WORKING PAPER · NO. 2026-11

The Economics of Scaling Early Childhood Programs: Lessons from The Chicago School

John A. List

JANUARY 2026

THE UNIVERSITY OF CHICAGO
BECKER FRIEDMAN INSTITUTE FOR ECONOMICS

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

The Economics of Scaling Early Childhood Programs: Lessons from The Chicago School
John A. List
January 2026
JEL No. C9, C91, C93, H10, H4, H41, J24

ABSTRACT

Many ideas show remarkable returns in small-scale trials but often disappoint when scaled to broader populations and contexts. Using early childhood investment as a case study, this study develops a dynamic human capital formation model that integrates complementary skill investment with "Option C thinking" on scaling challenges. The model is stylized in the Chicago tradition: micro-founded with optimizing agents, dynamic skill production, and a policymaker evaluating scaling decisions. It formalizes how naive extrapolation from pilot studies systematically overestimates policy efficacy by ignoring "voltage drops," declining treatment effects due to unrepresentativeness at scale. The model demonstrates that optimal scaling policy requires mechanism-based design that anticipates these failures through backward induction from implementation realities. The scientific insights from a set of recent studies provide valuable perspectives on the model.

John A. List
University of Chicago
Department of Economics
and Australian National University
and also NBER
jlist@uchicago.edu

The Economics of Scaling Early Childhood Programs: Lessons from The Chicago School

John A. List*

January 5, 2026

Abstract

Many ideas show remarkable returns in small-scale trials but often disappoint when scaled to broader populations and contexts. Using early childhood investment as a case study, this study develops a dynamic human capital formation model that integrates complementary skill investment with "Option C thinking" on scaling challenges. The model is stylized in the Chicago tradition: micro-founded with optimizing agents, dynamic skill production, and a policymaker evaluating scaling decisions. It formalizes how naive extrapolation from pilot studies systematically overestimates policy efficacy by ignoring "voltage drops, " declining treatment effects due to unrepresentativeness at scale. The model demonstrates that optimal scaling policy requires mechanism-based design that anticipates these failures through backward induction from implementation realities. The scientific insights from a set of recent studies provide valuable perspectives on the model.

1 Introduction

The economics of human capital, the notion that individuals can invest in themselves to enhance their productive capacity, stands as one of the Chicago School's most enduring contributions to economic science. Beginning with Schultz (1961) presidential address to the American Economic Association, which challenged the

*University of Chicago, NBER, and Australian National University, jlist@uchicago.edu

This special issue was a direct result of the generous funding from the Griffin Incubator. Thanks to the hundreds of participants in the inaugural Griffin Incubator who deepened our understanding of early childhood investment and scaling. We are in the early innings of a science that promises to flourish: the science of using science. This study was enhanced by comments from Brent Hickman and Juanna Joensen. The assistance of Nour Wael Abdelbaki, Franco Daniel Albino, and Faith Fatchen improved the study. All errors remain my own.

prevailing view that education was merely consumption, Chicago economists developed a coherent framework for understanding how individuals and societies invest in skills, health, and knowledge across the life cycle. Becker (1964, 1993) seminal treatises formalized this intuition, modeling human capital investment decisions as rational economic choices where individuals weigh current costs against future earnings gains, just as firms evaluate physical capital investments.

Mincer (1958, 1974) provided the empirical foundation, demonstrating through his famous earnings function how years of schooling and labor market experience translate systematically into wage differentials—a relationship that has proven remarkably stable across time and space. Ben-Porath (1967) extended this framework dynamically, showing how optimal investment strategies evolve over the life cycle as individuals face changing opportunity costs and remaining working horizons. This human capital research program revolutionized labor economics, education policy, and development economics, earning Becker the Nobel Prize in 1992 and establishing Chicago as the intellectual center for understanding how economies transform raw human potential into productive capability.

Within this broader human capital tradition, Heckman and his collaborators redirected attention to a critical yet underexplored phase: early childhood. Building on insights from developmental psychology and neuroscience, Heckman's research program demonstrated that the foundational years from conception through age five represent a uniquely productive period for human capital formation, characterized by neuroplasticity, dynamic complementarities, and self-productivity that generate returns far exceeding later-life investments (Cunha and Heckman 2007, 2008; Cunha, Heckman, and Schennach 2010). The technology of skill formation framework formalized how "skills beget skills" and "learning begets learning"—early cognitive and socio-emotional capabilities raise the productivity of subsequent investments, creating dynamic multiplier effects across the life cycle.

Evidence from landmark interventions, such as the Perry Preschool Project and the Abecedarian Project, documented extraordinary returns to high-quality early childhood programs, with benefit-cost ratios approaching 10:1 through mechanisms including improved educational attainment, higher earnings, better health, and reduced criminal justice involvement (Heckman et al. 2010; Campbell et al. 2014; García et al. 2020). This work was extended by several scholars, including Pedro Carneiro, Flavio Cunha, Jorge Luis García, and many others, who established early childhood investment as perhaps the most cost-effective strategy for reducing inequality while simultaneously enhancing aggregate productivity, a rare alignment of equity and efficiency objectives.

Yet a troubling pattern emerged as policymakers attempted to translate these findings into large-scale programs: interventions proven effective in carefully controlled pilots frequently disappointed when scaled to broader populations. Head Start, the nation's largest early childhood program serving nearly one million children annually at costs exceeding $10 billion, showed modest cognitive gains that largely faded by third grade despite being inspired by the Perry Preschool success (Puma et al. 2012).[1] State pre-K expansions generated heterogeneous results far below experimental benchmarks (Durkin et al. 2022). This efficacy-effectiveness gap extends far beyond early childhood education. From health interventions to financial products to behavioral nudges, researchers have documented systematic "voltage drops" when promising pilots reach scale (List 2022). Indeed, it is estimated that across disciplines, ranging from software development to medicine to education and beyond, between 50 to 90 percent of results will lose voltage at scale (List 2022).

The voltage effect phenomenon challenges the core premise of evidence-based policymaking: that rigorous empirical evidence could reliably guide policy decisions. Working with Al-Ubaydli, Suskind, and others, List developed a systematic framework for understanding why interventions fail at scale across diverse domains—identifying common threats including false positives, representativeness failures, spillover effects, supply-side constraints, and cost structures that change with implementation scope (Al-Ubaydli, List, and Suskind 2017, 2019, 2020; Al-Ubaydli et al. 2021; List 2022, 2024).

What began as pragmatic concerns about implementation has matured into a rigorous research program with formal theoretical foundations, systematic empirical documentation, and practical design principles. This transformation elevates scaling from a policy challenge to a fundamental question in economic science: under what conditions do cost-effectiveness estimates generalize, and how should researchers design studies to maximize policy-relevant inference? The economics of scaling has become a bona fide object of scientific study (see, e.g., Al-Ubaydli, List, and Suskind 2017; Al-Ubaydli et al. 2017; Al-Ubaydli, List, and Suskind 2019, 2020; Al-Ubaydli, Lai, and List 2023; Brandon et al. 2022; Muralidharan and Singh 2025; Mobarak 2022; Angrist, Bergman, and Matsheng 2022; Angrist et al. 2023; Vivalt 2020; Fatchen, List, and Pagnotta 2025; Dougan, García, and Polovnikov 2025).[2]

Most importantly for the purposes of this study, what emerged from this literature is a scientific approach to avoid voltage drops. List (2022, 2024) introduced "Option C thinking," a framework that reorients the research design process. Rather than simply adding a treatment arm to augment standard A/B testing, Option C thinking transforms how researchers approach initial discovery. It shifts focus from purely establishing efficacy under controlled conditions to anticipating the constraints, moderators, and implementation realities that interventions will face at scale. The key insight is that evidence about scalability should be gener-

1. For a recent study on fadeout using the Chicago Heights Early Childhood Center initiative (CHECC) see (List and Uchida 2024).

2. The emerging "science of scaling" literature makes a crucial distinction often conflated in policy discussions: external validity versus scaling. External validity concerns whether treatment effects generalize across populations and settings, while scaling addresses whether the benefit-cost profile maintains when moving from pilots to population-level implementation (List 2024).

ated alongside the efficacy test, not after it. Option C thinking asks: if I want to scale up this idea, what extra information do I need beyond demonstrating that it works? This includes examining what constraints the intervention will face at scale, what key factors can impact scaling success, and whether the mediation paths and moderators observed in controlled settings persist under realistic implementation conditions. By embedding these scaling considerations into the original experimental design, researchers produce policy-based evidence that the science of scaling demands; evidence that reveals not just whether an intervention works, but under what conditions and for whom it will continue working when deployed broadly.

The contribution of this study is to formalize Option C thinking within a micro-founded model of skill production, extending Heckman's framework to explicitly incorporate the mechanisms that undermine scaled interventions. Given that few policy challenges carry greater economic and social stakes than childhood inequality, the integration of the human capital formation framework with lessons from the scaling literature represents a natural synthesis within the Chicago tradition: rigorous theoretical foundations combined with deep attention to implementation realities, all grounded in the belief that well-designed policy can enhance both economic efficiency and social equity.

The model includes heterogeneous children with dynamic skill formation, where early investments exhibit complementarity with initial endowments and later-life outcomes. Policymakers who adopt traditional A/B testing approaches evaluate programs based on efficacy in controlled settings, then face scaling decisions that often lead to voltage drops. In contrast, Option C thinking requires designing programs from the outset with scalability constraints embedded in the intervention architecture, anticipating implementation failures, cost dynamics, and real-world behavioral responses. This formalization clarifies when programs designed for scalability succeed, identifies which design features prevent voltage drops, and demonstrates how mechanism-based design improves policy outcomes.

The theoretical model generates several key insights. First, I derive conditions under which initial disadvantage creates dynamic complementarities that favor targeted over universal interventions, formalizing Heckman's intuition about equity-efficiency alignment. Second, I show that traditional research approaches systematically overestimate benefits by ignoring "voltage drops": declining effects from unrepresentative samples and situations, rising marginal costs, supply-side quality degradation, and general equilibrium spillovers. Third, the comparative statics reveal that optimal program scale balances marginal benefits against these scaling frictions, with the solution depending critically on the voltage drop magnitude and policymakers' ability to mitigate it through design adjustments. I then examine how the model's empirical predictions align with several recent empirical studies. The evidence reveals broad support for the framework's core mechanisms across diverse domains, from the microdynamics of skill formation to family investment decisions, peer effects, targeted interventions, and explicit analyses of scaling challenges.

The remainder of this study proceeds as follows. Section 2 introduces a simple dynamic skill formation model specifying how investments transform into skills through complementarity and self-productivity. Section 3 augments this simple model by introducing a scaling framework that serves as a theoretical bridge, providing formal economic scaling foundations to offer actionable guidance for preventing the voltage drops that plague scaled interventions across health, education, and social policy domains. Section 4 derives the policymaker's optimization problem with explicit first-order conditions and comparative statics, demonstrating how initial disadvantage, investment efficiency, and cost structures determine optimal targeting and scale under Option C design principles. Section 5 examines empirical evidence across multiple domains, illustrating how Option C thinking has succeeded in practice and where traditional approaches have failed. Section 6 concludes with policy implications and directions for future research.

2 Dynamic Skill Formation: Simple Model Setup

Mounting evidence across neuroscience, psychology, and public health converge on a powerful insight: the earliest years represent an unparalleled opportunity for intervention. During this critical developmental window, neuroplasticity peaks as synaptic connections form at rates exceeding one million per second (Phillips and Shonkoff 2000), skill formation exhibits strong complementarities through self-productivity (Cunha and Heckman 2007; Knudsen et al. 2006), and environmental inputs become biologically embedded through epigenetic mechanisms with lasting effects on stress response systems and executive function (Shonkoff et al. 2012; Willoughby and Blair 2016; McEwen and McEwen 2017).

Investments during this sensitive period yield returns that compound dynamically over the life cycle, affecting not only cognitive and socio-emotional development but also physical health, social relationships, and economic productivity decades later (Campbell et al. 2014; García et al. 2020; Elango et al. 2015). Well-designed early childhood programs can disrupt intergenerational transmission of disadvantage while simultaneously enhancing economic efficiency—a rare alignment of equity and productivity objectives supported by evidence from developmental economics, experimental psychology, and program evaluation research (Doyle 2020; Duncan and Magnuson 2013).

This section develops a stylized model in the Chicago tradition—micro-founded with optimizing agents, explicit functional forms amenable to analytical results, and clear mechanisms linking primitives to outcomes. The model captures three essential features of early childhood interventions: heterogeneity in initial conditions, dynamic complementarity in skill formation, and resource constraints facing policymakers.

2.1 Environment and Timing

Consider a simplified lifecycle with two childhood periods followed by adulthood. Period 0 corresponds to early childhood (ages 0-5), period 1 to later childhood and adolescence (ages 6-18), and period 2 to adult labor market outcomes. This stark periodization abstracts from continuous development while capturing the key insight that early investments affect both immediate skills and the productivity of future investments.

The population consists of a continuum of children indexed by $i \in [0,1]$, each with initial skill endowment $\theta_{0,i}$ drawn from distribution F_0 with density f_0, mean μ_0, and variance σ_0^2. For notational convenience, I integrate over the skill distribution using measure $f_0(\theta_{0,i})d\theta_{0,i}$. This initial endowment captures all pre-intervention factors affecting child development: parental education and income, neighborhood quality, prenatal health, genetic endowments, and early home environment quality. Disadvantaged children—those from low-income families, unstable households, or under-resourced communities—are characterized by lower realizations of $\theta_{0,i}$.

2.2 Technology of Skill Formation

Skills evolve according to a production function that embeds two core mechanisms from Heckman's framework: self-productivity and dynamic complementarity. The skill production function is given by:

$$\theta_{t,i} = A \cdot \theta_{t-1,i}^{\beta} \cdot (1 + I_{t-1,i})^{\delta} + \varepsilon_{t,i} \tag{1}$$

where $t \in \{1,2\}$ indexes childhood periods and i indexes individual children, with θ_0 representing initial endowment, θ_1 skills after early childhood (period 0 to 1), and θ_2 skills entering adulthood (period 1 to 2). In what follows, I suppress individual subscripts when describing the production technology generically, reintroducing them when analyzing population distributions.

In equation (1), each parameter has a clear economic interpretation:

Self-productivity ($\beta \in (0,1)$). The term θ_{t-1}^{β} captures "skill begets skill"—children with higher skills in period $t-1$ develop more skills in period t. The parameter $A > 0$ governs the overall productivity of the skill formation process, while the exponent $\beta < 1$ implies diminishing returns to existing skills.

Investment technology ($\delta \in (0,1)$). The term $(1 + I_{t-1})^{\delta}$ represents how program investments translate into skill gains, where $I_{t,i} \geq 0$ denotes investment intensity measured in resource units (e.g., thousands of dollars per child annually). We use $1 + I$ rather than I to ensure the function is well-defined when $I = 0$. The exponent $\delta < 1$ implies diminishing marginal returns to investment intensity.

Dynamic complementarity. Critically, the multiplicative form delivers:

$$\frac{\partial^2 \theta_t}{\partial \theta_{t-1} \partial I_{t-1}} = A\beta\delta \cdot \theta_{t-1}^{\beta-1} \cdot (1 + I_{t-1})^{\delta-1} > 0 \tag{2}$$

This positive cross-partial formalizes dynamic complementarity: investments are more productive for children with stronger baseline skills. The marginal product of investment increases with existing skill levels:

$$\frac{\partial}{\partial \theta_{t-1}}\left[\frac{\partial \theta_t}{\partial I_{t-1}}\right] = A\beta\delta \cdot \theta_{t-1}^{\beta-1} \cdot (1 + I_{t-1})^{\delta-1} > 0 \tag{3}$$

Equation (3) restates (2) with the order of differentiation reversed, underscoring the symmetry in how prior skills enhance the productivity of investments and vice versa, a property that follows from the commutativity of mixed partial derivatives in the skill production function.

Dynamic complementarity creates a tension between equity (targeting low-θ_0 children who are most disadvantaged) and efficiency (investing where marginal returns are highest). The resolution of this tension depends on the degree of diminishing returns (β) and the social planner's welfare function.

Stochastic shocks ($\varepsilon_{t,i}$). The shock $\varepsilon_{t,i} \sim N(0, \sigma_\varepsilon^2)$ captures idiosyncratic factors affecting development—health shocks, family disruptions, peer influences, or measurement error in skills. These shocks introduce uncertainty but, being mean-zero and independent across children, wash out in aggregate welfare calculations.

Investment costs. Each unit of investment costs $c_t > 0$, representing the opportunity cost of resources. In the baseline model, c_t is constant (implying perfect supply elasticity). Below, I relax this assumption by introducing $c_t(q)$ with $c_t'(q) > 0$ to capture supply constraints that raise marginal costs at scale.

2.3 Outcomes and Welfare

Adult outcomes depend on accumulated skills at the end of childhood. I specify a linear earnings function:

$$W_i = \phi\theta_{2,i} \tag{4}$$

where $\phi > 0$ translates skills into labor market productivity. This reduced-form specification subsumes a competitive labor market where wages equal marginal products. The linearity assumption finds support in log-linear Mincer equations relating skills to earnings (Heckman, Stixrud, and Urzua 2006).

The policymaker maximizes a utilitarian social welfare function:

$$SW = \int [W_i - c_0 I_{0,i} - c_1 I_{1,i}]\, f_0(\theta_{0,i}) d\theta_{0,i} \tag{5}$$

This specification weights all children equally (utilitarian welfare) and aggregates earnings net of program costs across the skill distribution, with children weighted by their population frequency $f_0(\theta_{0,i})$. Importantly, with heterogeneous treatment effects from the multiplicative production function, the utilitarian welfare function treats absolute skill gains equivalently across all children. A child moving from

$\theta_2 = 5$ to $\theta_2 = 6$ generates the same welfare gain ($\phi \cdot 1$) as a child moving from $\theta_2 = 10$ to $\theta_2 = 11$, even though the former represents a 20% gain while the latter is only 10%.

Alternative specifications could incorporate inequality aversion—for instance, prioritizing percentage improvements for disadvantaged children or using a concave transformation of earnings such as $u(W_i) = W_i^{1-\gamma}/(1-\gamma)$ with $\gamma > 0$. Such modifications would strengthen the case for targeting disadvantaged children beyond what my baseline specification implies. The utilitarian baseline thus provides a conservative benchmark for redistribution.

Three additional features merit discussion. First, the model ignores fiscal externalities and spillovers. In reality, investments that boost skills reduce future social costs through lower crime, public assistance, and healthcare utilization while increasing tax revenues (Heckman et al. 2010). Including these externalities would increase the net benefits of intervention, raising optimal investment levels. Second, I abstract from discounting by normalizing to present value terms. Third, I abstract from political economy constraints, budget limitations, and implementation frictions beyond those explicitly modeled below. The normative benchmark I characterize represents what an unconstrained social planner would choose.

2.4 Baseline Counterfactual

Understanding how skills evolve without intervention provides the essential benchmark against which to measure program effects. This baseline counterfactual characterizes the inequality dynamics that motivate early childhood investment: how initial advantages compound through self-productivity, potentially widening gaps over time even when proportional differences compress.

Without intervention ($I_{t,i} = 0$ for all t, i), skills evolve purely through self-productivity. The production function (1) simplifies considerably when the investment term $(1 + I_t)^\delta$ equals unity, yielding:

$$\begin{aligned}
\theta_{1,i} &= A\theta_{0,i}^{\beta} + \varepsilon_{1,i} \\
\theta_{2,i} &= A(A\theta_{0,i}^{\beta} + \varepsilon_{1,i})^{\beta} + \varepsilon_{2,i} \\
&\approx A^{1+\beta}\theta_{0,i}^{\beta^2} \quad \text{(suppressing shocks)}
\end{aligned}$$

This baseline evolution exhibits complex inequality dynamics that depend critically on whether gaps are measured in relative or absolute terms. The distinction proves essential for understanding when and why interventions can reduce inequality and provides several insights.

A first insight is that relative inequality compresses over time. Consider two children with initial endowments $\theta_0^H > \theta_0^L$ (high versus low). After period 1, their skill ratio becomes:

$$\frac{\theta_1^H}{\theta_1^L} = \frac{A(\theta_0^H)^\beta}{A(\theta_0^L)^\beta} = \left(\frac{\theta_0^H}{\theta_0^L}\right)^\beta \tag{6}$$

Since $\beta < 1$ by assumption (diminishing returns in self-productivity), we have $(\theta_0^H/\theta_0^L)^\beta < \theta_0^H/\theta_0^L$, meaning the skill ratio compresses: children who start behind catch up proportionally. This compression continues into period 2, with the skill ratio declining to $(\theta_0^H/\theta_0^L)^{\beta^2}$. The multiplicative structure with $\beta < 1$ creates convergence in percentage terms—a child starting with half the skills of another will end up closer to parity over time purely through diminishing returns.

However, a second insight is that the absolute skill gaps tell a different story. The absolute difference after period 1 evolves:

$$\theta_1^H - \theta_1^L = A(\theta_0^L)^\beta \left[\left(\frac{\theta_0^H}{\theta_0^L}\right)^\beta - 1\right] \tag{7}$$

Defining $r \equiv \theta_0^H/\theta_0^L > 1$ as the initial skill ratio, the absolute gap can be written as:

$$\theta_1^H - \theta_1^L = A(\theta_0^L)^\beta[r^\beta - 1]$$

The absolute gap widens (i.e., $\theta_1^H - \theta_1^L > \theta_0^H - \theta_0^L$) when:

$$A(\theta_0^L)^\beta[r^\beta - 1] > \theta_0^L[r - 1]$$

Rearranging yields the condition:

$$A > \frac{r - 1}{(\theta_0^L)^{\beta-1}(r^\beta - 1)}$$

This expression reveals that absolute gaps widen when A is sufficiently strong relative to initial disadvantage θ_0^L. The economic intuition proves subtle: even with diminishing returns creating proportional convergence, the advantaged child's higher skill level generates larger absolute skill increments each period. When the baseline θ_0^L is not too small and A is strong, these larger increments dominate the compression effect from diminishing returns, causing absolute gaps to expand.

Importantly, since $\beta < 1$ implies $r^\beta - 1 < r - 1$ for $r > 1$, the denominator $(r^\beta - 1)$ is smaller than the numerator $(r - 1)$, raising the threshold A required for gap widening. As initial inequality r increases, satisfying this threshold becomes progressively harder—very large initial gaps tend to compress even in absolute terms unless self-productivity is extremely strong.

Interestingly, the gap dynamics continue and potentially amplify as children age. The absolute skill difference entering adulthood becomes:

$$\theta_2^H - \theta_2^L = A^{1+\beta}(\theta_0^L)^{\beta^2} \left[\left(\frac{\theta_0^H}{\theta_0^L}\right)^{\beta^2} - 1\right] = A^{1+\beta}(\theta_0^L)^{\beta^2}[r^{\beta^2} - 1] \tag{8}$$

The exponent $\beta^2 < \beta$ creates even stronger relative compression by adulthood—a child starting with half the skills of another will close more than half the proportional gap by period 2. However, the multiplicative factor $A^{1+\beta}$ scales up absolute skill levels across both periods of development. Because the advantaged child starts from a higher base, this scaling produces larger absolute skill increments at each stage. The result: relative gaps compress (proportional catch-up continues) while absolute gaps can simultaneously widen (the disadvantaged child falls further behind in skill units). Children starting behind can fall progressively further behind in absolute terms when self-productivity is strong (A large) and initial disadvantage is moderate (not too small θ_0^L).

These baseline dynamics provide the rationale for early intervention targeting disadvantaged children. The goal is to boost θ_1 for low-θ_0 children through investment $I_0 > 0$, breaking the pattern where initial disadvantages compound through self-productivity. Early investments prove particularly powerful because they operate before the β^2 exponent in period-2 skills: raising a disadvantaged child's θ_1 creates a stronger foundation for subsequent development, with the gains magnifying through both self-productivity and enhanced productivity of any later investments $I_1(\theta_1)$, discussed more fully below.

The model thus formalizes the tension between two forces. Self-productivity with $\beta < 1$ creates automatic compression in relative terms, suggesting gaps may close without intervention. But absolute gaps can widen when advantaged children's higher skill bases generate larger period-by-period increments, and complementarity means these children also receive more subsequent investment, potentially locking in or expanding initial disadvantages. Well-designed interventions aim to counteract this dynamic by providing compensatory investments early in the lifecycle, when their effects compound most powerfully through the remaining periods of skill formation.

3 Early Childhood Meets Scaling

The model setup demonstrates why early interventions generate high returns: the multiplicative production function creates dynamic complementarities, and diminishing returns ($\beta < 1$) favor compensatory investment in disadvantaged children. However, translating small-scale efficacy into population-level impact requires understanding how benefit-cost profiles systematically change during implementation and scaling.

Recent research on the economics of scaling (Al-Ubaydli, List, and Suskind 2017, 2019, 2020; Al-Ubaydli et al. 2021; Al-Ubaydli, Lai, and List 2023; List 2022, 2024) identify five distinct mechanisms that generate "voltage drops" in benefit-cost profiles when translating pilot successes to population-level implementation: (1) false positives, where publication bias and specification searching inflate pilot effect sizes;

(2) sample unrepresentativeness, where pilot participants differ systematically from target populations in motivation or baseline characteristics; (3) situational unrepresentativeness, where controlled pilot conditions diverge from real-world implementation environments; (4) supply-side constraints, where scaling strains provider capacity and dilutes intervention quality; and (5) general equilibrium effects, where population-level implementation triggers behavioral responses absent in small pilots—such as crowding out private investments, generating spillovers to untreated populations, or shifting relative prices.[3]

This framework resonates with implementation science's focus on translating efficacious interventions into routine practice (Fixsen 2005; Chambers and Azrin 2013; Glasgow et al. 2019). Recognizing these parallels, the National Institutes of Health convened interdisciplinary teams to bridge economics, psychology, and public health approaches to scaling. Complementing this interdisciplinary synthesis, Supplee et al. (2022) argue that open science practices—including preregistration, transparency in reporting negative results, specification of core components and mechanisms, and rigorous testing under realistic implementation conditions—can address the underlying incentives in research that contribute to scaling failures.

The economic scaling framework thus serves as a theoretical bridge, providing formal economic foundations for implementation science's empirical observations while offering actionable guidance for preventing the voltage drops that plague scaled interventions across health, education, and social policy domains. This section formalizes List's (2022, 2024) "Option C" framework by explicitly modeling how benefit-cost profiles change when moving from controlled pilots to broad deployment.

3.1 Pilot Studies and the Scaling Illusion

Consider a targeted pilot, analogous to Perry Preschool, Abecedarian, or CHECC, conducted on n disadvantaged children with $\theta_{0,i} < \bar{\theta}$. Treated children receive investment $I_0 = I^*$ in period 0, while controls receive $I_0 = 0$. No subsequent investments occur ($I_1 = 0$), allowing observation of dynamic effects through self-productivity.

From the multiplicative production function (1), early investment generates treatment effects:

$$\Delta\theta_{2,i} = (A^{pilot})^{1+\beta}\theta_{0,i}^{\beta^2}\left[(1+I^*)^{\delta\beta} - 1\right] \tag{9}$$

Two features merit emphasis. First, complementarity generates heterogeneous treatment effects: absolute gains increase with baseline skills $\theta_{0,i}$. The proportional

3. Defined broadly, such effects include changes in relevant counterfactuals. For example, (Kline and Walters 2016) show that test score gains from Head Start are much larger if the counterfactual is no preschool rather than another similar quality preschool program. A more recent study in this spirit is Dougan, García, and Polovnikov (2025). For recent work exploring fadeout effects of early childhood see

gain in period-1 skills is constant across children—each experiences $(1 + I^*)^\delta - 1$ proportional increase in θ_1. However, because these gains compound through self-productivity in period 2, the *absolute* adult skill gains $\Delta\theta_2$ increase with θ_0. Second, dynamic amplification through self-productivity magnifies early advantages, with the exponent β^2 determining how much baseline heterogeneity affects adult outcomes.

The pilot's average treatment effect and benefit-cost ratio become:[4]

$$\text{ATE} = (A^{pilot})^{1+\beta}\left[(1+I^*)^{\delta\beta} - 1\right] \cdot \mathbb{E}[\theta_{0,i}^{\beta^2} | \theta_{0,i} < \bar{\theta}] \tag{10}$$

$$\left(\frac{B}{C}\right)^{pilot} = \frac{\phi \cdot (A^{pilot})^{1+\beta}\left[(1+I^*)^{\delta\beta} - 1\right] \cdot \mathbb{E}[\theta_{0,i}^{\beta^2} | \theta_{0,i} < \bar{\theta}]}{c_0 \cdot I^*}$$

Classic studies report impressive ratios: Perry Preschool achieved a B/C ratio roughly between 7 and 10, while Abecedarian generated internal rates of return exceeding 10%. However, these estimates systematically overstate scaled returns, as I turn to now.

3.2 Five Sources of Voltage Drops

As discussed above, pilot benefit-cost ratios often overestimate scaled returns through five distinct mechanisms (Al-Ubaydli et al., 2017; 2020; List, 2022; 2024). Table 1 translates each mechanism into the early childhood context, formalizing how quality degradation, sample composition, cost inflation, compliance, and general equilibrium effects operate through specific parameters in the skill production framework:

Quality voltage ($A^{\textbf{scale}} = (1 - v_A)A^{\textbf{pilot}}$). Exceptional pilot implementation typically features master's-level teachers, 4:1 child-to-staff ratios, and intensive oversight from researchers committed to the program's success. These conditions prove unsustainable at scale, where teachers have standard credentials, ratios increase to budget-constrained levels (often 8:1 or higher), and monitoring intensity declines as interventions transition from research projects to routine programming. The degradation in implementation quality directly reduces the productivity parameter A in the skill formation technology. With $\beta = 0.7$, a 30% quality voltage ($v_A = 0.3$) causes benefits to overstate scaled effects by factor $(1 - 0.3)^{-(1+0.7)} = (0.7)^{-1.7} \approx 2.1$—pilot studies suggest effects more than twice as large as what scaling will achieve.

Sample composition voltage ($\mathbb{E}[\theta_{0,i}^{\beta^2}|\textbf{pilot}] > \mathbb{E}[\theta_{0,i}^{\beta^2}|\textbf{scale}]$). Volunteer families who opt into pilot programs systematically differ from populations reached

4. The benefit-cost ratio (BCR), derived from cost-benefit analysis (CBA), is a tool that compares the monetary value of a policy's benefits to its costs to determine if it is worthwhile (BCR > 1 indicates net positive). While it is widely used in areas such as environmental regulation, infrastructure, and public health, it has several significant limitations that can lead to flawed decisions. Even so, I make use of it here for parsimony.

through automatic enrollment or mandatory participation. Even within disadvantaged populations, pilot participants exhibit higher baseline skills, more stable home environments, and greater organizational capacity—factors captured by higher realizations of θ_0. Because treatment effects scale with baseline skills through complementarity (equation 9), the same intervention generates smaller absolute gains when applied to populations with lower average θ_0. This composition effect operates multiplicatively with the β^2 exponent from dynamic self-productivity, amplifying initial skill differences across the lifecycle.

Cost inflation ($c_0^{\text{scale}} = (1 + \kappa)c_0^{\text{pilot}}$). Pilots benefit from below-market resource contributions that disappear at scale. University partners may donate facilities, graduate students provide labor at trainee stipends rather than market wages, and development costs for curriculum and assessments get amortized across research budgets. When programs scale, they face true economic costs: market rents for facilities, competitive wages for staff, and full development cost recovery. Additionally, pilots often benefit from volunteer contributions—parents provide transportation, staff work unpaid overtime, community organizations donate space—that cannot be sustained when interventions become routine services. The cost multiplier κ can range from 0.5 to 2.0 (List, 2022), meaning scaled programs cost 50-200% more per child than pilot implementations.

Compliance voltage ($I^{\text{eff}} < I^*$). Volunteer participants attend more regularly and engage more intensively than populations enrolled through automatic or mandatory mechanisms. Perry Preschool achieved approximately 85% attendance rates, while Head Start averages closer to 65%. Since skills evolve as $\theta_1 = A\theta_0^{\beta}(1 + I_0)^{\delta}$, reduced dosage I^{eff} directly lowers skill gains by factor $(I^{\text{eff}}/I^*)^{\delta\beta}$. With $\delta = 0.6$ and $\beta = 0.7$, a 25% reduction in effective dosage ($I^{\text{eff}}/I^* = 0.75$) reduces treatment effects by $(0.75)^{0.42} \approx 0.88$, a 12% voltage drop operating purely through reduced program contact hours.

General equilibrium effects. Population-level implementation triggers behavioral responses absent in small pilots. Supply-side constraints emerge as scaling strains provider capacity—the pool of master's-level early childhood teachers is finite, causing quality to decline as programs expand beyond this constraint. Public programs crowd out private investments as families reduce their own spending when government provides services. Peer composition changes as interventions alter neighborhood demographics and school populations. Skill price effects emerge when population-wide skill distributions shift, potentially reducing returns to skills through changes in relative wages. These general equilibrium channels prove difficult to formalize in a simple reduced-form voltage parameter, as they operate through multiple margins and often exhibit threshold effects where small programs generate no response but large programs trigger substantial adjustments.

The true scaled benefit-cost ratio incorporates all five voltage mechanisms:

$$\left(\frac{B}{C}\right)^{\text{scale}} = \frac{\phi \cdot (A^{\text{scale}})^{1+\beta}\left[(1+I^{\text{eff}})^{\delta\beta}-1\right] \cdot \mathbb{E}[\theta_{0,i}^{\beta^2}|\text{scale}]}{c_0^{\text{scale}} \cdot I^{\text{eff}}} \tag{11}$$

Defining total voltage drop as $v_{\text{total}} = 1 - (B/C)^{\text{scale}}/(B/C)^{\text{pilot}}$, empirical evidence documents voltage drops of 50-90% across domains (List, 2022). A pilot showing $B/C = 8$ may correspond to a true scaled ratio B/C between 2 and 4. Understanding which specific mechanisms drive voltage drops in particular contexts proves essential for Option C design strategies that provide evidence on how to mitigate these effects.

Table 1: Sources of Voltage Drops in Early Childhood Interventions

Mechanism	**Formal Parameter**	**Description**
Quality voltage	$A^{scale} = (1 - v_A)A^{pilot}$	Exceptional pilot implementation (master's teachers, 4:1 ratios, intensive oversight) exceeds sustainable scaled quality. Benefits overstate by $(1 - v_A)^{-(1+\beta)}$; with $\beta = 0.7, v_A = 0.3$, this equals $2.1\times$.
Sample composition	$\mathbb{E}[\theta_{0,i}^{\beta^2}\|pilot] > \mathbb{E}[\theta_{0,i}^{\beta^2}\|scale]$	Volunteer families within disadvantaged populations have higher baseline skills than automatic enrollment reaches due to selection on parental motivation and organization.
Cost inflation	$c_0^{scale} = (1 + \kappa)c_0^{pilot}$	Pilots benefit from donated facilities, below-market labor, and amortized development costs. True economic costs rise at scale.
Compliance voltage	$I^{\text{eff}} < I^*$	Volunteer participants attend more regularly (85% vs. 65%), reducing effective dosage at scale by factor $(I^{\text{eff}}/I^*)^{\delta\beta}$.
General equilibrium	Multiple channels	Supply constraints strain provider capacity; public programs crowd out private investments; peer composition changes; skill price effects emerge population-wide.

3.3 Three Policy Options

Upon receiving promising experimental evidence on a program, policymakers face three distinct strategic approaches, each with different implications for ultimate policy outcomes.

Plan A (No Intervention). Plan A is the most conservative response because it maintains the baseline counterfactual with $I_{t,i} = 0$ for all t, i. Social welfare equals

$$SW_A = \int \phi \theta_{2,i}^{baseline} f_0(\theta_{0,i}) d\theta_{0,i} = \int \phi \cdot A^{1+\beta} \theta_{0,i}^{\beta^2} f_0(\theta_{0,i}) d\theta_{0,i}, \tag{12}$$

foregoing potential gains but avoiding implementation costs and the risk of scaled program failures. This strategy dominates when voltage drops are sufficiently severe that scaled programs fail to cover their costs, or when political economy constraints make sustained implementation infeasible. The opportunity cost of Plan A—forgone benefits from effective interventions—must be weighed against the resource costs and potential harms from poorly implemented scaled programs.

Plan B (Naive Scaling). Plan B represents the strategy where policymakers extrapolate pilot results without anticipating systematic degradation in the benefit cost profile. Under this view, the standard policy response mechanically extrapolates pilot results to population-level implementation, assuming average treatment effects generalize: $\mathbb{E}[\Delta\theta_{2,i}|\text{scale}] = \mathbb{E}[\Delta\theta_{2,i}|\text{pilot}]$. This equality potentially fails due to a variety of reasons, including both quality ($A^{scale} < A^{pilot}$) and composition effects ($\mathbb{E}[\theta_{0,i}^{\beta^2}|\text{scale}] < \mathbb{E}[\theta_{0,i}^{\beta^2}|\text{pilot}]$). Expected welfare from naive scaling becomes:

$$SW_B = \int_{\theta_{0,i} < \bar{\theta}_B} \left[\phi(\theta_2^{baseline} + (1 - v_{total}) \Delta\theta_2^{pilot}) - c_0^{scale} I^{pilot} \right] f_0(\theta_{0,i}) d\theta_{0,i} \tag{13}$$

When total voltage v_{total} is large, programs fail to cover costs: $\phi(1 - v_{\text{total}})\Delta\theta_2 < c_0^{\text{scale}} I^{\text{pilot}}$, implying $SW_B < SW_A$. The intervention destroys value by consuming resources while generating insufficient benefits. This explains disappointing implementations like Head Start's modest and fading cognitive effects despite being inspired by Perry Preschool's extraordinary success. Plan B compounds errors by maintaining pilot investment intensity I^{pilot} under degraded implementation quality A^{scale}, wasting resources on over-investment given the actual productivity of scaled programs. The fundamental mistake lies in treating experimental treatment effects as policy parameters that remain invariant across implementation contexts. This is precisely the assumption that voltage effects systematically violate: effect sizes change when moving from pilots to scaled implementation.

Option C (Smart Scaling). The alternative approach uses economic theory and mechanism testing to anticipate voltage drops, then designs interventions explicitly to mitigate them. Rather than viewing scaling as simply implementing the pilot protocol on more subjects, Option C treats scaling as a distinct design problem requiring forward-looking analysis of how effects, costs, and implementation quality change with scope. In the case of early childhood interventions, the approach follows a three-stage implementation sequence:

Stage 1: Mechanism testing under realistic conditions. Before committing to large-scale rollout, conduct experiments testing interventions with average implementers (not exceptional pilot staff), representative participants (automatic enrollment rather than volunteers), realistic costs (market wages and facilities), and varied compliance mechanisms. These "stress tests" reveal true voltage parameters v_A, v_C, and cost multipliers κ, providing the empirical foundation for subsequent design decisions. Crucially, mechanism testing identifies *which* components of multi-faceted interventions drive effects, enabling designers to preserve essential elements while eliminating costly features that contribute little to outcomes. For instance, testing might reveal that weekly home visits generate 80% of treatment effects at 40% of cost compared to more intensive daily center-based programs.

Stage 2: Targeted scaling based on heterogeneity. Using estimates from Stage 1, calculate heterogeneous treatment effects: $\Delta\theta_{2,i} = (A^{scale})^{1+\beta}\theta_{0,i}^{\beta^2}[(1 + I^{eff})^{\delta\beta} - 1]$. Individual net benefits $NB_i = \phi\Delta\theta_{2,i} - c_0^{scale}I^{eff}$ vary systematically across children. Optimal targeting enrolls all children with $NB_i > 0$, potentially using observable characteristics correlated with $\theta_{0,i}$, such as parental education, family income, and neighborhood quality, to implement means-testing or risk-based eligibility. This targeting exploits the same complementarity structure that generates voltage through sample composition, but inverts it: rather than allowing selection to bias pilot estimates upward, policymakers deliberately concentrate resources on populations where the intervention generates positive net value. When β is sufficiently small (strong diminishing returns in self-productivity), optimal targeting reaches disadvantaged children despite complementarity favoring investment in higher-θ_0 individuals, because the utilitarian welfare function combined with budget constraints creates higher marginal welfare per dollar invested in initially disadvantaged populations.

Stage 3: Supply-side investments to reduce voltage. If Stage 1 testing reveals that voltage stems primarily from quality degradation (v_A is large), invest in teacher training programs, curriculum designs that maintain fidelity with less intensive oversight, technology platforms supporting implementation quality, and career ladders attracting quality staff to the early childhood sector. Formally, incorporate knowledge gain k reducing voltage: $A^{\text{OptionC}} = [1 - v_A + k(\text{training, design})]A^{\text{pilot}}$. The optimization problem weighs direct service costs against quality improvement investments: while spending resources on training and systems rather than serving additional children reduces scale q holding budget fixed, the quality improvements may increase net benefits per child sufficiently to offset the coverage reduction. This trade-off proves especially favorable when voltage is severe and quality improvements are relatively cost-effective.

Critically, this is a sequential process where learning propagates forward: Stage 1 testing informs Stage 2 targeting decisions and Stage 3 quality investments, which together determine optimal scale and resource allocation. The policymaker chooses optimal scale q^* and quality investments k^* by solving:

$$\max_{q,k} SW_C = \int_{\theta_{0,i} \in S(q)} \phi \left[\theta_2^{baseline} + \Delta\theta_2(A^{OptionC})\right] f_0(\theta_{0,i}) d\theta_{0,i} - C(q,k) \tag{14}$$

where $S(q)$ denotes the set of enrolled children and $C(q,k) = c_0(q,k) \cdot I \cdot q + F(k)$ includes both per-child costs that may vary with scale and quality, and fixed costs $F(k)$ for quality improvement infrastructure. Unlike Plan B, which mechanically replicates pilot protocols at fixed intensity I^{pilot}, Option C optimizes both scale q and quality investments k, generating a complex cost structure captured by the reduced form $C(q,k)$.

The fundamental insight of Option C thinking is that scaling represents a design problem distinct from demonstrating efficacy. Pilot studies answer "does this work under ideal conditions?" while Option C asks "how should we modify the intervention to maintain effectiveness at scale given realistic implementation constraints?" This reframing transforms voltage drops from threats to be ignored (Plan B) into predictable challenges to be addressed through mechanism-based design (Option C).

4 Optimal Policy: Characterization and Implications

This section characterizes the policymaker's optimal investment strategy under the dynamic skill formation technology, deriving conditions under which targeted interventions maximize social welfare. The analysis proceeds in two steps: first solving the social planner's optimization problem to characterize optimal investment paths, then examining how these optimal policies respond to scaling challenges including voltage drops and heterogeneity across children.

4.1 The Social Planner's Problem

Consider a benevolent social planner who chooses investment trajectories to maximize aggregate social welfare, defined as the sum of adult earnings net of investment costs across all children. The planner faces the skill formation technology specified in equation (1), which constrains how investments transform into skills through self-productivity and dynamic complementarity. The optimization problem balances the marginal productivity of investments against their opportunity costs, accounting for how early investments affect not only immediate skill development but also the returns to subsequent investments through complementarity.

Consider first the optimization problem for a representative child with initial skills θ_0. I suppress individual subscripts i in this section for notational simplicity, reintroducing them in Section 4.3 when analyzing heterogeneous populations. Formally, the planner solves:

Formally, the planner solves:

$$\max_{I_0,I_1} SW = \phi\theta_2 - c_0 I_0 - c_1 I_1 \tag{15}$$

subject to the technology of skill formation:

$$\theta_1 = A \cdot \theta_0^{\beta} \cdot (1 + I_0)^{\delta}$$
$$\theta_2 = A \cdot \theta_1^{\beta} \cdot (1 + I_1)^{\delta}$$

The objective function (15) captures three elements. First, $\phi\theta_2$ represents adult productivity valued at wage ϕ per skill unit, assuming competitive labor markets where wages equal marginal products. Second, $c_0 I_0$ and $c_1 I_1$ represent the opportunity costs of early childhood and later childhood investments respectively, where c_t denotes the resource cost per unit of investment in period t. The utilitarian specification weights all children equally, treating a one-unit increase in any child's adult skills as generating identical welfare gains ϕ regardless of that child's position in the skill distribution. This provides a normatively neutral baseline; modifications incorporating inequality aversion would strengthen the case for targeting disadvantaged children beyond what our results demonstrate.

The production function constraints embed the model's key economic mechanisms. Self-productivity parameter $\beta \in (0, 1)$ captures how existing skills enhance future skill development, with $\beta < 1$ implying diminishing returns—children with higher skills develop faster, but at a decreasing rate. Investment technology parameter $\delta \in (0, 1)$ governs how resource inputs translate into skill gains, with $\delta < 1$ implying diminishing marginal returns to investment intensity. Critically, the multiplicative form creates dynamic complementarity: $\partial^2\theta_t/\partial\theta_{t-1}\partial I_{t-1} > 0$, meaning investments prove more productive for children entering each period with higher skills.

I solve this problem through backward induction. The approach first characterizes optimal later-childhood investment I_1^* conditional on entering skills θ_1, then works backward to determine optimal early-childhood investment I_0^* accounting for how it affects both immediate skills θ_1 and the optimal response $I_1^*(\theta_1)$ in the subsequent period. This recursive structure reveals how early investments create value through two distinct channels: direct effects on intermediate skills and indirect effects on the productivity of future investments through complementarity.

Initially, I treat the problem as deterministic, suppressing the stochastic shocks ε_t that add realism but complicate exposition without changing fundamental insights. I also begin by analyzing a representative child with initial skills θ_0 before extending to heterogeneity across children indexed by i with different initial endowments $\theta_{0,i}$ drawn from distribution F_0. This simplification clarifies the core economic mechanisms before introducing distributional considerations essential for targeting policy.

4.2 Key Scaling Results

4.2.1 Optimal Investment Path

The first-order conditions characterize optimal investments in each period by equating marginal benefits to marginal costs. These conditions reveal how complementarity creates path dependence: optimal investments today depend on inherited skills, and create skills tomorrow that influence optimal future investments.

Later childhood investment. Consider first the simpler problem of choosing I_1 to maximize $\phi\theta_2 - c_1 I_1$ conditional on entering period 1 with skills θ_1. From the production function $\theta_2 = A\theta_1^\beta(1+I_1)^\delta$, the marginal product of period-1 investment is:

$$\frac{\partial\theta_2}{\partial I_1} = A\delta\theta_1^\beta(1+I_1)^{\delta-1}$$

Setting marginal benefit $\phi\cdot\partial\theta_2/\partial I_1$ equal to marginal cost c_1 yields the first-order condition:

$$\phi \cdot A\delta\theta_1^\beta(1+I_1^*)^{\delta-1} = c_1$$

Solving for optimal investment gives:

$$I_1^*(\theta_1) = \left[\frac{\phi A\delta\theta_1^\beta}{c_1}\right]^{\frac{1}{1-\delta}} - 1 \tag{16}$$

This policy function reveals several key properties. First, optimal investment increases with entering skills θ_1 due to complementarity—the partial derivative is:

$$\frac{\partial I_1^*}{\partial\theta_1} = \frac{\beta}{1-\delta}\left[\frac{\phi A\delta}{c_1}\right]^{\frac{1}{1-\delta}}\theta_1^{\frac{\beta}{1-\delta}-1} > 0 \tag{17}$$

This can be rewritten more intuitively as:

$$\frac{\partial I_1^*}{\partial\theta_1} = \frac{\beta}{1-\delta}\cdot\frac{I_1^*+1}{\theta_1},$$

The elasticity of optimal investment with respect to entering skills equals $\beta/(1-\delta)$, combining the complementarity parameter β (measuring how much baseline skills raise investment productivity) with the investment technology curvature δ (measuring diminishing returns to investment intensity). When complementarity is strong (β large) or returns to investment are highly diminishing (δ small), the investment gradient steepens sharply—high-skill children should receive substantially more resources than low-skill children.

Second, optimal investment increases with the skill price ϕ, the technology parameter A, and decreases with costs c_1. These comparative statics reflect standard economic logic: when returns to skills rise or investment becomes more productive or cheaper, optimal investment increases. The specific functional form, however, reveals non-linear responses: a 10% increase in skill prices raises optimal investment by $1/(1-\delta)$ times 10%, approximately 2.5 times as much when $\delta = 0.6$. This amplification reflects how higher returns justify pushing further into the region of diminishing marginal returns.

Third, unlike models with separable production functions where optimal investment depends only on cost-benefit ratios independent of existing skills, here complementarity creates state-dependence: the same child receives different optimal investments depending on skills carried forward from earlier periods. This path-dependence proves central to understanding why early interventions matter—they shape not just immediate outcomes but the entire subsequent trajectory of optimal investment and skill development.

Early childhood investment. The problem of choosing I_0 proves more complex because early investments affect welfare through two channels: directly boosting θ_1, and indirectly raising the productivity of subsequent investments through the complementarity $\partial I_1^*/\partial\theta_1 > 0$. The planner maximizes $\phi\theta_2(\theta_1(I_0), I_1^*(\theta_1(I_0))) - c_0 I_0 - c_1 I_1^*(\theta_1(I_0))$ by choosing I_0, where we make explicit that both θ_1 and optimal later investment I_1^* depend on the early investment choice.

Applying the chain rule to account for these indirect effects, the first-order condition becomes:

$$\phi A^2 \beta\delta \cdot \theta_0^{\beta}\theta_1^{\beta-1} \cdot (1+I_0^*)^{\delta-1}(1+I_1^*)^{\delta} = c_0 \tag{18}$$

This condition has a natural interpretation. The left side represents the marginal benefit of early investment, which factors as: ϕ (value per unit adult skill) times A^2 (productivity compounded across two periods) times $\beta\delta$ (elasticities of production with respect to skills and investment) times the skill terms $\theta_0^{\beta}\theta_1^{\beta-1}$ (reflecting how baseline skills and resulting intermediate skills combine through complementarity) times the investment terms $(1+I_0^*)^{\delta-1}(1+I_1^*)^{\delta}$ (capturing diminishing returns in both periods). The multiplier $A\beta\theta_1^{\beta-1}$ captures dynamic returns: early investments boost θ_1, which increases period-2 skills both directly through self-productivity ($A\theta_1^{\beta}$) and indirectly by raising the productivity of later investments ($\partial I_1^*/\partial\theta_1 > 0$ combined with complementarity).

The presence of I_1^* in the first-order condition creates an implicit relationship between I_0^* and θ_0 that I analyze through comparative statics. Note that I_1^* depends on $\theta_1 = A\theta_0^{\beta}(1+I_0)^{\delta}$, creating a feedback loop where early investments influence their own returns by shaping the productivity of subsequent investments. This dynamic linkage—absent in static models—proves central to understanding optimal targeting in the presence of complementarity.

4.2.2 Targeting Disadvantaged Children

A critical policy question asks whether to target disadvantaged (low θ_0) or advantaged (high θ_0) children. The answer proves non-trivial: equation (3) demonstrates that investments have higher marginal products for high-skill children (complementarity), yet optimal policy may still favor targeting the disadvantaged. This section resolves the apparent tension.

Applying the implicit function theorem to the first-order condition (equation 18) and using the relationship $\theta_1 = A\theta_0^{\beta}(1 + I_0)^{\delta}$, I find an optimality targeting condition: when diminishing returns in self-productivity are sufficiently strong (β sufficiently below 1) and the utilitarian welfare function places equal weight on all children, optimal early investment decreases with initial skills:

$$\frac{dI_0^*}{d\theta_0} < 0 \tag{19}$$

From the FOC (18), the relationship between optimal early investment and initial skills can be derived by taking the total derivative with respect to θ_0.[5] The sign of $dI_0^*/d\theta_0$ depends on whether $\delta(1+\beta) > 1$. When investment technology is sufficiently productive relative to self-productivity—as holds for empirically plausible values $\beta \approx 0.7$, $\delta \approx 0.6$—we obtain $dI_0^*/d\theta_0 < 0$, implying optimal policy targets disadvantaged children.

The condition $\delta(1+\beta) > 1$ balances three forces: (i) Complementarity ($\partial^2\theta_2/\partial\theta_0\partial I_0 > 0$) favors investing in high-θ_0 children; (ii) Diminishing returns in self-productivity ($\beta < 1$) mean initial skill differences compress through the nested structure $\theta_2 \propto \theta_0^{\beta^2}$, reducing the importance of initial heterogeneity; (iii) Investment productivity (δ) determines how efficiently resources can be deployed to disadvantaged children. When investment technology is sufficiently productive (δ high enough given β), the gains from targeting disadvantaged children overcome the complementarity effect that would otherwise favor the advantaged. This reconciles the apparent tension between dynamic complementarity and progressive targeting.

Beyond the formal mathematics, three complementary mechanisms explain why the optimal policy targets disadvantaged children when $\delta(1 + \beta) > 1$:

1. **Welfare per dollar spent**: While both high- and low-θ_0 children experience similar proportional gains in period-1 skills, the social planner evaluates investments based on absolute lifetime earnings gains. With linear earnings ($W = \phi\theta_2$), the welfare gain from moving θ_2 from 5 to 5.5 equals the gain from moving it from 10 to 10.5. When investment productivity δ is sufficiently high, achieving these equal absolute gains for disadvantaged children requires less investment than for already-advantaged children, making disadvantaged children more cost-effective targets.

5. The implicit function theorem yields $dI_0^*/d\theta_0 = \beta^2(1 + I_0)/[\theta_0(1 - \delta - \beta\delta)]$.

2. **Compounding from a lower base**: The multiplicative structure $\theta_2 = A\theta_1^\beta(1+I_1)^\delta$ means that investments lifting a disadvantaged child to moderate skills create strong foundations for period-2 development. A child moving from $\theta_0 = 0.3$ to $\theta_1 = 0.5$ gains more dynamically than one moving from $\theta_0 = 0.8$ to $\theta_1 = 1.0$, even though both represent similar proportional changes, because the former creates greater scope for productive later-life investments. When δ is sufficiently high, this compounding effect dominates complementarity.

3. **Budget constraints and marginal returns**: When resources are limited (formalized in Section 4.3), the social planner faces diminishing marginal welfare returns. Complementarity means that achieving large absolute skill gains for high-θ_0 children becomes increasingly expensive. When investment productivity δ satisfies $\delta(1+\beta) > 1$, resources are better allocated spreading moderate improvements across many disadvantaged children rather than concentrating on large gains for few advantaged children.

This result reconciles complementarity with progressive targeting: optimal human capital policy targets disadvantaged children not *despite* complementarity, but because when investment technology is sufficiently productive relative to self-productivity—formalized as $\delta(1+\beta) > 1$—the utilitarian welfare function and diminishing returns in skill production create stronger welfare gains per dollar invested in disadvantaged populations. This formalizes Heckman's insight about equity-efficiency alignment in early childhood investment.[6]

4.2.3 Voltage Drops Amplify Through Dynamics

The comparative statics directly connect to scaling challenges. If scaling reduces investment efficiency from A^{pilot} to $A^{scale} = (1-v_A)A^{pilot}$, optimal investment declines. To characterize this relationship, I consider a log-linear approximation around the optimal investment path. Taking logs of the first-order condition (18) and differentiating with respect to $\log A$ (accounting for the indirect effect through θ_1) yields:[7]

6. Note that this result depends on parametric assumptions. The condition $\delta(1+\beta) > 1$ is satisfied for a range of empirically plausible values: with $\beta \in [0.6, 0.8]$ and $\delta \in [0.5, 0.7]$, compensatory targeting emerges under utilitarian welfare. With inequality-averse social welfare functions that place extra weight on disadvantaged children, the targeting recommendation would strengthen and hold under weaker conditions. The key insight is that even under neutral utilitarian welfare, the structure of skill formation can justify progressive targeting when investment productivity is sufficiently high.

7. To see this, note that log-transforming both sides of the FOC (18) yields:

$$\log\phi + 2\log A + \log\beta + \log\delta + \beta\log\theta_0 + (\beta-1)\log\theta_1 + (\delta-1)\log(1+I_0^*) + \delta\log(1+I_1^*) = \log c_0.$$

Since $\theta_1 = A\theta_0^\beta(1+I_0)^\delta$, we have $\log\theta_1 = \log A + \beta\log\theta_0 + \delta\log(1+I_0)$. Differentiating with respect to $\log A$ and using the chain rule for θ_1 yields the approximation, where the denominator includes $\beta\delta$ from the indirect effect through θ_1 and I_1^*. The approximation assumes changes in I_0

$$\frac{d \log I_0^*}{d \log A} \approx \frac{2}{1 - \delta + \beta\delta} \tag{20}$$

For typical parameter values ($\beta \approx 0.7$, $\delta \approx 0.6$), this gives:

$$\frac{d \log I_0^*}{d \log A} \approx \frac{2}{1 - 0.6 + 0.7(0.6)} \approx \frac{2}{0.82} \approx 2.44$$

Therefore, a voltage drop v_A reduces optimal investment by approximately:

$$\frac{I_0^{\text{scale}}}{I_0^{\text{pilot}}} \approx (1 - v_A)^{2.44} \tag{21}$$

This implies voltage drops have *strongly amplified* effects on optimal investment. A 30% quality reduction ($v_A = 0.3$) reduces optimal investment by approximately $(0.7)^{2.44} \approx 0.42$, or a 58% decline—substantially more than the direct 30% efficiency loss.

Why does amplification occur? The exponent exceeds 1 for two reasons: (1) efficiency A directly enters the marginal product of investment, and (2) lower A reduces θ_1, which further decreases the marginal product through complementarity. These effects compound, making optimal investment highly sensitive to implementation quality. Plan B (naive scaling) compounds this error by maintaining pilot investment levels $I^{pilot} = I_0^*(A^{pilot}) > I_0^*(A^{scale})$ under degraded efficiency A^{scale}, wasting resources. Treatment effects realize as:

$$\Delta\theta_2^{scale} = \left(\frac{A^{scale}}{A^{pilot}}\right)^{1+\beta} \Delta\theta_2^{pilot} = (1 - v_A)^{1+\beta} \Delta\theta_2^{pilot} \tag{22}$$

The exponent $1+\beta$ arises because voltage affects both periods of skill formation. With $\beta \approx 0.7$, voltage drops are amplified by factor $(1 - v_A)^{1+\beta} \approx (1 - v_A)^{1.7}$ for treatment effects—compounding across periods but less than the 2.44 factor for optimal investments, which include endogenous behavioral adjustments through the complementarity mechanism. This distinction matters for policy: while scaled programs may see moderated effect sizes, the amplified investment reductions could exacerbate equity gaps if not addressed through Option C designs.

Option C prevents these failures by (1) testing under realistic conditions to estimate A^{scale} accurately, then (2) setting optimal investment $I_0^*(A^{scale})$ rather than mechanically using I^{pilot}, and (3) investing in quality improvements k to reduce

---

dominate elasticity calculations. The approximation in equation (20) holds exactly when $\beta = 1$ (linear self-productivity), where it simplifies because indirect effects through θ_1 fully offset when the multiplicative and power structures align, and provides a good approximation for $\beta \in [0.6, 0.9]$, typical of empirical estimates in the skill formation literature. The exact elasticity is $2/(1-\delta+\beta\delta)$, accounting for indirect effects through θ_1 and I_1^*.

voltage: $A^{adj} = (1 - v_A + k)A^{pilot}$. The amplification result underscores why accurate estimation of scaled efficiency is crucial—small errors in measuring v_A translate into large errors in optimal investment levels.

Taking stock of the insights thus far reveals the crucial interaction between targeting and voltage drops. The targeting result from Section 4.2.2 shows that disadvantaged children should receive more investment when β is sufficiently small. However, voltage drops disproportionately undermine this optimal targeting strategy. Since the rationale for compensatory investment depends on achieving sufficient treatment effects to overcome initial disadvantage, quality voltage (v_A) has particularly severe consequences for disadvantaged populations. When $A^{\text{scale}} < A^{\text{pilot}}$, the optimal threshold θ_0^* rises, implying fewer disadvantaged children clear the benefit-cost hurdle. This interaction explains why naive scaling (Plan B) often fails most dramatically for the populations that would benefit most from well-implemented programs—a troubling pattern observed across Head Start implementations and state pre-K expansions.

4.3 Heterogeneity and Budget-Constrained Targeting

The analysis thus far has characterized optimal investment for a representative child with initial skills θ_0. I now extend to the realistic setting where policymakers face a heterogeneous population of children with different initial endowments and must allocate a finite budget across this distribution. This extension transforms the abstract characterization of optimal investment into concrete guidance for program design: who should receive services, how much should each child receive, and when do budget constraints bind such that some children optimally receive zero investment despite potentially positive returns?

Consider a continuum of children indexed by $i \in [0, 1]$ with initial skill endowments $\theta_{0,i}$ drawn from distribution F_0 with density f_0, mean μ_0, and variance σ_0^2. This distribution captures all pre-intervention heterogeneity stemming from parental education and income, neighborhood quality, prenatal health, genetic endowments, and early home environment quality. Disadvantaged children—those from low-income families, unstable households, or under-resourced communities—concentrate in the left tail of F_0 with low realizations of $\theta_{0,i}$.

The policymaker faces a realistic budget constraint $B > 0$ limiting total expenditures across both investment periods and all children:

$$\max_{I_{0,i}, I_{1,i}} \int \phi\theta_{2,i} f_0(\theta_{0,i}) d\theta_{0,i} \quad \text{subject to} \quad \int (c_0 I_{0,i} + c_1 I_{1,i}) f_0(\theta_{0,i}) d\theta_{0,i} \leq B \tag{23}$$

This constrained optimization problem incorporates three key features distinguishing it from the unconstrained representative-agent analysis. First, the objective integrates welfare across the entire skill distribution, weighting each child by the density $f_0(\theta_{0,i})$ with which they appear in the population. Second, the budget constraint couples investment decisions across children—resources allocated to one child reduce availability for others, creating opportunity costs absent in the single-child problem. Third, heterogeneity in $\theta_{0,i}$ generates heterogeneous treatment effects through complementarity, making the targeting decision economically meaningful rather than simply distributional.

Introducing multiplier $\lambda \geq 0$ on the budget constraint, I formulate the Lagrangian as follows:

$$\mathcal{L} = \int \phi\theta_{2,i} f_0(\theta_{0,i}) d\theta_{0,i} - \lambda \left[\int (c_0 I_{0,i} + c_1 I_{1,i}) f_0(\theta_{0,i}) d\theta_{0,i} - B \right] \tag{24}$$

Taking first-order conditions with respect to $I_{0,i}$ and $I_{1,i}$ for each child i yields the optimality conditions:

$$\begin{aligned} \phi \frac{\partial \theta_{2,i}}{\partial I_{0,i}} &= \lambda c_0 \\ \phi \frac{\partial \theta_{2,i}}{\partial I_{1,i}} &= \lambda c_1 \end{aligned} \tag{25}$$

These conditions have a natural economic interpretation: the shadow value of the budget constraint λ represents the marginal welfare gain from relaxing the budget by one dollar. At the optimum, this marginal welfare gain must equal the marginal product of investment (in welfare units) divided by the per-unit cost for every child receiving positive investment. If any child had marginal product above λc_t, reallocating a dollar toward that child would increase welfare. If any child had marginal product below λc_t, reallocating away from that child would increase welfare.

When the budget constraint binds ($\lambda > 0$), optimal policy exhibits three distinct features that characterize realistic program targeting:

Property 1: Incomplete coverage. There exists a threshold skill level θ_0^* such that children receive positive early investment if and only if their initial skills fall below this threshold:

$$I_{0,i} > 0 \quad \Leftrightarrow \quad \theta_{0,i} < \theta_0^* \tag{26}$$

This threshold emerges because complementarity creates a skill gradient in marginal products: children with very high θ_0 have such high marginal products that achieving the same welfare gains requires prohibitive resource expenditures given diminishing returns $\delta < 1$. Below some critical θ_0^*, the marginal benefit per dollar spent exceeds λ, justifying positive investment. Above this threshold, marginal products remain positive but insufficient to justify resource allocation given opportunity costs.

The threshold θ_0^* is determined endogenously by two conditions: (i) the first-order condition must hold with equality for the marginal child at θ_0^*, meaning their marginal product equals λc_0; (ii) the budget constraint must bind with equality, meaning total expenditures across all served children exactly exhaust available resources B. These two equations jointly determine both the threshold θ_0^* and shadow price λ.

Property 2: Declining investment within treated population. Among children receiving intervention ($\theta_{0,i} < \theta_0^*$), optimal investment intensity decreases with initial skills:

$$\frac{\partial I_{0,i}}{\partial \theta_{0,i}} < 0 \quad \text{for all } i \text{ with } \theta_{0,i} < \theta_0^* \tag{27}$$

This within-program gradient reflects the same forces driving the targeting condition in Section 4.2.2: when β is sufficiently small, disadvantaged children offer higher welfare returns per dollar invested. The budget constraint sharpens this gradient by introducing explicit opportunity costs—every additional dollar allocated to a moderate-skill child means one less dollar available for the most disadvantaged children, and the planner chooses the allocation that maximizes aggregate welfare given these trade-offs.

The declining investment schedule $I_{0,i}(\theta_{0,i})$ takes a specific functional form determined by the first-order condition (25). For children below the threshold, investment solves:

$$\phi \cdot A^2 \beta\delta \cdot \theta_{0,i}^{\beta} \theta_{1,i}^{\beta-1} \cdot (1 + I_{0,i})^{\delta-1} (1 + I_{1,i}^*)^{\delta} = \lambda c_0 \tag{28}$$

As $\theta_{0,i}$ rises toward θ_0^*, the left side increases (marginal product rises with skills through complementarity), so $I_{0,i}$ must decline to maintain equality given diminishing returns $(1 + I_{0,i})^{\delta-1}$ decreasing in $I_{0,i}$. At the threshold itself, investment drops discontinuously to zero because the marginal benefit per dollar just equals opportunity cost λ at infinitesimal investment, but any finite investment would yield marginal product below λ given diminishing returns.

Property 3: Budget exhaustion determines threshold. The optimal threshold θ_0^* satisfies the budget constraint with equality:

$$\int_0^{\theta_0^*} [c_0 I_{0,i}^*(\theta_{0,i}) + c_1 I_{1,i}^*(\theta_{0,i})] f_0(\theta_{0,i}) d\theta_{0,i} = B \tag{29}$$

This condition embeds comparative statics revealing how the threshold responds to policy parameters. Increasing the budget B raises θ_0^*, expanding coverage to moderately disadvantaged children as resources permit serving more of the distribution. Improving technology parameter A or reducing costs c_t also raises θ_0^* because each dollar generates more welfare, stretching the budget to cover more children. Conversely, voltage drops that reduce A^{scale} below A^{pilot} lower θ_0^*, contracting optimal coverage as degraded implementation cannot justify serving as many children given fixed budget.

These three properties jointly characterize optimal means-tested programs. Programs like Head Start implement precisely this structure: eligibility criteria target families below income thresholds (incomplete coverage based on observables correlated with θ_0), and within eligible populations, more intensive services often reach the most disadvantaged families (declining investment schedule), with coverage expanding or contracting as federal appropriations vary (budget-determined threshold).

The heterogeneity analysis clarifies three critical insights connecting individual optimization to population-level program design and therefore hold key scaling implications:

Insight 1: Sample composition matters enormously. Treatment effects follow the multiplicative structure:

$$\Delta\theta_{2,i} = (A)^{1+\beta}\theta_{0,i}^{\beta^2}[(1+I)^{\delta\beta} - 1] \tag{30}$$

Average treatment effects therefore depend critically on the distribution of θ_0 in the treated sample. A pilot enrolling volunteers with mean baseline skills $\mathbb{E}[\theta_{0,i}|\text{pilot}] = \mu_{\text{pilot}}$ generates different average effects than scaled implementation reaching automatic enrollees with mean $\mathbb{E}[\theta_{0,i}|\text{scale}] = \mu_{\text{scale}}$. If $\mu_{\text{pilot}} > \mu_{\text{scale}}$ due to positive selection, the pilot ATE overstates scaled effects even if implementation quality and compliance remain identical. The bias operates through the $\theta_{0,i}^{\beta^2}$ term, which amplifies baseline differences through dynamic self-productivity.

More subtly, the entire distribution F_0 matters, not just the mean. With concave treatment effects $\Delta\theta_2(\theta_0)$ when $\beta^2 < 1$, Jensen's inequality implies $\mathbb{E}[\Delta\theta_2] < \Delta\theta_2(\mathbb{E}[\theta_0])$—average effects fall below what we would predict from mean skills. If pilot samples have lower variance $\sigma^2_{\text{pilot}} < \sigma^2_{\text{scale}}$ through screening processes that exclude both very high and very low performers, scaled implementation reaching a wider distribution will generate systematically different effects through the nonlinearity of $\Delta\theta_2(\theta_0)$.

Insight 2: Optimal targeting requires estimating heterogeneity. Option C designs should include subgroup analyses estimating the treatment effect function $\Delta\theta_2(\theta_0)$ rather than reporting only average effects. This requires either: (i) pre-specified analyses stratifying by baseline skills measured through assessments, with sufficient sample sizes in each stratum; (ii) interaction terms between treatment assignment and baseline characteristics in regression specifications; or (iii) machine learning approaches that flexibly estimate conditional average treatment effects as functions of covariates.

Without heterogeneity estimates, policymakers cannot implement the optimal targeting rule characterized in equations (26)-(29). They face a stark choice: universal coverage (serving all children regardless of θ_0) or crude categorical eligibility (serving all below some income threshold regardless of how θ_0 varies within low-income families). Both approaches sacrifice efficiency relative to the optimal policy that conditions investment intensity on predicted θ_0 using all available information. The welfare losses from ignoring heterogeneity can be substantial—targeting resources toward children with θ_0 just below the optimal threshold wastes money on populations generating minimal net benefits while underfunding children far below the threshold who offer the highest returns.

Insight 3: Universal programs face efficiency-equity tradeoffs. While targeted programs align efficiency and equity by concentrating resources on disadvantaged children who simultaneously face the largest skill deficits and offer the highest marginal returns per dollar invested (when β is small), universal programs covering all children face a fundamental tension. Complementarity means universal implementation generates larger absolute gains for advantaged children: $\Delta\theta_2$ increases with θ_0 even though welfare per dollar $\Delta\theta_2/I$ may decrease.

This creates a trilemma for universal programs. If investment intensity varies optimally across children ($I_i^*(\theta_0)$), advantaged children receive more resources through complementarity, creating political sustainability challenges ("why does the rich child get more than mine?"). If investment is uniform across all children ($I_i = \bar{I}$), the program wastes resources on advantaged children who need less while underfunding disadvantaged children who need more. If the program attempts to invert the gradient by giving disadvantaged children more resources, it fights against complementarity and sacrifices aggregate efficiency.

Targeted programs avoid this trilemma by focusing resources within the region $\theta_0 < \theta_0^*$ where optimal investment naturally decreases with baseline skills. The incomplete coverage removes precisely those children (high θ_0) for whom complementarity would otherwise demand high investment levels in an efficiency-maximizing allocation, eliminating the tension between optimal resource allocation and progressive distributional objectives.

Translating these theoretical targeting rules into operational program design raises several implementation challenges. First, θ_0 is not directly observable—it is a latent variable capturing all pre-intervention factors affecting development. Practical targeting must rely on observable proxies: family income, parental education, neighborhood characteristics, or assessment scores. The quality of targeting depends critically on how well these observables predict θ_0, with measurement error or weak correlations causing mistargeting that reduces program effectiveness.

Second, categorical eligibility based on income thresholds creates discontinuities that may not align with the smooth optimal allocation $I^*(\theta_0)$. A child just above the income cutoff might have lower θ_0 than one just below due to other disadvantages, yet they receive no services while the higher-skill child receives full program benefits. Continuous eligibility sliding scales based on multiple risk factors could better approximate optimal targeting, but administrative complexity and compliance costs often favor simpler categorical rules.

Third, means-testing generates stigma and take-up barriers that create additional voltage through reduced participation. If targeted programs induce shame or signal low status, eligible families may decline enrollment, systematically selecting away the most disadvantaged children who might benefit most. Universal programs avoid these participation barriers but sacrifice targeting efficiency. The optimal choice depends on the relative magnitudes: how much does targeting improve benefit-cost ratios versus how much do take-up effects and stigma reduce coverage of the most disadvantaged?

5 Putting the Chicago School Theory to the Test

The theoretical framework developed in this paper formalizes how naive extrapolation from pilot studies systematically overestimates policy efficacy by ignoring voltage drops—declining treatment effects due to quality degradation, sample unrepresentativeness, cost inflation, and general equilibrium spillovers when moving from controlled pilots to population-level implementation. This section provides summaries of ten recent studies, across diverse domains of human capital investment, which each provide crucial evidence on the model's scaling mechanisms.

The studies are ordered to build a coherent narrative: I begin with foundational work on the micro-dynamics of skill formation and measurement, progress through family investment decisions and intergenerational transmission, examine peer effects and social context, explore targeted interventions with heterogeneous effects, and conclude with explicit analyses of scaling challenges and implementation realities.

Collectively, these papers validate the Option C framework while revealing new dimensions of complexity—from endogenous technology parameters that evolve with investment histories to multi-generational complementarities that delay returns across decades to general equilibrium responses that fundamentally alter treatment effects at scale. The empirical richness documented here both confirms the theoretical predictions and suggests important extensions, particularly regarding heterogeneity in production function parameters, discrete stage-specific technologies, and feedback loops between investment activity and future productivity that create path dependence in skill formation.

5.1 Study Summaries

5.1.1 Heckman and Zhou: "A Study of the Microdynamics of Early Childhood Learning"

Heckman and Zhou (2026) exploit unique weekly measurement data from a scaled Chinese home-visiting program to investigate the micro-level mechanics of skill formation that aggregate models necessarily obscure. Their novel contribution challenges the standard skill production function $\theta_{t+1} = f(\theta_t, I_t, G_t)$ by demonstrating that skills nominally classified as identical across developmental stages in fact do not share a common unit scale—a mathematical complication with profound substantive implications.

The estimated microdynamics reveal that different skill levels are characterized by different production functions, rejecting the standard assumption of a single time-invariant technology $f(\cdot)$ operating across the entire lifecycle. Instead, learning occurs through "skill-lifecycle-stage-specific learning processes" that align more closely with reinforcement learning mechanisms from cognitive psychology than with the continuously differentiable CES functions typically employed in economics. This finding directly addresses a persistent puzzle: the ubiquitous "fadeout" phenomenon where measured treatment effects decline over time, which researchers typically attribute to depreciation (skills decay without continued investment) or catch-up growth (control group receives compensatory investments). The estimated model provides an alternative explanation: fadeout reflects measurement artifact arising from arbitrary test score scales that change meaning across ages combined with genuine forgetting processes operating at different rates for different skills.

Speaking to the theoretical framework advanced in this study, this microdynamic analysis reveals that the voltage parameter v_A (quality degradation in scaled implementation) varies not just across programs but across developmental stages and skill domains within the same program. The Chinese home-visiting intervention—designed to promote parenting through curriculum delivered by village health workers—shows heterogeneous effects depending on parental baseline capabilities and home environments, with the intervention effectiveness varying substantially across the skill distribution. This suggests that optimal targeting criterion θ_0^* (the skill threshold below which intervention generates positive net benefits) is not a scalar but a vector varying by skill domain and developmental stage, complicating the simple targeting rule derived in equation (19) (and the properties in section 4.3).

The methodological innovation of bypassing input endogeneity issues and lack of comparable measures of skills that plague previous studies through careful experimental design and item response theory measurement provides a roadmap for Option C thinking: rather than assuming treatment effects generalize across populations and contexts, researchers should measure the micro-mechanics of how interventions affect learning processes, then use structural models to aggregate these mechanisms into predictions about scaled implementation. In addition, the finding that family environments mediate treatment effects, with "different effects for children with different types of parents and home environments," formalizes the interaction between program quality (A) and household characteristics (θ_0, I_{family}) that generates composition voltage: average treatment effects in pilots with selected families systematically overstate effects when scaling to broader populations where complementarity between program inputs and family inputs proves weaker.

Most profoundly, the paper demonstrates that the standard specification treating skills as cardinally measured stock variables evolving smoothly through time represents an oversimplification that masks critical discontinuities, state dependencies, and forgetting processes that become visible only with high-frequency measurement. This suggests that our theoretical understanding of skill formation, including the dynamic complementarity and self-productivity mechanisms central to my theoretical framework, requires refinement to accommodate the actual discrete, stage-specific, and stochastic learning processes operating at micro timescales.

5.1.2 Cotton, Hickman, List, Price, and Roy: "Why Don't Struggling Students Do Their Homework? Disentangling Motivation and Study Productivity as Drivers of Human Capital Formation"

Cotton and colleagues (2026) provide identification of a two-dimensional model separating unobserved student motivation (willingness to study) from unobserved productivity (conversion rate of study time into skill), using field-experimental data combining real-time web-based activity tracking with randomized incentive variation across 1,676 middle school students in diverse Chicago-area districts. Their structural estimates reveal a counterintuitive finding with interesting implications: many struggling students actually exhibit higher motivation than their higher-performing peers, but face severe productivity disadvantages: they cannot efficiently convert study time into completed work and skill gains.

This finding directly challenges the standard interpretation of low investment in my theoretical framework, which typically attributes underinvestment to either budget constraints or preference parameters, but the productivity decomposition reveals a third channel: students with low A (productivity parameter) rationally choose low I (investment) because their skill production function $\theta_{t+1} = A \cdot \theta_t^\beta \cdot (1 + I)^\delta$ offers poor returns to effort. The distributional analysis proves particularly striking: Black and Hispanic students demonstrate higher willingness to study but face "substantial productivity disadvantages, largely attributable to school quality differences," suggesting the achievement gap operates primarily through $A_{\text{minority}} < A_{\text{majority}}$ rather than through differential investment choices conditional on productivity.

This mechanism has deep implications for Option C thinking about intervention design: programs providing incentives or information about returns to schooling (attempting to increase motivation) will fail because they target the wrong margin—struggling students already want to succeed but lack the foundational skills, instructional support, or learning environments to convert effort into achievement. Effective interventions must address the productivity constraint directly through remediation, scaffolding, or school quality improvements that raise the A parameter. The estimated feedback loop between investment activity and productivity—where sustained engagement in learning improves future learning efficiency—creates dynamic complementarity operating through the technology itself rather than simple skill stocks: $A_{t+1} = h(A_t, I_t, \text{age})$, with $\partial A_{t+1}/\partial I_t > 0$. This generates path dependence where early productivity disadvantages compound: low-A students rationally disengage, which prevents the productivity-enhancing feedback from operating, creating "poverty traps" in skill formation.

The finding that dynamic skill complementarities "arise mainly from children's aging and from a feedback loop between investment activity and productivity, rather than from carrying forward past skill stocks" challenges the standard CES specification where θ_t enters directly but A remains fixed, suggesting production functions should model technology parameters as evolving endogenously through accumulated learning experiences. For scaling theory, the productivity-motivation decomposition reveals why promising interventions often fail: pilots typically attract selected samples with higher baseline productivity (higher A), meaning treatment effects $\mathbb{E}[\theta_{t+1} - \theta_t|\text{pilot}]$ overstate effects when scaled to populations with lower productivity who cannot effectively utilize the intervention—a composition voltage operating through heterogeneity in production function parameters rather than initial skill endowments.

5.1.3 Caucutt, Lochner, Mullins, and Park: "Child Skill Production: Accounting for Parental and Market-Based Time and Goods Investments"

Caucutt and colleagues (2026) tackle a fundamental identification challenge in skill formation: distinguishing complementarity in the production function from selection in investment decisions. Their innovation lies in developing a relative demand estimation strategy that exploits intratemporal optimality conditions to identify substitutability parameters between parental time, home goods, and market-based childcare while separately estimating intertemporal dynamics governing how skills evolve.

Using PSID-CDS data on children ages 5-12, they find moderately strong complementarity between all inputs (refuting perfect substitution) but little difference in production technology by parental education—a result with profound implications for my framework's targeting conclusions in section 4.2.2. If high-educated parents invest more primarily through income effects and preference parameters rather than having fundamentally more productive investment technologies (higher A), then the rationale for compensatory targeting of disadvantaged children strengthens: optimal investment $I_0^*(\theta_0)$ decreases with initial skills not because of production function curvature alone, but because disadvantaged children face similar production possibilities yet receive fewer investments due to budget constraints.

The paper's methodological contribution to Option C thinking emerges from decomposing total voltage into component mechanisms: scaling a program that provides free childcare will generate different effects than one subsidizing home goods or incentivizing parental time, because these inputs enter the production function $\theta_{t+1} = A \cdot f(\text{time}, \text{goods}, \text{childcare}; \theta_t)$ with different substitution elasticities. The estimated complementarity (rejecting perfect substitution) implies that voucher programs allowing parents to substitute freely between time and market childcare will achieve lower skill gains than integrated programs providing both inputs in fixed proportions—the voltage drop operates through substitution patterns not captured in pilot studies where researchers control input bundles.

The counterfactual simulations prove that "estimated input complementarity has important implications for policies that subsidize inputs or provide free child care," with effect sizes varying substantially depending on which margin families adjust. This formalizes a key Option C principle: understanding mechanism heterogeneity across populations and contexts requires estimating structural parameters (substitution elasticities, productivity differences) rather than simply scaling average treatment effects. The finding of little technology difference by parental education further challenges models assuming skill-skill complementarity necessarily implies skilled parents are more productive investors—instead, the socioeconomic gradient in outcomes arises primarily from differential investment levels $I_{\text{rich}} > I_{\text{poor}}$ under similar production possibilities, suggesting policies removing budget constraints for disadvantaged families could substantially narrow skill gaps.

5.1.4 Del Boca, Flinn, Verriest, and Wiswall: "Parenting with Patience: Parental Incentives and Child Development"

Del Boca and colleagues (2026) construct a clever Markov Perfect Equilibrium framework where both forward-looking parents and adolescent children endogenously choose investments in the child's cognitive development, with the critical innovation being that parental incentive provision $I_{\text{incentives}}$ today affects the child's discount factor δ_{child} tomorrow. This creates an intertemporal trade-off formalized as a time-inconsistency problem: extrinsic motivation boosts short-run cognitive skill accumulation θ_1 but reduces the child's future intrinsic motivation to self-invest by lowering their discount factor, thereby reducing θ_2 through decreased future self-investment.

Their model provides micro-foundations for dynamic complementarity that operate through preference formation rather than production function curvature. In the framework's notation, excessive use of extrinsic incentives reduces the child's future willingness to self-invest (lowering their own $I_{\text{child,future}}$) by damaging intrinsic motivation, creating an endogenous voltage where interventions succeed at raising θ_1 while undermining the self-productivity mechanism that would compound those gains into θ_2. While this operates through preference formation rather than the production technology parameter A directly, the effect manifests as reduced future investment that mimics a voltage drop in the technology itself.

This mechanism importantly explains a puzzling empirical regularity: interventions that successfully boost test scores in the short run often show fadeout effects, which researchers typically attribute to depreciation of skills $\delta_{\text{depreciation}} > 0$. But this paper suggests an alternative mechanism—that the intervention itself damaged the production technology by reducing children's intrinsic motivation, creating an endogenous voltage drop where $A^{\text{future}} = A^{\text{baseline}} - k(\text{incentives})$. The policy implication challenges the foundation of many accountability systems: testing-focused interventions may succeed on their explicit objective (raising θ_1) while undermining the deeper goal of fostering self-directed learning that drives θ_2 through skill self-productivity.

5.1.5 Dettmer, Heckman, Pantano, Ronda, and Suomi: "Effects of Multi-Generational Exposure to Early-Life Advantage: Lessons from a Primate Study"

Dettmer and colleagues (2026) leverage three decades of randomized rearing assignments in rhesus monkeys to identify multigenerational complementarities impossible to measure in human populations. The experimental design, which randomly assigns both mothers (generation t) and their offspring (generation $t+1$) to either mother-rearing (high-quality early environment) or nursery-rearing (adverse conditions), permits clean identification of how treatment effects depend on the previous generation's experience.

The central finding dramatically illustrates intergenerational skill complementarity: benefits of mother-rearing materialize only for offspring whose mothers were themselves mother-reared, with precisely zero treatment effects for offspring of nursery-reared mothers regardless of the offspring's own treatment assignment. This represents a pure interaction term in the production function: $\frac{\partial^2 \theta_{\text{child}}}{\partial I_{\text{mother}} \partial I_{\text{grandmother}}} > 0$, where investments in generation t only yield returns conditional on investments in generation $t-1$.

The theoretical implications for my model in this study prove profound. Standard skill production models specify $\theta_{t+1} = A \cdot \theta_t^\beta \cdot (1+I_t)^\delta$, assuming the technology parameter A is fixed or varies only with observable characteristics. But this primate evidence suggests A itself depends on ancestral investments: $A_{\text{child}} = A_{\text{baseline}} \cdot g(I_{\text{grandmother}})$, where $I_{\text{grandmother}}$ denotes the investment the grandmother made in the mother during her childhood, and $g(\cdot)$ is strongly increasing—potentially even step-functional ($A \approx 0$ if mother nursery-reared).

This creates a fundamental scaling challenge: interventions providing high-quality early environments to generation t will show disappointing effects when evaluated on contemporaneous outcomes, because the benefits require complementary investments in generation $t+1$ to materialize. The voltage drop takes an unusual form—not declining treatment effects with scale, but delayed treatment effects across generations, systematically biasing short-run benefit-cost calculations. The identified transmission mechanism—parenting quality as the primary pathway, with no effects detected through in-utero environment despite shared gestational exposure—reveals that interventions targeting parent-child interactions face voltage drops when moving from controlled pilots (where interaction quality is maintained through intensive oversight) to scaled implementation (where quality degrades).

The critical implication for Option C thinking: programs addressing intergenerational transmission require sustained multi-generation commitment, because the full returns to investing in disadvantaged mother at age t materialize only through enhanced productivity of investments in her children at $t + 20$. Naive extrapolation from single-generation pilots will systematically underestimate long-run benefits while overestimating short-run returns, creating political economy challenges where myopic policymakers abandon effective interventions before complementarities manifest across generations.

5.1.6 Agostinelli, Doepke, Sorrenti, and Zilibotti: "It Takes a Village: The Economics of Parenting with Neighborhood and Peer Effects"

Agostinelli and coauthors (2026) fascinating study provides perhaps the most direct empirical demonstration of the voltage effect operating through general equilibrium mechanisms. Their structural model, estimated on Add Health data, reveals that parents endogenously adopt authoritarian parenting styles, restricting children's peer selection, as a compensatory response to heterogeneous or low-quality peer environments. This behavioral response creates a fundamental tension in scaling: small-scale "Moving to Opportunity" style interventions generate substantial benefits by relocating individual disadvantaged children to affluent schools where the peer environment remains unchanged, but scaling these interventions triggers defensive parental responses that erode treatment effects by half.

The formal connection to my theoretical framework in this study operates through the unmodeled general equilibrium channel: the benefit-cost ratio $(B/C)^{\text{pilot}}$ fails to account for systematic changes in parents' equilibrium investment strategies $I_1(\theta_1, \text{peer composition})$. When one child moves, receiving community parents maintain permissive styles because their children's peer group remains high-quality. But when scale increases from one to forty children, the discontinuous jump in peer heterogeneity induces affluent parents to switch to authoritarian restrictions, directly counteracting the intervention's intended mechanism.

The theoretical contribution in Agostinelli et al. (2026) extends my simple scaling model by formalizing how interventions that change population-level distributions F_0 trigger behavioral responses that were not present in pilot studies. Most keenly, the paper demonstrates that successful scaling requires overcoming both homophily bias in children's preferences and strategic parental interference, barriers that strengthen rather than weaken with program scale. Unlike standard models where marginal costs remain constant or increase linearly with scale, these behavioral responses create convex cost functions $C(q, k)$ with $\frac{\partial^2 C}{\partial q^2} > 0$, making some interventions fundamentally unscalable without addressing the general equilibrium mechanisms.

5.1.7 Falk, Kosse, and Pinger: "Mentoring and Schooling Decisions: Causal Evidence"

Falk and colleagues (2026) provide rare causal evidence that a low-intensity mentoring intervention can address inequality of opportunity arising from Germany's early tracking system, where children from low-SES families are 31.1 percentage points less likely to enter academic tracks than high-SES children—a gap that persists at 21.7 percentage points even after conditioning on academic performance. The one-year mentoring program (cost: €1,000 per child) increased high-track attendance among treated low-SES children by 11 percentage points at the time of tracking (Grade 5), an effect that persisted with minimal fadeout at 10.3 percentage points five to six years later (Grade 9/10), closing roughly one-third of the unconditional gap.

This effect size proves particularly remarkable when interpreted through the targeting framework developed above: despite complementarity in the educational production function that might favor investing in already-advantaged children, the intervention succeeded by operating at a critical decision juncture, Germany's fourth-grade tracking system, where relatively modest investments can shift trajectories with lasting consequences. The minimal fadeout from 11 to 10.3 percentage points over five years suggests the intervention created dynamic advantages through changed educational placement rather than direct skill production, illustrating how targeted interventions at key junctures can generate persistent returns through trajectory changes.

The theoretical mechanism underlying this success deserves emphasis: the intervention operates primarily through information and beliefs, shifting both teacher perceptions (who observe mentored children and update recommendations) and parental expectations (who become familiar with academic-track education through mentor role models), rather than directly entering the skill production function $\theta = A \cdot \theta^{\beta} \cdot (1 + I)^{\delta}$. This suggests an important extension to the scaling framework: some of the most cost-effective interventions target decision-making under imperfect information at critical junctures, operating outside the production function through belief formation and information provision. Such interventions may exhibit different voltage characteristics than programs attempting direct skill production, with lower implementation costs but greater sensitivity to institutional context and information environments.

The mechanistic evidence reveals dual pathways operating through both supply and demand sides of the tracking decision. On the supply side, treated children received improved teacher recommendations, suggesting the mentoring enhanced either actual capabilities or teachers' perceptions of children's potential for high-track success. On the demand side, parents of treated children more frequently overruled low-track recommendations, indicating shifted beliefs about educational returns. The intervention targeted low-SES families one to two years before tracking, with mentors all having completed high track and enrolled in university. The mentors served as role models who introduced academic-track concepts into family contexts where such paths were often unfamiliar.

Yet, the paper's most profound contribution to scaling theory emerges from its cost-effectiveness: at €1,000 per child generating 11 percentage point effects on high-stakes educational outcomes, the benefit-cost ratio substantially exceeds many intensive early childhood programs. This suggests that targeting specific decision nodes, the tracking juncture, rather than attempting comprehensive skill production may represent an Option C approach that maintains efficacy while achieving scalability through modest cost structure $c_0(q)$ that remains approximately linear even as q increases substantially.

5.1.8 Cappelen, Charness, Ekström, Gneezy, and Tungodden: "Exercise Improves Academic Performance"

Cappelen and colleagues (2026) demonstrate a profound empirical validation of skill complementarity operating through an unexpected channel: physical health capital enhancing educational persistence and completion. Their randomized intervention, which provided free gym access to university students, generated 0.15 standard deviation improvements in completed study points through reduced course dropout and exam failure, without affecting grade performance on completed exams. Their study suggests the mechanism operates primarily through improved self-regulation and sustained effort rather than enhanced cognitive ability per se. The treatment effect heterogeneity proves most theoretically illuminating: the entire effect concentrates among students with poor baseline lifestyle habits and low self-control, yielding gains of approximately 0.5 standard deviations for students who scored below the median on lifestyle, self-control, happiness, and study hours at baseline (Cappelen et al., Table 7, column 1), while producing negligible effects for other students.

This pattern directly instantiates the targeting theorem from the theoretical framework above: when β (self-productivity) is sufficiently small, disadvantaged individuals become the optimal intervention targets. The mechanism of improved self-control functioning as an increase in the child's effective discount factor δ in skill production also speaks to dynamic complementarity. Students with initially compromised self-regulation cannot efficiently convert study time into learning (low productivity parameter A), but physical exercise removes this constraint, effectively increasing A for the treated group. The specificity of effects on course completion rather than grades reinforces that the intervention addresses a behavioral constraint (persistence, time-consistent decision-making) rather than enhancing raw cognitive capacity.

The study's deepest contribution to scaling theory lies in what it reveals about the interaction between targeting mechanisms and composition voltage. The pilot's volunteer participants likely exhibited higher baseline θ_0 than would characterize universal implementation, meaning the observed 0.15 SD average effect embeds positive selection bias through standard composition voltage. However, the strong heterogeneity by baseline self-control suggests a more nuanced scaling prediction: if the intervention specifically removes a binding constraint (self-control) that is more severe in non-volunteer populations, automatic enrollment might reach students for whom the constraint binds more tightly, potentially generating larger effects despite lower baseline θ_0. This would represent composition voltage working in reverse—with $\mathbb{E}[\theta_{0,i}^{\beta^2}|\text{scale}] < \mathbb{E}[\theta_{0,i}^{\beta^2}|\text{pilot}]$ predicting larger rather than smaller scaled effects. Testing this hypothesis requires actual scaling studies that measure both baseline self-control and treatment effects across volunteer versus automatic enrollment populations, demonstrating how Option C thinking identifies when standard voltage predictions may be inverted by mechanism-specific heterogeneity.

5.1.9 Mullins: "A Structural Meta-Analysis of Welfare Reform Experiments and Their Impacts on Children"

Mullins (2026) pioneers a structural approach to meta-analysis that directly implements Frisch's (1936) vision: using economic theory to coordinate accumulation of empirical evidence rather than simply averaging treatment effects. By estimating a unified model across three welfare reform experiments (Connecticut Jobs First, Florida's Family Transition Program, Minnesota Family Investment Program) spanning six sites, he identifies how variation in experimental design (changes in benefit formulae, mandatory services, time limits, childcare subsidies) maps to deep parameters governing maternal labor supply, childcare use, and skill formation technology.

This methodology provides precisely what the Option C framework requires: structural parameters invariant across policy regimes rather than reduced-form treatment effects that vary with implementation details. The model reveals that seemingly similar "welfare-to-work" interventions operate through distinct mechanisms—time limits reduce participation while mandatory services increase employment primarily by raising job arrival rates rather than imposing non-pecuniary costs, decomposing treatment effects into interpretable economic channels. The skill formation estimates prove troubling: increases in household income generate modest positive effects (1% SD per log-point increase), but the transition from unpaid to paid childcare associates with 6.9% SD losses in behavioral skills, with substantial heterogeneity across latent types suggesting complementarity between maternal quality and childcare quality.

The theoretical contribution to the framework herein operates through revealing how voltage drops vary systematically with latent individual characteristics: experimental populations disproportionately select individuals with lower labor market productivity and temporary negative shocks, implying $\mathbb{E}[\theta_0|\text{experiment}]$ differs from $\mathbb{E}[\theta_0|\text{population}]$ not just in mean but in the entire distribution of unobserved heterogeneity. The counterfactual applying estimated treatment effects to the representative SIPP sample demonstrates dramatic composition voltage—average treatment effects look "quite different" in the non-selected population, with concerning negative consequences for children's behavioral skills that were masked in the experimental samples.

This keen insight illustrates a subtle scaling failure: heterogeneous treatment effects combined with selection on latent productivity means pilot benefit-cost ratios systematically overstate scaled returns not because the intervention changes (quality voltage) but because the population changes (composition voltage) in ways that interact with treatment effect heterogeneity. The structural approach enables calculating optimal policy under realistic assumptions about population composition, demonstrating how the Option C thinking of anticipating heterogeneity and selection can improve policy design relative to naive extrapolation from experimental means.

5.1.10 Cuddy and Currie: "Rules vs. Discretion: Treatment of Mental Illness in U.S. Adolescents"

Cuddy and Currie (2026) cleverly investigate a fundamental tension in the science of scaling: how evidence from carefully controlled clinical trials translates into population-level treatment guidelines, and how practitioner adherence to those guidelines affects real-world outcomes. Using national insurance claims covering 45,000 adolescents, they create a natural experiment in scaling approaches by comparing three regulatory regimes: FDA-approved treatments (Plan A: strict evidence-based rules), professional association guidelines allowing off-label use (Option C: informed flexibility), and practitioner discretion inconsistent with any guidelines (Plan B: unguided scaling).

The results demonstrate that implementation quality, captured as adherence to evidence-based protocols, exhibits voltage drops with severe consequences: "red-flag" prescribing increases self-harm rates, emergency room visits, and healthcare costs over 24 months relative to guideline-consistent treatment. Linking to the proposed framework above, this careful empirical work maps directly to quality voltage where $A^{\text{red-flag}} < A^{\text{guidelines}} < A^{\text{FDA}}$, but with a critical insight: the optimal scaling strategy is not strict FDA approval (which would restrict treatment menus too severely) but professional guidelines that allow controlled experimentation while maintaining quality standards.

The empirical magnitudes underscore the practical importance of these voltage distinctions. Adolescents receiving red-flag prescriptions experience significantly elevated rates of adverse outcomes compared to those receiving guideline-consistent care, with effects persisting throughout the 24-month follow-up period. The mechanism through which professional guidelines maintain quality while preserving flexibility proves instructive: rather than restricting practitioners to a narrow formulary, guidelines synthesize evidence from the broader clinical literature (including off-label uses with empirical support) while explicitly flagging combinations that lack any evidence base. This approach addresses the fundamental information problem that FDA approval processes create systematic gaps where effective treatments remain unapproved for pediatric populations. The finding that guideline-consistent care achieves the lowest facility use and total costs while maintaining safety demonstrates how the Option C approach successfully navigates the tension between standardization and adaptation.

In this manner, the Cuddy and Currie (2026) study empirically validates the Option C framework by showing that guidelines represent a middle path—more flexible than rigid FDA protocols but more structured than complete practitioner discretion—that successfully addresses scaling challenges while preserving treatment efficacy. The theoretical contribution reveals that scaling mechanisms differ fundamentally between consumer goods (where quality might degrade gradually with supply constraints) and expert services (where practitioner judgment creates discrete quality tiers).

Further, the comparative statics prove that voltage parameter v_A varies discontinuously with regulatory regime: treatments consistent with some evidence base maintain quality, while deviations into pure discretion trigger sharp voltage drops. This suggests that optimal scaling policy for expert services requires mechanism-based design principles embedded in flexible but binding guidelines rather than either pure centralization (FDA only) or pure decentralization (practitioner discretion)—precisely the Option C approach of anticipating implementation realities while providing structured frameworks that preserve treatment integrity.

6 Synthesis, Implications, and Paths Forward

These ten studies collectively validate the core predictions of the scaling framework while revealing layers of complexity that demand theoretical extensions. Three fundamental insights emerge with particular force.

First, voltage drops operate through more diverse channels than the initial framework suggests. Beyond the standard mechanisms, such as quality degradation (v_A), composition effects ($\mathbb{E}[\theta_{0,i}^{\beta^2}|\text{scale}] \neq \mathbb{E}[\theta_{0,i}^{\beta^2}|\text{pilot}]$), and cost inflation, the empirical evidence reveals endogenous technology changes where the production function parameter A itself evolves with investment histories (Cotton et al. 2026), intergenerational complementarities where returns materialize only across multiple generations (Dettmer et al. 2026), and general equilibrium responses where populations strategically counteract interventions (Agostinelli et al. 2026).

The Del Boca et al. (2026) finding that extrinsic incentives reduce future intrinsic motivation represents perhaps the most troubling voltage mechanism: interventions can succeed on their explicit objectives while inadvertently degrading the very technology they aim to enhance. This result suggests equation (1) should be augmented to include technology parameters that depend on past intervention strategies: $A_{t+1} = A(I_0, I_1, ..., I_t)$, capturing how implementation choices create path dependence in production possibilities.

Second, heterogeneity in unobserved productivity parameters fundamentally alters optimal targeting and scaling predictions. The Cotton et al. (2026) decomposition of motivation versus productivity reveals that achievement gaps reflect primarily differences in the technology parameter A rather than in skill endowments θ_0 or investment levels I conditional on productivity. This implies that composition voltage operates not just through the distribution of initial skills but through the distribution of production function parameters themselves—a dimension of heterogeneity that standard models miss entirely. When pilots select on unobserved productivity (higher A individuals volunteer), treatment effects systematically overstate population effects because low-A individuals cannot effectively utilize the intervention.

The Mullins (2026) structural meta-analysis confirms this mechanism empirically: welfare reform experiments selected participants with temporarily depressed labor market opportunities, and counterfactual simulations on representative populations show dramatically different effects. This suggests equation (11) for scaled benefit-cost ratios requires an additional term capturing heterogeneity in A: $(B/C)^{\text{scale}} = \phi \cdot (A^{\text{scale}})^{1+\beta} \cdot [(1 + I^{\text{eff}})^{\delta\beta} - 1] \cdot \mathbb{E}[\theta_{0,i}^{\beta^2} | \text{scale}, A]/(c_0^{\text{scale}} \cdot I^{\text{eff}})$, where the expectation now conditions on productivity as well as skill levels.

Third, the studies reveal profound tensions between short-run and long-run evaluation horizons that challenge conventional benefit-cost analysis. The Dettmer et al. (2026) primate study demonstrates that returns to early interventions may require a full generation to materialize through improved parenting quality in the next generation, while Heckman and Zhou (2026) show that fadeout in measured effects reflects arbitrary test score scales and stage-specific production functions rather than genuine depreciation. These findings suggest that typical evaluation windows (2-5 years) systematically underestimate true returns while potentially overweighting short-run effects that prove ephemeral.

The tension becomes acute in the Del Boca et al. (2026) finding that interventions boosting short-run skills may damage long-run prospects by reducing intrinsic motivation—a dynamic that requires decades to fully manifest but would be invisible in standard evaluation horizons. This time-scale problem creates a fundamental challenge for Option C thinking: how can researchers conduct mechanism tests that reveal long-run effects without waiting decades for results? The Cappelen et al. (2026) study offers a potential solution by focusing on mechanisms (self-control, lifestyle habits) that have well-established connections to long-run outcomes, allowing immediate measurement of mediators rather than waiting for ultimate effects.

The synthesis across studies also reveals a more optimistic dimension: well-designed interventions can achieve remarkable cost-effectiveness when they target binding constraints at critical junctures. The Falk, Kosse, and Pinger (2026) mentoring program generated large, persistent effects at one-fortieth the cost of intensive early childhood programs by addressing the specific mechanism (parental beliefs and teacher recommendations) operating at Germany's tracking decision point. Similarly, Cappelen et al. (2026) achieved substantial gains by removing barriers to exercise for students whose self-control problems prevented effective studying—the intervention succeeded precisely because it targeted the constraint rather than attempting comprehensive skill enhancement. This suggests that Option C thinking should emphasize diagnostic precision: identifying which specific parameter in the production function equation constrains development for which subpopulations, then designing targeted interventions that address those specific parameters rather than implementing universal programs that attempt to raise all inputs for all children.

The Caucutt et al. (2026) finding that production technologies show little difference by parental education offers particular policy relevance when combined with evidence on investment complementarity. If disadvantaged families face similar production possibilities but provide fewer investments due to budget constraints, then policies removing financial barriers could substantially narrow achievement gaps—a finding that strengthens the case for progressive targeting derived in equation (19). However, the strong complementarity between inputs (time, goods, childcare) revealed in their estimates suggests that narrow interventions subsidizing single inputs may generate disappointing returns as families substitute away from other complementary inputs. This points toward integrated programs providing balanced input packages rather than categorical subsidies targeting isolated margins.

Finally, the Cuddy and Currie (2026) analysis of mental health treatment provides perhaps the clearest empirical validation of the Option C framework itself. By comparing three scaling approaches (strict evidence-based rules (FDA), flexible guidelines (professional associations), and unguided discretion (practitioner choice)), they demonstrate that the middle path balancing structure with adaptation achieves the best outcomes. This finding generalizes beyond mental health treatment: optimal scaling requires neither mechanical replication of pilot protocols (which fails when contexts differ) nor complete adaptation to local conditions (which permits quality degradation), but rather structured frameworks that maintain fidelity to core mechanisms while allowing controlled variation in implementation details. The paper thus validates the three-stage Option C approach: mechanism testing under realistic conditions (identifying which treatment components drive effects), targeted scaling based on heterogeneity (reaching populations most likely to benefit), and supply-side investments to reduce voltage (maintaining quality through professional standards and training).

These empirical insights suggest several theoretical extensions. The standard production function $\theta_{t+1} = A \cdot \theta_t^{\beta} \cdot (1 + I_t)^{\delta}$ requires augmentation to capture: (i) endogenous technology parameters $A_{t+1}(I_0, ..., I_t)$ that evolve with investment histories; (ii) stage-specific production functions $f^{(s)}(\cdot)$ that vary discontinuously across developmental stages rather than smoothly over time; (iii) intergenerational complementarities where A_{child} depends on $I_{\text{parent's childhood}}$; and (iv) general equilibrium feedbacks where population-level implementation alters the peer environment and parental responses that enter the production function. Incorporating these extensions while maintaining tractability for policy analysis represents an important agenda for future theoretical work. The empirical studies reviewed here provide rich guidance for how these extensions should be structured to capture the actual mechanisms operating in skill formation and program scaling.

References

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2026. "It takes a village: the economics of parenting with neighborhood and peer effects." *Journal of Political Economy.*

Al-Ubaydli, Omar, Chien-Yu Lai, and John A List. 2023. *A simple rational expectations model of the voltage effect.* Technical report. National Bureau of Economic Research.

Al-Ubaydli, Omar, Min Sok Lee, John A List, Claire L Mackevicius, and Dana Suskind. 2021. "How can experiments play a greater role in public policy? Twelve proposals from an economic model of scaling." *Behavioural Public Policy* 5 (1): 2–49.

Al-Ubaydli, Omar, John A List, Danielle LoRe, and Dana Suskind. 2017. "Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature." *Journal of Economic Perspectives* 31 (4): 125–144.

Al-Ubaydli, Omar, John A List, and Dana Suskind. 2017. "What can we learn from experiments? Understanding the threats to the scalability of experimental results." *American Economic Review* 107 (5): 282–286.

———. 2019. *The science of using science: Towards an understanding of the threats to scaling experiments.* Technical report. National Bureau of Economic Research.

———. 2020. "The Science of Using Science: Toward an Understanding of the Threats to Scalability." *International Economic Review* 61 (4): 1387–1409.

Angrist, Noam, Peter Bergman, and Moitshepi Matsheng. 2022. "Experimental evidence on learning using low-tech when school is out." *Nature human behaviour* 6 (7): 941–950.

Angrist, Noam, David K Evans, Deon Filmer, Rachel Glennerster, Halsey Rogers, and Shwetlena Sabarwal. 2023. "How to improve education outcomes most efficiently?"

Becker, Gary. 1964. "Human capita." *New York: N ational Bureau of Economic R esearch.*

———. 1993. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* 3rd. University of Chicago Press.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy* 75 (4, Part 1): 352–365.

Brandon, Alec, Christopher M Clapp, John A List, Robert D Metcalfe, and Michael Price. 2022. *The human perils of scaling smart technologies: Evidence from field experiments.* Technical report. National Bureau of Economic Research.

Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health" [in en]. *Science* 343, no. 6178 (March): 1478–1485. ISSN: 0036-8075, 1095-9203, accessed June 26, 2024. https://doi.org/10.1126/science.1248429. https://www.science.org/doi/10.1126/science.1248429.

Cappelen, Alexander W, Gary Charness, Mathias Ekström, Uri Gneezy, and Bertil Tungodden. 2026. "Exercise Improves Academic Performance." *Journal of Political Economy.*

Caucutt, Elizabeth, Lance John Lochner, Joseph Mullins, and Youngmin Park. 2026. "Child Skill Production:Accountigfor Parental and Market-Based Time and Goods Investments." *Journal of Political Economy.*

Chambers, David A, and Susan T Azrin. 2013. "Research and services partnerships: partnership: a fundamental component of dissemination and implementation research." *Psychiatric Services* 64 (6): 509–511.

Cotton, Christopher, Brent Hickman, John A List, Joseph Price, and Sutanuka Roy. 2026. "Why Don't Struggling Students Do Their Homework? Disentangling Motivation and Study Productivity as Drivers of Human Capital Formation." *Journal of Political Economy.*

Cuddy, Emily, and Janet Currie. 2026. "Rules vs. discretion: Treatment of mental illness in us adolescents." *Journal of Political Economy.*

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.

———. 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of human resources* 43 (4): 738–782.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931. https://doi.org/10.3982/ECTA6551.

Del Boca, Daniela, Christopher Flinn, Ewout Verriest, and Matthew Wiswall. 2026. "Parenting with Patience: Parental Incentives and Child Development." *Journal of Political Economy.*

Dettmer, Amanda M, James J Heckman, Juan Pantano, Victor Ronda, and Stephen J Suomi. 2026. "Effects of Multi-Generational Exposure to Early-Life Advantage: Lessons from a Primate Study." *Journal of Political Economy.*

Dougan, William R., Jorge Luis García, and Illia Polovnikov. 2025. "Reexamining a Multisite Early Childhood Education Program: Gradual Scale-Up, Birth Spacing, and Parental Investment." Forthcoming, *Journal of Political Economy Microeconomics.*

Doyle, Orla. 2020. "The first 2,000 days and child skills." *Journal of Political Economy* 128 (6): 2067–2122.

Duncan, Greg J, and Katherine Magnuson. 2013. "Investing in preschool programs." *Journal of economic perspectives* 27 (2): 109–132.

Durkin, Kelley, Mark W Lipsey, Dale C Farran, and Sarah E Wiesen. 2022. "Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade." *Developmental Psychology* 58 (3): 470.

Elango, Sneha, Jorge Luis García, James J Heckman, and Andrés Hojman. 2015. "Early childhood education." In *Economics of means-tested transfer programs in the United States, volume 2,* 235–297. University of Chicago Press.

Falk, Armin, Fabian Kosse, and Pia Pinger. 2026. "Mentoring and Schooling Decisions: Causal Evidence." *Journal of Political Economy.*

Fatchen, Faith, John A List, and Francesca Pagnotta. 2025. *Using AI to Generate Option C Scaling Ideas: A Case Study in Early Education.* Technical report. National Bureau of Economic Research.

Fixsen, DL. 2005. "Implementation research: A synthesis of the literature." *University of South Florida, Louis de la Parte Florida Mental Health Research Institute, The National Implementation Research Network.*

Frisch, Ragnar. 1936. "Annual survey of general economic theory: the problem of index numbers." *Econometrica: Journal of the Econometric Society,* 1–38.

García, Jorge Luis, James J. Heckman, Duncan Leaf, and María José Prados. 2020. "Quantifying the life-cycle benefits of an influential early-childhood program." *Journal of Political Economy* 128 (7): 2502–2541. https://doi.org/10.1086/705718.

Glasgow, Russell E, Samantha M Harden, Bridget Gaglio, Borsika Rabin, Matthew Lee Smith, Gwenndolyn C Porter, Marcia G Ory, and Paul A Estabrooks. 2019. "RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review." *Frontiers in public health* 7:64.

Heckman, James J, Seong Hyeok Moon, Rodrigo Pinto, Peter A Savelyev, and Adam Yavitz. 2010. "The rate of return to the HighScope Perry Preschool Program." *Journal of public Economics* 94 (1-2): 114–128.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor economics* 24 (3): 411–482.

Heckman, James J, and Jin Zhou. 2026. "A study of the microdynamics of early childhood learning." *Journal of Political Economy.*

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Quarterly Journal of Economics* 131 (4): 1795–1848. https://doi.org/10.1093/qje/qjw027.

Knudsen, Eric I, James J Heckman, Judy L Cameron, and Jack P Shonkoff. 2006. "Economic, neurobiological, and behavioral perspectives on building America's future workforce." *Proceedings of the national Academy of Sciences* 103 (27): 10155–10162.

List, John A. 2022. *The voltage effect: how to make good ideas great and great ideas scale.* London: Penguin Business. ISBN: 978-0-241-55684-9.

———. 2024. "Optimally generate policy-based evidence before scaling." *Nature* 626 (7999): 491–499.

List, John A., and Haruka Uchida. 2024. *Here Today, Gone Tomorrow? Toward an Understanding of Fade-out in Early Childhood Education Programs.* Working Paper 33027. National Bureau of Economic Research, October. https://doi.org/10.3386/w33027.

McEwen, Craig A, and Bruce S McEwen. 2017. "Social structure, adversity, toxic stress, and intergenerational poverty: An early childhood model." *Annual Review of Sociology* 43 (1): 445–472.

Mincer, Jacob. 1958. "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy* 66 (4): 281–302.

———. 1974. *Schooling, Experience, and Earnings.* NBER. Book. https://www.nber.org/books-and-chapters/schooling-experience-and-earnings.

Mobarak, Ahmed Mushfiq. 2022. "Assessing social aid: the scale-up process needs evidence, too." *Nature* 609, no. 7929 (September 29, 2022): 892–894. ISSN: 0028-0836, 1476-4687. https://doi.org/10.1038/d41586-022-03039-2. https://www.nature.com/articles/d41586-022-03039-2.

Mullins, Joseph. 2026. "A structural meta-analysis of welfare reform experiments and their impacts on children." *Journal of Political Economy.*

Muralidharan, Karthik, and Abhijeet Singh. 2025. *Adapting for scale: Experimental Evidence on Technology-aided Instruction in India.* Technical report. National Bureau of Economic Research.

Phillips, Deborah A, and Jack P Shonkoff. 2000. "From neurons to neighborhoods: The science of early childhood development."

Puma, Mike, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer. 2012. "Third Grade Follow-Up to the Head Start Impact Study: Final Report. OPRE Report 2012-45." *Administration for Children & Families.*

Schultz, Theodore W. 1961. "Investments in human capital." *American Economic Review* 51 (1): 1–17.

Shonkoff, Jack P, Andrew S Garner, Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, Section on Developmental and Behavioral Pediatrics, Benjamin S Siegel, Mary I Dobbins, et al. 2012. "The lifelong effects of early childhood adversity and toxic stress." *Pediatrics* 129 (1): e232–e246.

Supplee, Lauren H, Robert T Ammerman, Anne K Duggan, John A List, and Dana Suskind. 2022. "The role of open science practices in scaling evidence-based prevention programs." *Prevention Science* 23 (5): 799–808.

Vivalt, Eva. 2020. "How much can we generalize from impact evaluations?" *Journal of the European Economic Association* 18 (6): 3045–3089.

Willoughby, Michael T, and Clancy B Blair. 2016. "Measuring executive function in early childhood: A case for formative measurement." *Psychological assessment* 28 (3): 319.