de Carvalho Andrade, Eduardo; de Castro, Luciano I.

Working Paper

Tougher educational exam leading to worse selection

Economics Discussion Papers, No. 2011-2

Provided in Cooperation with:
Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: de Carvalho Andrade, Eduardo; de Castro, Luciano I. (2011) : Tougher educational exam leading to worse selection, Economics Discussion Papers, No. 2011-2, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:
https://hdl.handle.net/10419/44432

Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
Nr. 2011-2 | January 25, 2011 | http://www.economics-ejournal.org/economics/discussionpapers/2011-2

Tougher Educational Exam Leading to Worse Selection

Eduardo de Carvalho Andrade
Insper São Paulo

Luciano I. de Castro
Kellogg School of Management, Northwestern University

Abstract A parallel of education with transformative processes in standard markets suggest that a more severe control of the quality of the output will improve the overall quality of the education. This paper shows a somehow counterintuitive result: an increase in the exam difficulty may reduce the average quality (productivity) of selected individuals. Since the exam does not verify all skills, when its standard rises, candidates with relatively low skills emphasized in the test and high skills demanded in the job may no longer qualify. Hence, an increase in the testing standard may be counterproductive. One implication is that policies should emphasize alignment between the skills tested and those required in the actual jobs, rather than increase exams' difficulties.

Paper submitted to the special issue
Quasi Markets in Education

JEL I2, J24

Keywords School standards; signaling model; cognitive skills; non-cognitive skills

Correspondence Luciano I. de Castro**,** Kellogg School of Management, Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208, United States;
e-mail: decastro.luciano@gmail.com

The authors are grateful to Flavio Cunha and Ann Spears for very helpful suggestions

1 Introduction

The importance of education and its special characteristics as a market environment have attracted increasing interest of economists.[1] Although those economists have been extremely careful in considering the peculiarities of this market, it is always tempting to rely on our developed intuitions to draw conclusions. For example, if we establish an analogy between education and a transformative process, it seems appropriate to use the intuition that a more severe control of the quality of the output will lead to an improving performance of the process. From this parallel, we would infer that raising the standards in exams leads to better quality. Is this an appropriate reasoning when applied to education services?

Although we can only conjecture the above intuition as a possible motivation for raising exams' standards, it *is a fact* that they are raising in number and importance in recent years, both to teachers and to students. In the teaching side, by 1999 forty-three states required applicants to pass some sort of standardized test such as the National Teacher Examination or Praxis examinations published by the Educational Testing Service. Currently, forty-seven states and the District of Columbia have alternative teacher certification methods as well as special programs formulated to facilitate the employment of individuals who would otherwise be uncertified (Kane, Rockoff and Staiger, 2008).

A similar trend exists for students.[2] In 2001 the federal No Child Left Behind Act mandated all states to adopt some form of "high-stakes" testing, where their results in these standardized tests have consequences for students, staff members, or the school. In particular, 20 states require high school

[1] See de Fraja and Romano (2002) and Cunha et. al. (2006).

[2] Koretz (2002, p. 754) observes the following: "In several respects, then, the current wave of high-stakes testing represents a continuation of trends in policy that extend back several decades. What the past decade has witnessed has been a rapid escalation in the application of high stakes, a greater emphasis on including nearly all students in large-scale assessments, and in many instances an increase in the difficult of tests that students are required to pass."

graduates to pass an exit exam and seven others plan to follow the same trend in the near future. Furthermore, in several states, the required exit exams have become more rigorous, moving from relatively modest requirements to tests linked to state academic standards.[3]

Notwithstanding the simple intuition sketched above, education specific papers haver argued that an increase in the standard may lead to an improvement in the quality of teachers/students. First, because less competent candidates are prevented from passing the standard, the overall quality of successful candidates should raise.[4] Second, the standard creates a greater incentive for individuals to invest in more occupation-specific human capital: they can recoup the full returns of their investment by signaling that they are better qualified.[5] Obviously, these arguments are not restricted to tests of teachers and students. They are also valid to other workers who are affected by occupational licensing such as nurses, engineers, accountants, auditors, lawyers and judges.

The objective of this paper is to show that the above intuition may be wrong: we prove that an increase in the standard may reduce the average quality (productivity) of those individuals who pass it. This apparently counterintuitive fact arises because tests do not emphasize all abilities that are important for job performance. A large number of papers show that noncognitive skills not tested in exams are important determinants of the performance in the labor market.[6] When the standard rises, at the margin candidates with relatively low cognitive skills but high noncognitive skills decide not to make the effort to meet the new standard. Candidates who succeed display more cognitive skills but the average level of noncognitive

[3]For more information, see Dee and Jacob (2006).

[4]This is the argument developed in Kleiner (2000), when he analyses the effects of occupational licensing on the quality of output, which is also valid for the case of standard or certification discussed here. He also compares occupational licensing and certification.

[5]This argument is developed in Shapiro (1986). Costrell (1994) and Betts (1998) develop a model with this feature.

[6]See, for instance, Edwards (1976), Klein, Spady and Weiss (1991), Cawley et. al. (1996), Bowles and Gintis (2001), de Fraja, Oliveira and Zanchi (2010), Heckman and Rubinstein (2001) and Heckman, Stixrud and Urzua (2006).

skills falls. As all skills contribute the workers' productivity in the market, the net effect may be a reduction on the average quality (productivity) of those individuals who pass the standard.

One important implication is that the difference between the skills demanded in the test and those required by the job implies that a rise in the standard may be counterproductive. In the case of teachers, certification requirement may prevent some with low knowledge from entering the profession. However, it also excludes others who would be quite effective in the classroom.[7] The net effect may be lower quality teachers. In the case of students, cognitive skills are prioritized in detriment to non-cognitive skills.[8] The final effect may be less productive workers.

More formally, a model is developed in which cognitive and noncognitive skills (or abilities) characterize individuals. There is a learning technology that improves individual's cognitive skills, and a production function that depends on both types of skills. The standard concerns only the cognitive abilities. When the standard rises, successful candidates make more effort to acquire knowledge, but the ones with relatively less cognitive skills and greater noncognitive skills may no longer pass the standard. The net effect may be a reduction in the average productivity of those who now pass, depending on the relative importance of the knowledge skill vis-à-vis the noncognitive skills in the production function. This result occurs in spite of the fact that both skills enter positively in the learning technology and in the production function. Also, our assumptions allow even for cases where cognitive skills are more important than noncognitive skills for the productivity. What drives the result is the fact that cognitive abilities are more emphasized in the exam than in the actual job.

The result in this paper may explain some facts previously uncovered by scholars. On the teacher side, Hanushek and Rivkin (2006) present a

[7]The possibility of this trade-off is discussed in Hanushek and Rivkin (2006). See Stixrud (2008) for some evidence that teachers' quality felt after a rising of standards.

[8]Costrell and Betts (2000) argues that the reason why high-achievers are against standard-based reform is that higher-order skills are de-emphasized by teachers.

review of the literature and conclude that there is mixed evidence on the effects of certification on teacher quality. Kane, Rockoff and Staiger (2008) examines the New York City public schools. They find that, on average, the certification status of a teacher has at most only small impacts on student test performance. Angrist and Guryan (2003) show that state-mandated teacher testing increases teacher wages, due to the barrier to entry associated with this requirement. However, they find that there is no corresponding increase in teachers' quality. More recently Stixrud (2008) conducted a supply and demand analysis to evaluate US teacher quality from 1960 until today and concluded that there has been a decline in female teacher quality.

On the student side, Dee and Jacob (2006) analyze the impact of exit exams on earnings. They made the analysis separately by race and concluded that more difficult exit exams tended to reduce the subsequent earnings of white and hispanic students while increasing those of black students. Costrell and Betts (2000) indicate that the most vocal opposition to a rise in standards comes from high achievers, which are the likely beneficiaries of the change.[9] Heckman and Rubinstein (2001) analyze the effects of the General Educational Development (GED) exam, which is taken by high school dropouts to certify that they have equivalent knowledge to high school graduates. They find that the GED is a mixed signal. Dropouts who take the GED are smarter (have higher cognitive skills) than other high-school dropouts and yet at the same time have lower levels of noncognitive skills. They also observe that both types of skills are valued in the market. Heckman, LaFontaine and Rodriguez (2008) investigate the effects of raising the difficult of obtaining the GED, either through increasing passing requirements or restricting access to young adults, and conclude that it reduces estimated dropout rates. In other words, high school students substitute alternative exam-based GED credentials for traditional graduation when the latter becomes a more difficult option.

Our results also offer a testable implication: the test is more effective

[9]Our results in section 4 can explain this, because the fall in productivity will imply in a fall in the wages for the qualified candidates.

in enhancing productivity when the mix of skills tested is closer to the set of skills needed in the job. This may remain true even if the production function heavily depends on the skills tested in the exam. Of course, it is not straightforward to verify this implication since it requires measurement of the alignment of the tested and the job-relevant skills. However, the implication is clear.

Our findings suggest that the widespread attempt to improve education by enhancing performance in pre-established standards may not be the correct objective. From this, we derive the following policy recommendation: it is more important to design the exam in order to test skills directly relevant to the jobs than to raise the standard. The adoption of this recommendation may have an important impact in the efforts to improve educational standards, measured with respect to the productivity of workers.

Although our point could be made in a slightly simpler model, we prefer to base our analysis on the well known model developed by Betts (1998), which is a combination of human capital and signalling model.[10] The key difference from our model and Bett's is the use of two types of skills (cognitive and noncognitive) in the production function instead of only one. In Betts, when the standard rises, there is an increase in the average productivity of those students who pass and fail the exam. Here, the average productivity (and wages) of those who pass the higher standard may decrease, as explained above.

Araujo, Gottlieb and Moreira (2007) also consider two dimensional skills and observe that "a test that places a stronger emphasis on noncognitive ability would be a more effective signal" (p. 1021). However, they arrive at this conclusion from a set of different assumptions. First, they propose a model to analyze the screening power of the GED exam, in which "for any pool of workers, the one with higher perseverance/lower intelligence is the most productive" (p. 1035). Thus according to their model, "if firms could identify the most intelligent individuals in a pool of workers, they would

[10]In fact, Betts (1998) used a formulation initially developed by Costrell (1994) and incorporated heterogeneity in the skills of students.

offer them lower wages" (p. 1036). In contrast, in our model a higher cognitive skill is not perceived as negative, even in equilibrium. Second, the authors assume that the individual may choose not to reveal the result of the test. Although this assumption may be realistic in their setting, this is inappropriate in the cases we are considering. For instance, the teacher's performance in the test is directly used in the hiring decision by the state. From their assumptions, they conclude that the exam is a neutral signal, that is, it does not affect education and wages (see their Proposition 5, p. 1036). In sum, their conclusion seems to be driven by the fact that the exam is unable to improve the screening of the individuals: hence their recommendation is directed to enhance the screening power of the test. In our model the exam allows a useful screening of the individuals, because the individuals who pass the standard have, on average, a higher productivity than the ones who fail it. Our result comes from a comparative statistics exercise and from the analysis of what happens with a raise in the standard. Moreover, since Araujo, Moreira and Gottlieb are primarily interested in the signaling game, their results are technically demanding. Our model is simpler to analyze and deliver a testable implication and policy recommendation that they do not.

Another relevant paper is Angrist and Guryan (2003), which focuses on the effects of the standards on teacher supply. If testing is seen as very costly (for example, if the opportunity cost of time invested in test preparation and test taking is high), the net teacher wage in the testing regime is lower than the teacher wage without testing. The rise of the standard reduces the average quality of those who pass the standard because the best qualified candidates opt for a different job.

This paper is divided in five sections, including this introduction. The next section presents the model. Section 3 and 4 present, respectively, the results with exogenous and endogenous wages. The last section concludes.

2 Model

We develop a hybrid human capital/signaling model based on Betts (1998). Individuals have two types of skills, cognitive (knowledge) and noncognitive (other abilities or characteristics).[11] In the case of teachers, for instance, these abilities may be represented by the subject knowledge and the didactic, respectively. Employers cannot observe directly the employees' skills; they can only observe whether the candidate meets the standard. For example, schools can check if a teacher has the certification. Firms can identify those who graduate from high school and those who do not. This information signals to employers the expected productivity of each worker; they obviously pay a higher wage to those who pass it. However, the standard verifies only one type of skill, which we call knowledge or cognitive skill.[12] Candidates may make an effort to augment this specific skill, which increases their perceived productivity, because it allows them to recoup the returns of their investment.

Each kind of skill may assume two possible values. The knowledge skills may be $\overline{k}$ (high level) or $\underline{k}$ (low level), where $\overline{k} > \underline{k}$. The combination of the noncognitive skills (s) can also assume two possible values: $\overline{s}$ (high level) and $\underline{s}$ (low level), where $\overline{s} > \underline{s}$. Hence, candidates are characterized by the pair (k, s) and there are four types: $a = \left(\overline{k}, \overline{s}\right)$, $b = \left(\overline{k}, \underline{s}\right)$, $c = \left(\underline{k}, \overline{s}\right)$ and $d = \left(\underline{k}, \underline{s}\right)$. The proportion of each type in the population is equal to p_i, $p_i > 0$, $\forall i$, $i \in T \equiv \{a, b, c, d\}$, where $\sum_{i \in T} p_i = 1$.

Candidate j maximizes utility, which depends on leisure (L_j) and lifetime earnings (w_j):

$$U = U\left(L_j, w_j\right) \ = L_j^{\alpha} w_j^{1-\alpha},$$

where $L_j \in \left[0, \overline{L}\right]$ and $\alpha \in (0, 1)$.

A candidate can improve his knowledge skill. His new knowledge level

[11] The existence of two types of skills is the main difference between this model and the one in Betts (1998). In his model, the educational achievement depends on effort and ability (only one type).

[12] We discuss this assumption in Remark 1.

(k_1) is defined by a learning technology (f) which depends on both skills (s and k) and effort $\left(e = \overline{L} - L \leq \overline{L}\right)$ in test preparation. Formally, this technology is the following:[13]

$$k_1 = f(k, s, e),$$

where f is continuously differentiable and increasing in all variables. In other words, this technology indicates that both skills (s and k) are important in the production of more knowledge.[14] It is assumed that $f(k, s, 0) = k$, which means that without effort the individual maintains his initial level of knowledge, i.e., there is no depreciation.

We make the reasonable assumption that the knowledge or cognitive skill is relatively more important vis-à-vis noncognitive skills in the process of acquiring more knowledge.[15] In other words, the learning technology is more intensive in k than in s. For example, it means that candidate type $b = \left(\overline{k}, \underline{s}\right)$ can acquire more knowledge than type $c = (\underline{k}, \overline{s})$ if they make the same effort, that is, for all $e \in \left[0, \overline{L}\right]$,

$$f\left(\overline{k}, \underline{s}, e\right) > f\left(\underline{k}, \overline{s}, e\right). \tag{1}$$

The focus of the test is on only one skill. It defines the minimum level of knowledge necessary to pass the standard, which is denoted by π. This feature of the models tries to capture the way the standards are frequently set. For example, for students, the exams tend to focus on cognitive skills. For teachers, the certification tests the candidates' subject knowledge. In order to succeed, the candidate's new knowledge level (k_1) has to be greater or equal to π. For each individual type $t = (k, s)$, the minimum effort necessary

[13] We suppress the student index j to save notation.

[14] Cunha and Heckman (2007) uses a similar production function in which the production of cognitive ability depends on investments and initial levels of cognitive and non-cognitive abilities.

[15] Cunha, Heckman and Schennach (2010) find evidence that the initial cognitive ability is relatively more important than the initial non-cognitive ability in the production of more cognitive ability.

for passing the standard π is denoted $e_t(\pi)$ and endogenously determined by $f(k, s, e_t(\pi)) = \pi$.

A worker i with final skills k_1 and s has the following productivity in the labor market $P_i = k_1^{1-\gamma} s^{\gamma}$, where $\gamma \in [0, 1]$. In other words, both skills affect positively his productivity in the labor market. In the next sections, two possibilities are analyzed; in both the employers do not observe directly the productivity of the workers. They only observe whether the worker meets the standard or not.

Remark 1 *Although we are referring to k as the cognitive ability and we assumed that the exam only tests this ability, another interpretation would reveal that such assumptions are without loss of generality. Indeed, one could have defined k as the ability that is tested in the standard and s as all other skills that are not tested, but may also be relevant in the job and, therefore, enter in the production function. Since the candidates only make effort to pass the standard, the assumption that s is not affected by the preparation for the exam is reasonable. This interpretation is useful in understanding the model and is consistent with most papers that use the terms cognitive and noncognitive skills. However, we do not insist in this interpretation because it is not essential for the results or clearly established in the literature.*

In section 3 we assume that wages are exogenous. The ones who pass the standard receive a high wage $\left(w^h\right)$ and the others a low wage $\left(w^l\right)$. Obviously, $w^h > w^l$. In section 4 wages are endogenously defined: perfect competition in the labor market ensures that workers are paid their expected productivity conditional on whether they have met the standard. As in Betts (1998), workers belong either to the group of individuals who pass the standard π or to the one who do not. Wages will be constant within each group due to the inability to observe each individual's skills. Given these wages, each worker maximizes his utility by choosing the optimum effort level.[16]

[16]Note that it is not in the individuals' interest to invest time to acquire other skills, as they are not going to be perceived as having more productivity because the standard verifies only the knowledge skill.

Before turning to the analysis, it is worth obtaining one important result that will be used in the next sections.

Lemma 2 *For each* π, $e_a(\pi) < e_b(\pi) < e_c(\pi) < e_d(\pi)$.

Proof. *Let's prove the second inequality. If* $e_b(\pi) \geq e_c(\pi)$, *then (1) implies that* $\pi = f\left(\overline{k}, \underline{s}, e_b(\pi)\right) > f\left(\underline{k}, \overline{s}, e_b(\pi)\right) \geq f\left(\underline{k}, \overline{s}, e_c(\pi)\right) = \pi$, *which is a contradiction. The other inequalities are proved similarly (using just the monotonicity of* f*).* ■

The above lemma indicates the order of the minimum effort necessary to pass any given standard among the different types of individuals. The necessary effort is decreasing with types in the natural order: a, b, c, d. This fact is a consequence of the assumption that the learning technology is more intensive in k than in s (i.e, condition (1) holds). As it is going to be clear in the next sections, our main results depend upon the order established in this lemma, but condition (1) is used only in this proof. Thus, instead of assuming (1), we could just have required directly the order established in the lemma.

3 Exogenous Wages

In this section we perform a comparative statistics for an increase in the standard, assuming that the wages are exogenously fixed. Our objective is to examine what happens with the average productivity of those individuals who pass the standard. Those who pass the standard receive a higher wage than those who do not and these wages are not correlated to the average workers' productivity.

The assumption of exogenously fixed wages is reasonable when the exam is used in the selection of workers in the public sector. In general, they have to pass an exam (or to obtain a certification) and their salaries are not frequently based on their productivity. Personnel systems that govern the pay and promotion practices are quite bureaucratic and rigid, such as is found

in the educational sector. As pointed out in Neal (2002), "given the results from the 'teachers effects' literature, it is quite striking that public schools pay teachers in a given subject the same wage, conditional on seniority and credentials, regardless of past job performance.(...) What is striking is the fact that persistent individual differences in teaching performance do not affect compensation among public school teachers even when principals are aware of these differences."

Formally, given π, individuals have to decide whether to make or not the necessary effort to pass the standard. For those who decide to make the effort, they need to set e such that $k_1 \geq \pi$. Employees who pass and do not pass the standard receive, respectively, wages equal to w^h and w^l $\left(w^h > w^l\right)$. We have the following:

Proposition 1 *There are $\pi_a > \pi_b > \pi_c > \pi_d$ such that if $\pi \leq \pi_d$, all types pass the exam; if $\pi \in (\pi_d, \pi_c]$, types a, b and c pass the standard; if $\pi \in (\pi_c, \pi_b]$, types a and b pass the standard; if $\pi \in (\pi_b, \pi_a]$, only type a passes the standard; and no type passes the standard if $\pi > \pi_a$.*

The above proposition indicates that for each standard level interval, there is a set of types in the population that decide to make the necessary effort to pass the standard. As expected, the higher the standard, the lower the number of individuals who are inclined to pay the cost to succeed. For example, when the standard is relatively high ($\pi \in (\pi_b, \pi_a]$), only the most qualified type (type a) passes the standard. The others, with relatively low skills, would have to make a too large effort to increase their knowledge skill; making such an effort would reduce their utility. Thus, they are better off receiving a lower wage with more leisure.

Note that there is no overlap among the standard level intervals. This fact implies that there is no possibility of multiple equilibria. As we discuss in the next section, this uniqueness of equilibrium does not hold in general when wages are endogenously defined.

We now analyze how the average productivity of those individuals who pass the standard changes when the standard rises from $\pi_0 \in (\pi_d, \pi_c]$ to

$\pi_1 \in (\pi_c, \pi_b]$. With this change, the equilibrium moves from one in which types a, b and c pass the standard to the one in which only types a and b do. The average productivity is given by the sum of the productivity in the labor market of all individuals who pass the standard divided by the number of these individuals. For example, in the case that types a and b pass the standard π_1, the average productivity (P_{ab}) is equal to:[17]

$$P_{ab} = \pi_1^{1-\gamma} \left[\frac{p_a \overline{s}^{\gamma} + p_b \underline{s}^{\gamma}}{p_a + p_b} \right].$$

The next proposition presents the main result of this section. It says that if the standard increases from a level where a, b and c pass to a level where only a and b pass, then this change implies a decrease in the average productivity if the non-cognitive skills are sufficiently important to the job productivity. More formally:

Proposition 2 *Assume that $\pi_c \geq \overline{k}$, that is, all candidates have to make an effort to pass the standard*[18] *and that the standard level changes from $\pi_0 \in (\pi_d, \pi_c]$ to $\pi_1 \in (\pi_c, \pi_b]$. Then, there exists $\gamma^* \in (0,1)$ such that for all $\gamma \in [\gamma^*, 1]$, a rise in the standard causes a fall in the average productivity of those individuals who pass it.*

The intuition behind this result is the following: as the focus of the standard is on the knowledge skill, an increase encourages the selection of workers with relatively higher knowledge (types a and b) and eliminates the ones with relatively higher noncognitive skills (type c). Note that when the parameter γ is greater than the threshold γ^*, it indicates that the noncognitive skills are relatively more important for the workers' productivity. The combination of

[17] This calculation is under the assumption that both types have to make an effort to pass the standard, which is the assumption made in the next proposition. Hence, $k_1 = \pi_1$, as the individuals do not make any effort beyond the minimum necessary to pass the standard.

[18] This assumption is not essential for the result, but the proof becomes much more complex without it.

these two facts leads to a decline in the average productivity of the individuals who pass the standard vis-à-vis the one in the previous equilibrium (now types a and b; before types a, b and c). Another implication of the result is that the most-able candidates can actually experiment a drop in utility even when the standard rises marginally. Indeed, they have to make a greater effort to succeed, but their wages do not change.[19]

It is important to note that the selection of the workers based on the standard is effective, because the productivity of those who pass the standard is always greater than the productivity of those who do not pass the standard. In other words, given the available instrument of selection based on knowledge, it is better to use it than to disregard it. However, the most appropriate instrument is the one that reaches a closer connection between the skills required in the standard and the ones in fact used in the production process.

Using a numerical example, we now show that the γ^* given in the proposition can be small. This means that the fall in the productivity may occur even if the cognitive skill is more important than the noncognitive skill for the production function, that is, even if $\gamma < 1/2$. Without loss of generality, we set $\pi_0 = 1 = \underline{s}$. Then, in table 1, we obtain the value of γ^* for different combinations of π_1 and $\overline{s}$, assuming $p_a = p_b = p_c = p_d = 1/4$. We consider small changes in the standard for two reasons. First, most changes in the standard are typically not dramatic. Second, a large change in the standard may harm the society in different ways, by abruptly pushing a large number of potential workers to the non-certified category.

The first column in table 1 presents different levels of the new standard π_1, as a percentage of the old standard π_0. A sufficient change that moves the equilibrium from the one in which types a, b, and c pass the standard to the one in which types a and b do. The change levels are: 0.1%, 0.2%, 0.5%,

[19] A similar observation is also valid when the wages are endogenous. However, the decrease in utility may be valid only for candidates near the threshold, because it is possible that high type agents receive higher wages without making too large effort to pass the new standard.

1% and 2%. The first line shows different values for the high knowledge skill $\overline{s}$, as a function of the low knowledge skill $\underline{s}$. The values of $\overline{s}$ are: 5%, 10%, 20% and 100% above that of $\underline{s}$. An example illustrates how to read table 1. When $\pi_1 = 1.002$ and $\overline{s} = 1.1$, then $\gamma^* = 0.11$. In other words, a rise in the standard by 0.2% leads to a fall in the average productivity of those individuals who pass the standard when $\gamma > \gamma^* = 0.11$.

$\pi_1 \backslash \overline{s}$	5%	10%	20%	60%	100%
0.1%	0.11	0.06	0.04	0.02	0.01
0.2%	0.19	0.11	0.06	0.03	0.02
0.5%	0.38	0.24	0.15	0.06	0.04
1%	0.55	0.39	0.25	0.11	0.08
2%	0.71	0.56	0.39	0.20	0.15
4%	0.83	0.71	0.57	0.34	0.26
6%	0.87	0.79	0.66	0.43	0.34

Table 1 - values of the threshold γ^* with respect to the % increase of the test difficult (π_1) and the spread of the noncognitive ability $(\overline{s})$.

It is worth making some comments about the values of γ^*. First, for a given π_1, γ^* decreases with $\overline{s}$. When $\overline{s}$ is relatively high, there is an important drop in the level of noncognitive skills among the individuals who pass the standard. It occurs because type c has skill $\overline{s}$ and no longer passes the standard. Hence, there is a drop in the average productivity of those who continue passing the standard even if γ^* is very low.

Second, for a given $\overline{s}$, γ^* increases with π_1. The intuition behind this result is the following: the greater the rise in the standard, the greater will be the final level of knowledge skill of those who pass the standard. This effect contributes to an increase in the workers' average productivity. As a consequence, in order for a fall to occur in this productivity, it is necessary to place greater importance on noncognitive skills, that is, a greater γ. Finally, the results in table 1 confirm the point made above that γ^* can assume low values. For most of the combinations presented, it is lower than 0.5.

The main result in this section may explain some facts mentioned in the introduction, such as the mixed evidence on the effects of certification on teacher quality or on student test performance.

4 Endogenous Wages

In this section, we relax the assumption that wages are fixed and exogenously determined. As in Betts (1998), perfect competition guarantees that workers are paid their expected productivity conditional on whether they have met the standard.[20] That is, wages are endogenously determined. In contrast with the previous section, there is no difference between wages and expected productivity. This assumption is more likely to hold in the private sector. Depending on their efforts, individuals separate themselves in two groups: those who pass and those who fail the standard. Wages will be the same within each group due to the inability to observe skills.

The objective here is to show that an increase in the standard may reduce the average productivity and the wages of those who succeed. This result contrasts with Betts (1998), who shows that there is a rise in the wages of those who pass and do not pass the standard.

Given lemma 1, an equilibrium is characterized by the following possible partitions of types who pass the standard: a, ab, abc and $abcd$. For each equilibrium, there are high wages (for those who pass the standard) and low wages (for those who fail the standard). The notation used is the following. When types a and b pass the standard (partition ab), w_{ab}^h and w_{ab}^l are, respectively, the wages of those who pass and do not pass the standard. In other words, the superscript represents high or low wages and the subscript indicates the possible partitions.

As in the previous section, given π, each type decides whether to make or not the necessary effort to pass the standard and to receive wages accordingly.

[20] Only a binary credential (to pass or not pass the standard) is relevant. This feature of the model is used in Costrell (1994) and Betts (1998) and tries to mimic the real world. For example, employers of high-school graduates rely almost exclusively on the diploma.

For those who decide to make the effort, they need to set e such that $k_1 \geq \pi$. The following proposition shows that the type of equilibrium, that is, the partition of types of individuals who pass or not the standard, depends on the standard level.

Proposition 3 *There are π_a, π_b, π_b^*, π_c, π_c^* and π_d such that: if $\pi \leq \pi_d$, all students pass the standard; if $\pi \in (\pi_d, \pi_c]$, students types a, b and c can pass the standard; if $\pi \in (\pi_c^*, \pi_b]$, students types a and b can pass the standard; if $\pi \in (\pi_b^*, \pi_a]$, only students type a pass the standard; and nobody passes the standard if $\pi > \pi_a$.*

The next proposition is the main result of this section.

Proposition 4 *Assume that $\pi_c \geq \overline{k}$. Then, there exists $\gamma^* \in (0,1)$, such that for all $\gamma \in [\gamma^*, 1]$: (i) $w_{abc}^h > w_{ab}^h$, $w_{abc}^l < w_{ab}^l$ and (ii) there are multiple equilibria when $\pi \in [\pi_c^*, \overline{\pi}]$, where $\overline{\pi} = \min\{\pi_c, \pi_b\}$.*

The first point of this proposition shows what happens to wages when the equilibrium moves from one in which types a, b and c pass the standard to one in which only types a and b pass it. w_{abc}^h is higher than w_{ab}^h for the same reason that there is a fall in the average productivity of those individuals who pass the standard in the previous section. As the standard focuses on knowledge skill, when it rises, workers with relatively high noncognitive skills (type c) no longer succeed, with negative consequences to the workers' productivity in the labor market when γ is greater than the threshold γ^*. w_{abc}^l is lower than w_{ab}^l because when type c moves to the group of individuals who do not pass the standard (type d), they increase its average productivity.

This result differs in an important way from the one in Betts (1998): here the wages of those workers who pass a higher standard may fall. This finding may explain some facts mentioned in the introduction, such as the opposition of high achievers from a rise in the standard or the reduction of subsequent earnings from more difficult exit exams.

The second point of the proposition is that there is the possibility of multiple equilibria when wages are endogenous. The explanation is the following. Consider an equilibrium in which types a, b and c pass the standard π. Given the skill premium associated with this equilibrium $\left(\frac{w^h_{abc}}{w^l_{abc}}\right)$, when the standard rises to $\pi = (\pi_c + \varepsilon)$, it is not in the interest of type c to make efforts to pass such a relatively high standard. When this point is reached, the skill premium falls as only types a and b pass the standard and $w^h_{abc} > w^h_{ab}$ and $w^l_{abc} < w^l_{ab}$. With this lower skill premium, even a fall in the standard to $\pi = (\pi_c - \varepsilon)$ will not be sufficient to make type c to pay the price to pass it. When faced with the skill premium $\left(\frac{w^h_{ab}}{w^l_{ab}}\right)$, individuals type c do not make the effort to pass the standard $\pi = (\pi_c - \varepsilon)$ that they would have passed if the skill premium were $\left(\frac{w^h_{abc}}{w^l_{abc}}\right)$. Hence, the possibility of two equilibria arises. This result contrasts with the previous section, as there is no possibility of multiple equilibria when wages are exogenous.

5 Conclusion

Since education services form a special kind of market, with a lot of government participation and severe regulation, but also a unique production function, it requires special care in the application of economic intuitions. Thus, even a seemingly straightforward idea as that the increase of standards lead to enhancement of the quality requires serious consideration.

Indeed, this paper shows that a rise in the standard may not produce such desired enhancement. Actually, higher standards may lead to less qualified labor force. This counter-intuitive result is more likely to occur when changes in the standard level are small, which is conceivably the more natural situation. The result follows from the fact that exams evaluate only some types of skills (which we call cognitive skills), while there are also other skills affecting the productivity in the labor market (non-cognitive skills). The driving element of the result is the misalignment between the set of skills to perform well in the test and to succeed in the job.

From this simple result, we obtain a policy recommendation and a testable implication. The policy recommendation is to concentrate the efforts on the alignment of the skills tested in the exams and those needed in the job.[21] The testable implication is that when the format of the test is better aligned with skills needed for the job then the selected candidates are more productive on average. For example, if teachers are selected based in their performance in the classroom and not only on their knowledge of the subject, the quality of teachers will be higher.

There is also a more subtle implication. Suppose that the exam is just a mechanism to signalize the level of education in a region or country and tests only cognitive skills. An example is the Programme for International Student Assessment (PISA), which is coordinated by the Organization for Economic Co-operation and Development (OECD). From this, a country could set as a goal to obtain higher scores in the PISA. This policy goal may appear desirable, but the findings of this paper suggest that this goal may not be the best choice for the future productivity of the country's labor force. Instead, it would be better to orientate the education to the development of the correct balance of skills needed in the future's jobs.[22]

This conclusion raises an important question: what are the most desirable set of skills to emphasize? Although some papers argue that cognitive skills are not the most important part, it seems that the literature lacks a general procedure to determine what is more important.[23]

Although our paper is inspired and primarily concerns educational markets, it should be noted that our conclusions are not restricted to the educational system. If a government or a firm plan to select new employees through an examination, then each should prepare an exam closely linked to

[21] Of course, such recommendation rests on the assumption that the education aims to enhance productivity. This recommendation is less important if the education has other objectives, such as the maintainence of culture, religion or the pure transmission of knowledge by itself.

[22] Of course, the OECD could align the PISA test to such set of skills, and this would be an important improvement.

[23] See, for instance, Heckman, Stixrud and Urzua (2006).

the actual work done by the candidates. A hiring decision based on the test of skills directly related to those needed in the job will select more productive workers. Although this is a simple and sensible recommendation, it seems that it has been overlooked in many real world cases.

Appendix

Proposition 1: *There are $\pi_a > \pi_b > \pi_c > \pi_d$ such that if $\pi \leq \pi_d$, all types pass the exam; if $\pi \in (\pi_d, \pi_c]$, types a, b and c pass the standard; if $\pi \in (\pi_c, \pi_b]$, types a and b pass the standard; if $\pi \in (\pi_b, \pi_a]$, only type a passes the standard; and no type passes the standard if $\pi > \pi_a$.*

Proof. Type t decides to pass the standard if:

$$\left(\overline{L} - e_t(\pi)\right)^{\alpha} \left(w^h\right)^{1-\alpha} \geq \left(\overline{L}\right)^{\alpha} \left(w^l\right)^{1-\alpha}. \tag{2}$$

For $t \in T = \{a, b, c, d\}$, define π_t endogenously as follows:

$$\left(\overline{L} - e_t(\pi_t)\right)^{\alpha} \left(w^h\right)^{1-\alpha} = \left(\overline{L}\right)^{\alpha} \left(w^l\right)^{1-\alpha};$$

Since $e_a(\pi) < e_b(\pi) < e_c(\pi) < e_d(\pi)$ by Lemma 1 and $e_t(\cdot)$ is increasing, it is easy to see that $\pi_a > \pi_b > \pi_c > \pi_d$. Moreover, if $\pi \leq \pi_d$, then (2) is satisfied for all types, again using Lemma 1. If $\pi \in (\pi_d, \pi_c]$, types a, b and c pass the standard. The other intervals come in similar fashion. ■

Proposition 2: *Assume that $\pi_c \geq \overline{k}$ and that the standard level changes from $\pi_0 \in (\pi_d, \pi_c]$ to $\pi_1 \in (\pi_c, \pi_b]$. Then, there exists $\gamma^* \in (0, 1)$ such that for all $\gamma \in [\gamma^*, 1]$, a rise in the standard causes a fall in the average productivity of those individuals who pass it.*

Proof. Because $\pi_c \geq \overline{k}$, all candidates have to make a positive effort to pass the standard. Let $P_{abc}(\gamma)$ be the average productivity when a, b, and c pass the exam. Similarly, let $P_{ab}(\gamma)$ denote the average productivity when

types a and b pass the exam. Since all types who pass the standard have final knowledge level equal to the standard, we have:

$$\begin{aligned} P_{abc}(\gamma) &= \pi_0^{1-\gamma}\left[\frac{(p_a+p_c)\,\overline{s}^{\gamma}+p_b\underline{s}^{\gamma}}{p_a+p_b+p_c}\right]; \\ P_{ab}(\gamma) &= \pi_1^{1-\gamma}\left[\frac{p_a\overline{s}^{\gamma}+p_b\underline{s}^{\gamma}}{p_a+p_b}\right]. \end{aligned}$$

Since $\overline{s}^{\gamma} > \underline{s}^{\gamma}$ and $p_c > 0$, it is clear that

$$P_{abc}(1) = \frac{(p_a+p_c)\,\overline{s}+p_b\underline{s}}{p_a+p_b+p_c} > \frac{p_a\overline{s}+p_b\underline{s}}{p_a+p_b} = P_{ab}(1).$$

Since $P_{abc}(\cdot)$ and $P_{ab}(\cdot)$ are continuous functions, the inequality remains true for sufficiently high γ. This concludes the proof. ■

Proposition 3: *There are π_a, π_b, π_b^*, π_c, π_c^* and π_d such that: if $\pi \leq \pi_d$, all students pass the standard; if $\pi \in (\pi_d, \pi_c]$, students types a, b and c can pass the standard; if $\pi \in (\pi_c^*, \pi_b]$, students types a and b* can *pass the standard; if $\pi \in (\pi_b^*, \pi_a]$, only students type a pass the standard; and nobody passes the standard if $\pi > \pi_a$.*

Proof. In order to have an equilibrium in which types a, b and c pass the standard, the standard π must be such that the following conditions hold:

$$\left[\overline{L} - e_d(\pi)\right]^{\alpha}\left(w_{abc}^h\right)^{1-\alpha} < \overline{L}^{\alpha}\left(w_{abc}^l\right)^{1-\alpha} \tag{3}$$

$$\left[\overline{L} - e_c(\pi)\right]^{\alpha}\left(w_{abc}^h\right)^{1-\alpha} \geqslant \overline{L}^{\alpha}\left(w_{abc}^l\right)^{1-\alpha} \tag{4}$$

Defining the standards π_d and π_c such that, respectively, (3) and (4) hold with equality, the equilibrium in which types a, b and c pass the standard occurs when π is in the interval $\pi \in (\pi_d, \pi_c]$. Given Lemma 1 and the fact that $w_{abc}^h > w_{abc}^l$, this interval exists.

In order to have an equilibrium in which types a and b pass the standard, the standard must be such that the following conditions hold:

$$\left[\overline{L} - e_c(\pi)\right]^{\alpha}\left(w_{ab}^h\right)^{1-\alpha} < \overline{L}^{\alpha}\left(w_{ab}^l\right)^{1-\alpha} \tag{5}$$

$$\left[\overline{L} - e_b(\pi)\right]^{\alpha}\left(w_{ab}^h\right)^{1-\alpha} \geqslant \overline{L}^{\alpha}\left(w_{ab}^l\right)^{1-\alpha} \tag{6}$$

Defining the standards π_c^* and π_b such that, respectively, (5) and (6) hold with equality, the equilibrium in which types a and b pass the standard occurs in the interval $\pi \in (\pi_c^*, \pi_b]$. Given Lemma 1 and the fact that $w_{ab}^h > w_{ab}^l$, this interval exists. The proofs for the other equilibria are analogous. ■

Proposition 4: Assume that $\pi_c \geq \overline{k}$. Then, there exists γ^*, $0 < \gamma^* < 1$, such that for all $\gamma \in [\gamma^*, 1]$: (i) $w_{abc}^h > w_{ab}^h$, $w_{abc}^l < w_{ab}^l$ and (ii) there are multiple equilibria when $\pi \in [\pi_c^*, \overline{\pi}]$, where $\overline{\pi} = \min\{\pi_c, \pi_b\}$.

Proof. Throughout this proof, we use the fact that, since $\pi_c \geq \overline{k}$, all candidates have to make a positive effort to pass the standard and all types who pass the standard have final knowledge level equal to the standard.

Let π_0 and π_1 be the standards, respectively, that types $\{a, b, c\}$ and $\{a, b\}$ pass. Then, we have:

$$w_{abc}^h(\gamma) = \pi_0^{1-\gamma}\left[\frac{(p_a + p_c)\,\overline{s}^{\gamma} + p_b\underline{s}^{\gamma}}{p_a + p_b + p_c}\right],$$

$$w_{ab}^h(\gamma) = \pi_1^{1-\gamma}\left[\frac{p_a\overline{s}^{\gamma} + p_b\underline{s}^{\gamma}}{p_a + p_b}\right].$$

Since $\overline{s}^{\gamma} > \underline{s}^{\gamma}$ and $p_c > 0$, if $\pi_0 = \pi_1 = \pi$, then $w_{abc}^h(\gamma) > w_{ab}^h(\gamma)$ for all $\gamma > \gamma^* = 0$. Since $\overline{s}^{\gamma} > \underline{s}^{\gamma}$ and $p_c > 0$, if $\pi_0 < \pi_1$, it is clear that:

$$w_{abc}^h(1) = \frac{(p_a + p_c)\,\overline{s} + p_b\underline{s}}{p_a + p_b + p_c} > \frac{p_a\overline{s} + p_b\underline{s}}{p_a + p_b} = w_{ab}^h(1).$$

As $w_{abc}^h(\gamma)$ and $w_{ab}^h(\gamma)$ are continuous functions, the inequality remains true for sufficiently high γ.

Note that:

$$w_{abc}^l = \frac{p_d\underline{k}^{1-\gamma}\underline{s}^{\gamma}}{p_d} < \frac{p_c\underline{k}^{1-\gamma}\overline{s}^{\gamma} + p_d\underline{k}^{1-\gamma}\underline{s}^{\gamma}}{(p_c + p_d)} = w_{ab}^l$$

for all γ as $\underline{s} < \overline{s}$.

Using the definitions of π_c and π_c^*, we have:

$$\left[\overline{L} - e_c(\pi_c)\right]^{\alpha}\left(w_{abc}^h\right)^{1-\alpha} = \overline{L}^{\alpha}\left(w_{abc}^l\right)^{1-\alpha} \tag{7}$$

$$\left[\overline{L} - e_c(\pi_c^*)\right]^{\alpha}\left(w_{ab}^h\right)^{1-\alpha} = \overline{L}^{\alpha}\left(w_{ab}^l\right)^{1-\alpha} \tag{8}$$

If $\frac{\left(w^h_{abc}\right)}{\left(w^l_{abc}\right)} > \frac{\left(w^h_{ab}\right)}{\left(w^l_{ab}\right)}$, then $\pi^*_c < \pi_c$. Note that this is the case because, for a given π, we have the following:

$$\frac{w^h_{abc}}{w^l_{abc}} = \frac{\pi^{1-\gamma}\left[\frac{(p_a+p_c)\overline{s}^\gamma+p_b\underline{s}^\gamma}{p_a+p_b+p_c}\right]}{\underline{k}^{1-\gamma}\underline{s}^\gamma} > \frac{\pi^{1-\gamma}\left[\frac{p_a\overline{s}^\gamma+p_b\underline{s}^\gamma}{p_a+p_b}\right]}{\underline{k}^{1-\gamma}\left[\frac{p_c\overline{s}^\gamma+p_d\underline{s}^\gamma}{p_c+p_d}\right]} = \frac{w^h_{ab}}{w^l_{ab}}.$$

This concludes the proof. ■

References

[1] Angrist, J. and Guryan, J. (2003): "Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements," NBER Working Paper # 9545.

[2] Araujo, A., Gottlieb, D. and Moreira, H. (2007): "A model of mixed signals with applications to countersignaling," Rand Journal of Economics, 38, 1020-1043.

[3] Betts, J. (1998): "The Impact of Educational Standards on the Level and Distribution of Earnings," The American Economic Review, 88, March, 266-275.

[4] Cawley, J., Conneely, K., Heckman, J. and Vytlacil, E. (1996): "Measuring the Effects of Cognitive Ability," NBER Working Paper #5645.

[5] Cunha, F. and Heckman, J. (2007): "The Technology of Skill Formation," The American Economic Review, 97, 31-47.

[6] Cunha, F., Heckman, Lochner, J. L. and Masterov, D. (2006): Interpreting the evidence on life cycle skill formation. Handbook of the Economics of Education, 1, 697-812.

[7] Cunha, F., Heckman, J. and Schennach, S. (2010): "Estimating the Technology of Cognitive and Non-Cognitive Skill Formation," Econometrica, 78, 883–931.

[8] Costrell, R. (1994): "A Simple Model of Educational Standards," The American Economic Review, 84, September, 956-971.

[9] Costrell, R. and Betts, J. (2000): "Incentives and Equity Under Standards-Based Reform," Discussion Paper 2000-20, Department of Economics, University of California, San Diego, August.

[10] Dee, T. and Jacob, B. (2006): "Do High School Exit Exams Influence Educational Attainment or Labor Market Performance?," NBER Working Paper #12199.

[11] De Fraja, G., Oliveira, T. and Zanchi, L. (2010): "Must try harder: evaluating the role of effort in educational attainment," The Review of Economics and Statistics, 92, 577–597.

[12] De Fraja, G. and Romano, R. (2002): "The Economics of Education: Editors Introduction". Bulletin of Economic Research, 54, 205–208.

[13] Edwards, R. (1976): "Individual Traits and Organizational Incentives: What Makes a Good Worker?" Journal of Human Resources, 11, 51-68.

[14] Hanushek, E. and Rivkin, S. (2006): "Teacher Quality" in Handbook of the Economics of Education, Chapter 18, Vol. 2, Elsevier.

[15] Heckman, J. and Rubinstein, Y. (2001): "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," AEA Papers and Proceedings, 91, 145-9.

[16] Heckman, J., LaFontaine, P. and Rodriguez, P. (2008): "Taking the Easy Way Out: How the GED Testing Program Induces Students to Drop Out," NBER Working Paper # 14044.

[17] Heckman, J. Stixrud, J. and Urzua, S. (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Ouctomes," NBER Working Paper #12006.

[18] Kane, T., Rockoff, J. and Staiger, D. (2008): "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City, Economics of Education Review, 27, 615 - 631.

[19] Klein, R. Spady, R. and Weiss, A. (1991): "Factors Affecting the Output and Quit Propensities of Production Workers," Review of Economic Studies, 58, 929-954.

[20] Kleiner, M. (2000): "Occupational Licensing," Journal of Economic Perspectives, 14, 189-202.

[21] Koretz, D. (2002): "Limitations in the use of achievement tests as measures of educators productivity," Journal of Human Resources, 37, 752–777.

[22] Neal, D. (2002): "How Vouchers Could Change the Market for Education," The Journal of Economic Perspectives, 16, 25-44.

[23] Shapiro, C. (1986): "Investment, Moral Hazard and Occupational Licensing," Review of Economics Studies 53, pp. 843-62.

[24] Stixrud, J. (2008): "U.S. Teacher Quality from 1960 until Today: A Supply and Demand Analysis," Mimeo.