

A Simple Model of Learning Styles*

Gervas Huxley[†]
University of Bristol
and CMPO

Mike W. Peacey[‡]
University of Bristol
and University of Bath

[Discussion Paper MAY, 2014]

Abstract

Much of the economic literature on education treats the process of learning as a ‘black box’. While such models have many interesting uses, they are of little use when a college seeks advice about reallocating resources from one input to another (e.g. from lecture hours to seminars). Commenting on such questions requires us to ‘open up’ the black box.

This paper shows what one such model would look like by explicitly modelling how students vary in their ‘learning styles’. We apply this framework to investigate how reforms to higher education (e.g. MOOCs) would affect students with different learning styles.

Keywords: Human Capital, Education Production Function, Learning Style, Independent Learner, MOOC

JEL: I20, I23, J24

*We are grateful to Ken Binmore, Nigel Duck, Bill Geoff, Paul Grout, Maija Halonen-Akatwijuka, Geraint Johnes, Ian Jewitt, Alastair Muriel, In-Uck Park, Helen Simpson, Alasdair Smith and Anna Vignoles for comments on earlier drafts. Earlier versions of this paper were presented at the XXII Meeting of AEDE in Spain and the 2013 EEA conference in Sweden, and we are grateful for comments from participants at both. This research was supported by an Economics and Social Research Council UK PhD studentship to Peacey

[†]E-mail: gervas.huxley@bristol.ac.uk

[‡]E-mail: m.peacey@bath.ac.uk

[In a production function with more than one input] “The i^{th} person has more ability if $f_i > f_j$... If sometimes $f_i > f_j$ and sometimes $f_j > f_i$, there is no unique ranking of their abilities.”

- Gary Becker (1975)¹

“In analyzing ability, society needs to recognize its multiple facets.”

- James Heckman (2010)

“Here is a bet and a hope that the next quarter century will see more change in higher education than the last three combined.”

- Lawrence Summers (2012)

1 Introduction

This paper is about how individuals invest in human capital and about the nature of the choices we face when we make these investments. Much of the economic literature on education treats the actual process of learning as a ‘black box’. The inputs provided by schools, colleges or parents (teachers, study materials, reading at home, etc.) are often lumped together into a single input (‘effort’, ‘investment’ or ‘expenditure’), which is assumed to affect student outcomes. Meanwhile students vary according to a single parameter (usually called ‘ability’) which affects their responsiveness to the learning input from educators.

While these ‘black box’ models have many interesting uses, they are of little help when a school or college seeks advice about reallocating resources from one input to another (e.g. from lecture hours to seminars, or from individual study time to group activities). Commenting on such questions requires us to ‘open up’ the black box capturing the variety of inputs available to educators, and also the heterogeneity of students in the ways that they respond to different inputs.

¹Footnote 107, page 110

There is a large educational literature on pedagogy and a closely related literature on learning that explores the relationship between how teachers teach and how students learn. We follow this literature by emphasizing the distinction between teaching style, which refers to the mix of inputs used in the classroom, and learning style which refers to how students respond to those inputs (see Bransford et al. 1999).

An obvious starting point for our model is Gardner's theory of multiple intelligences (Gardner (1993)), where intelligence is defined in terms of learning style. The best known version is the VAK model, in which learners are visual, auditory or kinaesthetic (Fleming (1995)). The implication of the model is that, in the classroom, teachers should include material suited to each kind of learner (Bransford et al. (1999), part III). This approach has been severely criticized: firstly, because the strong version of the theory denies any role for ability; secondly, because it is not supported by the evidence (Pashler et al. (2008)).

We do not believe these ideas provide a satisfactory motivation for a theory of learning styles. However it remains the case that educationists and economists have increasingly come to understand that students are heterogeneous across a variety of personality and non-cognitive dimensions. In the Roy model (Roy (1951)) heterogeneous individuals choose between occupations based on comparative advantage. In the Heckman extension of this model (Heckman, Stixrud, and Urzua (2006)) the cost of completing tasks depends on personality traits. These tasks provide different economic benefits, and individuals choose the tasks in which they have a comparative advantage. If tasks could be achieved through multiple methods, one could imagine the cost of each method depending on personality. The analogous result would then be that individuals choose the means best suited to their personality (i.e. cost minimization).

We consider a two input production function where students face meaningful choices about how they produce a given level of education. The result is that in our model students are heterogeneous, not just in ability but also in what we will refer to as learning style (Bransford et al. (1999), part II). In these circumstances the attempt to develop pedagogic strategies that achieve

the best match between teacher and student makes sense.

The approach that comes closest to the one we adopt in this paper emphasizes the importance of ‘personalized learning’ (Candy (1991)). Personalized learning is a centerpiece goal of the United States 2010 National Education Plan (Group et al. (2010)).

All of the literature mentioned so far either explicitly or implicitly distinguishes between studying and being taught as mechanisms for acquiring education. We follow this approach by viewing investment in human capital as involving a choice between self study and tuition. As well as modeling the impact of each input separately, our approach emphasizes that the interactions between them will be crucial. The education literature usually ignores problems arising from scarcity, however the choice of learning technology will have implications for, and be influenced by, the allocation and pricing of scarce educational inputs.

Calculating the cost minimizing bundle requires knowledge of input prices and the underlying production function. In the last analysis, it is students and not colleges who must optimize. In our model, we assume perfectly informed, fully rational students who can effortlessly solve this problem and we explore the comparative statics of the solution. Even with these very strong assumptions we show that cost minimization cannot be taken for granted.

Technology can affect the cost of existing pedagogy² or result in novel pedagogies, and thus change the way students cost minimize. For example, Powerpoint might increase lecture quality and therefore the productivity of tuition. Whilst e-learning, which promotes user generated content (e.g, blogs or wikis), might increase the productivity of study. Thus technological change can be study-augmenting, tuition-augmenting, or both. The introduction of dictaphones in the 1980s (and more recently podcasts), which permitted students to review lectures easily, has increased the direct productivity of both study and tuition and the complementarity between the two.

We apply our model to the current debate about the future of higher education. Although millions of students have enrolled on MOOCs, the

²See Section 4.

full significance of massive open online courses (MOOCs) remains uncertain (Waldrop (2013)). It has been suggested that MOOCs could replace some traditional HE degrees, particularly if accreditation by MOOCs comes to be seen as comparable to that offered on a traditional degree program (Friedman (2012), Barber et al. (2013)). In this paper we assume that MOOCs offer an alternative to the traditional model and investigate the consequences for different kinds of learners.

As different pedagogic techniques are increasingly being subjected to rigorous, controlled trials, and the quantitative literature regarding different educational inputs (class sizes, teacher quality, etc.) continues to grow, we provide a framework for integrating these findings into a broader model of teaching and learning. Such a framework will hopefully lay the groundwork for increasingly fruitful interaction between developmental psychologists and educational economists, helping to translate findings from psychological and pedagogical research more directly into concrete policy developments.

This paper has three key results. Firstly, for students of a given ability, the graduate premium will depend on learning style. Heterogeneity in learning styles results in different benefits from a given amount of tuition. Secondly, the possibility that MOOCs will replace some traditional institutions depends on the distribution of learning styles within the population. Thirdly, the current ‘one size fits all’ model is inefficient and we characterize the welfare gains that would arise from unbundling.

1.1 Economic Literature

In Becker (1962) ability is a synonym for productivity. Since there is only one input, different means of investing in human capital are not considered. This ignores the possibility that, for example, students might make large investments in human capital without significant expenditure on education.

In Becker (1965) individuals allocate time between two goods: leisure and work. By combining Becker’s work on human capital and time allocation (Becker (1962), Becker (1965)) Grossman (1972) was able to model investment in health human capital. In his model, time is allocated between three

goods: leisure, work and time spent investing in health (e.g. exercise³). Bidle and Hamermesh (1990) also allocate time between three goods: leisure, work and time spent investing in sleep.

In our model, since tuition is paid for through earnings, which involve an investment of time, the choice is ultimately one about the allocation of time between three goods: self study, tuition and work. Our paper also focuses on how various pricing structures affect students, and how students' decisions and attainment depend on learning style.

In addition to the literature on human capital, the work which comes closest to ours is on education production functions. We view this literature as motivated largely by econometric considerations (for early and more recent surveys are Hanushek (1979) and Machin and Vignoles (2005)). This work measures the impact of a variety of inputs such as class size, peer effects and teacher quality. From our perspective these relate to teaching as opposed to learning (see Section 4).

A more recent literature, which focuses on early childhood development, seeks to understand the role of parental characteristics and home environment in developing the cognitive skills of young children (e.g. Todd and Wolpin (2003)). In our model these variables would be parameters not inputs. Using just study and tuition, we create a micro-foundation for different learning styles by modeling a learning production function.

The 'technology of skill formation' provides a theoretical foundation for the early childhood literature (Cunha and Heckman (2007), Cunha, Heckman, and Schennach (2010)). In these production functions individuals must decide on investments at different stages of childhood. Although educational investments remain one dimensional, each intertemporal investment is an input in the education production function. Heckman is interested in how these inputs determine educational attainment and how the ratio used changes with respect to interest rates.

We are interested in the degree of complementarity between study and

³The comparative statics we are interested in are not explored by Grossman. For example, Grossman does not explore how a given health status can be achieved using different input combinations nor does he address the influence of price on the choice of inputs.

tuition, just as Heckman is interested in the degree of complementarity between investments made at different stages of the life cycle. In this work, early childhood investments enhance non-cognitive ability and this increases the returns to investments made at later stages of the life cycle. This emphasis on non-cognitive ability has resulted in an interest in the psychology of personality formation and the relationship between personality, skill formation, and occupational choice (Almlund et al. (2011)). We are interested in how differences in personality might influence how students learn.

2 The Model

An individual i lives for N periods⁴. In each period she will make choices which maximise her lifetime utility. Her lifetime utility, U , is increasing in consumption of a numéraire good, m_t :

$$U = \sum_{t=1}^N u(m_t) \quad (1)$$

She faces choices about investment in education. In any period, t , it is possible to obtain e_t education in a number of ways according to her education production function:

$$e_{i,t}(S_t, \tau(T_t)) \quad (2)$$

where S_t is time spent on study in period t , T_t is time spent receiving tuition in period t and τ is an intensity adjusted measure of tuition (see section 4). Education is strictly increasing in both S_t and T_t .

$$\frac{\partial e_{i,t}}{\partial S_t} > 0, \quad \frac{\partial e_{i,t}}{\partial T_t} > 0 \quad (3)$$

In period t it is possible to obtain a “level t diploma” subject to:

- (1) meeting the prerequisite (having a level $t-1$ diploma).

⁴Discounting of utility, depreciation of education and the role of interest rates are not of interest in our model. Hence we do not include these in Equations 1, 4 and 6.

(2) attaining a level of education of at least e_t^* in this period⁵.

Education is cumulative and as such i 's stock of education in period t is given by:

$$E_{i,t} = \sum_{s=1}^t \mathbb{1}(e_{i,s}^*) \quad (4)$$

where $\mathbb{1}(e_s^*)$ denotes if a diploma was acquired in period s .

The wage, w_t , is an increasing function of this education stock:

$$w_t = W(E_{t-1}) \quad (5)$$

The individual's decision problem is to choose a stream of education and work to maximize consumption. She is subject to an intertemporal money constraint (6), and N intratemporal time constraints (7). The first constraint implies that the lifetime value of the stream of expenditure (on tuition or consumption) cannot exceed the lifetime value of the stream of earned income:

$$\sum_{t=1}^N (w_t H_t - m_t - p T_t) = 0 \quad (6)$$

where H_t is hours worked in period t and p is the price of tuition. p is assumed an exogenous positive constant. The second set of constraints implies the amount of time devoted to each activity must add up to the endowment of time, Ω , in each period:

$$\Omega = S_t + T_t + H_t \quad t = 1 \dots N \quad (7)$$

We define the ‘‘Graduate Premium’’ as the net increase in consumption achieved through education: the difference between consumption available as a graduate minus the consumption that could have been achieved without education⁶.

⁵We assume that a single diploma cannot be obtained over multiple periods.

⁶The literature on returns to education considers tuition fees relative to the wage differential of graduates. In our model it can be shown that if the total amount of learning time (i.e. $S + T$) is different to the hours worked in a non-graduate job, then estimated returns may be biased. I.e. if a graduate wage is \bar{w} , non-graduate wage is \underline{w} , and tuition

The utility maximization problem involves choosing the *level* of education to acquire (E_t) and *how* it is acquired (S_t, T_t).

2.1 Two Period Model

Rather than looking at how much to invest over the life cycle (Grossman (1972)), we focus on how investments are made during a period. Focusing on one period of investment is equivalent to considering a two period model, since investment only occurs in the first period. Hence, after solving this model, we remove the time period indexing⁷.

$$E_1 = \begin{cases} 1 & \text{if } e_{i,1}(S_1, \tau(T_1)) \geq e_1^* \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

We adopt the following notation for the wage function:

$$W(E_1) = \begin{cases} \bar{w} & \text{if } E_1 = 1 \\ \underline{w} & \text{if } E_1 = 0 \end{cases} \quad (9)$$

Since the individual only lives for 2 periods, it is clear that no investment will occur in period 2. If the individual chooses $e_1 = e_1^*$ she becomes a “graduate”, and her consumption is:

$$m^{E1} = \Omega \bar{w} + \underline{w}(\Omega - S_1 - T_1) - pT_1 \quad (10)$$

If she chooses $e_1 = 0$ and becomes a “non-graduate” her consumption is:

$$m^{E0} = 2\Omega \underline{w} \quad (11)$$

The graduate premium is $G = (m^{E1} - m^{E0})$. The individual maximizes G fees are pT , the classic approach would suggest $Return = (\bar{W} - \underline{w}) - \underline{w} - pT$ (difference in future earnings minus the forgone wage minus the cost of tuition). In general, the return would be $Return = (\bar{w} - \underline{w}) - \underline{w} + (h - S - T)(\underline{w}) - pT$, where h is the hours worked at the non-graduate job. If $S + T > h$ the student loses leisure time and if $S + T < h$ the student could work part time whilst at college.

⁷In the Appendix we present the details of an N period model.

subject to:

- (1) The education production function (Equation 8)
- (2) The wage function (Equation 9)
- (3) The time constraint: $\Omega \geq S_1 + T_1$

Clearly this will result in an optimal choice of $e_1 \in \{0, e_1^*\}$. If education is chosen, the optimal inputs can be found by minimizing $c_1(S_1, T_1) = \underline{w}(T_1 + S_1) + pT_1$, subject to the above constraints.

2.2 C.E.S. Education Production Function

We use a CES education production function to express and interpret the differences in learning styles set out in Section 3:

$$e_i(S, \tau) = A(\alpha_i S_i^\rho + (1 - \alpha_i)\tau^{\rho_i})^{\frac{1}{\rho_i}} \quad (12)$$

where τ denotes the value of $\tau(T)$, $\rho \leq 1$, $0 \leq \alpha \leq 1$ and $A = 1$.

Using this function, an individual's learning style is stationary and independent of prices, education level and wealth⁸. In the Appendix we discuss how learning styles change during the education lifecycle.

The CES share parameter, α , is the *independence parameter* (see Section 3.1). The *flexibility parameter*, ρ , represents the degree of complementarity between study and tuition in producing education (see section 3.2)⁹. In the CES production function, the parameters are interdependent and must be carefully interpreted (see Temple (2012)).

In a two-input model it is natural to use total factor productivity (A) as a measure of what Becker called ability (see Section 5.1.1). Because we are interested in how differences in learning style affect returns, in our results (Section 5) we hold A constant across individuals, thus giving no individual an unambiguous advantage in producing E .

⁸The optimal ratio of inputs used will be independent of education, but will depend on price and learning styles.

⁹The elasticity of substitution is related by $\sigma = \frac{1}{1-\rho}$.

2.3 The Wage Function

Human capital is the only variable in our wage function, we ignore the influence of all other variables that might determine income. This specification has two important implications:

- (A1) The wage function is independent of ability;
- (A2) The wage function is independent of the mix of inputs and the individual's learning style.

A1 assumes that ability only affects the wage indirectly via its influence on educational attainment¹⁰. Becker (1962) does not make this assumption, at least in principal, he allows the wage to depend on both ability and education¹¹.

A2 becomes relevant in a two input model. Employers have no preferences about how human capital is acquired. For example how an individual learns a language does not matter, what matters is whether she can speak it. For this reason, we assume that study and tuition affect the wage only through education. Hence choices about study and tuition will depend on both the wage elasticity of education and an individual's learning style.

In neoclassical theories of production, consumers have no preferences over how goods are produced. There are good reasons to believe that this assumption may not hold in the case of education¹². If employers want workers to replenish their human capital on the job they will seek out 'life long learners', in this case employers might have preferences over both the level of education and the learning style¹³.

¹⁰It is of course true that in equilibrium there will be a relationship between ability and wages.

¹¹It turns out that this assumption also has implications for the discussion of ability. See Appendix.

¹²In this case the mix of inputs affects the wage through both the indirect effect on education and the direct effect on productivity.

¹³Students might want to signal ability and learning style. We briefly consider this in the Appendix.

3 Learning Styles and Education Production Functions

In our model differences in learning style correspond to different education production functions. We propose that the fundamental distinction can be modeled in terms of two variables: study and tuition. Although the resources suitable for different learning styles vary, we suggest that many of these differences can be captured in two dimensions: time and money.

A two-input model provides a very different rationale for the belief that one learning style is not necessarily better than another (Riener and Willingham (2010)). Since the optimal mix of inputs is determined by both prices and learning style, ability is not independent of the vector of prices (see Section 5.1.1).

We define two aspects of learning style: *independence* and *flexibility*. Independence is measured by the ratio of tuition to study. The degree of flexibility refers to the extent to which learners can substitute the two inputs.

Since higher education typically bundles study and tuition, the relevance of these learning style parameters is usually ignored. We view the introduction of MOOCs as representing a change in technology that unbundles these inputs and this will have implications for demand.

3.1 Independent Learner

Independence measures the weight given to each input in the production function. For independent learners, the output elasticity for study will be higher than for tuition.

The marginal rate of technical substitution measures the trade-off between study and tuition:

$$MRTS_{(S,T)} = \frac{\frac{\partial e}{\partial S}}{\frac{\partial e}{\partial T}} \quad (13)$$

Independent learners will buy a less tuition-intensive bundle than their

less independent counterparts (whom we will refer to as directed learners). Although independent learners will choose to make most of their human capital investments in the form of study, these students need not have higher returns than directed learners. Total investments will not necessarily be greater for one type of learner than another. If low-return students are directed learners, the relationship between ability and expenditure set out in the literature would be reversed.

3.2 Flexible Learner

Flexibility is a measure of how a student can adapt to different combinations of the two inputs. This implies that we are interested in whether study and tuition are substitutes or complements. In the standard model, it is implicit that study and tuition are perfect complements. We believe that this assumption is inconsistent with both the traditional pedagogic literature and the newer cognitive-science literature (Bransford et al. (1999), part II, Ambrose et al. (2010)).

The elasticity of substitution (Equation 14) measures how much individuals can substitute inputs conditional on how much of each input they are already using.

$$\sigma_{(S,T)} = \frac{d \ln(\frac{S}{T})}{d \ln(MRTS_{T,S})} \quad (14)$$

If, as seems likely, investments in study and tuition are complementary then high investment in study will increase the return to investment in tuition and vice versa. Almost everybody would agree that if you do not study there is little point in investing in tuition. However this statement can be reversed; if you work hard it will be worthwhile investing in tuition.

Consider a parallel with sport: talented athletes make large investments in exercise (study). However, precisely because they are so talented it is also worth their while to purchase large amounts of training (tuition). It is not uncommon for world class athletes to employ a full-time trainer.

In our model this translates into the statement that provided study and tuition are complementary the higher your ability *and* the more time you

spend studying, the higher should be your expenditure on tuition.

4 The Tuition Production Function

Tuition is dissemination of knowledge by a teacher and is an input in our education production function. The input T will always correspond to one hour of tuition, however this tuition may vary in quality: a student in a one hour class of five students is likely to obtain more from this hour than if she were in a larger class¹⁴. We use the function $\tau(T)$ to adjust tuition time, T , to account for quality. In addition to class size τ will depend on parameters such as content, classroom resources and teacher effects (Vignoles et al. (2000) and Hanushek and Rivkin (2006)).

Decisions about quality will depend on how it varies with demand (which will depend on learning style - see Section 3) and costs. We illustrate this idea with class size.

We distinguish between material that is either ‘core’ or ‘discursive’. With core material class size has little impact on quality. In contrast, discursive material is harder to provide in large classes. A MOOC requires students to follow a syllabus, view videos and participate in online forums and thus delivers core material. However the delivery of discursive material in this way is much less effective, and therefore the amount of this type of tuition provided by a MOOC is strictly limited (Bowen (2012)).

We model traditional HE as maintaining quality when discursive material is taught ($\tau(T) = T$). This is possible because a traditional university provides this type of material in small group seminars. Our model of traditional HE is one in which a generous amount of tuition is available for a fixed fee (see Section 5.2).

We believe that with current technology, one hour of online interaction with peers is lower in quality than a traditional one hour small group tutorial with an experienced professor¹⁵. To simplify we assume that MOOCs can

¹⁴In principal almost all study requires at least some tuition. A learner who uses a text book is receiving a small amount of tuition.

¹⁵This is disputed (see Barber et al. (2013)).

only provide core material, and they do so for free¹⁶. Therefore they can deliver \bar{T}^{core} with the same quality as a traditional HE institution:

$$\tau(T) = \begin{cases} T, & \text{if } T \leq \bar{T}^{core} \\ \bar{T}^{core}, & \text{if } T > \bar{T}^{core} \end{cases} \quad (15)$$

In general, cost minimization will result in colleges teaching large groups for core material and small groups for discursive material¹⁷. Overall, a traditional university has constant returns to scale (Cohn, Rhine, and Santos (1989) and Getz, Siegfried, and Zhang (1991)). The technological change represented by MOOCs implies that core material may be provided, by natural monopolies, to ever larger audiences (Bowen (2012)).

5 Results

This section considers the delivery of higher education. We investigate how an individual’s Graduate Premium is affected by her learning style and the pricing structure, holding ‘ability’ constant¹⁸. We use this framework to address three questions: firstly, who would choose to acquire education through a MOOC compared to traditional HE; secondly, how behavior changes when tuition is unbundled, and how these changes depend on learning style; thirdly, whether classes should be compulsory.

An algebraic solution to our model, showing the optimal inputs (S^* , T^*) which would be chosen under a *variable fee* (see Section 5.3), is given in the Appendix. In our results we make no attempt to quantify the distribution of learning styles present in a population¹⁹.

¹⁶In the future providers of MOOCs may charge. Provided pricing is much lower than traditional HE, our results will still hold.

¹⁷It may also involve providing different class sizes for different learner types (see section 3).

¹⁸See Section 5.1.1.

¹⁹Therefore we cannot address such questions as “What proportion of school leavers progress to HE?”.

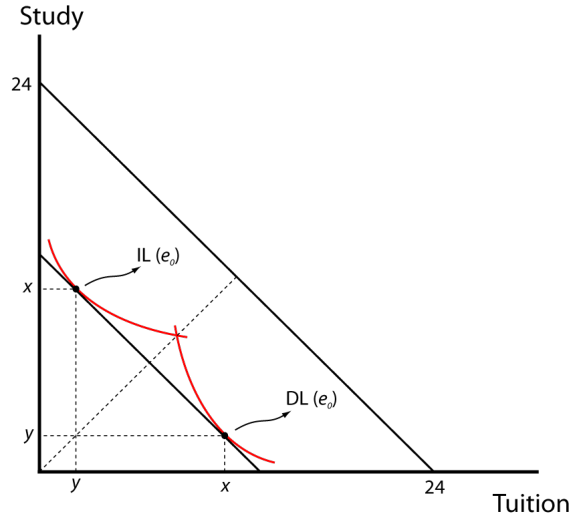


Figure 1: Optimal inputs when $p = 0$

5.1 Graduate Premium and Learning Style

Proposition (1). *Graduate Premium depends on learning style*

In Propositions 1.1 and 1.2 we explain this relationship.

Proposition (1.1). $\frac{\partial GP}{\partial \alpha} > 0$ iff $\frac{w}{w+p} < \frac{\alpha}{1-\alpha}$.

This proposition allows us to state the conditions that must be met in order for independence to unambiguously increase productivity.

First compare symmetric learners: (i.e. $\alpha_i = A$ and $\alpha_j = 1 - A$). From Proposition 1.1 if $A > \frac{1}{2}$, and $p > 0$, i will obtain a higher Graduate Premium than her directed counterpart. Figure 1 shows these symmetric learners when $p = 0$. The independent learner is labeled IL and the directed learner is labeled DL. Symmetry and $p = 0$ imply the individuals use symmetric bundles and obtain the same Graduate Premium.

Now consider the asymmetric case: Consider $\alpha_i = A$ and $A > \frac{1}{2}$. First, consider $\alpha_j < 1 - A$. In this case i is more independent than j is directed. Proposition 1.1 again shows that i will always do better.

However there can be an advantage to being directed. If $\alpha_j > 1 - A$, i will receive a lower Graduate Premium than j if the price of tuition is sufficiently

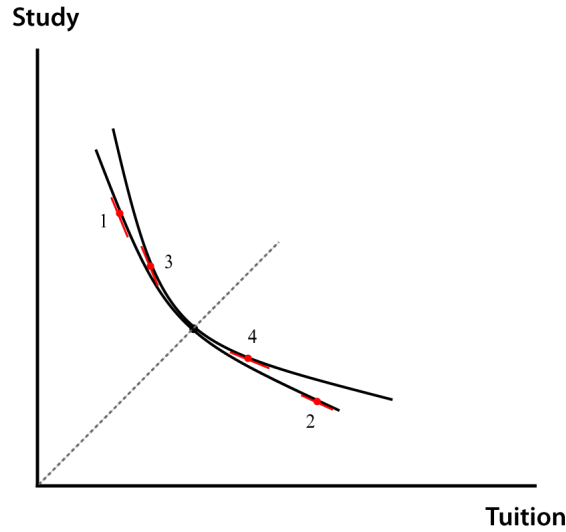


Figure 2: Flexibility and cost

low. In this case the benefit from being a highly-specialized directed learner outweighs the fact that tuition is the more expensive input.

Proposition (1.2). $\frac{\partial GP}{\partial \rho} > 0$.

Learners with high ρ are at an advantage. Figure 2 shows isoquants of two learners who produce the same e . For both high (points 1 and 3) and low (points 2 and 4) prices, the less flexible learner faces a higher cost. Illustrating the result that as complementarity increases the cost of producing a given level of e increases.

Proposition (2). *Increases in p reduce the Graduate Premium for all learners (i.e. $\frac{\partial c}{\partial p} > 0$).*

To illustrate the importance of learning styles in determining the Graduate Premium consider the case where, apart from opportunity cost, education is free and unlimited ($p = 0$). In this case, we might expect all individuals to obtain the same Graduate Premium since they have the same total factor productivity²⁰. However individuals with different learning styles use inputs

²⁰See Section 5.1.1.

differently and therefore have different opportunity costs²¹.

When students pay for tuition, their choice of S and T will change and this has differential effects on the Graduate Premium. In Propositions 2.1 and 2.2 we explain how these effects depend on learning style.

Proposition (2.1). $\frac{\partial^2 c}{\partial p \partial \alpha} < 0$

Independence mitigates the reduction in the Graduate Premium. The individuals who lose most when price increases are directed learners because of their heavy reliance on tuition.

Proposition (2.2). $\frac{\partial^2 c}{\partial p \partial \rho} < 0$ iff $\frac{w}{w+p} < \frac{\alpha}{1-\alpha}$.

In general, flexibility mitigates the reduction in the Graduate Premium²². This is because higher complementarity between S and T means that when price increases, substitution away from T is increasingly expensive. This can be seen in Figure 2. At a high price of tuition (points 1 and 3) the flexible learner is more S-intensive, whereas at a lower price of tuition (points 2 and 4) the flexible learner is more T-intensive.

5.1.1 Ability

Ability is defined by the ratio of output to input. When considering more than one input this definition is no longer straightforward. Inputs must be aggregated and thus weighted by price. Because we allow marginal returns for each input to vary between individuals the focus shifts to relative price, and who can make better use of the cheaper input.

Definition (Ability) Where two individuals choose to produce e^* , if individual i has a lower cost than j , then i will have a higher level of utility and is said to be of higher ability.

$$e_i(T^{*i}, S^{*i}) = e_j(T^{*j}, S^{*j}) \quad (16)$$

²¹Individuals only obtain the same Graduate Premium if they are all forced to use bundles which are symmetric. This is the only case where Graduate Premium is independent of learning style.

²²However if $\frac{w}{w+p} < \frac{\alpha}{1-\alpha}$, this result will reverse (e.g. if a learner continued to acquire education solely from tuition).

and

$$c(T^{*i}, S^{*i}) < c(T^{*j}, S^{*j}) \quad (17)$$

This is equivalent to saying that if the cost of e^* is fixed for the two individuals, then the individual with the highest ability will acquire more education. We believe this definition of ability is intuitive and more consistent with the pedagogical literature²³.

Corollary (1). *Ability depends on price*

Since Ability is defined by cost, Proposition 2 describes this dependence. This does not imply ability is meaningless - as Becker understood²⁴. We refer to an ability ranking that is independent of prices as a “Becker ability ranking”. Figure 3 illustrates this result. At $p = 0$, IL and DL have the same ability. At a higher price DL is lower ability than IL because the increase in cost for DL exceeds the increase in cost for IL.

In this respect Heckman’s 2 period, 1 input model is similar to our 1 period, 2 input model. It follows that the same ambiguity must arise. The cost of investment I_0 is the opportunity cost of investment in the next period, $(1+r)I_1$. If the interest rate changes, the relative price of the two investments will change. The higher ability individual is now whoever makes better use of the cheaper investment²⁵.

5.2 Traditional HE vs MOOCs

Result (1). *To benefit from a MOOC students must be sufficiently Independent and Flexible.*

To illustrate this result we simulated 1,000,000 individuals, with the same level of Becker ability but heterogeneous learning styles. We compare the

²³In general ability will also depend on e^* . Thus ability rankings can change through the education life cycle. The highest ability 10 year old may no longer be the highest by the time she enrolls at university. In a CES production function (Section 2.2) this problem does not arise.

²⁴See Becker (1975), page 110 footnote 107.

²⁵In the Appendix we show that this problem arises even in a one-input model.

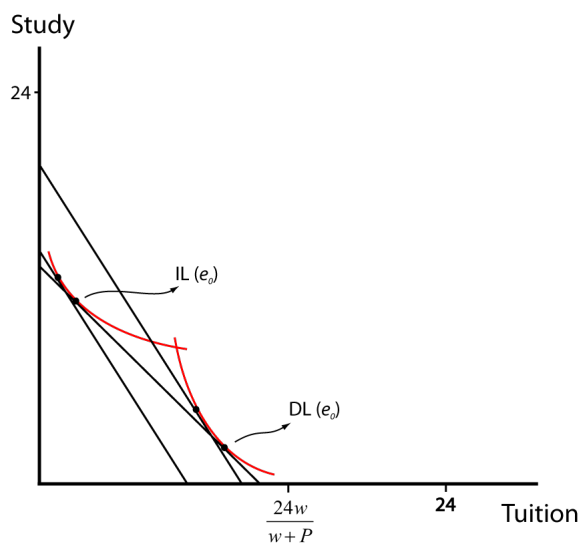


Figure 3: The impact of learning style on cost minimization

Graduate Premium under traditional HE and MOOCs. Individuals maximize their Graduate Premium by choosing the system which best suits their learning style (Figure 4). In this figure different points represent individuals with specific learning styles, with independence and flexibility measured on the vertical and horizontal axes respectively. The diagram plots all individuals who obtain a positive graduate premium, individuals who choose MOOCs colored blue and individuals who choose traditional HE are colored red²⁶.

The MOOC allows some students who would attend a traditional university to complete their education for a lower cost and hence receive a larger Graduate Premium. In addition some students, discouraged by the high cost of traditional university, now find acquiring education worthwhile. The students who choose to attend a MOOC tend to be both independent and flexible (Figure 4, top right). These learners receive little benefit from the face to face contact provided in small groups by traditional universities. However, there will also be some directed, flexible learners who find the price reduction justifies the extra study required by a MOOC (shown in Figure 4 by the blue

²⁶We generate 1,000,000 individuals with characteristics distributed in $\alpha \in (0, 1) \times \rho \in (-10, 1)$. Inputs are measured in units of 0.25 hours.

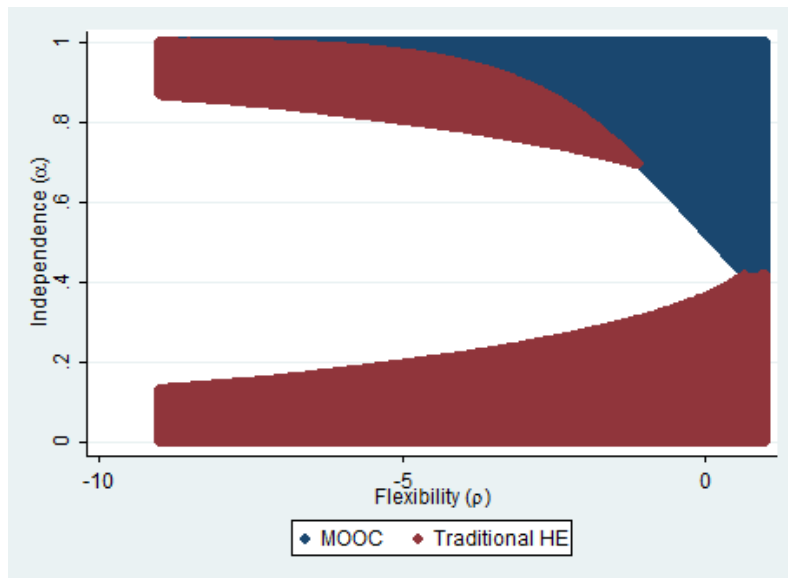


Figure 4: MOOC vs Traditional HE

area continuing below $\alpha = 0.5$).

Independent inflexible learners will choose traditional universities (Figure 4, top left). For these students, study and tuition are complementary. Therefore these students benefit from the more intensive contact offered by a traditional university. Finally, directed learners choose traditional universities (Figure 4, bottom). These learners require more tuition than can be provided by a MOOC because the returns to study are too low.

The recent development of MOOCs represents a radical change in the proportions in which S and T are bundled. We believe that the potential for MOOCs to transform HE depends in part on the importance of learning style, and also the extent to which current arrangements provide an inefficient bundle.

5.3 Unbundling (Fixed vs Variable Fees)

The HE model we have adopted is one in which the university charges a fixed fee and provides optional lectures and classes. This means study and tuition are implicitly bundled (Wang (1975), Adams and Yellen (1976), Nor-

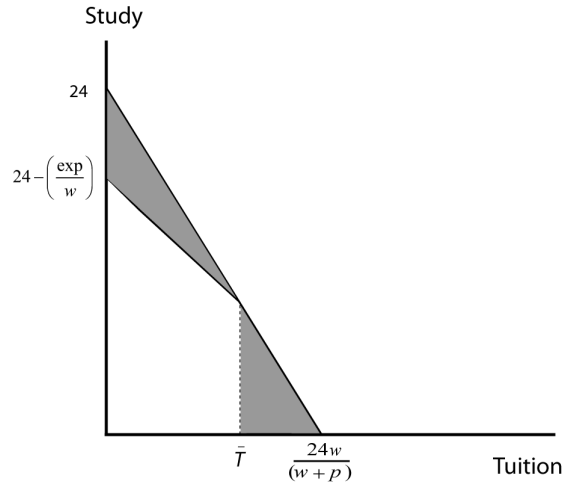


Figure 5: Unbundling relaxes the constraint

ton (2013)). In this section we attempt to investigate the welfare implications of unbundling on students with different learning styles.

In practice students are almost never offered a variable fee structure, rather they must choose S given T . If learning styles are heterogeneous at least some students must be purchasing a sub-optimal bundle. Unbundling would allow students to purchase their desired quantity of T , resulting in hourly paid tuition.

To compare the welfare effects of variable fees relative to fixed fees, we set the price of tuition equal to $p = \frac{\text{fixed fee}}{T}$. This means any bundle available under the fixed fee is available under the variable fee²⁷. Relaxing the constraint in this way ensures that all learners will gain from unbundling (see Figure 5).

Result (2). *Unbundling benefits everyone, with the most distorted learners benefiting most.*

Unbundling gives rise to both price and wealth effects. These depend on the student's initial bundle, how distorted she was, and her learning style.

²⁷For a university this change is unlikely to be resource neutral. Unbundling will change the demand for tuition, and therefore affect revenue.

Learners who choose zero tuition will not change their behavior when price increases, and only benefit from a wealth effect. Students who choose an interior solution with a fixed fee will change their behavior when unbundling occurs.

The most distorted learners will change their behavior the most and these changes can operate via both learning style parameters. Flexible learners initially located at the kink will experience large gains. Directed learners will also purchase more tuition after unbundling. Replacing x hours of study (opportunity cost wx) with y hours of tuition (cost $y(w+p)$), will generate a benefit of $y(w+p)wx$. Independent learners value the tuition they currently receive $\in (w, w+p)$. Thus when the price rises, they substitute away from tuition. How much study is required for this substitution will determine the size of the gain.

5.4 Optional vs Compulsory Class Attendance

In this section we expand on our tradition HE model by investigating the implications of compulsory class attendance. Thus rather than students choosing $T \leq \bar{T}$ we set $T = \bar{T}$, with the same fixed fee paid. We discuss the empirical work on this question (starting with Romer (1993)) in the Appendix.

Result (3). *Compulsory attendance reduces the Graduate Premium for all students*²⁸.

With compulsory attendance, the Graduate Premium is now determined by the minimum number of hours (of S) needed to reach the required education level and therefore what matters is the opportunity cost of the compulsory class.

The reduction in the Graduate Premium is greatest for independent learners, as for these students the ‘unnecessary’ tuition imposes the greatest cost. Figure 6 illustrates these asymmetric effects. With optional classes (a) IL

²⁸This result is especially dependent on the strong assumption that students are rational and perfectly informed. In future work we aim to build a richer model with students who, for example, face self-control problems.

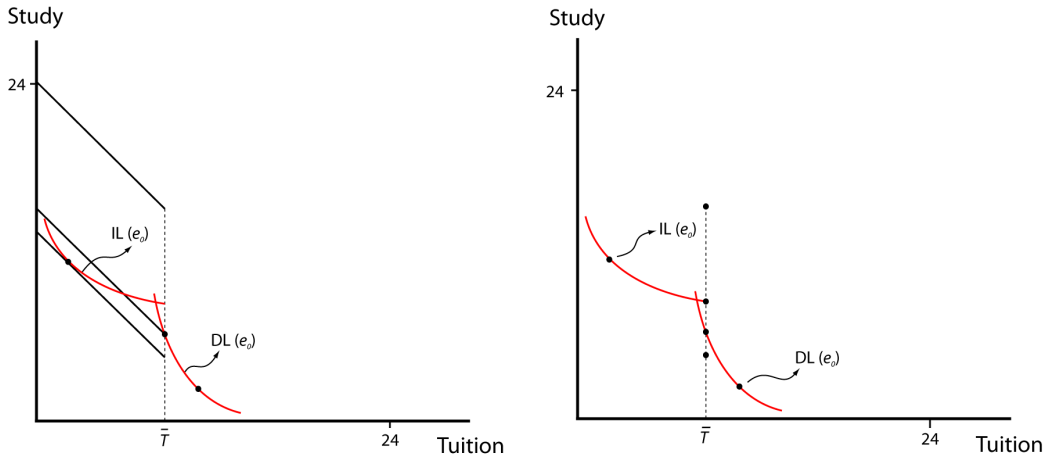


Figure 6: Optimal inputs under fixed fee with (a) optional classes and (b) compulsory classes

has a higher graduate premium than DL. Compulsory classes (b) increase the cost of achieving e^* for both learners. However, the increase in cost to IL exceeds the increase to DL and therefore DL now has a higher graduate premium.

These effects are ameliorated by flexibility - when tuition is compulsory the benefits of being flexible outweigh the advantages that can accrue from complementarity. Inflexible learners cannot use the ‘extra’ tuition to substitute for study, whilst flexible learners can do so.

5.5 Summary of Results

These results point to a number of conclusions. The graduate premium will depend on how higher education is delivered and the heterogeneity of the population in terms of learning styles. MOOCs have the potential to provide large gains - but only to learners who are independent and flexible. More generally, unbundling the inputs will benefit all learners.

Combining MOOCs with unbundling would provide the most efficient outcome - the delivery of core material via a MOOC and discursive material via a more flexible higher education sector. Thus different kinds of learners

can choose the bundle of study and tuition that best suits their individual learning style, whilst benefiting from the economies of scale provided by a MOOC.

6 Conclusion

How teachers teach and students learn have consequences for the efficient allocation of scarce resources. A two input model with tuition and study captures critical differences in teaching and learning discussed in the education literature.

Because the current economic literature has yet to explore this distinction, important efficiency considerations are ignored. We have shown that, even if students have the same Becker ability, the graduate premium depends on learning style and the mix of inputs matters for efficiency (Section 5.1). For a given cost structure tuition may be over or under supplied. If learning styles are heterogeneous it becomes important to ensure that resources are allocated in accordance with individual needs. In practice education markets offer students very little choice about the amount of tuition to purchase, rather they must choose how much to study given a fixed amount of tuition.

These results hold because we specify a price for each type of investment: The price of study comes in the form of an opportunity cost whereas the price students pay for tuition is a market price. Independent learners should pay less for their education even if they are high ability and make large overall investments in human capital.

If high ability students are more independent than lower ability students it is possible that those colleges that recruit the highest ability students should charge lower fees than colleges that recruit lower ability students, at least to achieve a given level of education. If study and tuition are complementary the paradox that prestigious colleges should charge less than lower ranked colleges is reversed. Recent evidence suggests that there is considerable complementarity between study and tuition²⁹.

²⁹Bandiera, Larcinese, and Rasul (2010) presents evidence that shows, in an unnamed research-intensive university, class size is important - particularly for high ability students.

We have used our model to investigate how changes in technology and the introduction of MOOCs will change the way higher education is delivered, the types of learners who participate, and the Graduate Premium (Section 5.2).

We have shown that the importance of these changes depends on how much students vary in their learning styles. We propose a distinction between core and discursive material, and predict that the future of traditional HE institutions will depend on their ability to deliver discursive material.

We have used our model to show how unbundling the core and discursive dimensions of tuition increases the potential for exploiting the scale economies and cost reduction potential of MOOCs (Section 5.3). If this is combined with ‘private tuition’ different kinds of learners can choose the bundle of study and tuition that best suits their learning style.

References

- Adams, William James and Janet L Yellen. 1976. “Commodity bundling and the burden of monopoly.” *The quarterly journal of economics* :475–498.
- Almlund, Mathilde, Angela Lee Duckworth, James J Heckman, and Tim D Kautz. 2011. “Personality Psychology and Economics.” *NBER working papers series* 2011 (online):w16822–w16822.
- Ambrose, Susan A, Michael W Bridges, Michele DiPietro, Marsha C Lovett, and Marie K Norman. 2010. *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons.
- Arrow, Kenneth J. 1962. “The economic implications of learning by doing.” *The review of economic studies* 29 (3):155–173.
- Bandiera, Oriana, Valentino Larcinese, and Imran Rasul. 2010. “Heterogeneous Class Size Effects: New Evidence from a Panel of University Students*.” *The Economic Journal* 120 (549):1365–1398.

- Barber, Michael, Katelyn Donnelly, Saad Rizvi, and Lawrence Summers. 2013. "An Avalanche is Coming." *Higher Education and the Revolution Ahead, IPPR2013, Pearson Publications* .
- Becker, Gary S. 1962. "Investment in human capital: A theoretical analysis." *The journal of political economy* 70 (5):9–49.
- . 1965. "A Theory of the Allocation of Time." *The economic journal* 75 (299):493–517.
- . 1975. *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.
- Biddle, Jeff E and Daniel S Hamermesh. 1990. "Sleep and the Allocation of Time." *Journal of Political Economy* :922–943.
- Bowen, William G. 2012. "The cost disease in higher education: is technology the answer?" *The Tanner Lectures, Stanford University (October 2012)* .
- Bransford, John D, Ann L Brown, Rodney R Cocking et al. 1999. *How people learn*.
- Bratti, Massimiliano and Stefano Staffolani. 2013. "Student time allocation and educational production functions." *Annals of Economics and Statistics* 111:103–140.
- Candy, Philip C. 1991. *Self-Direction for Lifelong Learning. A Comprehensive Guide to Theory and Practice*. ERIC.
- Cohn, Elchanan, Sherrie LW Rhine, and Maria C Santos. 1989. "Institutions of higher education as multi-product firms: Economies of scale and scope." *The Review of Economics and Statistics* :284–290.
- Correspondence. 1994. "Correspondence: Should Class Attendance Be Mandatory?" *Journal of Economic Perspectives* 8 (3):205–215. URL <http://www.aeaweb.org/articles.php?doi=10.1257/jep.8.3.205>.

- Cunha, Flavio and James Heckman. 2007. “The Technology of Skill Formation.” *The American Economic Review* 97 (2):31–47.
- Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. “Estimating the technology of cognitive and noncognitive skill formation.” *Econometrica* 78 (3):883–931.
- Fleming, Neil D. 1995. “I’m different; not dumb. Modes of presentation (VARK) in the tertiary classroom.” In *Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA)*, HERDSA, vol. 18. 308–313.
- Friedman, Thomas. 2012. “Come the revolution.” *The New York Times* 15.
- Gardner, H. 1993. *Frames of mind: the theory of multiple intelligences*. BasicBooks. URL http://books.google.co.uk/books?id=_vLmG9qER0gC.
- Getz, Malcolm, John J Siegfried, and Hao Zhang. 1991. “Estimating economies of scale in higher education.” *Economics Letters* 37 (2):203–208.
- Grossman, Michael. 1972. “On the concept of health capital and the demand for health.” *The Journal of Political Economy* 80 (2):223–255.
- Group, National Educational Technology Plan Technical Working et al. 2010. “Transforming American education: Learning powered by technology.”
- Hanushek, Eric A. 1979. “Conceptual and empirical issues in the estimation of educational production functions.” *Journal of human Resources* :351–388.
- Hanushek, Eric A and Steven G Rivkin. 2006. “Teacher quality.” *Handbook of the Economics of Education* 2:1051–1078.
- Heckman, James J. 2010. “A Research Agenda for Understanding the Dynamics of Skill Formation.” *American Economics Association White Paper, October* 4.

- Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior.” *Journal of Labor Economics* 24 (3):411–482.
- Klump, Rainer, Peter McAdam, and Alpo Willman. 2012. “The normalized CES production function: theory and empirics.” *Journal of Economic Surveys* 26 (5):769–799.
- Machin, Stephen and Anna Vignoles. 2005. *What’s the Good of Education?: The Economics of Education in the UK*. Princeton University Press.
- Norton, Andrew. 2013. “The Unbundling and Re-Bundling of Higher Education.” *Grattan Institute* .
- Pashler, Harold, Mark McDaniel, Doug Rohrer, and Robert Bjork. 2008. “Learning styles concepts and evidence.” *Psychological science in the public interest* 9 (3):105–119.
- Riener, Cedar and Daniel Willingham. 2010. “The myth of learning styles.” *Change: The magazine of higher learning* 42 (5):32–35.
- Romer, David. 1993. “Do students go to class? Should they?” *The Journal of Economic Perspectives* 7 (3):167–174.
- Roy, Andrew Donald. 1951. “Some thoughts on the distribution of earnings.” *Oxford economic papers* 3 (2):135–146.
- Spence, Michael. 1973. “Job market signaling.” *The quarterly journal of Economics* 87 (3):355–374.
- Summers, Lawrence H. 2012. “What you (really) need to know.” *New York Times* .
- Temple, Jonathan. 2012. “The calibration of CES production functions.” *Journal of Macroeconomics* 34 (2):294–303.
- Thaler, Richard H and Cass R Sunstein. 2003. “Libertarian paternalism.” *The American Economic Review* 93 (2):175–179.

Todd, Petra E and Kenneth I Wolpin. 2003. “On the specification and estimation of the production function for cognitive achievement*.” *The Economic Journal* 113 (485):F3–F33.

Vignoles, Anna, Rosalind Levacic, James Walker, Stephen Machin, and David Reynolds. 2000. *The relationship between resource allocation and pupil attainment: a review*. Centre for the Economics of Education, London School of Economics and Political Science.

Waldrop, MM. 2013. “Campus 2.0.” *Nature* 495 (7440):160–163.

Wang, William KS. 1975. “The unbundling of higher education.” *Duke Law Journal* :53–90.

A Mathematical Appendix

A.1 Algebraic Results

The results of the model are found by solving the following Lagrangian equation:

$$L = wS + (w + P_t)T + \lambda((\alpha S^\rho + (1 - \alpha)T^\rho)^{\frac{1}{\rho}} - e^*) \quad (18)$$

Assuming the individual chooses a skilled job the following first order conditions hold:

$$\frac{\partial L}{\partial S} = w + \lambda \alpha S^{1-\rho} (\alpha S^\rho + (1 - \alpha)T^\rho)^{\frac{1-\rho}{\rho}} \quad (19)$$

$$\frac{\partial L}{\partial T} = (w + p) + \lambda (1 - \alpha) T^{\rho-1} (\alpha S^\rho + (1 - \alpha)T^\rho)^{\frac{1-\rho}{\rho}} \quad (20)$$

$$\frac{\partial L}{\partial \lambda} = (\alpha S^\rho + (1 - \alpha)T^\rho)^{\frac{1}{\rho}} - e^* \quad (21)$$

Solving:

$$S^* = \frac{e^*}{(\alpha + (1 - \alpha) \left(\frac{\alpha(w+p)}{(1-\alpha)w}\right)^{\frac{\rho}{\rho-1}})^{\frac{1}{\rho}}} \quad (22)$$

$$T^* = \frac{e^* \left(\frac{\alpha(w+p)}{(1-\alpha)w}\right)^{\frac{\rho}{\rho-1}}}{(\alpha + (1 - \alpha) \left(\frac{\alpha(w+p)}{(1-\alpha)w}\right)^{\frac{\rho}{\rho-1}})^{\frac{1}{\rho}}} \quad (23)$$

It is convenient to discuss these results in terms of a ratio, $\frac{S^*}{T^*}$:

$$R^* = \frac{S^*}{T^*} = \left(\frac{\alpha(w+p)}{(1-\alpha)w} \right)^{\frac{1}{1-\rho}} \quad (24)$$

A.2 Proofs of Propositions

Proposition (1). *Graduate premium depends on learning style*

Proof. The GP works in the opposite direction to the cost of acquiring education. Thus individuals who achieve e^* with lower costs will have higher graduate premiums. The cost of obtaining e^* for individual i is:

$$c_i(w, w+p, e^*) = (\alpha_i^{\sigma_i} w^{1-\sigma_i} + (1-\alpha_i)^{\sigma_i} (w+p)^{1-\sigma_i})^{\frac{1}{1-\sigma_i}} e^* \quad (25)$$

The cost function clearly depends on α and ρ and hence cost, and graduate premium, depend on learning style. □

Proposition (1.1). $\frac{\partial GP}{\partial \alpha} > 0$ iff $\frac{w}{w+p} < \frac{\alpha}{1-\alpha}$.

Proof.

$$\frac{\partial \ln(c_i(w, w+p, e^*))}{\partial \alpha} = \frac{\sigma(-w(w+P)^\sigma \alpha^{\sigma-1} + w^{1+\sigma}(1-\alpha)^{\sigma-1} + w^\sigma(1-\alpha)^{\sigma-1}p)}{(\sigma-1)(\alpha^\sigma w(w+p)^\sigma + (1-\alpha)^\sigma w^{1+\sigma} + (1-\alpha)^\sigma p w^\sigma)} \quad (26)$$

The denominator has the sign of $\sigma-1$, since the rest is clearly positive. Since $\sigma > 0$, the numerator takes the sign of:

$$-w(w+P)^\sigma \alpha^{\sigma-1} + w^{1+\sigma}(1-\alpha)^{\sigma-1} + w^\sigma(1-\alpha)^{\sigma-1}p \quad (27)$$

This is negative iff:

$$w^{\sigma-1}(1-\alpha)^{\sigma-1} < (w+p)^{\sigma-1} \alpha^{\sigma-1} \quad (28)$$

If $\sigma-1 > 0$ then $\frac{w}{(w+p)} < \frac{\alpha}{(1-\alpha)}$. If $\sigma-1 < 0$ then $\frac{w}{(w+p)} > \frac{\alpha}{(1-\alpha)}$. Hence, in either case, $\frac{\partial c_i(w, w+p, e^*)}{\partial \alpha} < 0$ iff $\frac{w}{(w+p)} < \frac{\alpha}{(1-\alpha)}$. □

Proposition (1.2). $\frac{\partial GP}{\partial \rho} > 0$

Proof.

$$\frac{\partial c_i(w, w+p, e^*)}{\partial \sigma} = (\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma)^{\sigma-1} (A+B) \quad (29)$$

where;

$$A = -\frac{\ln(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma)}{\sigma^2} \quad (30)$$

and;

$$B = \frac{\alpha^{1-\sigma} w^\sigma (\ln(w) - \ln(\alpha)) + (1-\alpha)^{1-\sigma} (w+p)^\sigma (\ln(w+p) - \ln(1-\alpha))}{\sigma(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma)} \quad (31)$$

which simplifies to;

$$\frac{\partial c_i(w, w+p, e^*)}{\partial \sigma} = -\frac{QR}{P} \quad (32)$$

where;

$$Q = (\alpha^{1-\sigma} w^\sigma + (w+p)^\sigma (1-\alpha)^{-\sigma} - (w+p)^\sigma (1-\alpha)^{-\sigma} \alpha)^{\sigma-1} \quad (33)$$

$$R = (1-\alpha)^\sigma w^\sigma \alpha (\sigma(\ln(\alpha) - \ln(w)) + \ln(z)) + (1-\alpha)(w+p)^\sigma \alpha^\sigma (\sigma(\ln(1-\alpha) - \ln(w+p)) + \ln(z)) \quad (34)$$

where $z = \alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma$, and;

$$P = \sigma^2 (\alpha w^\sigma (1-\alpha)^\sigma + (w+p)^\sigma \alpha^\sigma - (w+p)^\sigma \alpha^{1+\sigma}) \quad (35)$$

Since, $P > 0$ and $Q > 0$, $\frac{\partial c_i(w, w+p, e^*)}{\partial \sigma}$ clearly has the opposite sign to R . R is always negative if $w > 1$ (verified by computer).

Note the sign of $\frac{\partial c_i(w, w+p, e^*)}{\partial \rho}$ is the same as $\frac{\partial c_i(w, w+p, e^*)}{\partial \sigma}$ since $\sigma = \frac{1}{1-\rho}$ and $\rho \leq 1$.

□

Corollary (1). *Ability depends on price*

Proof. This follows from our definition of ability and Proposition 1. \square

Proposition (2). *Increases in p reduce the GP for all learners (i.e. $\frac{\partial c}{\partial p} > 0$).*

Proof.

$$\frac{\partial \ln(c)}{\partial p} = \frac{(1-\alpha)^\sigma (w+p)^{1-\sigma}}{(w+p)(\alpha^\sigma w^{1-\sigma} + (1-\alpha)^\sigma (w+p)^{1-\sigma})} \quad (36)$$

Which is clearly positive for everyone. \square

Proposition (2.1). $\frac{\partial^2 c}{\partial p \partial \alpha} < 0$

Proof.

$$\frac{\partial \frac{\partial \ln(c)}{\partial p}}{\partial \alpha} = -(A + B) \quad (37)$$

$$\text{Where } A = \frac{(1-\alpha)^{1-\sigma} (1-\sigma) (w+p)^\sigma}{(1-\alpha)(w+p)(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma)}$$

$$\text{and } B = \frac{(\frac{\alpha^{1-\sigma} (1-\sigma) w^\sigma}{\alpha} - \frac{(1-\alpha)^{1-\sigma} (1-\sigma) (w+p)^\sigma}{1-\alpha})(1-\alpha)^{1-\sigma} (w+p)^\sigma}{(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma)^2 (w+p)}$$

Multiplying out, to give a (positive) common denominator of $(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma)^2 (w+p)(1-\alpha)$, gives a numerator:

$$-((1-\alpha)^{1-\sigma} (1-\sigma) (w+p)^\sigma)((\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma)$$

$$-(1-\alpha) \left(\left(\frac{\alpha^{1-\sigma} (1-\sigma) w^\sigma}{\alpha} - \frac{(1-\alpha)^{1-\sigma} (1-\sigma) (w+p)^\sigma}{1-\alpha} \right) (1-\alpha)^{1-\sigma} (w+p)^\sigma \right) \quad (38)$$

Which can be re-written as:

$$-(1-\alpha)^{1-\sigma} (1-\sigma) (w+p)^\sigma (\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma} (w+p)^\sigma + (1-\alpha) \left(\frac{\alpha^{1-\sigma} w^\sigma}{\alpha} - \frac{(w+p)^\sigma}{(1-\alpha)^\sigma} \right)) \quad (39)$$

Which simplifies to:

$$-(1-\alpha)^{1-\sigma} (1-\sigma) (w+p)^\sigma \alpha^{-1} \alpha^{1-\sigma} w^\sigma \quad (40)$$

Which is clearly negative. \square

Proposition (2.2). $\frac{\partial^2 c}{\partial p \partial \rho} < 0$ iff $\frac{w}{w+p} < \frac{\alpha}{1-\alpha}$.

Proof.

$$\begin{aligned}
\frac{\partial^2 \ln(c)}{\partial p \partial \sigma} = & \frac{(1-\alpha)^{1-\sigma} \ln(1-\alpha)(w+p)^\sigma}{(w+p)(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma}(w+p)^\sigma)} + \frac{(1-\alpha)^{1-\sigma}(w+p)^\sigma \ln(w+p)}{(w+p)(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma}(w+p)^\sigma)} \\
& - \frac{(1-\alpha)^{1-\sigma}(w+p)^\sigma (-\alpha^{1-\sigma} \ln(\alpha) w^\sigma + \alpha^{1-\sigma} w^\sigma \ln(w))}{(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma}(w+p)^\sigma)^2 (w+p)} \\
& - \frac{(1-\alpha)^{1-\sigma}(w+p)^\sigma (-(1-\alpha)^{1-\sigma} \ln(1-\alpha)(w+p)^\sigma + (1-\alpha)^{1-\sigma}(w+p)^\sigma \ln(w+p))}{(\alpha^{1-\sigma} w^\sigma + (1-\alpha)^{1-\sigma}(w+p)^\sigma)^2 (w+p)}
\end{aligned} \tag{41}$$

Which can be simplified and then factorized as:

$$\frac{\partial^2 \ln(c)}{\partial p \partial \sigma} = [A + B][\ln(\alpha) - \ln(w) - \ln(1-\alpha) + \ln(w+p)] \tag{42}$$

Where $A = (w+p)^\sigma \alpha^\sigma (1-\alpha)^{2\sigma} w^{2\sigma-1} (p^2 + 2wp + w^2)$ and $B = w^{\sigma+2} (w+p)^{2\sigma} (1-\alpha)^\sigma \alpha^{2\sigma} (w+p)$.

Since A and B are both > 0 , the sign of $\frac{\partial^2 \ln(c)}{\partial p \partial \sigma}$ is given by $[\ln(\alpha) - \ln(w) - \ln(1-\alpha) + \ln(w+p)]$. Which is just a log transformation of $\frac{\alpha}{1-\alpha} - \frac{w}{w+p}$.

Note the sign of $\frac{\partial^2 \ln(c)}{\partial p \partial \rho}$ is the same as $\frac{\partial^2 \ln(c)}{\partial p \partial \sigma}$ since $\sigma = \frac{1}{1-\rho}$ and $\rho \leq 1$. \square

A.3 N Period Model

This appendix follows from Section 2. Since U is increasing in m , maximizing utility is achieved through maximizing consumption. Assuming the individual chooses to graduate with a *k-diploma*, her graduate consumption is given by:

$$m^{Ek} = (N-k)\Omega w_k + \sum_{j=1}^k ((\Omega - T_j - S_j)w_j - pT_j) \tag{43}$$

If she chooses no education and becomes a non-graduate her consumption is:

$$m^{E0} = N\Omega w_0 \quad (44)$$

The graduate premium is $G = (m^{Ek} - m^{E0})$. The individual maximizes G subject to:

- (1) The education production function (Equations 2 and 4)
- (2) The wage function (Equation 5)
- (3) Time constraints (Equation 7)

We note there is an equivalent cost minimization problem: Minimize $C^k = \sum_{j=1}^k c_j(T_j, S_j) = \sum_{j=1}^k ((T_j + S_j)w_j - pT_j)$, subject to the same constraints.

Lemma (1). *The individual would always minimize the cost of whatever education is undertaken in each period.*

*Proof.*³⁰

It is clear that the education levels chosen in each period will either be 0 or e_i^* . This is because, for $0 < e_i < e_i^*$, the marginal benefit from education is zero and the individual would be wasting resources (which could be used to increase her graduate premium) by not choosing $e_i = 0$. If $e_i > e_i^*$, then the gain would come from decreasing education to e_i^* .

Moreover, if it is optimal to choose no education in period i , then for $j = i + 1$, it will also be optimal to choose no education. In both periods the education production function and forgone wage are identical and the individual faces a similar decision. However, in period j the benefit of the increased wage would last for one less period.

Suppose that $S_1^*, S_2^*, \dots, S_N^*$ and $T_1^*, T_2^*, \dots, T_N^*$ are the amounts of study and tuition undertaken in each period to maximize the graduate premium. Since $\frac{\partial e_i}{\partial S_i} > 0$ and $\frac{\partial e_i}{\partial T_i} > 0$, it is clear that there is only one way to produce

³⁰We are grateful to Alasdair Smith for suggesting this approach.

$e_i = 0$, which is to use $S_i = T_i = 0$. Hence S_i^* and T_i^* are also the cost minimizing inputs.

If $e_i = e_i^*$, we use proof by contradiction to show that S_i^* and T_i^* satisfy this period's cost minimization problem. Suppose using $(S'_i, T'_i) \neq (S_i^*, T_i^*)$ is the least costly way for an individual to produce e_i^* in period i .

By changing to (S'_i, T'_i) but keeping all other variables unchanged in all other periods would only affect the graduate premium through the payoff in this period (since e_i^* is still produced in period i). This switch to (S'_i, T'_i) would decrease her cost by $(p + w_i)(T_i^* - T'_i) + w_i(S_i^* - S'_i) > 0$, and hence increase her graduate premium by this amount. \square

B Non-Mathematical Appendix

B.1 Ability³¹

This appendix follows from Section 5.1.1. Even in a one input model this problem can arise, although it can be avoided with a homogeneous forgone wage.

Consider the following case. In the standard model if two individuals, i and j , have the same number of years schooling and i attains a higher mark she is said to be higher ability than j .

$$\frac{y}{a_i} < \frac{y}{a_j} \tag{45}$$

Correspondingly, if two individuals attain the same mark, and if one has had fewer years of schooling, she is said to be higher ability.

In the standard model of human capital there is a role for opportunity cost. This is because the forgone wage is included in calculations of the optimal level of investment in education. However, opportunity cost does not enter into the standard definition of ability.

In a model where there are only two wages: w (a non-graduate wage) and \bar{w} (a graduate wage) the inclusion of forgone costs is not problematic. This

³¹We are grateful to Paul Grout for suggesting this approach.

is because when ‘years of schooling’ are considered then the opportunity cost for all individuals is the same (everyone needs to give up $n\underline{w}$).

Thus the ratios become:

$$\frac{y + n\underline{w}}{a_i} < \frac{y + n\underline{w}}{a_j} \quad (46)$$

This can also be extended away from a binary case, so long as wages are only a function of education (and not ‘ability’).

There is slightly more to consider when effort is used, if this translates to time. But essentially, the problem is made insignificant by fixing wages.

However in reality individuals face different opportunity costs. In particular, opportunity cost is likely to increase with ability. In this instance, the introduction of heterogeneous forgone wages causes ambiguity in the definition of ability. If the individual can obtain a forgone wage of $\underline{w}_i > \underline{w}_j$, then the input/output ratio becomes:

$$\frac{y + n\underline{w}_i}{a_i} <? > \frac{y + n\underline{w}_j}{a_j} \quad (47)$$

Which of these ratios is larger is ambiguous, since $a_i > a_j$ but $\underline{w}_i > \underline{w}_j$.

B.2 Education Lifecycle

The education production function may change as the level of education increases, and therefore during the education lifecycle. The relevant measure in this case is how the cost changes as education increases, which is similar to Arrow’s learning by doing progress ratio³².

As Heckman and his coauthors have emphasized (e.g. Cunha and Heckman (2007)), obtaining education is a cumulative process; to produce any education greater than e_0 , the individual must first obtain e_0 . We are interested in two mechanisms through which learning style might change:

- (1) Exogenously with education level; and

³²See Arrow (1962).

(2) Endogenously via path dependencies.

By the latter we mean the possibility that current learning style depends how education at earlier stages of the life cycle was produced.

To address these questions we propose combining Heckman's two period model with our two input model. This would result in a two period, two input education production function that contains four inputs and four parameters;

$$e(S_0, T_0, S_1, T_1; \alpha_0, \alpha_1, \rho_0, \rho_1) \quad (48)$$

Exogenous variation in the learning style parameters means that investment in period 2 depends on the education level in period 1. Endogenous variation in the learning style parameters implies that the learning style in the second period depends on the inputs used in the first period. This means that, like Heckman, we are interested in the inputs' second order cross-partial derivatives. I.e. Heckman is therefore interested in the degree of substitutability between investments at different stages of the education life cycle (Cunha and Heckman (2007)).

For example a high proportion of study, relative to tuition in early childhood, might increase the independence later in life. Thus individuals can 'learn' to become independent learners. If the mix of inputs used in early childhood influences the learning style adopted later in life there may be a trade off between static and dynamic efficiency.

A production function that captures some of these properties would be one in which the period 1 independence parameter is a function of the proportion of inputs (γ) used in period 0; $\alpha_1(\alpha_0, \rho_0)$. Using a simple linear relationship we can express total education attainment as:

$$e_1(S_0, T_0, S_1, T_1) = ((\alpha\gamma_0)T_1^\rho + \frac{(1-\alpha)}{\gamma_0}S_1^\rho)^{\frac{1}{\rho}} \quad (49)$$

where $\gamma_0 = \frac{S_0}{S_0+T_0}$ s.t, $E(S_0, T_0) = e_0$.

If resources cannot be shifted across periods, we can simply solve using backwards induction. If resources can be shifted across periods we will have

to equate the marginal products, including externalities, of each input.

B.3 Signaling³³

The most important alternative to human capital as an explanation for education is the signaling model, where education is used to signal unobserved ability (Spence (1973)). If employers are interested in learning styles, individuals may wish to signal these unobservables³⁴.

In this case employers would like to observe how individuals learn. However we think it is unlikely that either of the parameters in the education production function are directly observable. Although study is especially hard to observe, tuition might be observable. If it was common knowledge that different universities (providing the same E) offer different T ³⁵, employers might be able to infer learning style from the choice of university.

For example if employers are willing to pay an independent learner a higher wage but α is not observable, employees will have to signal their learning styles using the observables E and T .

The separating equilibrium is one where independent learners choose a university which gives very little tuition, and directed learners choose a university where there is more tuition. While both universities differ in the amount of tuition offered, the content of the education received is identical³⁶.

An employer would offer a wage which is increasing in E (education attainment), but decreasing in tuition. The various costs and benefits would be captured in the relevant incentive compatibility constraints.

The need to signal learning style in addition to ability will increase the

³³ S, T are inputs into production function - their observability addresses similar questions to those of effort in a moral hazard problem. (α, ρ) are parameters of the production function - their observability addresses similar questions to types in an adverse selection model.

³⁴If an employer cares about learning styles, A2 must be relaxed. This is discussed in Section 2.3.

³⁵In order to infer S from E and T , one would have to make several strong assumptions about the returns to scale exhibited by the production function.

³⁶For example, two students graduating *summa cum laude* from Harvard and William & Mary have the same educational attainment.

total cost of the signal, therefore increasing the welfare loss relative to perfect information. In order to signal independence students will purchase less than their cost minimizing bundle of tuition and thus education is obtained inefficiently. This is true whether or not education itself is productive.

B.4 Empirical Testing

Data on S is hard to observe, and so α and ρ are hard to identify. However, the CES production function we use has been identified in many contexts and we believe that with a rich enough dataset our model can be tested (see Klump, McAdam, and Willman (2012)). Moreover, there is a small but growing literature that addresses many of the issues raised by our model.

Here we discuss question of whether attendance in class should be compulsory. In Section 5.4 we show that in our model compulsory attendance reduces welfare for all students. Romer (1993) asks “Do students attend class? Should they?”. He finds that in three elite universities “absenteeism is rampant”. Attendance is effectively voluntary, which he states was not the case a generation ago. He finds that after controlling for motivation and ability attendance in class is strongly correlated with performance. It is also correlated with class size and teacher quality.

The paper generated an unusually large correspondence (JEP 1994). Many of these disagreed with Romer’s conclusion; firstly because of doubts about his identification strategy, and secondly because he ignores the opportunity cost associated with attending class, which is at the core of our paper.

If the correlation is causal Romer’s result implies that students are not optimizing. A behavioral explanation seems to be the most likely and suggests that a policy based on paternalism might be justified (Thaler and Sunstein (2003)).

If the students are maximizing, and therefore are substituting study for tuition, our model implies a negative correlation between ability and independence. Perhaps the most natural interpretation in terms of our model is that students have not understood the complementarity between study and tuition.

Since Romer (1993), the case for compulsory attendance has probably been weakened since changes in technology mean the delivery of core material no longer requires face to face interaction (see Section 4). For a recent contribution to this topic see Bratti and Staffolani (2013).