Skill Ladders

Cameron Taylor*

March 30, 2020

Abstract

This paper presents a model of skills and derives properties of the optimal investment into educational skills. In the model students can acquire basic and advanced skills at a cost to a policymaker who is budget-constrained. The optimal policy is very sensitive to the structure of the returns to skill - even when advanced skills give unbounded marginal returns, it may be optimal to invest more in basic skills if skills represent a “skill ladder”. These results offer new interpretations on the existing empirical evidence on education interventions. There is a single object that determines whether to invest more in basic or advanced skills and whether the skill ladder model applies. I develop a methodology to estimate the returns to skills and this object and apply it to mathematics (advanced skill) and self-esteem (basic skill) in the NLSY. The results show that the returns to skill reflect that the true state of the world is between the two stark viewpoints and that there is substantial racial heterogeneity in the returns to skills from the lens of the model, suggesting that there may be benefits to focusing more on basic skills in educational policy making and that optimal skill targeting may differ by race.

1 Introduction

A major approach to combatting inequality in the US is through public interventions in education. Programs such as No Child Left Behind aim to use test-taking to give school children certain skills to reduce skill-gaps with the intent to reduce labor market and social inequality.

Recently there has been a large debate on the way policymakers design education curriculums. The recently instated Common Core targets“English Language Arts” and“Mathematics” with the goal that ”students be ready to success academically in credit-bearing, college-entry courses and in workforce training programs”.[1]Education curriculums provide a clear example of a plan to invest

*Stanford University Graduate School of Business. Email contact: cntaylor@stanford.edu. Thank you to Paul Oyer and Isaac Sorkin for helpful discussions and advice and to classmates in the Stanford Econ 248 class for helpful comments.

[1]Source: Common Core Standards PDF.

in certain skills for students. This paper analyzes the problem of curriculum design as an optimal investment problem emphasizing heterogeneity in skills and develops and provides an empirical test using NLSY data.

Economists have long been interested in understanding what education policies work well to improve outcomes for students and combat inequality. While there is a large empirical literature on the impacts of education interventions[2] relatively few theoretical models exist to understand the properties of optimal interventions through understanding the education production function.[3] One exception is the technology of skill formation models pioneered by Cunha and Heckman (2007) (CH) and estimated by Cunha et al. (2010).

This paper analyzes a model that treats student's skills as multidimensional and derives the optimal way for policymakers to invest in education interventions. The main insight is that the structure of how skills map into labor market returns starkly determines how policymakers should allocate their resources to providing skills to students. While advanced skills such as advanced mathematics or computer programming are important to labor market outcomes, skills can build on each other. If more basic skills such as discipline and self-esteem are required to realize returns from these advanced skills, what this paper calls the "skill ladder model", optimal investment requires more investment in these basic skills, no matter how much additional gain one gets from the advanced skills. This insight builds off of Cunha and Heckman (2007) and the literature on the student skill production function and education policy.

The model and analysis contributes to the literature on the theory of education interventions by analyzing the skill dimension of policy-making. The analysis shows that the optimal policy differs between the two models, highlighting the nuance required in designing good policy. Other than timing of investments, papers in the theoretical literature on education interventions focus mostly on incentives (e.g. Lazear, 2006; Barlevy and Neal, 2012) and so this paper highlights a new dimension over which we should be thinking carefully about policy. In particular, this paper emphasizes the importance of connecting the skill production function to casual empirical observations about important labor market skills.

A concrete policy targeted by this paper is curriculum design. These policies are generally designed to target skills of successful labor market participants. For example, in today's world, the ability to do computer programming is clearly a high return skill, when observed in isolation. However, skills may be dependent on one another as in the "skill ladder" model, and policies that target improving children's programming skills may give very little gains.

These ideas are relevant to the current debate surrounding the largest education policy in recent

[2]E.g. Chetty et al. (2011), Neal and Schanzenbach (2010), Angrist et al. (2013), Dobbie and Fryer (2013), Fryer (2014)

[3]Some exceptions are Lazear (2006) and Barlevy and Neal (2012) who focus on incentives in policy design.

history, the Common Core. Many of the criticisms around the Common Core policy involve the specific skills they target. In particular, there is much debate about the type of standards set by the Common Core policy. One parent's take nicely summarizes this sentiment: "Schools should have standards. States should have standards. But they've got to be good standards, and they have to be realistic standards".[4] This paper provides a theoretical justification that these types of investments that aim "too high" could be inefficient.

Another important dimension that is brought up clearly through the lens of the model and tested in the empirical work is that demographic groups may have different skill production functions and returns to skills. This implies that policies targeting important racial skill gaps will not benefit from one of the key tenants of the Common Core, universal standards. The empirical evidence shows that the returns to skills and production functions are quite different across races, and so finds evidence that universal skill targeting may be inefficient.

The model emphasizes the multiplicity of skills that can be targeted by a policy, the fact that policy-making is budget constrained, and how these skills map into labor market outcomes. It compares two specific skill and reward technologies: the "skill independence" and "skill-ladder" models. While the skill technology analyzed in this paper is a special case of the fully general CH model, the approach allows one to gain insight from the importance of multi-dimensional skills. In particular, the workhorse examples that CH use to explore timing of investment mainly consist of one-dimensional skill worlds so that there is no concern over *which skills* policymakers are choosing to invest in. Instead most of the theoretical focus of CH and related papers is on timing of interventions.

To illustrate the intuition behind the main result in words consider a policymaker considering investing in two skills for children in a school, a basic skill B and an advanced skill A. It is costly for the policymaker to invest in each and they have a budget for their total investment.

Consider one case which I call "skill independence": achieving the advanced skill alone is sufficient for achieving the basic skill *or* the advanced skill is rewarded on the labor market independent of the basic skill. In this case, it is optimal for the policymaker to invest more heavily in the advanced skill because skills are substitutes and the advanced skill will reward students more.

The other case is what I call a "skill ladder": one does not necessarily receive the basic skill when gaining the advanced skill *and* the basic skill is necessary to be rewarded for the advanced skill. I call this a skill ladder because the skills are ordered and if the student is missing one of the rungs on the ladder, they fall to the bottom. In this case it is optimal for the policymaker to invest more heavily in the basic skill because the basic skill provides a guaranteed return and augments the advanced skill. This is true even as the marginal returns of the advanced skill over the basic skill become unbounded.

[4] Source: CBS News.

The important policy insight of this paper is that these two models both seem empirically plausible but their policy implications are dramatically different. This paper suggests a specific method for connecting educational policies focused on childhood skills and the skill production function and labor market return function.

The insights from the model allow for a novel interpretation of many disparate findings in the education intervention literature. I examine the model's implications for No Child Left Behind, charter schools, the GED and Head Start. In general, the evidence seems to point towards the skill ladder model being an empirically relevant model and provides a potential explanation for many of these important policy results.

I use the model to conduct empirical analysis to understand what optimal investments look like in a specific case. Importantly, the model suggests a very simple "sufficient statistic" to understand properties of optimal investments. I study an empirical application of the model to mathematics skills and self-esteem in the NLSY79. The exercise looks at a case-study in which mathematics is the advanced skill and self-esteem is the basic skill. This exercise is relevant to current policy as the emphasis on mathematics and similar skills has been a major emphasis of recent educational policies.

The empirical exercise shows that neither the stark viewpoints of the skill ladder or skill independence model are appropriate descriptions of the how skills are rewarded on the labor market in this sample for these two skills. The data do seem slightly more consistent with the skill independence model where mathematics skills are the advanced skill, but the reasoning of the model suggests that only investing in mathematics skills is sub-optimal. Statistical imprecision causes the implications to be somewhat limited.

Racial heterogeneity in the skills production function is also explored. The data reveals important heterogeneity in the returns to skills by race, suggesting that the design of policies could be improved by carefully tailoring interventions to the intervention population. In particular, the results suggest that the skill ladder model is a better description of the returns to skills for Black and Hispanic children than for White children. This is at odds with one of the major founding tenants of the Common Core:

> The central concept [...] is that the nation's 40 million K-12 students should be offered the same high-standard education no matter where they go to school; a child in Mississippi, say, should finish each grade with the same general proficiencies as one in Maine - and ready to compete in an increasingly competitive global marketplace.[5]

The organization of the paper is as follows. First, I set up the model and make formal the intuition provided in the introduction (Sections 2.1-2.2). I will examine a parametric example to

[5]Source: https://www.gse.harvard.edu/news/ed/14/09/what-happened-common-core.

make the ideas more concrete (Section 2.3). A more general version of the model is analyzed in Section 3 which helps motivate better the empirical analysis, Then, new interpretations of existing empirical evidence is offered through the lens of the model (Section 4). I then look at the empirical application of mathematics skills and self-esteem in the NLSY79, formulating an appropriate sufficient statistic for the normative implications of the model (Section 5). Finally I conclude (Section 6).

2 Basic Model

2.1 Setup

Consider a population of students of mass 1. Suppose that there are two skills that students can achieve while in school: basic skills B and advanced skills A. Every student either has each skill or not. These skills translate into future wages according to the variables $W_0 = 0$ (for no skill), $W_A \geq 0$ (for advanced skill only), $W_B > 0$ (for basic skill only), and $W_{BA} > 0$ (for both skills).

The assumptions above allow for the possibility that advanced skills alone are worthless while basic skills have a guaranteed positive payoff. The basic idea for this, explored more below in setting up the two cases of interest, is that without basic skills, advanced skills may be worthless. For example, if a student can solve advanced mathematics problems but does not know how to handle their own mental health or communicate properly, they may not be able to be rewarded for those math skills.

Suppose for simplicity that the population of students have no skills and is homogeneous. A policymaker is considering the optimal way for the school to invest in these skills to maximize the wages of the students. Their goal is potentially motivated by a desire to reduce wage inequality. They are budget-constrained in their resources (time, effort, money) and must allocate these resources across the different skill. In particular, they have a budget of T and investment in each skill is measured as t_B and t_A so that the budget constraint is $t_B + t_A \leq T$. Investments must be non-negative. Investment of t_i in skill $i \in \{B, A\}$ translates into a proportion of $q(t_i)$ students receiving that skill where q is continuous, strictly increasing and satisfies $q(0) = 0$ and $q \leq 1$.

Consistent with empirical evidence on skill wage gaps, I assume that $W_{BA} >> W_B$. That is, students who have both basic and advanced skills are rewarded substantially more on the labor market than students who have only the basic skill.

I illustrate the basic setup of the model in Figure 1. Policymakers decisions partition the mass of students into 4 different groups based on receiving different skills according to investments and the technology $q(\cdot)$. These skills then translate into labor market wages.

There are two special cases of the model I focus on in this paper. The first is what I call

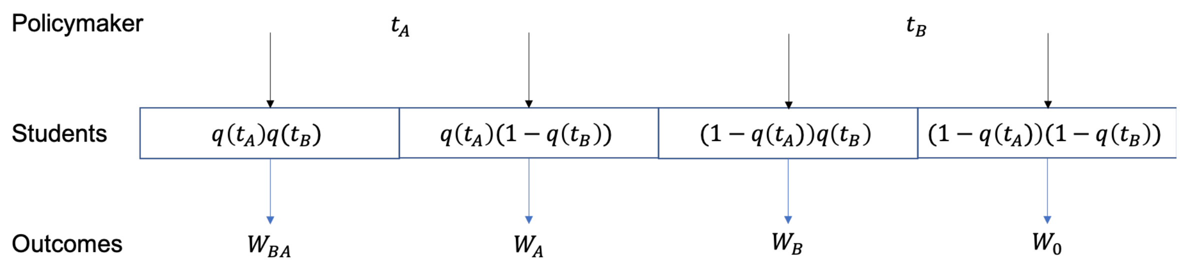

Figure 1: Model Illustration

skill independence. Skill independence has the property that either (a) achieving skill A implies achieving skill B for any student or (b) skill A is rewarded on the labor market regardless of the presence of skill B. In the model this translates to $W_A = W_{BA}$.

The second model I focus on is what I call the **skill ladder** model. In this model achieving the advanced skill A does not imply that a student achieves the basic skill B *and* the high rewards to the advanced skill require the basic skill. In the model this translates to $W_A = W_0 = 0$. The motivation for such a possibility is given above, and works off the possibility that students are only rewarded for their "worst" skill.

Given the setup, the problem for solving for the optimal educational policy in the skill independence model is given by:

$$\max_{t_B,t_A} q(t_B)W_B + q(t_A)W_{BA} - q(t_B)q(t_A)W_B \text{ s.t. } t_B + t_A \leq T \tag{1}$$

while in the skill ladder case the corresponding problem is given by

$$\max_{t_B,t_A} q(t_B)W_B + q(t_B)q(t_A)(W_{BA} - W_B) \text{ s.t. } t_B + t_A \leq T \tag{2}$$

These can be derived by looking at the cases of probabilities of receiving each skill and some basic algebraic manipulation. For example, with mass $q(t_B)q(t_A)$ of the students receive both skills, mass $q(t_B)(1 - q(t_A))$ only receive the basic skill, etc.

2.2 Main Theoretical Results

First it is important to establish that (1) and (2) actually have solutions.

Proposition 1. Both (1) and (2) have solutions.

Proof. The set of choice variables $\{(t_A, t_B) : t_A \geq 0, t_B \geq, t_B + t_A \leq T\}$ is compact. Moreover, since $q(\cdot)$ is continuous and the wage variables are real numbers, the objective functions are continuous. Thus a solution exists. □

The following two propositions that describe the different properties of solutions to (1) and (2) comprise the main theoretical results of this paper. I present the propositions and then the proofs in succession since they contain very similar content.

Proposition 2. At any solution to (1) (t_B^*, t_A^*), it must be that $t_B^* \leq t_A^*$.

Proposition 3. At any solution to (2) (t_B^*, t_A^*), it must be that $t_B^* \geq t_A^*$.

Proof of Proposition 2. Suppose not, i.e. that $t_B > t_A$ at some proposed solution. Consider an alternate solution $t'_B = t_A$ and $t'_A = t_B$. This clearly satisfies the budget constraint. The value of the objective function at the new values is

$$\begin{aligned} q(t'_B)W_B + q(t'_A)W_{BA} - q(t'_B)q(t'_A)W_B &= q(t_A)W_B + q(t_B)W_{BA} - q(t_B)q(t_A)W_B \\ &> q(t_B)W_B + q(t_A)W_{BA} - q(t_B)q(t_A)W_B \end{aligned}$$

since $W_B < W_{BA}$ and $q(t_B) > q(t_A)$ by $q(\cdot)$ strictly increasing. Thus the new value of the objective function is strictly higher, contradicting that the original was a proposed solution. □

Proof of Proposition 3. Suppose not, i.e. that $t_B < t_A$ at some proposed solution. Consider an alternate solution $t'_B = t_A$ and $t'_A = t_B$. This clearly satisfies the budget constraint. The value of the objective function at the new values is

$$\begin{aligned} q(t'_B)W_B + q(t'_B)q(t'_A)(W_{BA} - W_B) &= q(t_A)W_B + q(t_B)q(t_A)(W_{BA} - W_B) \\ &> q(t_B)W_B + q(t_B)q(t_A)(W_{BA} - W_B) \end{aligned}$$

since $W_B > 0$ and $q(t_A) > q(t_B)$. Thus we have a contradiction as required. □

Thus we see that the properties of optimal policies are different across these two plausible models, with only a single assumption difference between them. Importantly these results hold for all W_{BA} and W_B as long as $W_{BA} > W_B > 0$. Even if the return to the advanced skill over the basic skill grows larger and larger ($W_{BA} - W_B \to \infty$) the skill ladder model calls for at least as much investment in the basic skill.

The intuition for these results is most easily seen when assuming that $q(t_i) = t_i$ and $T < 1$. In this case the cross-partial derivative in the arguments in (1) is $-W_B < 0$ and so the investment choices are substitutes. Because $W_{BA} >> W_B$, the advanced skill has a higher payoff and so it is optimal to shift more investment to that skill. In (2) the skills are not substitutes and the marginal benefit of the basic skill is higher due to the guaranteed payoff of $W_B > 0$ and how it reinforces higher payoffs through advanced skills W_{BA}. The advanced skill in (2) only reinforces higher payoffs W_{BA} complementary to investment in B, t_B.

There are two senses in which the model and Propositions 2 and 3 above may not appear very convincing. The first reason is due to the model assumptions: the skill technology functions $q(t_A)$ and $q(t_B)$ are assumed to be identical. It is instructive to consider what relaxing these functions to $q_A(t_A)$ and $q_B(t_B)$ might imply about the robustness of the results.

First note that in the skill independence model if $q_A(t) > q_B(t)$ then Proposition 2 holds, and in the skill ladder world if $q_A(t) < q_B(t)$ then Proposition 3 holds. Thus by assuming that schools are uniformly better at providing one skill or another, one of the propositions holds in one of the models. It is also not hard to see that $q_A(t) > q_B(t)$ pushes against Proposition 3 in the skill ladder model and vice versa for Proposition 2 in the skill independence model. Thus the important intuition is that the skill transmission technology pushes back against each model in a way that makes sense for which skill they more strongly target.

Another reason that these propositions may not appear convincing is the chance for equality $t_A^* = t_B^*$. If in many cases we have $t_A^* = t_B^*$ in both (1) and (2), this result is not as important. I now explore some parametric examples that suggest, with natural and sensible assumptions on $q(\cdot)$, the solutions are quite different.

2.3 Parametric Examples

For the parametric examples I look at two specifications of $q(\cdot)$: $q(t) = t$ with $T \leq 1$ so that probabilities are well-behaved, and $q(t) = 1 - e^{-t}$ to add some (concave) curvature to the probability function and allow T to be unrestricted. For both assume that $W_{BA} >> W_B$ so that the difference is sufficiently large (how large is needed will become clear when the analytical solutions are shown).

First suppose that $q(t) = t$ and consider the skill independence problem (1). Then since the objective function is strictly increasing in each t_i for all interior t_{-i} we must have that the budget constraint binds. Thus, we can write the problem as

$$\max_{t_A} (T - t_A)W_B + t_A W_{BA} - (T - t_A)t_A W_B$$

and the second derivative is $2W_B > 0$ so that this function is strictly convex. Thus, the solution is at a corner. So we compare $t_A = T$ which produces value TW_{BA} and $t_A = 0$ which produces value $TW_B < TW_{BA}$. Thus the optimal solution in skill independence world under this technology is

$$t_A = T.$$

Now consider the skill ladder problem (2) with the same q technology. We can write the problem as

$$\max_{t_A}(T - t_A)W_B + (T - t_A)t_A(W_{BA} - W_B).$$

The second derivative in this case is $-2(W_{BA} - W_B) < 0$ so that the function is strictly concave and thus an interior maximum is optimal. The first-order condition in this case is

$$-W_B + (W_{BA} - W_B) - 2t_A(W_{BA} - W_B) = 0$$

which is rearranged to give

$$t_A = \frac{T}{2} - \frac{W_B}{W_{BA} - W_B}.$$

The differences are striking: skill independence has a corner solution in which policymakers should be completely investing in advanced skills whereas the skill ladder setup has a solution that involves at least half of their time spent on creating basic skills at the skill. Importantly as $W_{BA} \to \infty$ the optimal policy requires $t_B^* \geq t_A^*$.

Working with $q(t) = 1 - e^{-t}$ and performing the same optimization yields that the solutions are respectively

$$t_A = T$$

in (1) and

$$t_A = \frac{T}{2} - \log(\frac{W_{BA}}{W_{BA} - W_B}).$$

in (2).

The solutions for both technologies $q(\cdot)$ display the stark differences in policy recommendations. Moreover, in both cases of the skill ladder model, t_A becomes closer to t_B as $W_{BA} - W_B$ grows (though at different rates). However, even as $W_{BA} - W_B \to \infty$, the skill ladder model requires that at least as much investment is put into basic skills as advanced skills so that there is a sizable discontinuity between the two optimal policies.

3 General Model

The more general model allows for interpolation between the two extreme cases considered above and adds student heterogeneity. In particular, consider the addition of three more parameters: θ, λ_B and λ_A where θ is the interpolation parameter between the skill independence and skill ladder worlds, and λ_j measures the proportion of students in the population that already have skill j.

In particular the more general model states that

$$W_A = \theta W_{BA},\ \theta \in [0, 1] \tag{3}$$

so that when $\theta = 1$ skill independence exists and when $\theta = 0$ skills form a ladder. In this more general model, the interpretation of the θ parameter is that, in this model A skills by themselves are useless, and θ measures the proportion of students who, when they learn A, also inherit B automatically. That is $(1-\theta)$ fall down the ladder and essentially have no skills. Basically skills are "risky".

As well, the model allows students to have heterogeneous skills so that λ_A proportion already have advanced skills and λ_B proportion already have basic skills. These skills could have been inherited through earlier educational programs or could be because the students inherit these skills genetically or through investments made in their own homes.

Then the investment problem can be defined by

$$\max_{t_A, t_B} \lambda_B q(t_A)(W_{BA} - W_B) + (1 - \lambda_A - \lambda_B) f(t, \theta) \tag{4}$$
$$f(t,\theta) = \theta q(t_A)(1 - q(t_B))W_{BA} + (1 - q(t_A))q(t_B)W_B + q(t_A)q(t_B)W_{BA}$$
$$\text{s.t. } t_A + t_B \le T$$

Note that with $\lambda_A = \lambda_B = 0$ and $\theta = 1$ we recover (1) above and with $\lambda_A = \lambda_B = 0$ and $\theta = 0$ we recover (2) above.

There is a solution for (4) using essentially the same proof as above.

Proposition 4. (4) has a solution for any $\theta \in [0, 1]$.[6]

Moreover, the intuition from the independence and ladder models above transfers to the more general model. Basic skills are more valuable as θ decreases and advanced skills are more valuable as θ increases.

Proposition 5. Consider solutions $t_A^*(\theta)$ and $t_B^*(\theta)$ to (4) above. $t_A^*(\theta)$ is increasing in θ (in the strong set order) and $t_B^*(\theta)$ is decreasing in θ (in the strong set order).

Proof. The goal is to apply the Topkis Monotone Comparative Static theorem by showing that the objective function has increasing differences. This is sufficient for the optimal solution to be increasing in the strong set order. I show the proof for $t_A^*(\theta)$; the proof for $t_B^*(\theta)$ is similar.

To reframe (4) as a simple maximization problem in t_A note that at any solution it must be that $t_A^* + t_B^* = T$, thus we can treat the optimization problem as one over a single variable t_A and replace $t_B = T - t_A$. Then the objective function is

$$F(t_A; \theta) = \lambda_B q(t_A)(W_{BA} - W_B) + (1 - \lambda_A - \lambda_B) f(t_A; \theta)$$

[6] Ideally, there exists a unique solution so that the application of Topkis' theorem below is not about solutions that are "set-increasing" but univariate functions. However it is hard to give conditions to guarantee such a unique solution.

To check increasing differences let $\bar{t}_A \geq \underline{t}_A$ and $\bar{\theta} \geq \underline{\theta}$. Then

$$\begin{aligned} F(\bar{t}_A, \bar{\theta}) - F(\underline{t}_A, \bar{\theta}) = \bar{\theta} W_{BA} \Big[q(\bar{t}_A)(1 - q(T - \bar{t}_A)) - q(\underline{t}_A)(1 - q(T - \underline{t}_A)) \Big] \\ + W_B \Big[(1 - q(\bar{t}_A)) q(T - \bar{t}_A) - (1 - q(\underline{t}_A)) q(T - \underline{t}_A) \Big] \\ + W_{BA} \Big[q(\bar{t}_A q(T - \bar{t}_A) - q(\underline{t}_A) q(T - \underline{t}_A) \Big] \end{aligned}$$

Similarly,

$$\begin{aligned} F(\bar{t}_A, \underline{\theta}) - F(\underline{t}_A, \underline{\theta}) = \underline{\theta} W_{BA} \Big[q(\bar{t}_A)(1 - q(T - \bar{t}_A)) - q(\underline{t}_A)(1 - q(T - \underline{t}_A)) \Big] \\ + W_B \Big[(1 - q(\bar{t}_A)) q(T - \bar{t}_A) - (1 - q(\underline{t}_A)) q(T - \underline{t}_A) \Big] \\ + W_{BA} \Big[q(\bar{t}_A q(T - \bar{t}_A) - q(\underline{t}_A) q(T - \underline{t}_A) \Big] \end{aligned}$$

and so taking these differences yields

$$(\bar{\theta} - \underline{\theta}) W_{BA} \Big[q(\bar{t}_A)(1 - q(T - \bar{t}_A)) - q(\underline{t}_A)(1 - q(T - \underline{t}_A)) \Big]$$

Now, $\bar{\theta} \geq \underline{\theta}$ and $q(\bar{t}_A) \geq q(\underline{t}_A)$ since q is strictly increasing. Also $q(T - \underline{t}_A) \geq q(T - \bar{t}_A)$ again by q strictly increasing. Thus $1 - q(T - \bar{t}_A) \geq 1 - q(T - \underline{t}_A)$. Thus, this expression is non-negative and thus we have increasing differences.

Thus by the appropriate Topkis Monotone Comparative static theorem, $t_A^*(\theta)$ is increasing in θ in the strong set order. □

It is also informative to explore the implications of student heterogeneity in skills in the more general model. λ_A reduces the overall payoff from investing: clearly the value to redesigning the curriculum in this model is lower if students already have the skills of interest. The intuition for how λ_B impacts the optimal solution is simple: λ_B makes all basic skill investing less profitable since students already have basic skills.

Proposition 6. $t_B^*(\lambda_B)$ is decreasing.

The more general model is particularly useful for allowing the ideas from the specific examples in the previous section to be taken to the data. It is challenging to figure out ways to empirically distinguish the skill ladder and skill independence model above without some concrete way to connect them. In this case, θ provides a parameter that interpolates between the two. Propositions 2, 3 and 5 make clear that θ is an important policy-relevant parameter in the model and so the empirical application will be devoted to exploring how to estimate θ in a transparent and credible way.

Before moving onto the empirical application, it is informative to view the existing empirical evidence on educational interventions through the lens of the model.

4 Interpreting Existing Empirical Evidence

In this section I interpret the empirical evidence from the education intervention literature in the context of the model. The goals are (1) to show how the model offers new interpretations on existing empirical results in a unified framework and (2) provide suggestive evidence between the two major models analyzed.

To do this I analyze the empirical literature on four major policy relevant education reforms. I focus on No Child Left Behind, Charter Schools, the GED and Head Start.

4.1 No Child Left Behind

No Child Left Behind (NCLB) has been studied by economists for its incentive problems related to "teaching to the test" (Lazear, 2006) and how it allocates teacher's time across students (Neal and Schanzenbach, 2010). This paper adds to the economics literature on NCLB by providing a way to interpret how the policy performs on skill targeting, as opposed to incentives.

One interpretation of policies that target testing is that they target cognitive skills associated with being able to answer questions on the tests. If NCLB is interpreted as targeting more "advanced" cognitive testing skills so that $t_A = T$, the appropriate question for policymakers about whether there are returns to shifting towards more basic skills depends on whether (a) the market requires basic skills to reward these test skills and (b) these basic skills are not achieved by having students pass tests.

If the the answer to both of these questions is yes, then the model suggests this policy lives in the skill ladder world and allocating resources to target more "basic" skills would be more efficient. Some examples of these more basic skills might include non-cognitive skills, but the models shows that the key property is they are any skills that is required in the labor market reward function for students to achieve the rewards of having the testing skills.

Thus, skill targeting provides another potential critique of NCLB. However, whether or not this aspect of NCLB is optimally designed is fundamentally an empirical question as highlighted by the model.

Note that another interpretation of NCLB is that it targets basic skills, as being able to pass a test only requires basic effort and discipline. If this is the case, then the relevant policy question is what the returns to more advanced investment activities, for example computer programming, are in relation to the basic skills that these tests give.

4.2 Charter Schools

As cited in the introduction, there is a large empirical literature in economics that looks at the heterogeneous impacts of charter schools. Angrist et al. (2013) show that charter schools that adopt a "No Excuses" approach to their education seem to have the most beneficial impacts on students. Angrist et al. (2013) say that No Excuses schools emphasize "discipline and comportment, traditional reading and math skills" and feature "strict discipline, uniforms, and cold calling".

My preferred interpretation of these features within the framework of this model is that these notions target basic non-cognitive and cognitive skills. Under this interpretation, the relative success of charter schools in achieving better outcomes by shifting investment towards basic skills is consistent with the skill ladder model in this paper.

While the success of charter schools is unlikely to be driven entirely by a clean cut in what skills are targeted, the skill framework of this paper does provide an interpretation of this success, and more importantly, the treatment effect heterogeneity.

4.3 GED

Heckman and Rubinstein (2001) find that the returns to the GED (meaning the returns to passing the exam and receiving the accreditation) are extremely low. The evidence and main conceptual argument is that while the GED signals and assesses cognitive skills, a lack of non-cognitive skills makes these skills meaningless.

This empirical evidence provides an example of the importance of the skill ladder model: cognitive skills from the GED (equivalent to high school graduate test taking abilities) are not rewarded without more "basic" non-cognitive. Thus this allows one to interpret the shortcomings of the GED program directly as attempting to assess and train a more advanced skill. It also suggests policy interventions to improve the program - investing more in basic and non-cognitive skills.

4.4 Head Start

There is a large literature estimating effects of the Head Start program on children (Garces et al., 2002; Ludwig and Miller, 2007). A particularly interesting finding in the literature is the "fade-out" of cognitive skills associated with Head Start (Deming, 2009).

If Head Start targets both cognitive and non-cognitive skills, and the cognitive skills fade out, then its measured benefits can be assumed to be derived from the non-cognitive skills. The fadeout and subsequent labor market returns could be consistent with a skill ladder model of the world where Head Start targets both basic and advanced skills, the advanced skill targeting is insufficient without basic skills (hence the fadeout), yet students still benefit from improved basic skills.

However the Head Start evidence is challenging to interpret in the context of this model. Suppose we are in the skill ladder world and that Head Start targets both basic and advanced skills. What is somewhat confusing in this story is that students do not retain advanced skills because of lack of basic skills, but do still benefit from some basic skills. This is a bit difficult to rationalize within the model since the fundamental problem is that I do not observe the Head Start production or investment function. In particular, this is also consistent with mostly targeting non-cognitive skills which then boost early test scores (for some reason) and not targeting cognitive skills - then the empirical observations have no content within the model since both have $W_B > 0$.

5 Empirical Application: Mathematics Skills and Self Esteem

5.1 Setup and Motivation

Many education interventions target concrete test-taking skills as opposed to other "softer interventions".[7] Would there be a gain from shifting resources towards these investments?

To answer this question this paper approaches it within the specific context of mathematics skills (A), a staple of test-taking, and self-esteem (B) as modeled above. Self-esteem is an important non-cognitive skill that could, in theory, alter the efficacy of mathematics skills according to the skill ladder model or the corresponding model above where $\theta < 1$.

While these are very specific skills, this question mirrors a policy story that is common throughout school districts across the US, as there are trade-offs and limited resources in what teachers and schools can teach children. In this case, suppose a school district is deciding whether to invest in an advanced mathematics program or a counseling program This is not unlike a trade-off between real programs that school districts face in allocating funds and interventions for low-skill communities. For example, Colorado recently implemented a program to counsel low-income youth in hopes of increasing their employment chances (Gonser, 2018).

5.2 Empirical Strategy and Data

An important part of the model's theoretical analysis is that it provides a sufficient statistic for when policymakers are deciding to shift investment between two skills and the key mechanism of the model. The two modeling approaches make clear that the crucial policy object is θ. As θ interpolates between 0 and 1, we interpolate between the skill ladder and skill independence models, leading to changes in the relative ratios of t_A^* and t_B^*. Following the ideas of Chetty (2009),

[7] One example of an important intervention that has been studied that focuses on "soft-skills" is in crime (Heller et al., 2017).

the model structurally directs the empirical exercise to focus on the single sufficient statistic object. I now develop a strategy to estimate θ convincingly.

The empirical exercise I aim to perform is closely related to previous empirical work that aims to uncover the returns to different types of skills including cognitive and non-cognitive skills (Heckman et al., 2006) and social skills (Deming, 2017) along with studies already mentioned that structurally estimate the technology of skill formation (Cunha et al., 2010).

This literature primarily utilizes the National Longitudinal Survey of Youth 1979 (NLSY79) sample as it includes a rich set of test scores and labor market outcomes. I follow the literature in drawing my raw measures of skills and outcomes from this sample.

Any estimate of θ must depend on returns to skills which presents two classic challenges. The first is to find a way to measure both A and B convincingly in the data. My model differs from previous empirical models in that I treat skills discretely. I treat skills discretely and continuously in the empirical implementation to make the analysis more comparable to previous work in the literature. The second important factor is the need to separate out W_A from W_{BA}. Thus taking a mean of wages of workers with skill A in the data is insufficient since the observed wage for workers with skill A includes workers with and without skill B.

To develop a parsimonious estimation methodology, first consider the following linear model for determining wages w in the context of the model where skills are treated as binary variables:

$$w = \beta_0 + \beta_A \text{Skill A} + \beta_B \text{Skill B} + \beta_{BA} \text{Skill A} \times \text{Skill B} + \epsilon. \tag{5}$$

This specification mirrors common specifications seen in the literature (Deming, 2017) and is what Heckman et al. (2006) the "conventional approach". However, a subtle but important difference between this specification and the canonical specification is that skills are allowed to depend on one another. To see the importance of this change consider the model without the interaction term. If we find that we cannot reject $\beta_B = 0$ or that β_B is very small and that we can reject β_A and we estimate that β_A is positive, this might lead us to believe that $W_{BA} = W_A$ and thus favor the skill independence model. However, the model places no restriction on W_{BA} and W_B. Importantly, even if W_B is very small, if in adding an interaction between Skill A and Skill B the return to Skill A goes to 0, then this is consistent with the skill ladder model. In other words, W_{BA} allows for complementarities between skills that is not captured in (5) without an interaction term.

If the skills are exogenous and exactly measured through test scores, then regardless of how they are related to one another in the model, we get that an appropriate estimator for W_A is $\hat{\beta}_A$ which can be estimated by OLS in (5), as it measures the sole return to A for people with no B skills. We also care about $\hat{\beta}_A$ in relationship to the estimate for W_{BA} which here can be seen to be

$\hat{\beta}_A + \hat{\beta}_B + \hat{\beta}_{BA}$.

The parameter θ is an important parameter, and can be estimated simply from this routine using the plug-in principle based on (3). In particular an appropriate plug-in estimator is

$$\hat{\theta} = \hat{W}_A/\hat{W}_{BA} = \hat{\beta}_A/(\hat{\beta}_A + \hat{\beta}_B + \hat{\beta}_{BA})$$

This estimate of θ provides an estimate of the relative distance between the skill ladder and skill independence models in the context of mathematics skills and self-esteem. The model highlights that as θ varies, the optimal policy varies, too.

Heckman et al. (2006) highlight two major problems with such procedures using data with variables similar to the NLSY. The first issue is the reverse causality problem, particularly related to the individual's schooling. An individual's schooling determines their test scores, and their test score results determine their own inferences about their abilities and thus subsequent schooling decisions.

Second, skills are challenging to measure directly. The NLSY and related datasets contain test score information which researchers often use as proxies for certain skills. This is problematic because it relies on the quality of the tests to precisely extract information about individuals skills, and then for these skills to map cleanly into the specified model. I call this overall issue the "measurement error" issue.

The way that Heckman et al. (2006) deal with these issues is to structurally model cognitive and non-cognitive skills as latent factors that are affected by the test score measurements in the NLSY and specify how these latent factors affect schooling and employment decisions. I take a different approach since the major innovation in my measurement problem and conceptual problem is to add the interaction term in (5). This makes the estimation problem more similar to Deming (2017) who also investigates some aspects of skill complementarity in the NLSY. The way I deal with the endogeneity issues in (5) is more similar to Deming (2017).

To deal with the first issue, I aim to reduce this type of schooling endogeneity as much as possible by focusing on the cohort of children whose skills were measured when they were age 18 at the latest, so that effects of college decision cannot drive the results. This attempts to more plausibly isolate the variation coming from skills in determining wages. Including extra controls related to subsequent outcomes for these children such as employment and schooling would yield the skill return estimates to be inconsistent due to the "bad control" problem (Angrist and Pischke, 2009). Thus, my empirical strategy only consists of utilizing controls for individuals of things that were determined before skills assigned, such as race, gender and age of the child. I also look at income 20 years later to increase the chance that everyone is "at-risk" of being employed. Since I evaluate wages in the same year for all individuals in my basic empirical strategy, I do not require time fixed effects.

The second issue is much more challenging and cannot be dealt with easily using existing machine learning or reduced-form methodologies. For mathematics skills, I take the Armed Services Vocational Aptitude Battery (ASVAB) exam mathematics knowledge (standardized) score. This test assesses individuals mathematics knowledge through questions related to algebra, geometry and fractions. It is one of the measures used by Heckman et al. (2006) in constructing their cognitive skill measures and one of the inputs into the AFQT score often used in the literature as a measure of cognitive skills (e.g. Neal and Johnson, 1996). To assess self-esteem, I use the Rosenberg Self-Esteem scale which measures perceptions of self-worth (Rosenberg, 1965). Positive returns to improved self-esteem have been examined empirically in the economics and psychology literature (e.g. Murnane et al., 2001; Groves, 2005; Waddell, 2006). For both skills, I normalize these within the sample to have a mean of 0 and a standard deviation of 1.

I allow my model to treat skills both continuously and discretely. I say that an individual "has a skill" if their raw score is above average, and they do "not have a skill" if their raw score is below average. I present this discretization because it is line with the conceptualization in the model but the results are not sensitive to it specifically.

The actual estimation undertaken consists of estimating the returns to skills coefficients β through estimating

$$w_i = \beta_0 + \beta_A \text{Skill A}_i + \beta_B \text{Skill B}_i + \beta_{BA} \text{Skill A}_i \times \text{Skill B}_i + X_i \gamma + \epsilon_i \qquad (6)$$

where X_i is a vector that includes the child's race, sex, and age at the time of the test. As noted above, this is similar to Deming (2017)'s strategy without a panel or time-varying element. I treat w_i as income instead of wage for simplicity.

5.3 Results

First I plot the correlation between the math and non-cognitive skill in Figure 2. The skills are positively correlated. We can clearly reject the model that treats math skills as sufficient for self esteem skills in this sample - there are many points below average in the Rosenberg score that are well above average in the math score. This appears to remove one of the possibilities of the skill independence model that the advanced skill is sufficient for the basic skill. However, it is still possible that the advanced skill is rewarded without the basic skill.

The results for estimating (6) are contained in Table 1. I include both continuous and discrete measures of the skills. Columns (1) and (2) include no other controls while columns (3) and (4) add standard sex-race-age NLSY controls so that the skills are estimated within these cells. The coefficient estimates on the skills do not differ much when adding these controls.

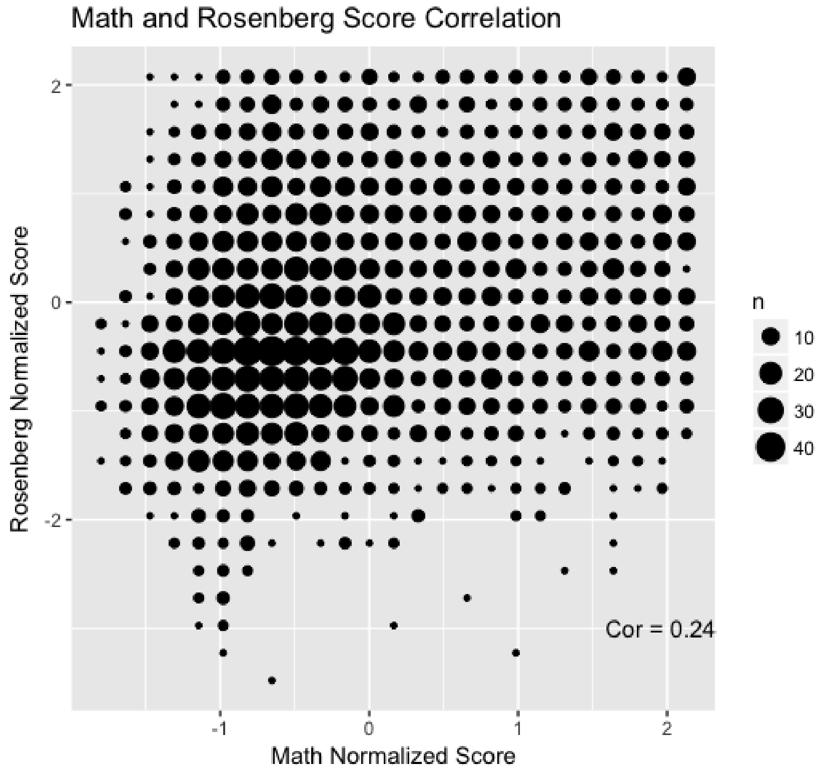

Figure 2: Math and Rosenberg Skills
Notes: This is a plot of the subsample of the NLSY 1979 that I use to estimate the returns to skill of the Math skill from the ASVAB and the Rosenberg skill from the Rosenberg test. The correlation between the test scores is also labeled in the figure and is approximately 0.24.

The first result that stands out is that math scores are rewarded on the labor market independent of non-cognitive self worth skills in all columns. The estimated $\hat{\beta}_A$ is estimated to be substantially large and precise. If we take the discrete specification to measure the presence of a skill as in the model, the parameter estimates suggest that having the mathematics skill in 1979 leads to about a \$14k increase in income in 2000, almost half of the mean income level in this subsample (\$30,560).

The fact that $\hat{\beta}_A$ is both economically and statistically significant suggests that we can reject the pure skill ladder view that advanced math skills are not rewarded by the labor market without the ability to manage self-esteem.

Can we similarly reject the pure skill independence model? Recall that this states that $W_{BA} = W_A$. In this case, this amounts to a statistical test of the returns from having both skills is equal to the returns from having only skill A, the mathematics skill. In the model the appropriate test has the null hypothesis $\hat{\beta}_B + \hat{\beta}_{BA} = 0$. Simply eye-balling Table 1 suggests that we will easily reject this. The F-statistic from this test is about 20.4 with a p-value far smaller than 0.01, suggesting that we can easily reject this.

Thus, both extreme models suggested by the theory are rejected. I form an estimate for $\hat{\theta}$ based

	Dependent variable:			
	Income			
	(1)	(2)	(3)	(4)
Math (Normalized)	10,093***		9,918***	
	(523)		(545)	
Rosenberg (Normalized)	3,814***		3,857***	
	(521)		(505)	
Math × Rosenberg (Cts)	2,198***		1,824***	
	(506)		(486)	
Math Discrete		14,381***		13,634***
		(1,523)		(1,514)
Rosenberg Discrete		3,930***		4,290***
		(1,413)		(1,363)
Math × Rosenberg (Discrete)		4,860**		4,261**
		(2,131)		(2,044)
Observations	3,591	3,591	3,591	3,591
R^2	0.137	0.085	0.209	0.162
Adjusted R^2	0.136	0.084	0.207	0.160

Table 1: Skill Return Estimates

Notes: Parameter estimates from estimating (6) with OLS on the NLSY79 subsample. This subsample includes all children who were 18 or younger in 1979. The tests were administered in 1980. The dependent variable is total non-military income for each individual. The discrete measures are simply indicators for positive scores and either test. *p<0.1; **p<0.05; ***p<0.01

on the the estimates from Table 1 in Column (4). To get standard estimates I bootstrap with 1,000 bootstrap simulations. The results are in Table 2 below. The estimate for θ is between 0.6 and 0.65. The statistical results reject a θ of less than about 0.45 and more than about 0.75. These results suggest that the skill independence model is a better description of the data in this specific application, although statistical imprecision does not allow for a sharp conclusion.

5.4 Heterogeneous Returns to Skills

Many education interventions are targeted at reducing specific inequality gaps. If skill production functions and the returns to skill differ by race, then optimal interventions may differ across races. This is particularly important to national curriculum design which often entails uniform standards across communities. The Common Core has this property.

To assess whether these types of heterogeneity are important, I estimate (6) on different races. To summarize the results I report $\hat{\theta}$ for each sub-group along with bootstrapped standard errors.

	Continuous Skills	Discrete Skills
$\hat{\theta}$	0.636	0.615
Std Error	0.0446	0.0725

Table 2: Estimates of θ

Notes: Estimated using results from Table 1 Column (3) and Column (4). Standard errors are computed by the bootstrap with 1,000 bootstrap samples. *p<0.1; **p<0.05; ***p<0.01

		Continuous Skills	Discrete Skills
Black	$\hat{\theta}$	0.55	0.471
	Std Error	0.072	0.107
Hispanic	$\hat{\theta}$	0.408	0.48
	Std Error	0.06	0.149
White	$\hat{\theta}$	0.725	0.664
	Std Error	0.067	0.102

Table 3: Estimates of θ by race

Notes: Estimated using results from estimating (6) on different subsamples of race. Standard errors are computed by the bootstrap with 1,000 bootstrap samples. *p<0.1; **p<0.05; ***p<0.01

The results are contained in Table 3. There are two striking results from this table. The first is that the estimated θ is smaller for Black and Hispanic populations, suggesting that with respect to mathematics and self-esteem, the skill ladder model is a better description of the data than for White populations. The second is that many of the estimates suggest $\hat{\theta}$ is less than 0.5 for Black and Hispanic populations. Taking θ as a literal measure of the distance between the skill ladder and skill independence model, this result suggests that the skill ladder model may be a better description of the relationship between mathematics and self-esteem in minority populations than skill independence. However, the model cannot reject larger values of $\hat{\theta}$. Further conclusions on the exact optimal policies by race require more data and modeling efforts, but these results suggest that taking into account the impact of self-esteem and other basic skills in minority populations on their labor market outcomes could be crucial in designing effective education interventions.

6 Conclusion

In this paper I analyzed a model of education interventions that emphasizes the space of skills and how skill production functions and the returns to these skills fundamentally alters optimal policy. I show that if skills form a "skill ladder" then investment in basic skills relative to other skills is optimal. However, if skills are independent in any of the two senses discussed in the paper, more advanced skill should be targeted.

I interpret the existing empirical evidence on education interventions using the model. The evidence suggests that the skill ladder model could be a useful way to interpret many of these results. It also suggests that the skill ladder model is empirically relevant for how we assess how skills map into labor market outcomes.

I examine an application using the NLSY79 examining the differential returns to mathematics skills and self-esteem. I develop a simple but credible way to estimate the returns to these skills. The estimates suggest that neither stark model is correct. The data on mathematics and self-esteem seem to be more consistent with a skill independence world in which mathematics skills get relatively good returns without self-esteem. However, self-esteem still has an important role. Moreover, there is important heterogeneity in the racial returns to skills.

References

Angrist, J. D., Pathak, P. A., and Walters, C. R. (2013). Explaining Charter School Effectiveness. *American Economic Journal: Applied Economics*, 5(4):1–27.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, 1 edition edition.

Barlevy, G. and Neal, D. (2012). Pay for Percentile. *American Economic Review*, 102(5):1805–1831.

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics*, 1(1):451–488.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Cunha, F. and Heckman, J. (2007). The Technology of Skill Formation. *AEA P&P*, 97(2):47.

Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78(3):883–931.

Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3):111–134.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Dobbie, W. and Fryer, R. G. (2013). Getting Beneath the Veil of Effective Schools: Evidence From New York City. *American Economic Journal: Applied Economics*, 5(4):28–60.

Fryer, R. G. (2014). Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments. *The Quarterly Journal of Economics*, 129(3):1355–1407.

Garces, E., Thomas, D., and Currie, J. (2002). Longer-Term Effects of Head Start. *American Economic Review*, 92(4):999–1012.

Gonser, S. (2018). School counselors keep kids on track. Why are they first to be cut?

Groves, M. O. (2005). How important is your personality? Labor market returns to personality for women in the US and UK. *Journal of Economic Psychology*, 26(6):827–841.

Heckman, J., Stixrud, J., and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

Heckman, J. J. and Rubinstein, Y. (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review*, 91(2):145–149.

Heller, S. B., Shah, A. K., Guryan, J., Ludwig, J., Mullainathan, S., and Pollack, H. A. (2017). Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago*. *The Quarterly Journal of Economics*, 132(1):1–54.

Lazear, E. (2006). Speeding, Terrorism, and Teaching to the Test. *The Quarterly Journal of Economics*, 121(3):1029–1061.

Ludwig, J. and Miller, D. L. (2007). Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *The Quarterly Journal of Economics*, 122(1):159–208.

Murnane, R. J., Willett, J. B., Braatz, M. J., and Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY. *Economics of Education Review*, 20(4):311–320.

Neal, D. and Schanzenbach, D. W. (2010). Left Behind By Design: Proficiency Counts and Test-Based Accountability. *The Review of Economics and Statistics*, 92(2):21.

Neal, D. A. and Johnson, W. R. (1996). The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy*, 104(5):869–895.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton University Press. Google-Books-ID: YR3WCgAAQBAJ.

Waddell, G. R. (2006). Labor-Market Consequences of Poor Attitude and Low Self-Esteem in Youth. *Economic Inquiry*, 44(1):69–97.