Targeting Skills in Education Interventions

Cameron Taylor*

September 6, 2019

Abstract

In this paper I explore optimal investments in different skills in education interventions. I analyze two closely related and plausible models and show that the policy implications starkly differ between them. I apply my results to interpret the existing empirical evidence on education interventions. I show that my model suggests a "sufficient statistic" approach to guiding skill-targeting in education policy. I then develop a methodology to utilize this approach and apply it to mathematics (advanced skill) and self-esteem (basic skill) in the NLSY. The results show that the returns to skill likely reflect that the true state of the world is likely between the two stark viewpoints.

1 Introduction

There is an important economics literature at the intersection of labor and education devoted to understanding effective policy interventions to reduce labor market inequality. While there is a large empirical literature on the impacts of education interventions[1] relatively few theoretical models exist to understand the properties of optimal interventions.[2] One exception is the technology of skill formation models pioneered by Cunha and Heckman (2007) and estimated by Cunha et al. (2010).

Cunha and Heckman (2007) (CH from here on out) emphasize dynamic complementarities in their model, anchoring it in empirical facts from the child development literature. The main insight of CH is that there is an important efficiency-equity trade-off in late stage interventions that is not present in early stage interventions. This insight forms the basis of the authors' argument that

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

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

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

early interventions are critical. The question that motivates this paper is the following: can we make more important policy recommendations by analyzing other aspects of this type of model?

In this paper I analyze properties of optimal policies that target different skills under two special cases of CH's general model with a budget-constrained policymaker. Instead of emphasizing how policies should optimize their timing, I explore how policies should optimize their choice of skills to focus on in a single static investment.

I contribute to the literature on the theory of education interventions by analyzing the skill dimension of policy-making. I show 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 my paper highlights a new dimension over which we should be thinking carefully about policy.

My 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. I compare two specific skill and reward technologies that I call the skill independence and "skill-ladder" models. While a special case of the fully general CH model, my approach allows one to gain insight from the importance of multi-dimensional skills.[3] 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.

The mathematical and economic intuition behind my main result is very close to that of the important CH insight on dynamics but is distinct in its policy recommendations. Thus, one take on this paper's contribution is that it applies the CH insight across the skill dimension instead of the time dimension.

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 certain 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

[3]CH are able to define the appropriate notions in the multi-skill setup of the model in their online Appendix but do not illustrate the intuition that I illustrate in this paper across skills as they do across time.

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.

The important policy insight of this paper is that these two models both seem empirically plausible but their policy implications are dramatically different. This suggests that skill-targeting in policy should be very closely tied to the skill production function and labor market return function.

I use these positive and normative insights from the model to interpret some of the findings in the education intervention literature related to No Child Left Behind, charter schools, the GED and Head Start. In general, the evidence seems to point towards the skill ladder model being a potential explanation for many of these policy results.

The model suggests a very simple "sufficient statistic" method that could help empirical researchers and policymakers in deciding how to target skills through policies. To illustrate this method I study an empirical application of the model to mathematics skills and self-esteem in the NLSY79. The empirical methodology show how the transparency of the model inform a very straightforward derivation of the sufficient statistic. 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 would likely be sub-optimal.

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 make the ideas more concrete (Section 2.3). Then, I discuss existing empirical evidence, how my model informs these results and vice versa (Section 3). 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 4). Finally I conclude (Section 5).

2 Model and Analysis

2.1 Setup

The main purpose of this section is to give the details of the intuition given in the introduction. The general model setup is as follows. 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),

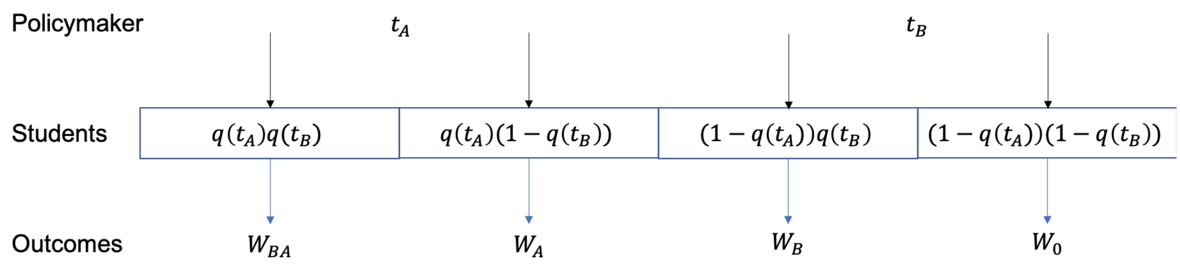

Figure 1: Model Illustration

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 **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 the optimal policy in 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 is

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

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.

The main analysis of this paper comes from studying the problems (1) and (2) and comparing their solution properties to real education policies and the empirical literature on education interventions.

2.2 Main 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 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. $\square$

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. $\square$

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.

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 Interpreting Existing Empirical Evidence

In this section I interpret the empirical evidence from the education intervention literature in the context of the model. The goal is to both show how the model clarifies and nuances the empirical results in a unified framework but also 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.

3.1 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.

3.2 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.

3.3 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.

3.4 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 $W_B > 0$.

4 Empirical Application: Mathematics Skills and Self Esteem

4.1 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 W_A. As W_A goes from 0 to W_{BA}, 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), the model structurally directs the empirical exercise to focus on the single sufficient statistic object. I now develop a strategy to estimate W_A convincingly.

In my approach I follow a long literature on non-cognitive and cognitive skills as interpreted in the previous section. I treat the basic skill B as basic cognitive skills and the advanced skill A as advanced cognitive skills. I discuss more how I measure and operationalize these below.

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.

The value of the model is that the major policy implications come only from estimating W_A and W_{BA}, and so the empirical strategy aims only to estimate the return to Skill A. There are two important issues to address to estimate W_A. The first is to find a way to measure both A and B in the data. My model differs from previous empirical models in that I treat skills discretely. I continue to treat skills discretely in the empirical implementation. I deviate from the literature in using a data driven machine learning approach to measure A and B instead of using latent factors of skills (e.g. Heckman et al., 2006).

The second important factor is that I 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:

$$w = \beta_0 + \beta_A \text{Skill A} + \beta_B \text{Skill B} + \beta_{BA} \text{Skill A} \times \text{Skill B} + \epsilon. \tag{3}$$

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 3 without an interaction term.

If the skills are exogenous and exactly measured, 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 (3), 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 $\hat{\beta}_A + \hat{\beta}_B + \hat{\beta}_{BA}$.

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 incredibly 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 (3). 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 (3) 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. My approach is to reduce the dimensionality of the problem and focus on two very specific measures of cognitive and non-cognitive skills so that the measurements are transparent. For cognitive 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). For non-cognitive skills, 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.

The focus on a single skill measure is also useful in that it makes the empirical exercise more clear within the context of the model, and the resulting empirical results suggestive of more concrete policy implications. Suppose, for example, a school district is considering how to optimally allocate their funding across two programs: a mathematics program and a counseling program. The mathematics program provides the advanced mathematics skills while the counseling program provides the basic skills required to help students deal with self-esteem issues. This is not unlike a trade-off between real programs that school districts face in allocating funds and interventions for low-skill communities.[4]

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.

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

$$w_i = \beta_0 + \beta_A \text{Skill A}_i + \beta_B \text{Skill B}_i + \beta_{BA} \text{Skill A}_i \times \text{Skill B}_i + X_i\gamma + \epsilon_i \tag{4}$$

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.

4.2 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 idea that having the ability to do math well is sufficient for the ability to have high self worth inn 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 (4) 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 sex-race-age controls so that the skills are estimated within these cells. The coefficient estimates on the skills do not differ much when adding these controls.

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

[4] Colorado recently implemented a program to counselor low-income youth in hopes of increasing their employment chances. (Gonser, 2018).

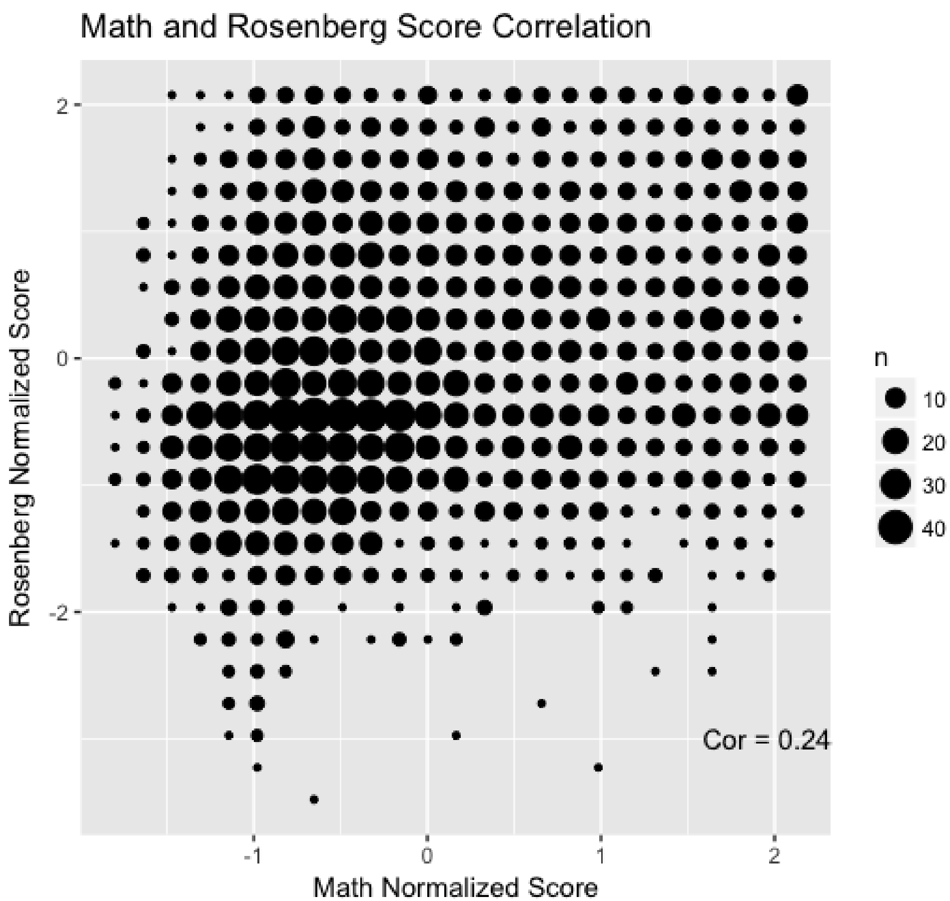

Figure 2: Math and Rosenberg Skills

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.

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. A natural intermediate ques-

	Dependent variable:			
	Income			
	(1)	(2)	(3)	(4)
Math (Normalized)	10,093***		9,918***	
	(523)		(545)	
Rosenberg (Normalized)	3,814***		3,857***	
	(521)		(505)	
Math $\times$ 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 $\times$ 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 (4) 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

tion is where the parameter estimates tell us on the scale of the different models. The way to connect the two extreme models is to specify that returns to skill A are

$$W_A = p \cdot W_{BA} \tag{5}$$

where $p \in [0, 1]$ and $p = 0$ is the pure skill ladder model whereas $p = 1$ is the pure skill independence model. The goal is to obtain an estimate for p. I utilize the plug-in or analogy principle. In this case re-arranging (5) and writing in estimate form, a reasonable estimator of p given estimates for W_A and W_{BA} is

$$\hat{p} = \hat{W}_A / \hat{W}_{BA}. \tag{6}$$

I form an estimate for $\hat{p}$ based 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 p is between 0.6 and 0.65. I can reject a p of less than about 0.45 and more than about 0.75. This suggests that the skill independence model may be a better description of the data in

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

Table 2: Estimates of p

Notes: Estimated using (6) and 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

this specific application, although the parameter p should not be taken as a literal interpretation of the distance between the two models.

5 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.

In this specific instance, the data and model might suggest that interventions putting more emphasis on mathematics skills as opposed to self-esteem issues for this population and time period would be beneficial. However, the model and data makes clear that the optimal policy is probably not some extreme one in which only one of these skills is specifically targeted, but to target both of these skills together.

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.

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.