexibility in scheduling — but

the fundamental structures remained. The pandemic opened a policy window, but the alternative solutions were not ready for implementation (Kingdon's policy stream was underdeveloped), and the political demand was for a return to normalcy, not for innovation.

The COVID experience confirms several principles from institutional theory. First, crisis alone is insufficient for reform; the alternative must be prepared in advance. Second, reversion to the mean is a powerful institutional force — disruption is temporary, the grammar is permanent. Third, the grammar of schooling is so deeply embedded in social expectations that even a global pandemic could not dislodge it.

However, COVID may have longer-term effects that are not yet visible. The pandemic normalized the idea that education can happen outside traditional school buildings. It created a population of parents who experienced their children's schooling directly and found it wanting. It accelerated the growth of microschools, learning pods, and homeschooling. These effects may represent the beginning of ideational change — a shift in shared assumptions about what education must look like — even if the grammar of the existing system remains intact.

4.8   TEACHER AGENCY AND THE LIMITS OF POLICY

One of the most important findings from the reform literature is the central role of teacher agency — the capacity of teachers to act purposefully within and upon the institutional structures that shape their work. Priestley et al. (2012), in a study cited 656 times, found that teacher agency in curriculum-making is not simply a matter of individual disposition; it is shaped by the institutional context in which teachers work. Teachers who operate in supportive environments with clear goals, adequate resources, and professional trust exercise more agency than teachers in controlling environments, regardless of their individual beliefs or capabilities.

Biesta et al. (2015), in a paper cited 1,139 times, extended this analysis by examining how teachers' beliefs mediate their responses to reform demands. Teachers do not simply implement policies; they interpret them through their existing beliefs about teaching, learning, and the purpose of education. A teacher who believes that the purpose of education is to transmit knowledge will interpret a "student-centered learning" mandate very differently from a teacher who believes the purpose is to develop agency. The implication is that reform cannot succeed by changing policy alone; it must also engage with the beliefs and professional culture of the teaching profession.

This finding connects to a broader insight about the relationship between institutional structure and individual agency. Argyris and Schön (1978) distinguished between "single-loop learning" — adjusting actions to achieve existing goals — and "double-loop learning" — questioning and changing the goals themselves. Most educational reform demands single-loop learning from teachers: do the same thing differently. Genuine institutional change requires double-loop learning: question whether the current goals and structures are the right ones. But double-loop learning is threatening because it challenges the assumptions on which the institution is built. Schools, like most organizations, have powerful defenses against double-loop learning — what Argyris called "defensive routines" that protect existing assumptions from examination.

Senge (1990) argued in *The Fifth Discipline* that organizations must become "learning organizations" capable of continuous self-examination and adaptation. A learning organization does not merely implement external mandates; it develops the internal capacity to identify problems, test solutions, and learn from results. The gap between the learning organization ideal and the reality of most schools is vast. Most schools are organized for efficiency and compliance, not for learning — the institution's own learning, that is, not just the students'.

4.9   THE NO CHILD LEFT BEHIND LESSON

The No Child Left Behind Act (NCLB) of 2001 and its successor, the Every Student Succeeds Act (ESSA) of 2015, represent the most ambitious top-down education reform efforts in American history. NCLB mandated annual testing in reading and math for all students in grades 3–8 and once in high school, required schools to demonstrate "adequate yearly progress" toward proficiency for all students, and imposed escalating sanctions on schools that failed to meet targets.

NCLB's effects illustrate the dynamics of institutional response to reform pressure. Schools responded to the testing mandate in predictable ways: they reallocated instructional time toward tested subjects (reading, math) and away from untested subjects (science, social studies, arts, physical education). They focused instruction on students near the proficiency cutoff — where marginal effort would produce the greatest improvement in accountability metrics — while devoting less attention to students far above or far below the cutoff. They adopted test preparation strategies that improved test scores without necessarily improving learning. In the worst cases, they engaged in outright cheating on tests.

This is the decoupling phenomenon at work. Schools changed their formal behavior (what was measured) while their core practices adapted in ways that optimized for the metric rather than for the underlying goal. NCLB's designers assumed that measuring outcomes would create pressure for improvement. What they did not anticipate — but institutional theory would have predicted — is that organizations optimize for what is measured, not for what is intended.

The NCLB experience is a cautionary tale for any reform effort that relies on external accountability without engaging the instructional core. Measuring test scores does not improve teaching. It improves test scores — which are related to but not identical with learning. The distinction matters enormously. Goodhart's Law — "when a measure becomes a target, it ceases to be a good measure" — applies to education with particular force because educational outcomes are multidimensional and the easily measured dimensions (test scores) are the least important ones from the perspective of the competence stack.

4.10   TOP-DOWN VERSUS BOTTOM-UP REFORM

The evidence on reform direction is nuanced. Top-down mandates (No Child Left Behind, Common Core) produce formal compliance but rarely change the instructional core. Bottom-up innovation (individual teachers experimenting with new methods) produces excellent results in individual classrooms but rarely scales. The evidence suggests that effective reform requires *both* — top-down support structures that create conditions for change, combined with bottom-up professional capacity to implement change at the classroom level.

Bryk et al. (2015), in *Learning to Improve*, argue that education reform should adopt the improvement science model from healthcare: networked improvement communities that use Plan-Do-Study-Act cycles to test changes iteratively, measure results with disciplined inquiry, and spread successful practices through structured networks. This approach — which Bryk calls "learning to improve" — occupies a middle ground between top-down mandate and bottom-up innovation. It provides structure and accountability (top-down) while preserving professional judgment and local adaptation (bottom-up).

The improvement science approach is theoretically promising but empirically early. The Carnegie Foundation for the Advancement of Teaching, which Bryk led, has implemented networked improvement communities in several contexts (community college math, new teacher development), with initial results suggesting that structured improvement processes can produce faster, more

reliable improvement than either top-down mandates or unstructured innovation. But the evidence base is still thin relative to the ambition of the claims.

4.11 THE MARKET-BASED REFORM DEBATE

A significant strand of reform thinking — particularly prominent in the United States since the 1990s — argues that the solution to institutional stagnation is market competition. If parents can choose among schools (through vouchers, charter schools, or open enrollment), the argument goes, competitive pressure will force schools to improve or lose students. Good schools will grow; bad schools will close. The market will do what policy mandates cannot.

The evidence on market-based reform is mixed and politically charged. The charter school lottery studies reviewed in Section VI provide some of the strongest causal evidence for positive effects of institutional alternatives. But the broader market-based reform literature shows more modest results. Voucher programs in Milwaukee, Cleveland, and Washington, D.C. have produced small positive effects in some studies and null effects in others. The assumption that competition alone will drive improvement has not been consistently supported.

The institutional analysis suggests why market-based reform has limited power. DiMaggio and Powell's isomorphism framework predicts that competitive pressure (mimetic isomorphism) will cause schools to converge on similar structures rather than differentiate. Schools competing for students will adopt the features that parents recognize and expect — the grammar of schooling — rather than innovating in ways that parents may not understand or trust. A school that eliminates grades, adopts competency-based progression, and abandons age-grading may be pedagogically superior, but it faces a marketing problem: it does not look like a school. Parents, like all consumers, prefer the familiar.

Market-based reform also assumes that parents have the information needed to choose well. In practice, parents often choose schools based on reputation, location, demographics, and test scores — not based on instructional quality or evidence-based practice. The information asymmetry between education providers and education consumers is far greater than in most markets. A patient can ask their doctor what the evidence says about a treatment. A parent cannot easily determine whether a school's assessment practices align with the evidence.

The market-based approach, in sum, is insufficient as a sole reform strategy. Competition can create space for alternatives (as charter schools have done), but it cannot, by itself, produce the institutional change that the evidence requires. Structural reform — changing what schools are expected to do, how teachers are trained, how learning is assessed, and what counts as accountability — is necessary in addition to, not instead of, market mechanisms.

THE SCIENCE-PRACTICE GAP: WHY EDUCATION IGNORES ITS EVIDENCE BASE

5.1 THE SIZE OF THE GAP

The gap between what learning science has established and what educational practice does is not merely large; it is a defining feature of the field. The examples documented across multiple L1 investigations are striking:

- Students rely on highlighting and rereading — among the least effective learning strategies — while rarely employing retrieval practice and spaced repetition — among the most effective (Dunlosky et al., 2013; L1-003).
- Schools organize instruction by time rather than mastery, despite strong evidence that mastery-based approaches produce $d = 0.5–1.0$ effect sizes when time constraints are relaxed (Bloom, 1968; Kulik et al., 1990; L1-010).
- Assessment systems rely primarily on grades — which negate the benefit of feedback (Butler, 1988) and undermine intrinsic motivation (Deci et al., 1999) — despite decades of evidence supporting formative assessment (Black & Wiliam, 1998; L1-003).
- Instruction typically follows a uniform model regardless of learner expertise, despite robust evidence for the expertise reversal effect (Kalyuga et al., 2003; L1-004).
- Growth mindset interventions continue to be promoted despite meta-analytic evidence showing near-zero effects after correction for bias (L1-002).
- Nearly half of teachers endorse neuromyths about learning (Dekker et al., 2012; L0-001v2).

The gap is not a matter of individual teacher ignorance. Many teachers are aware of the evidence but operate in institutional contexts that prevent them from acting on it. A teacher who knows that formative assessment is more effective than grades cannot unilaterally eliminate grades when the school requires them for report cards, college transcripts, and accountability systems. A teacher who understands the expertise reversal effect cannot provide individualized instruction when class sizes are 30 and the schedule allows 50 minutes per period. The science-practice gap is a structural problem, not a knowledge deficit.

5.2 THE GAP ILLUSTRATED: WHAT L1 AGENTS FOUND VS. WHAT INSTITUTIONS DO

The evidence-practice disconnect becomes vivid when we juxtapose the findings of the lab's L1 agents with the institutional practices they describe:

Assessment (L1-003). Black and Wiliam (1998), in one of the most influential educational research papers ever published, demonstrated that formative assessment — assessment designed to inform teaching and learning rather than to evaluate and rank — produces some of the largest effect sizes in the educational literature. The testing effect (retrieval practice) is among the most replicated findings in cognitive psychology. L1-003 concluded that every learning session should include low-stakes retrieval practice, that feedback should be about the work rather than the person, and that grades should be separated from feedback.

What institutions actually do: most classrooms are dominated by summative assessment. Grades are the primary currency of the system. Feedback, when provided, is typically delivered alongside grades (which Butler, 1988, showed causes students to ignore the feedback and attend only to the grade). Retrieval practice is used incidentally, not systematically. The institutional infrastructure of education — transcripts, GPA calculations, class rank, college admissions — is built on a grading system that the evidence shows is counterproductive for learning.

Motivation (L1-002). Self-determination theory provides one of the strongest evidence bases in psychology. The undermining effect — tangible, expected, contingent rewards reliably undermine intrinsic motivation — has been replicated extensively (Deci et al., 1999). L1-002 concluded that curriculum design should minimize extrinsic rewards, support autonomy, calibrate challenge to competence, and foster genuine relationships.

What institutions actually do: schools are organized around extrinsic rewards and punishments. Grades, honor rolls, class rank, detention, and behavioral reward systems are the institutional mechanisms of motivation. Competition for limited rewards — valedictorian, college admission, scholarships — structures the social environment. The language of school is controlling rather than autonomy-supportive. The institutional design systematically thwarts the basic psychological needs that SDT identifies as essential for sustained motivation.

Instructional Design (L1-004). The expertise reversal effect — what works for novices harms experts and vice versa — is one of the most practically important findings in instructional design. L1-004 concluded that instruction should adapt to learner expertise, moving from explicit instruction for novices through productive failure for intermediates to independent inquiry for advanced learners.

What institutions actually do: instruction is uniform. All students in a class receive the same instruction regardless of their prior knowledge. The one-teacher, one-classroom, same-pace model makes differentiation extremely difficult. The schedule (50-minute periods) prevents the extended, flexible time blocks that expertise-adaptive instruction requires. The assessment system (same test for all students) reinforces the assumption that all students should learn the same material at the same pace.

These are not edge cases or unusual examples. They represent the standard operating procedure of mainstream education. The science-practice gap is not a gap at the margins; it is a gap at the center.

5.3   MEDICINE AS COMPARISON

The comparison between education and medicine is instructive because medicine has partially closed its science-practice gap while education has not. Understanding why illuminates the structural differences that make education harder.

Medicine's evidence-based practice movement began in the early 1990s with the explicit goal of closing the gap between research evidence and clinical practice. The infrastructure that developed over the following three decades includes: Cochrane systematic reviews that synthesize evidence across studies; clinical practice guidelines that translate evidence into specific treatment recommendations; regulatory mechanisms (FDA, NICE) that evaluate evidence before approving treatments; medical education curricula that teach evidence appraisal skills; and institutional mandates that require adherence to evidence-based guidelines.

This infrastructure does not exist in education at comparable scale or influence. Education has the What Works Clearinghouse (WWC) in the United States and the Education Endowment Foundation (EEF) in the United Kingdom, which are analogous to Cochrane reviews. But their influence

on practice is far weaker than Cochrane's influence on medical practice. There is no educational equivalent of the FDA — no regulatory body that evaluates evidence before approving educational interventions. There is no equivalent of clinical practice guidelines — no authoritative, evidence-based recommendations for classroom instruction that carry institutional force. Teacher education curricula typically do not include training in research appraisal. And there is no institutional mandate requiring evidence-based practice in classrooms.

Cairney and Oliver (2017) argue that "evidence-based policymaking is not like evidence-based medicine" because the policy environment is fundamentally different from the clinical environment. Medicine has relatively clear outcome measures (morbidity, mortality, symptom reduction), relatively clear causal pathways (biological mechanisms), and relatively strong professional consensus about what constitutes evidence. Education has contested outcome measures (test scores? well-being? long-term flourishing?), complex causal pathways (learning is affected by dozens of interacting factors), and deep professional disagreement about what constitutes evidence and even about what the goals of education should be.

These structural differences are not merely obstacles to be overcome; they reflect genuine differences in the nature of the two fields. Medical treatments can be standardized because biological mechanisms are relatively uniform across patients. Educational interventions cannot be standardized in the same way because learning is profoundly shaped by context — the learner's prior knowledge, motivation, cultural background, and the specific classroom environment. Joyce and Cartwright (2019) argue that even well-validated educational interventions may not "work locally" if the conditions of the research context differ from the implementation context.

5.4 WHY TEACHERS DON'T USE EVIDENCE

Research on barriers to evidence use in education identifies several structural factors, not just individual knowledge deficits.

Access. Most teachers do not read research journals, not because they are incurious but because journals are behind paywalls, written in inaccessible language, and designed for academic audiences. Weinstein, Madan, and Sumeracki (2018) demonstrated that it is possible to translate cognitive science into accessible strategies, but such translation efforts are rare.

Time. Teachers work in time-constrained environments. Finding, reading, evaluating, and translating research evidence requires time that is not allocated in the typical teacher's schedule. Professional development time, when available, is often consumed by administrative requirements or by programs that are themselves not evidence-based.

Incentives. Teachers are not rewarded for using evidence. They are rewarded (or at least evaluated) based on student test scores, classroom management, and compliance with administrative requirements. Evidence use is invisible in most evaluation systems.

Training. Teacher education programs typically do not train teachers to appraise research evidence. Connolly et al. (2018) found that the number of randomized controlled trials in education research increased substantially between 1980 and 2016, but this increase in evidence production has not been matched by increased capacity among practitioners to evaluate and use that evidence.

Institutional constraints. Even teachers who know the evidence face institutional barriers to implementing it. The grammar of schooling constrains what is possible in a classroom. You cannot implement mastery-based progression in a time-based system. You cannot replace grades with formative feedback when grades are required for transcripts. You cannot provide individualized instruction when class sizes are too large. The science-practice gap is maintained not by teacher ig-

norance but by institutional structures that make evidence-based practice impractical or impossible within the existing grammar.

5.5 THE RESEARCHER-PRACTITIONER DIVIDE

The science-practice gap is maintained not only by institutional constraints on practitioners but also by institutional constraints on researchers. Academic reward structures create their own contribution to the gap. Researchers are rewarded for publishing in prestigious journals, not for translating findings into practice. Tenure and promotion decisions are based on citation counts, h-indices, and journal impact factors — none of which measure whether research has influenced practice. A researcher who spends five years translating cognitive science into classroom-ready materials has produced nothing that "counts" in the academic reward structure. A researcher who publishes the same findings in abstruse journal articles has built a career.

This creates a systematic underproduction of translation. The people who produce the evidence are not rewarded for making it usable. The people who could use the evidence do not have the training or time to translate it themselves. The result is what might be called a "translation gap" — a structural failure in the knowledge production system that leaves evidence stranded in academic journals, inaccessible to the practitioners who could implement it.

The Education Endowment Foundation in the UK has attempted to address this gap by funding both research and implementation, and by producing practitioner-facing summaries of evidence. The What Works Clearinghouse in the US reviews evidence but has been criticized for a narrow focus on randomized controlled trials that excludes much of the evidence base. Neither institution has the authority or influence to mandate evidence-based practice in the way that Cochrane reviews influence clinical practice.

The most promising translation efforts have been bottom-up rather than top-down. Weinstein, Madan, and Sumeracki (2018) created "The Learning Scientists" — a set of freely available, accessible materials translating cognitive science into study strategies for students and teachers. Their work demonstrates that translation is possible, but it remains a cottage industry rather than a systematic enterprise. What is needed is institutional infrastructure for translation — organizations whose explicit purpose is to bridge the gap between research and practice, staffed by people who understand both worlds, and funded at a scale commensurate with the size of the gap.

5.6 THE PROBLEM OF CONTEXTUAL VARIATION

Joyce and Cartwright (2019) raise a fundamental challenge for evidence-based education that goes beyond the simple question of access and translation. Their argument is that even well-validated interventions may not work in new contexts because the causal mechanisms that produce the effect depend on contextual factors that may not be present in the implementation setting.

In medicine, this problem is less severe because biological mechanisms are relatively uniform across patients. A drug that lowers blood pressure in a clinical trial will lower blood pressure in a clinical setting, assuming the drug is administered correctly. In education, the analogous assumption often fails. A teaching method that produced learning gains in a controlled study with motivated students, a skilled teacher, and a supportive school culture may not produce the same gains in a different setting with different students, a less skilled teacher, and a hostile school culture. The "intervention" in education is not a pill that acts directly on biology; it is a complex social process that depends on relationships, context, and implementation quality.

This does not mean that educational evidence is useless — far from it. The principles of effective learning (retrieval practice, spaced repetition, formative assessment, expertise-adaptive instruction) are supported by such robust and varied evidence that they can be considered genuine principles, not just context-specific findings. But the *implementation* of these principles must be adapted to context. What looks different in every setting is not whether retrieval practice works, but how to implement retrieval practice in this classroom, with these students, within this schedule, given this assessment system.

This is why the improvement science approach (Bryk et al., 2015) is promising: it builds contextual adaptation into the implementation process through PDSA cycles that test changes locally and iteratively adjust them based on local results. It does not assume that what worked in one setting will work identically in another; it provides a structured process for adapting evidence-based principles to local conditions.

5.7 IMPLEMENTATION SCIENCE: BRIDGING THE GAP

The field of implementation science — developed primarily in healthcare — provides frameworks for understanding and closing science-practice gaps. Damschroder et al. (2009) proposed the Consolidated Framework for Implementation Research (CFIR), which has been cited over 14,000 times and identifies five domains that affect implementation success: intervention characteristics, outer setting, inner setting, characteristics of individuals, and the implementation process.

Nilsén (2015) reviewed implementation theories, models, and frameworks, identifying a rich literature on how to move evidence into practice. Greenhalgh et al. (2017) developed the NASSS framework (Non-Adoption, Abandonment, and challenges to Scale-Up, Spread, and Sustainability), which explains why health technologies are often not adopted or are abandoned after initial adoption.

These frameworks have been developed primarily for healthcare and have been only minimally applied to education. The adaptation is not straightforward — the organizational contexts differ significantly — but the core insight is transferable: closing a science-practice gap requires not just better evidence but better implementation infrastructure. Evidence alone does not change practice; evidence embedded in supportive institutional structures, with trained implementers, adequate resources, and aligned incentives, can change practice.

Bryk et al.'s (2015) improvement science approach represents the most promising adaptation of implementation science principles to education. Networked improvement communities bring together researchers and practitioners in structured collaborations focused on specific problems of practice. The approach uses Plan-Do-Study-Act cycles — borrowed from quality improvement in manufacturing and healthcare — to test changes iteratively, measure results, and spread what works. This approach addresses several barriers simultaneously: it creates time for evidence use (by building it into the improvement process), provides training in evidence appraisal (through participation in the community), creates incentives for evidence use (through shared accountability for results), and works within institutional constraints (by testing changes that are feasible in the local context).

Part III

ALTERNATIVES AND DESIGN

BUILDING AND SUSTAINING ALTERNATIVES

6.1 CHARTER SCHOOLS: WHAT THE EVIDENCE SHOWS

Charter schools represent the largest-scale attempt to create alternatives to conventional public schooling within the American education system. The charter school evidence base is now substantial and includes some of the strongest causal designs in education research — lottery-based studies that exploit the random assignment inherent in oversubscribed charter admissions.

Abdulkadiroğlu et al. (2011), in a study with an extraordinarily high FWCI of 136.45, used Boston's charter school lottery to estimate causal effects. They found that Boston's charter schools produced substantial achievement gains — roughly 0.4 standard deviations per year in math and 0.2 in reading — compared to traditional public schools. These are large effects by educational research standards. Dobbie and Fryer (2011) found similar results in New York City and identified five practices that characterized the most effective charter schools: increased instructional time, frequent teacher feedback, data-driven instruction, high-dosage tutoring, and a culture of high expectations.

However, the charter school evidence is more nuanced than either advocates or critics typically acknowledge. The positive results are concentrated in urban charter schools serving predominantly low-income, minority students — particularly "No Excuses" charter networks like KIPP, Success Academy, and Achievement First. Suburban charter schools and virtual charter schools show much weaker results, and some show negative effects. The charter school literature demonstrates that the *type* of charter matters enormously — a charter school license does not automatically produce better outcomes; specific institutional design features do.

The institutional design features that predict charter school effectiveness — extended instructional time, frequent feedback, data-driven instruction, high expectations, intensive tutoring — are features that the learning science evidence supports. They represent, in effect, a partial closing of the science-practice gap within a new institutional framework that is freed from some (but not all) of the grammar's constraints. Charter schools can extend the school day because they are not bound by district scheduling. They can provide frequent feedback because they hire and retain teachers who will do the work. They can use data-driven instruction because they build institutional infrastructure for it.

But charter schools also illustrate the limits of institutional alternatives. Many charter schools rely on teacher-centered, highly structured instructional models that, while effective at producing test score gains, may not develop upper-layer competencies (judgment, metacognition, epistemic character) as defined by the competence stack. The "No Excuses" model in particular has been criticized for its emphasis on behavioral compliance and its reliance on extrinsic rewards and punishments — features that the SDT evidence (L1-002) suggests will undermine long-term motivation. Some charter networks have acknowledged this tension and are evolving toward more autonomy-supportive models, but the evolution is uneven.

6.2 MICROSCHOOLS AND LEARNING COOPERATIVES

The post-COVID period has seen rapid growth in microschools — small, independent learning environments typically serving 5–15 students — and learning cooperatives organized by parent groups. This movement represents a grassroots attempt to build alternatives outside the conventional system.

The evidence base for microschools is almost nonexistent. A scholar search for microschool research returned only 9 results, mostly tangential. This is not surprising — the movement is too new and too heterogeneous for systematic research. What we know comes primarily from descriptive reports and practitioner accounts.

The microschool movement is theoretically interesting because it represents an attempt to escape the grammar of schooling entirely rather than reforming it from within. Microschools can implement competency-based progression, multi-age grouping, personalized pacing, and project-based curriculum because they are not subject to the institutional pressures (accreditation, state testing, Carnegie units) that constrain conventional schools. They can create the psychologically safe, feedback-rich environments that L1-009 and L1-010 identify as prerequisites for upper-layer competence development.

But microschools face their own challenges. Scalability is the most obvious — a learning model that requires one adult per 5–10 students is not economically viable for mass education. Equity is a related concern — microschools are currently accessible primarily to affluent families who can afford tuition or who have a stay-at-home parent available to organize and supervise learning. Quality assurance is difficult without external evaluation standards. And sustainability is uncertain — many microschools depend on the energy and commitment of a single founding family or teacher.

6.3 THE DRIFT PROBLEM

L1-006 identified a persistent pattern in alternative education: over time, alternative institutions tend to drift back toward conventional structures. Montessori schools gradually introduce more teacher-directed instruction, grades, and age-segregated classrooms. Progressive schools slowly reintroduce the structures they were founded to escape. Even charter schools, which were created to be laboratories of innovation, have largely converged on a narrow set of instructional models.

The drift problem is a direct consequence of institutional isomorphism. The forces identified by DiMaggio and Powell (1983) — coercive pressure from regulators and funders, mimetic pressure from successful models, normative pressure from professionally trained teachers — operate on alternative institutions just as they operate on conventional ones. A Montessori school that wants to attract families must communicate in terms those families understand — grade levels, transcripts, test scores. A charter school that wants to retain its charter must demonstrate results on state assessments. A progressive school that hires conventionally trained teachers absorbs their conventional assumptions.

The drift problem suggests that building alternatives is necessary but not sufficient for institutional change. Individual alternatives can escape the grammar of schooling temporarily, but unless the broader institutional environment changes — the regulatory framework, the accountability system, the teacher training system, the college admissions process — the gravitational pull of the grammar will eventually reassert itself.

6.4 THE HOMESCHOOLING SURGE AND ITS INSTITUTIONAL IMPLICATIONS

The growth of homeschooling — accelerated dramatically by the COVID pandemic — represents a different kind of institutional challenge. Homeschooling families are not building alternative institutions; they are opting out of institutions entirely. The research base, as L0-001v2 documented, is thin and plagued by self-selection bias (Gaither, 2017). But the institutional significance of the homeschooling movement extends beyond its measurable outcomes.

Homeschooling represents a rejection of the grammar of schooling at its most fundamental level — a refusal to accept that learning requires institutional mediation. For Applied Pedagogy, the homeschooling phenomenon raises a provocative question: if the grammar of schooling is the problem, and if reforming institutions is so difficult, is the solution to bypass institutions entirely?

The evidence suggests that bypassing institutions is not a viable solution for most families. Martin-Chang et al. (2011), as L1-006 documented, found that *structured* homeschooling outperformed conventional schooling while *unstructured* homeschooling underperformed it. The implication is that some organizational structure is necessary for effective learning — but the specific structures of the grammar of schooling are not. The question is what minimal institutional structure supports effective learning without importing the grammar's pathologies.

This question is directly relevant to Applied Pedagogy's institutional design. The answer from the evidence reviewed across all L1 agents is that the minimal necessary structure includes: clear competency standards (L1-010); expertise-adaptive instruction that moves from explicit to inquiry-based as competence develops (L1-004); formative assessment with rapid, specific feedback (L1-003); environments that support autonomy, competence, and relatedness (L1-002); and psychological safety that treats error as information (L1-009). What is *not* necessary — and is in many cases harmful — is the grammar's specific embodiments: age-graded classrooms, Carnegie units, letter grades, 50-minute periods, and competitive ranking.

6.5 WHAT COUNTERACTS DRIFT?

The evidence on what prevents institutional drift is thin but suggestive. Several factors appear to help:

Strong organizational identity. Institutions with a clear, distinctive mission and a strong organizational culture are more resistant to drift than those with vague or generic missions. The most successful Montessori schools maintain fidelity to Montessori principles through careful hiring, ongoing professional development, and explicit communication of the institutional identity.

Structural commitment devices. Some institutions build structural features that make drift difficult. Sudbury Valley, the prototype democratic school, has no curriculum, no classes, and no grades — these are not policies that can be gradually relaxed but structural features that would require a fundamental reconstitution of the institution to change.

Network support. Institutions embedded in networks of similar institutions resist drift better than isolated ones. Charter management organizations (CMOs) like KIPP maintain fidelity across their networks through shared training, shared curriculum, shared assessment, and mutual accountability. The Montessori movement's professional organizations provide a similar function, though less effectively.

Selection and socialization. Institutions that carefully select and socialize their members — both teachers and families — resist drift better than those that accept whoever comes. This has obvious equity implications — selective institutions are not a solution for mass education.

THE COMPETENCE STACK AND INSTITUTIONAL DESIGN

7.1 INSTITUTIONS AS COMPETENCE ENVIRONMENTS

The competence stack (COMPETENCE-TARGET.md) defines five layers of competence — domain knowledge, skill, judgment, metacognition, and character/disposition — and identifies the environmental multiplier as a first-order determinant of whether upper-layer competence develops or degrades. This framing transforms institutional design from an administrative question into a pedagogical one.

L1-009 found that environments that penalize honesty do not merely suppress truth-telling; they degrade the capacity to perceive truth over time. L1-010 found that training contexts that create psychological safety, error management cultures, and honest self-assessment norms produce better learning outcomes across all layers of the competence stack. The convergence is clear: institutional environments are not neutral containers for learning; they are active shapers of the kind of competence that can develop within them.

The standard grammar of schooling creates environments that are reasonably effective at layers 1–2 (domain knowledge and skill) but systematically hostile to layers 3–5 (judgment, metacognition, and character). The mechanisms are structural:

Judgment (layer 3) requires exposure to varied, ambiguous, consequential situations with feedback (Kahneman & Klein, 2009; L1-009). The standard classroom provides uniform, well-defined, low-consequence problems with delayed feedback. The structural features that would support judgment development — varied case studies, ambiguous problems, rapid feedback on judgment calls — require different scheduling (longer periods), different assessment (evaluating reasoning processes, not just answers), and different authority structures (teachers as coaches rather than evaluators).

Metacognition (layer 4) requires honest self-assessment and calibration (L1-009). The standard grading system punishes honest self-assessment by treating uncertainty as ignorance. A student who writes "I'm not sure about this" on an exam gets a lower grade than a student who guesses confidently. The structural features that would support metacognition — calibration exercises, prediction-first pedagogy, revision opportunities — require assessment systems that reward accurate self-monitoring rather than confident performance.

Character and disposition (layer 5) are shaped by the hidden curriculum (L1-009). An institution that talks about intellectual honesty while rewarding confident performance teaches that confidence matters more than honesty. An institution that espouses collaboration while grading on a curve teaches that learning is a zero-sum competition. Character is not taught; it is *cultivated or degraded by the environment*. The structural features that would support epistemic character — psychological safety, error tolerance, visible modeling of intellectual humility by authority figures — require fundamental changes in institutional culture, not just policy.

7.2 ORGANIZATIONAL LEARNING AND THE INSTITUTIONAL CAPACITY FOR SELF-CORRECTION

The competence stack's environmental multiplier — question 5: "Are they allowed to tell the truth in that system?" — has a direct institutional analog. Organizations, like individuals, can be more or less capable of honest self-assessment. Argyris (1977) identified "organizational defensive routines" — patterns of behavior that prevent organizations from examining their own assumptions and practices. These routines are pervasive in education.

When a school's test scores decline, the typical institutional response is to blame external factors (student demographics, family situations, funding) or to double down on existing approaches (more test prep, more discipline, more homework). The response that Argyris would call "double-loop learning" — questioning whether the school's fundamental assumptions about teaching and learning are correct — is rare because it is threatening. It implies that the institution itself may be part of the problem.

Edmondson's (2019) work on psychological safety, extensively cited in L1-009 and L1-010, applies to organizational self-assessment as well as individual self-assessment. An institution in which teachers and administrators feel safe reporting problems, questioning assumptions, and proposing changes is more capable of self-correction than one in which problems are concealed, assumptions are protected, and proposals for change are experienced as criticism. Psychological safety is not just a feature of the learning environment for students; it is a feature of the institutional environment for the adults who run the institution.

The implication for Applied Pedagogy is that the institution must be designed to learn — not just to deliver learning to students, but to learn about its own effectiveness and to improve continuously. This requires institutional structures that support honest self-assessment at every level: student self-assessment (L1-009), teacher reflection on practice (L1-010's debriefing), and institutional assessment of whether the design is working as intended. An institution that designs a beautiful curriculum but cannot detect when it is failing is an institution that will drift toward ineffectiveness without knowing it.

Van Dyck et al. (2005), cited in both L1-009 and L1-010, found that organizations with "error management cultures" — cultures that treat errors as opportunities for learning rather than as occasions for blame — outperform organizations with "error prevention cultures" — cultures that focus on preventing errors rather than learning from them. The distinction is subtle but consequential. Error prevention cultures create an environment in which errors are concealed, near-misses go unreported, and systemic problems accumulate until they produce catastrophic failures. Error management cultures create an environment in which errors are surfaced, analyzed, and used to improve the system. Applied Pedagogy should design for error management, not error prevention — at every level, from student learning to institutional operations.

7.3 THE TRAINING-EDUCATION CONTRAST

L1-010 documents a revealing contrast between training contexts (military, aviation, medicine) and educational contexts. Training contexts consistently produce better outcomes at upper layers of the competence stack. The structural reasons are illuminating:

The organization bears the cost of failure. In training contexts, when the learner fails to develop competence, the organization pays the price — in failed missions, crashed aircraft, patient harm. This creates powerful institutional incentives for effective learning. In education, when the learner fails to develop competence, the *learner* pays the price — in poor grades, limited opportunities,

reduced life outcomes. The organization (the school) pays nothing. This difference in who bears the cost of failure produces radically different institutional designs.

Assessment is diagnostic, not evaluative. Training contexts use assessment to identify gaps in competence so they can be addressed. They do not use assessment to rank or sort learners. A pilot who fails a simulator check does not get a failing grade; they get additional training. This structural difference — diagnostic versus evaluative assessment — is what L1-003 recommends but what the grammar of schooling prevents.

Time is flexible. Training contexts allow learners to take the time they need to achieve mastery. A medical resident who needs six months to master a procedure gets six months, not a failing grade after four. The Carnegie unit structure prohibits this flexibility in education.

Feedback is rapid, specific, and connected to consequences. The debriefing practices described in L1-010 — after-action reviews, structured reflection, immediate feedback on performance — create the tight feedback loops that support competence development at all layers. The typical educational feedback loop — submit an assignment, wait days or weeks for a grade, receive minimal specific guidance — is far too slow and too vague.

The training-education contrast suggests that the grammar of schooling is not just suboptimal for learning; it is structurally incompatible with the development of upper-layer competence. Institutions that need their members to develop judgment, metacognition, and epistemic character — and Applied Pedagogy argues that all educational institutions should need this — must adopt institutional structures that are fundamentally different from the current grammar.

Part IV

SYNTHESIS

IMPLICATIONS FOR APPLIED PEDAGOGY: WHAT KIND OF INSTITUTION SHOULD THIS BE?

8.1 THE STRATEGIC CHOICE

Applied Pedagogy faces a three-way strategic choice that every entity attempting to implement evidence-based education must face:

Option 1: Reform from within. Work within existing institutions to change practice. This is the approach of improvement science (Bryk et al., 2015), professional learning communities, and evidence-based practice advocacy. Its advantage is scale — it engages the existing system where millions of students already are. Its disadvantage is the grammar: the structural constraints of existing institutions limit what is achievable, and the forces of institutional isomorphism and decoupling work against sustained change.

Option 2: Build alternatives. Create new institutions free from the grammar of schooling. This is the approach of charter schools, microschools, and alternative education models. Its advantage is freedom — new institutions can be designed from the ground up to embody the evidence. Its disadvantage is the drift problem and the scalability problem: new institutions face isomorphic pressure to converge toward conventional structures, and small-scale alternatives cannot serve the majority of learners.

Option 3: Prepare for crisis. Develop ready-to-implement alternatives that can be deployed when crisis opens a policy window. This is the Kingdon strategy — have the solution ready when the problem becomes undeniable and the political conditions align. Its advantage is leverage — crisis-driven change can produce rapid, large-scale transformation. Its disadvantage is uncertainty — crises are unpredictable, policy windows close quickly, and crisis-driven change can be captured by interests other than learning (Klein's shock doctrine concern).

The evidence suggests that Applied Pedagogy should pursue all three strategies simultaneously, with emphasis determined by the current political and institutional landscape.

8.2 INSTITUTIONAL DESIGN PRINCIPLES

Based on the evidence reviewed in this investigation, an Applied Pedagogy institution should embody the following design principles:

1. Design the hidden curriculum first. Before specifying content or pedagogy, specify the institutional culture. What does the institution reward? What does it punish? What does it model? What behaviors do people in authority display? The hidden curriculum is the most powerful pedagogical tool the institution possesses. It must be designed as deliberately as the formal curriculum.

2. Align assessment with learning, not sorting. Use criterion-referenced assessment (L1-010) rather than norm-referenced grading. Separate feedback from evaluation (L1-003). Use assessment diagnostically to identify and address gaps, not evaluatively to rank learners. Allow flexible time-to-mastery where possible. This requires abandoning the Carnegie unit model and adopting a competency-based framework.

3. Create the conditions for upper-layer competence. Build psychological safety explicitly and structurally. Treat error as information. Use debriefing and after-action reviews after every

complex learning activity (L1-010). Use case-based and scenario-based instruction for judgment development (L1-009). Build calibration practice into regular assessment for metacognition (L1-009). Model intellectual humility from positions of authority for character development (L1-009).

4. Build structural resistance to drift. Anticipate the isomorphic pressures that will push the institution toward conventional structures. Build commitment devices — structural features that make drift difficult. Invest in strong organizational identity. Create or join networks of similar institutions for mutual support and accountability. Hire and socialize staff carefully.

5. Invest in teacher capacity. The Finnish lesson is clear: teacher quality is the single most important institutional variable. Invest in selection, training, ongoing professional development, and professional autonomy. Treat teaching as a profession that requires deep expertise, not a delivery mechanism for pre-packaged curriculum.

6. Plan for multiple timescales. Implement evidence-based practice in current contexts (short-term). Build alternative institutions that embody the evidence (medium-term). Develop ready-to-implement designs for crisis-driven change (long-term). Do not bet on any single strategy.

7. Close the feedback loop. Build institutional infrastructure for continuous improvement. Use structured inquiry to evaluate whether innovations are producing the intended effects. Share findings across the network. The improvement science model (Bryk et al., 2015) provides a practical framework.

8. Design for organizational learning. The institution itself must be a learning organization — capable of honest self-assessment, error management, and double-loop learning. This means building psychological safety for adults (not just students), creating structured reflection practices at the institutional level, and treating institutional errors as information rather than as failures. An institution that cannot learn about its own effectiveness cannot sustain evidence-based practice over time.

8.3 WHAT APPLIED PEDAGOGY SHOULD NOT DO

The institutional analysis also clarifies several approaches that the evidence suggests would be counterproductive:

Do not assume that producing better evidence will close the gap. The science-practice gap is structural, not informational. More research, better communicated, is necessary but not sufficient. Without institutional infrastructure to support evidence use — protected time, aligned incentives, training in evidence appraisal, improvement processes — better evidence will join the existing evidence in being ignored.

Do not assume that building a single exemplary institution will change the system. Individual alternatives, however excellent, do not create systemic change. They create proof of concept, which is valuable, but they do not change the isomorphic pressures, regulatory frameworks, or cultural expectations that maintain the grammar of schooling. Systemic change requires systemic strategies.

Do not underestimate the grammar. The grammar of schooling has absorbed, diluted, or defeated every reform effort of the past century. Applied Pedagogy's institutional design should be built with the understanding that the grammar is the default state toward which all educational institutions tend, and that deliberate, structural countermeasures are necessary to resist it. Optimism about the power of good ideas is not a substitute for structural analysis of institutional dynamics.

Do not ignore the equity dimension. The current grammar of schooling is inequitable — it produces systematically worse outcomes for disadvantaged students. But alternatives to the grammar are also inequitable — they are currently accessible primarily to affluent families. Any institutional design that serves only the privileged few is not a solution but a new form of the

problem. Applied Pedagogy's institutional design must be accessible, affordable, and scalable — or it must include a deliberate strategy for becoming so.

THE GRAMMAR AS PARADIGM: A KUHNIAN ANALYSIS

The lab's translations of Thomas Kuhn's *The Structure of Scientific Revolutions* provide a framework for understanding institutional change in education that is more illuminating than the reform literature itself typically offers.

Kuhn argues that scientific change is not cumulative but revolutionary — old paradigms are not gradually improved but replaced by new ones that are fundamentally incompatible. The replacement requires a crisis that makes the old paradigm's failures undeniable, followed by a period of extraordinary science in which competing frameworks are proposed and debated, followed by the adoption of a new paradigm that reorganizes the field.

The grammar of schooling is a paradigm in Kuhn's sense. It is not merely a set of organizational arrangements; it is a way of understanding what education *is*. It defines what counts as a school, what counts as teaching, what counts as learning, and what counts as evidence that learning has occurred. When someone imagines a "school," they imagine age-graded classrooms with a teacher at the front and students in rows. When they imagine "teaching," they imagine explanation followed by practice. When they imagine "learning," they imagine acquiring knowledge that can be demonstrated on a test. These are not neutral descriptions; they are paradigmatic assumptions that make certain arrangements seem natural and others seem deviant.

9.1 PRE-PARADIGM CHAOS AND THE CURRENT MOMENT

Kuhn describes the pre-paradigm period in a science as characterized by competing schools, each with its own methods, standards, and foundational texts, and none able to establish dominance. The current moment in education reform has some features of pre-paradigm chaos. Competing frameworks — traditional schooling, progressive education, Montessori, unschooling, competency-based education, project-based learning, direct instruction — coexist without any single framework establishing the kind of dominance that the grammar of schooling has held for a century.

But the analogy is imperfect. In science, the pre-paradigm period precedes the establishment of normal science. In education, the grammar of schooling *is* the established paradigm, and what we see is not pre-paradigm chaos but the kind of crisis that Kuhn describes as preceding a revolution: a proliferation of competing approaches that signals the existing paradigm's inability to solve its own problems.

The grammar of schooling's problems are increasingly visible. Motivational decline across schooling (L1-002). The assessment paradox (L1-003). The science-practice gap (this investigation). The failure to develop upper-layer competence (L1-009). The superior outcomes of training contexts with fundamentally different institutional structures (L1-010). These are anomalies in Kuhn's sense — problems that the existing paradigm has generated but cannot solve within its own framework.

9.2 THE CRISIS THAT ISN'T (YET)

Kuhn argues that crisis must reach a certain severity before a paradigm shift becomes possible. The existing paradigm must be perceived as failing at its *own* problems, not merely at problems defined by the alternative. By this standard, education is in a state of pre-crisis rather than full

crisis. The grammar of schooling continues to function tolerably well by its own standards — it produces measurable test scores, sorts students into hierarchies, and provides credentials that are accepted by employers and universities. The problems identified by learning science (motivational decline, failure at upper-layer competence, the science-practice gap) are not problems *within* the grammar's framework; they are problems visible only from *outside* it.

This is why the COVID pandemic did not produce a revolution in education. The crisis was external to the grammar — it disrupted the physical infrastructure of schooling but did not demonstrate that the grammar's *own* logic was flawed. When the physical disruption ended, the grammar reasserted itself because it had not been shown to fail at its own tasks. A true Kuhnian crisis in education would require the grammar of schooling to fail at something that the grammar itself considers important — producing measurable academic achievement, maintaining social order, providing credentials that open doors. If the credential system broke down (as AI may cause it to), or if standardized test scores consistently failed to predict adult outcomes (which they largely do, but this is not widely recognized), or if the grammar's products — graduates — were visibly and undeniably incompetent in ways that mattered to society, then a genuine crisis might develop.

9.3   PREPARING THE ALTERNATIVE PARADIGM

Kuhn observes that the new paradigm must be substantially developed *before* the crisis peaks. Copernicus had a detailed heliocentric model ready when the Ptolemaic crisis became acute. Lavoisier had the oxygen theory ready when the phlogiston crisis became undeniable. The alternative paradigm need not be perfect — Copernicus's model was not more accurate than Ptolemy's for many decades — but it must be articulate enough to attract adherents and demonstrate its potential.

Applied Pedagogy's role, from a Kuhnian perspective, is to develop the alternative paradigm — to articulate what evidence-based institutional design looks like in enough detail that it can be deployed when the crisis arrives. This means producing not just research findings (which the lab has done extensively) but institutional designs, operational specifications, and proof-of-concept implementations that demonstrate the alternative in practice. The Kingdon framework makes the same point in policy terms: the solution must be ready in the "policy stream" before the "problem stream" creates the political conditions for change.

The lab's research — across ten L1 investigations — has produced the intellectual foundation for an alternative paradigm. The competence stack defines the goal (five-layer competence, not just knowledge and skill). The expertise-adaptive model defines the instructional approach (matched to learner development, not uniform). The assessment research defines the feedback system (formative, diagnostic, separated from grades). The motivation research defines the relational environment (autonomy-supportive, competence-building, relational). The institutional analysis in this investigation defines the structural requirements (competency-based, psychologically safe, drift-resistant, equity-focused).

What remains is to assemble these components into a coherent institutional design and test it. This is the work that lies ahead.

10

10.1 HIGH CONFIDENCE FINDINGS

The following findings rest on robust evidence from multiple research traditions and have been consistently replicated:

1. **The grammar of schooling is real and powerful.** Schools resist structural change not primarily because of ideological opposition but because of deep organizational path dependence, institutional isomorphism, and the coordination costs of systemic change. (Tyack & Cuban, 1995; DiMaggio & Powell, 1983; Meyer & Rowan, 1977)

2. **The hidden curriculum is real and consequential.** Institutions teach through their structures at least as powerfully as through their content. The hidden curriculum can be made visible and deliberately shaped. (Jackson, 1968; Hafferty, 1998; Lempp & Seale, 2004)

3. **The science-practice gap in education is structural, not merely informational.** Teachers don't use evidence primarily because institutional structures prevent it, not because they are unaware of it. The gap will not be closed by better communication alone; it requires institutional change. (L1-002, L1-003, L1-004; Dekker et al., 2012; Dunlosky et al., 2013)

4. **The institutional environment is a first-order determinant of competence formation at layers 3–5.** This is the most consistently supported cross-cutting finding across the entire research lab. (L1-009; L1-010; Edmondson, 1999; Argyris, 1977)

5. **Reforms that do not reach the instructional core do not produce lasting change.** Policy changes that remain at the level of formal structure decouple from classroom practice. (Elmore, 2004; Cuban, 1993; Meyer & Rowan, 1977)

10.2 MEDIUM CONFIDENCE FINDINGS

These findings are supported by substantial evidence but have significant limitations or qualifications:

6. **Improvement science (networked improvement communities, PDSA cycles) is a promising approach to closing the science-practice gap.** The evidence base is growing but still limited compared to the ambition of the approach. (Bryk et al., 2015)

7. **Crisis opens policy windows for institutional change, but the window closes quickly and can be captured by interests other than learning.** The COVID experience confirms both the opportunity and the rapid reversion to the mean. (Blyth, 2002; Kingdon, 1984; Zhao, 2020)

8. **Charter schools with specific institutional design features (extended time, frequent feedback, data-driven instruction) produce substantial achievement gains for urban students.**

But the effects are concentrated in specific contexts and school types, and the relationship between achievement gains and upper-layer competence development is unclear. (Abdulka-diroğlu et al., 2011; Dobbie & Fryer, 2011)

9. **Finland and Singapore demonstrate that systemic education reform is possible over multi-decade timescales when focused on teacher capacity and instructional quality.** But both are small, distinctive systems, and scalability to larger, more diverse contexts is uncertain. (Sahlberg, 2011)

10.3   LOW CONFIDENCE / OPEN QUESTIONS

10. **Whether microschools and learning cooperatives represent a viable long-term alternative to conventional schooling.** The evidence base is almost nonexistent. The theoretical case is interesting but untested at scale.

11. **Whether institutional drift can be reliably prevented in alternative institutions.** The mechanisms of drift (isomorphic pressure) are well-understood; the countermeasures are suggestive but not rigorously evaluated.

12. **Whether the improvement science approach can produce systemic change in education or whether it will be absorbed by the grammar like previous reforms.** The approach is theoretically sound but faces the same institutional forces that have defeated previous reforms.

13. **What institutional design would allow Applied Pedagogy to implement its findings at scale without being captured by the grammar of schooling.** This is the lab's central practical question, and this investigation can frame it but not answer it.

Dissertation complete. Total length: approximately 22,000 words. All five guiding questions from the mandate have been addressed. Cross-references to all specified L1 agents have been integrated. The COMPETENCE-TARGET.md environmental multiplier has been treated as an institutional design problem throughout.

BIBLIOGRAPHY

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S., Kane, T. J., & Pathak, P. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics*, 126(2), 699–748.

Argyris, C., & Schön, D. A. (1978). *Organizational Learning*. Addison-Wesley.

Biesta, G., Priestley, M., & Robinson, S. (2015). The role of beliefs in teacher agency. *Teachers and Teaching*, 21(6), 624–640.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1–12.

Blyth, M. (2002). *Great Transformations: Economic Ideas and Institutional Change in the Twentieth Century*. Cambridge University Press.

Brown, M. E. L., Coker, O., Heybourne, A., & Finn, G. M. (2020). Exploring the hidden curriculum's impact on medical students: Professionalism, identity formation and the need for transparency. *Medical Science Educator*, 30, 1107–1121.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to Improve: How America's Schools Can Get Better at Getting Better*. Harvard Education Press.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58(1), 1–14.

Cairney, P., & Oliver, K. (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health Research Policy and Systems*, 15, 35.

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980–2016. *Educational Research*, 60(3), 276–291.

Cuban, L. (1993). *How Teachers Taught: Constancy and Change in American Classrooms, 1890–1990*. Teachers College Press.

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S., Alexander, J., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4(1), 50.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668.

Dekker, S., Lee, N. C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, 3, 429.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160.

Dobbie, W., & Fryer, R. G. (2011). Getting beneath the veil of effective schools: Evidence from New York City. *NBER Working Paper 17632*.

Domitrovich, C. E., Bradshaw, C. P., Poduska, J., et al. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*, 1(3), 6–28.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest*, 14(1), 4–58.

Edmondson, A. C. (2019). *The Fearless Organization*. Wiley.

Elmore, R. F. (2004). *School Reform from the Inside Out*. Harvard Education Press.

Fullan, M. (2007). *The New Meaning of Educational Change* (4th ed.). Teachers College Press.

Greenhalgh, T., Robert, G., Bate, P., Macfarlane, F., & Kyriakidou, O. (2005). *Diffusion of Innovations in Health Service Organisations*. BMJ Books.

Greenhalgh, T., Wherton, J., Papoutsi, C., et al. (2017). Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *Journal of Medical Internet Research*, 19(11), e367.

Hafferty, F. W. (1998). Beyond curriculum reform: Confronting medicine's hidden curriculum. *Academic Medicine*, 73(4), 403–407.

Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, 33(8), 16–30.

Jackson, P. W. (1968). *Life in Classrooms*. Holt, Rinehart & Winston.

Joyce, K. E., & Cartwright, N. (2019). Bridging the gap between research and practice: Predicting what will work locally. *American Educational Research Journal*, 57(3), 1045–1082.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64(6), 515–526.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.

Kingdon, J. W. (1984). *Agendas, Alternatives, and Public Policies*. Little, Brown.

Klein, N. (2007). *The Shock Doctrine*. Picador.

Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2), 265–299.

Lempp, H., & Seale, C. (2004). The hidden curriculum in undergraduate medical education: Qualitative study of medical students' perceptions of teaching. *BMJ*, 329(7469), 770–773.

Mehta, J. (2013). *The Allure of Order: High Hopes, Dashed Expectations, and the Troubled Quest to Remake American Schooling*. Oxford University Press.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363.

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, 10, 53.

Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using Evidence: How Research Can Inform Public Services*. Policy Press.

Priestley, M., Edwards, R., Priestley, A., & Miller, K. (2012). Teacher agency in curriculum making: Agents of change and spaces for manoeuvre. *Curriculum Inquiry*, 42(2), 191–214.

Sahlberg, P. (2011). *Finnish Lessons: What Can the World Learn from Educational Change in Finland?* Teachers College Press.

Scott, W. R. (2014). *Institutions and Organizations: Ideas, Interests, and Identities* (4th ed.). Sage.

Senge, P. M. (1990). *The Fifth Discipline*. Doubleday.

Serdyukov, P. (2017). Innovation in education: What works, what doesn't, and what to do about it? *Journal of Research in Innovative Teaching & Learning*, 10(1), 4–33.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15–21.

Tyack, D., & Cuban, L. (1995). *Tinkering Toward Utopia: A Century of Public School Reform*. Harvard University Press.

Tyack, D., & Tobin, W. A. (1994). The "grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal*, 31(3), 453–479.

van Dyck, C., Frese, M., Baer, M., & Sonnentag, S. (2005). Organizational error management culture and its impact on performance: A two-study replication. *Journal of Applied Psychology*, 90(6), 1228–1240.

Weinstein, Y., Madan, C. R., & Sumeracki, M. A. (2018). Teaching the science of learning. *Cognitive Research: Principles and Implications*, 3, 2.

Whitehead, A. N. (1929). *The Aims of Education and Other Essays*. Macmillan.

Zhao, Y. (2020). COVID-19 as a catalyst for educational change. *Prospects*, 49, 29–33.