A signalling model of school grades: centralized versus decentralized examinations

De Paola, Maria and Scoppa, Vincenzo

16 January 2008

A signalling model of school grades: centralized versus decentralized examinations

Maria De Paola and Vincenzo Scoppa
Dipartimento di Economia e Statistica, Università della Calabria
m.depaola@unical.it; v.scoppa@unical.it

Abstract: *In this paper we examine the signalling value for skills of different examination systems in relation to errors that may affect grades obtained by students. Firm use school grades as a signal of the effective skills of workers, taking into account that evaluation are effected by stochastic shocks. We show that more precise evaluation systems, being associated to a higher reactivity of wages to school grades, induce an higher level of student effort. However, the effect is heterogeneous: low ability students tend to react less compared to high ability students. Moreover, from our analysis, it emerges that individuals endowed with low abilities may prefer less accurate evaluation systems. Nevertheless, when productivity increases the convenience of these systems reduces and the number of individuals preferring them shrinks. Our analysis highlights an important trade-off between centralized and decentralized evaluation systems. Frequent evaluations, typical of decentralized systems, reduce the impact on grades of errors that influence student performance and in this way diminish signal noise, on the other hand, different teachers generally adopt different performance assessment standards, and this tends to produce noisier evaluations. Conversely, centralized systems use common evaluation standards, but their frequency is limited by relevant administration costs and then produce evaluations that are more affected by errors influencing student performance. In the final part of the paper we investigate the relationship between the optimal class size and evaluation systems. We show that under decentralized evaluation systems the class size also affects the signal noise, since larger classes may reduce the frequency of evaluations undertaken by teachers.*

JEL Classification: D02, H42, I28

1. Introduction

External exit examinations and high standards are often advocated as the appropriate device for evaluating students in order to improve their skills and produce positive effects on the school system. A number of authors emphasize the positive consequences that centralised exams may produce on agents involved in the educational process.

First of all, since grades on centralised exams provide a better signal of students' skills for employers compared to locally-graded examinations, wages tend to be higher where central exams are in place and students have greater incentives to learn (see Wößmann, 2005). Secondly, according to Bishop and Wößmann (2004), central exams provide parents with information on the performance of their children against an absolute standard and relative to other students in the educational system. This information allow parents to understand whether it is the whole class which is performing badly or the low performance is limited to their own child, and in this way enables them to exert pressure both on students and teachers (Bishop 1996; 1997; 2006). Thirdly, central exams also affect teachers' behaviour since student performance on standardized tests can be used to monitor teaching quality on a regular basis and to offer output based incentives schemes (Lavy, 2004; Glewwe and Kremer 2006; Jürges, Richter, and Schneider 2004). Finally, with centralized examinations the achievement of students becomes crucial for school reputation and for attracting good students.

Empirical studies conducted by Bishop (1997) and (2006) indicate that central exit examinations significantly improve student performance[1]. Similarly, Wößmann (2005), using the international TIMSS micro data, estimates the effect of central exams suggesting that students who take a central exam at the end of upper secondary education outperform students in states without central exams. Jurges, Buchel and Schneider (2005) find analogous results considering the German educational system.

According to some recent studies, the positive effects of central exit examinations are heterogeneous and vary according to student abilities. Wößmann (2005) shows that the positive effect of central exams on student performance is higher for high ability students, while low ability students react less. Dee and Jacob (2006) find that exit exams increase student performance, but significantly reduce the probability of completing high school, particularly for black students, and increase the dropout rate in urban and high-poverty school districts. Similar results emerge by a study conducted by Dean and De Cicca (2001).

In spite of the favour encountered by centralized exams among economists they are strongly criticized among teachers and pedagogical specialists, who question their efficacy, since they undermine educational freedom and the pedagogical discretion that is supposed to be necessary to deal with heterogeneity among students. According to their view, external exams induce educators to teach to the test and to ignore important areas of knowledge. Teachers are able to evaluate, through direct observations, aspects of the students knowledge that are difficult to measure through standardized exams, which usually consist in a number of very specific questions. Lazear (2006) examines this issue concluding that standardized predictable test should be used when learning and monitoring are very costly, but should not be used with high ability students. The relative advantage of this type of test depends also on the degree of intrinsic motivation characterizing teachers: when the intrinsic motivation is low they perform better, while with highly motivated teachers less predictable exams are preferable.

An important issue, which we analyse in this paper, is represented by the effect produced by different evaluation systems on measurement errors affecting students' evaluations

[1] See also Bishop and Mane (2001).

at exams and the use of these evaluations as a signal of effective skills. We consider two different types of errors: errors that influence student performance and errors deriving from different performance evaluation standards. According to the relevance of these errors, grades obtained by individuals during their educational career are a more o less reliable signal of their effective skills. In fact, we assume that firms are not able to observe neither individuals' abilities nor the effort they provide in the educational process, but they simply observe the grade obtained by each student at the end of his educational career and try to infer student effective skills from it.

Starting form these assumptions, we show that more precise evaluation systems are associated to a higher reactivity of wages to school grades and therefore induce a higher level of student effort. However, as shown in the appendix, if ability and effort are complements, the effect is heterogeneous among students, since the incentive effect of more accurate evaluation systems increases with students' ability.

From our analysis it emerges that individuals endowed with low abilities may prefer less accurate evaluation systems, since, due to the lack of effective signals of individual skills, they induce firms to pay a more egalitarian wage, based on average abilities and effort. Interestingly, the number of individuals preferring less precise evaluation systems reduces when productivity increases. In fact, these systems produce two effects: on the one hand, they lead to a more egalitarian pay structure, while, on the other hand, they reduce effort and, as a consequence, reduce the total output produced in the economy. When the productivity of skills is high the wage reduction deriving from the lower level of effort tends to counterbalance the positive effect that low ability agents obtain from the income redistribution. Then, the higher the productivity of skills the greater the support offered by society to evaluation systems that produce better signals for the labour market.

Using this framework, we compare advantages and disadvantages of centralized and decentralized evaluation systems. Decentralized evaluations carried out by teachers allow to evaluate students more frequently, since teachers and students interact on a daily basis, compared to centralized systems. Frequent evaluations reduce the impact of errors that influence student performance, such as being lucky enough to have studied the material relevant for the exam, or being in good health condition during the exam. However, different teachers may adopt different criteria and evaluation standards, introducing noise in student evaluation. On the other hand, centralized exams involve relevant costs and their frequency is limited, implying that evaluations emerging form them may be more affected by stochastic shocks affecting students' performance, but at the same time, since they refer to a common standard, generate more reliable and comparable evaluations.

Depending on the importance that these two type of errors play in distorting students' performance at exams (the evaluation they obtain during their educational career) compared to their effective skills, it may result efficient to adopt a centralized evaluation system instead of a decentralized one. We show that grades on centralized exams provide a better signal that locally-graded exams of student's competences to employers only when the variance of errors affecting student performance is lower compared to the variance of errors deriving from the adoption of different assessment methods.

In the final part of the paper we examine how evaluation systems affect the optimal class-size. In fact, even if evaluation carried out by teachers are generally considered costless, teachers may find it difficult to evaluate their students if their number is high. When class size increases, the cost to evaluate students increases for teachers and they may reduce the frequency of evaluations, making qualifications a more noisy signal of skills. In educational systems that rely on teachers' evaluations, the policy maker deciding the class size has to take into account this effect. On the other hand, when the evaluation system is at centralized level, class size is decided only in relation to its effects on student achievement and to its costs. In fact, class size does not produce any effect on the information provided by qualifications relatively to students' effective skills.

The paper is organized as follows. Section 2 presents a model showing how the signalling value of school grades is influenced by the accuracy of the evaluation systems and its effects on student effort and welfare. In section 3 we compare advantages and disadvantages respectively of centralized and decentralized evaluation systems. Section 4 is devoted to discuss the relationship between class size and evaluation systems. Section 5 offers some concluding remarks.

2. Measurement errors affecting student performance and the signalling value of evaluations

We assume that individuals are risk-neutral and live for two periods: in the first period they go to school, sustaining the cost of effort, and in the second period they enter the labor market, obtaining a wage W. There is no discounting. Individuals are identical in every respect except their ability that is distributed according to a probability density function with mean $\bar{a}$ and variance σ_a^2.

Students attend school and attain an educational qualification with an evaluation of their skills made by schools. This evaluation is denoted by v which is affected by student effort e and ability a and by two different types of errors: one related to factors that may influence the student performance, indicated by ε, with mean zero and variance σ_{ε}^{2}, and the other, indicated by η, with mean zero and variance σ_{η}^{2}, related to aspects concerning the teacher evaluation system. The first type of error depends on whether students were lucky enough to have studied the material precisely relevant for the exam, on how they felt that day, while the second is related to elements that affect teacher evaluation, for example the time he spends to correct the proof or whether he is more or less demanding. These errors are not correlated among them: it is plausible to think that stochastic variables affecting student behaviour are not be related with those affecting teachers' evaluations. Moreover, we assume that errors are not correlated to individual ability or to his effort. Formally: $Cov(e,\eta) = Cov(e,\varepsilon) = Cov(a,\eta) = Cov(a,\varepsilon) = Cov(e,\eta) = Cov(\varepsilon,\eta) = 0$.

We assume the following specific functional form for the qualification obtained by individual *i*, v_i[2]:

[1] $$v_i = e_i + a_i + \varepsilon_i + \eta_i$$

The expected value of the qualification is equal to: $E(v) = \overline{e}^* + \overline{a}$ and its variance is $Var(v) = \sigma_a{}^2 + \sigma_\varepsilon{}^2 + \sigma_\eta{}^2$.

The lifetime individual utility function – recalling that individuals live for two periods and that there is no discounting – takes the following simple form: $E(U_i) = E(W_i) - c(e)$. In order to avoid cumbersome analytical expressions, we use an explicit function to represent the cost of effort given by: $c(e) = \gamma e^2/2$, where γ is a parameter measuring the disutility of effort. This explicit function is common in the principal-agent literature (see, for example, Baker, Gibbons and Murphy, 1994; Prendergast, 1999).

Labour market

We assume that the output y_i produced by an individual in the labour market is related to his skills, deriving from his innate ability and on the effort provided during the period he was attending school. Therefore, we suppose that skills are equal to $s_i = e_i + a_i$. Output is then related to skills according to the following production function:

[2] $$y_i = \pi s_i$$

where π is a productivity parameter.

---

[2] In the Appendix we experiment with a different functional form in which ability and effort interact in a complementary way.

With perfect information and perfectly competitive labour markets, the wage W obtained by the individual will be equal to his output: $W_i = \pi(e_i + a_i)$.

However, we assume that firms are not able to observe neither individuals' abilities neither the effort they provided in the educational process, but only observe the evaluation v_i obtained by each student. Therefore, firms seek to infer the effective skills of workers on the basis of the evaluation v_i. This is a typical "signal extraction" problem and firms estimate skills on the basis of the following formulation $E(s_i \mid v_i)$. An important result in statistics is that if two variables are jointly normally distributed, $E(s_i|v_i)$ is a weighted sum of the unconditional mean $E(s_i)$ and the signal v_i. Formally we have that $E(s_i \mid v_i) = \beta_0 + \beta_1 v_i$, where the parameters β_0 and β_1 can be estimated using the standard OLS formulae. It follows that the two parameters β_0 and β_1 are given by the following expressions:

[3] $$\beta_1 = \frac{Cov(s,v)}{Var(v)} = \frac{Cov(e+a, e+a+\varepsilon+\eta)}{Var(e+a+\varepsilon+\eta)}$$

[4] $$\beta_0 = E(s) - \beta_1 E(v)$$

Given our assumptions on variables' variance and covariance, it is possible to show that $Cov(s,v) = Var(a) = \sigma_a^{\,2}$. Therefore:

[5] $$\beta_1 = \frac{\sigma_a^{\,2}}{\sigma_a^{\,2} + \sigma_\varepsilon^{\,2} + \sigma_\eta^{\,2}}$$

[6] $$\beta_0 = (\bar{e}^* + \bar{a})(1 - \beta_1)$$

It follows that the wage paid by employers to each employee depends on the expected skills according to the following function:

[7] $$W_i = \pi E(e_i + a_i \mid v_i) = \pi(\beta_0 + \beta_1 v_i)$$

This equation tells us how information on v_i updates firms expectation on individual skills and, then, how v_i affects wages.

Student behaviour

The expected utility of a student i with ability a_i who provides a level of effort e_i and obtains an evaluation v_i is equal to:

[8] $$E(U_i) = E(W_i) - c(e_i) = \pi[\beta_0 + \beta_1 E(v_i)] - \frac{\gamma e_i^2}{2} = \pi[\beta_0 + \beta_1 (a_i + e_i)] - \frac{\gamma e_i^2}{2}$$

Students decide the level of effort which maximizes their utility function, taking as given how the market rewards skills. By maximizing $E(U_i)$ with respect to effort, we obtain the following first order condition: $\pi\beta_1 - \gamma e_i = 0$, from which the optimal level of effort is:

[9] $$e^* = \frac{\pi\beta_1}{\gamma}$$

The level of effort provided by each student depends on the parameter representing productivity, the cost of effort, and on β_1, the reactivity of wages to skill evaluations. Since these factors are the same for all individuals, they provide an identical level of effort[3].

Substituting eq. [5] in eq. [9] we obtain:

[10] $$e^* = \frac{\pi}{\gamma}\left[\frac{\sigma_a^2}{\sigma_a^2 + \sigma_\varepsilon^2 + \sigma_\eta^2}\right]$$

The optimal level of effort increases when the variance of ε and η are lower (the evaluation system is less affected by stochastic variables) since employers, receiving a better signal of students' abilities, are willing to pay a higher wage premium on grades attained at school. When evaluation is not affected by errors, that is $(\sigma_\varepsilon^2 + \sigma_\eta^2) \to 0$, then $\beta_1 \to 1$ and the optimal effort is equal to the first best level, $e^* = \frac{\pi}{\gamma}$.

On the other hand, if measurement errors increase to infinity, that is σ_ε^2 or σ_η^2 tend to ∞, students' effort decreases, because the evaluation is less informative on individual skills and firms pay a wage based on average abilities, $W = \pi\bar{a}$.

Moreover, if the variance of abilities is higher, then β_1 increases, positively affecting the effort provided by students. This because evaluation is more important when the variability of abilities is higher. In addition, when heterogeneity in individual abilities is high the effect produced by shocks is less relevant and the signal provided by schools is more informative.

[3] In the Appendix we show that when effort and ability are complements the effort provided by individuals depends on their ability. When the evaluation system becomes more precise, effort especially increases for high ability subjects.

The effects of the evaluation system accuracy on student welfare

In this section we evaluate whether students' welfare improves when the evaluation system adopted by the school system becomes more precise.

Considering the optimal level of effort $e^* = \frac{\pi\beta_1}{\gamma}$, the student's expected utility given by eq. [8] can be written as:

[11] $$E(U_i) = \pi\left[\beta_0 + \beta_1\left(a_i + \frac{\pi\beta_1}{\gamma}\right)\right] - \frac{(\pi\beta_1)^2}{2\gamma}$$

Substituting β_0 in [11] we obtain the following:

[12] $$E(U_i) = \pi\left[\bar{a} + \beta_1\left(a_i - \bar{a} + \frac{\pi_1}{\gamma}\right)\right] - \frac{(\pi\beta_1)^2}{2\gamma}$$

We are now able to analyze the effect produced by an increase in the variance of error terms ε and η, respectively $\sigma_\varepsilon{}^2$ and $\sigma_\eta{}^2$, on students' utility. Let us denote the variance of errors ε and η with $\sigma^2 = \sigma_\varepsilon^2 + \sigma_\eta^2$, deriving the individual expected utility with respect to σ^2, we obtain:

[13] $$\frac{\partial E(U)}{\partial\sigma^2} = \left[\frac{\partial\beta_1}{\partial\sigma^2}\pi(a_i - \bar{a})\right] + \left(\frac{\pi^2}{\gamma}\frac{\partial\beta_1}{\partial\sigma^2}\right)\left(1 - \frac{1}{2}\beta_1\right) + \left(\frac{\pi^2}{\gamma}\beta_1\right)\left(-\frac{1}{2}\frac{\partial\beta_1}{\partial\sigma^2}\right)$$

which after some rearrangements becomes:

[14] $$\frac{\partial E(U)}{\partial\sigma^2} = \frac{\partial\beta_1}{\partial\sigma^2}\left[\pi(a_i - \bar{a}) + \left(\frac{\pi^2}{\gamma}\right)(1 - \beta_1)\right]$$

It is possible to show that [14] is negative, implying that individual utility reduces when σ^2 increases, when:

[15] $$a_i > \tilde{a} = \bar{a} - \left(\frac{\pi}{\gamma}\right)(1 - \beta_1) \qquad \text{since } \frac{\partial\beta_1}{\partial\sigma^2} < 0.$$

It follows that students whose ability is above the threshold level $\tilde{a}$ are negatively affected by less accurate evaluation systems, while students with abilities below this threshold are positively affected. In fact, a wage system based on average skills, redistributing resources from high skilled individuals to low skilled ones, tends to favour the latter category.

The threshold value $\tilde{a}$ decreases when the disutility of effort becomes lower. More interestingly, this value decreases when productivity increases, implying that in highly productive economic systems also individuals with relatively low abilities prefer more accurate evaluation systems. Less accurate evaluation systems produce, in fact, two effects. On the one hand, they lead to a more egalitarian pay structure, while on the other hand, they reduce effort and, as a consequence, reduce the total output produced in the economy. When π is high the wage reduction deriving from the lower level of effort tends to counterbalance the positive effect that low ability individuals obtain from the income redistribution deriving from a low value of β_1.

3. A comparison between centralized and decentralized evaluation systems

In this section we compare centralized and decentralized evaluation systems referring to the advantages deriving from repeated observations and costs due to heterogeneous assessment methods. While decentralized evaluations, carried out by teachers during their activity, often represents a by-product of teaching activity, deriving form the continuous interaction with students, centralized exams involve relevant costs and their frequency is limited.

During the educational process, teachers and students interact on daily basis and teachers have local knowledge regarding students' psychological and physical conditions. Both these facts help at reducing the effect of stochastic shocks hitting students on the evaluation they obtain when the examination system is based on grades awarded at local level. On the other hand, teachers may follow different measurement criteria and this introduces noise in student evaluation making less convenient for employers to base their pay systems on observed educational performance. Conversely, central exams adopt a common standard and provide a better signal. However, they are undertaken less frequently and shocks affecting student performance may play a more relevant role in shaping examinations' results.

To analyse this kind of trade-off we assume that with a centralized evaluation system students are evaluated according to a common standard and, as a consequence, the error term η deriving from heterogeneous evaluation methods is not relevant, implying that $\sigma_\eta^2 = 0$. On the other hand, we assume that due to very high administration costs, this type of exam is undertaken only at the end of the educational process and shocks affecting students may influence their performance at exams. It follows that the variance of evaluations awarded by the centralized system is equal to $Var(v) = \sigma_a^2 + \sigma_\varepsilon{}^2$.

When evaluation is at decentralized level, delegated to teachers, it is possible to evaluate student performance a large number of times, which we denote with n. Therefore, effects deriving from stochastic variables related to student performance are reduced, and the variance of this type of error is equal to $\frac{\sigma_\varepsilon{}^2}{n}$. On the other hand, since different teachers adopt different evaluation methods the variance of evaluations is equal to: $Var(v) = \sigma_a^2 + \frac{\sigma_\varepsilon{}^2}{n} + \sigma_\eta^2$.

It follows that the expected utility of individual *i* under a centralized examination system is equal to:

[16] $$E\left(U_i{}^C\right) = \pi\left[\bar{a} + \beta_1{}^C\left(a_i - \bar{a} + \frac{\pi}{\gamma}\right)\right] - \frac{\left(\beta_1{}^C\pi\right)^2}{2\gamma}$$

Instead, under a decentralized evaluation system, the individual expected utility takes the following form:

[17] $$E\left(U_i{}^D\right) = \pi\left[\bar{a} + \beta_1{}^D\left(a_i - \bar{a} + \frac{\pi}{\gamma}\right)\right] - \frac{\left(\beta_1{}^D\pi\right)^2}{2\gamma}$$

The centralized system produces a higher utility compared to the decentralized one when $E\left(U_i{}^C\right) > E\left(U_i{}^D\right)$, which corresponds to the following:

[18] $$\pi\left[\bar{a} + \beta_1{}^C\left(a_i - \bar{a} + \frac{\pi}{\gamma}\right)\right] - \frac{\left(\beta_1{}^C\pi\right)^2}{2\gamma} > \pi\left[\bar{a} + \beta_1{}^D\left(a_i - \bar{a} + \frac{\pi}{\gamma}\right)\right] - \frac{\left(\beta_1{}^D\pi\right)^2}{2\gamma}$$

Simplifying some terms we obtain the following condition:

[19] $$\left(\beta_1{}^C - \beta_1{}^D\right)\left[\left(a_i - \bar{a} + \frac{\pi}{\gamma}\right) - \frac{\pi}{2\gamma}\left(\beta_1{}^C + \beta_1{}^D\right)\right] > 0$$

For individuals with average ability $a_i = \bar{a}$, the previous equation becomes:

[20] $$\left(\beta_1{}^C - \beta_1{}^D\right)\frac{\pi}{\gamma}\left[1 - \frac{\beta_1{}^C + \beta_1{}^D}{2}\right] > 0$$

Since $\beta_1{}^C < 1$ and $\beta_1{}^D < 1$, the term in square brackets is always positive. Therefore, a centralized evaluation system is preferred when $\beta_1{}^C > \beta_1{}^D$, that is when $\frac{\sigma_\varepsilon^2}{n} + \sigma_\eta^2 > \sigma_\varepsilon^2$ (the variance of the decentralized system is greater than the variance of centralized one) and gains obtained by reducing the variance of ε do not compensate the loss deriving form adding in the evaluation the variance of the error term η, typical of the decentralized system ($\sigma_\eta^2 > \sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{n}$).

It is interesting to notice that the difference $\sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{n}$ increases when σ_ε^2 increases, implying that the advantage of decentralized systems is higher when shocks affecting student performance have a higher variance.

For individuals whose abilities are higher than the average ability, $a_i > \overline{a}$, the same result holds. On the other hand, for low ability individuals the term in square brackets in expression [20] may be negative and in this case individuals may prefer evaluation systems with a higher variance.

4. Class size and decentralized evaluation systems

In the previous sections we assumed the number of evaluations undertaken in each evaluation system as exogenously given. However, the frequency of examinations is usually decided in relation to costs and benefits. In this section we show that under decentralized examination systems there is an important relationship among class size and the frequency of evaluations students are required to undertake.

Generally teachers' evaluations are considered costless. Nevertheless if teachers face very large classes it may result difficult for them to judge students on the basis of daily interactions, participation to work-class, etc. When class size increases, the cost of evaluating students increases and teachers may decide to award grades to students on the basis of a lower number of evaluations.

In this section we consider a setting in which under a decentralized evaluation system a central authority is able to define class size, but it is not able to define the number of evaluations students have to undertake. This assumption is quite realistic since even when the authority is able to monitor the number of evaluations administrated by each teacher, he is still unable to ascertain their quality and accuracy. Then, we analyze the choice of the class-size by a policy maker who takes into account teachers' behaviour. We model this choice as a sequential game in which, in the first stage, the policy-maker sets the class-size, while in the second stage, teachers decide how many evaluations to undertake.

Let us assume a school system with N students, C classes and C teachers, where $S=\frac{N}{C}$ is the size of each class. Teachers maximize an objective function which depends positively on rewards that students obtain on the labour market and negatively from the cost of effort provided in teaching and evaluation activities that in our framework is represented by the number of evaluation undertaken for each student. When class size increases this cost increases, more precisely we assume that the evaluation cost suffered by each teacher teaching a class C of size S is equal to nS^2. Teachers decide the number of evaluations, taking as given the number of students they have in the class (class size).

They maximize the following utility function:

[21] $$U^T=\pi\left[\bar{a}+\frac{n\sigma_a{}^2}{n\sigma_a{}^2+\sigma_\varepsilon{}^2+n\sigma_\eta{}^2}\left(\frac{\pi}{\gamma}\right)\right]-nS$$

where the first term represents the wage that the representative student with ability $\bar{a}$ will receive on the labour market and the second term is the cost per student that the teacher suffers in relation to his evaluation activity.

Deriving [21] with respect to n we obtain the following FOC:

[22] $$\frac{\partial U^T}{\partial n}=\pi\left[\frac{\sigma_a{}^2\sigma_\varepsilon{}^2}{\left[n\sigma_a{}^2+\sigma_\varepsilon{}^2+n\sigma_\eta{}^2\right]^2}\left(\frac{\pi}{\gamma}\right)\right]-S=0$$

From which we obtain the optimal number of evaluations:

[23] $$n=\frac{\left(\sigma_a{}^2\sigma_\varepsilon{}^2\frac{\pi^2}{\gamma}\right)^{\frac{1}{2}}-S^{\frac{1}{2}}\sigma_\varepsilon{}^2}{S^{\frac{1}{2}}\left[\sigma_a{}^2+\sigma_\eta{}^2\right]}$$

From [23] it is possible to ascertain that *n* reduces when *S* increases. It is also interesting to notice that the effort teachers provide in their evaluation activity increases when the variance of shocks hitting students increases. In fact, since teachers take care of student welfare, they try to increase their effort in providing better evaluations when students' performance is highly affected by factors behind their control.

On the other hand, when the signalling value of the qualifications awarded by schools is diminished by the fact that schools adopt highly differentiated evaluation methods, teachers' incentives to provide effort in evaluating students is reduced. As a consequence, the positive effect of adopting decentralized evaluation systems, highlighted in the previous section, may result undermined if the variance of shocks concerning teachers and schools is very high.

Since, as shown above, the number of evaluations administrated by teachers to their students is related to the number of students they face, the policy maker in setting the class-size has to consider the negative effect that a large class-size produces on the precision of the evaluation system.

We analyze the choice of the optimal class size when the policy-maker aims to maximize the wage obtained by students on the labour market net of school costs. These costs, denoted with *D*, are due to the wages paid to teachers and to the rental value of the capital associated to each classroom. The cost per student is then given by $CT = \frac{DC}{N} = \frac{D}{S}$. The social welfare function, considering the wage obtained by the representative student with ability $\bar{a}$ net of cost per student, is given by:

[24] $$V = \pi\left[\bar{a} + \left(\frac{\sigma_a{}^2}{\sigma_a{}^2 + \frac{\sigma_\varepsilon{}^2}{n^*(S)} + \sigma_\eta{}^2}\right)\frac{\pi}{\gamma}\right] - \frac{D}{S}$$

Substituting [23] in the social welfare function and making some rearrangements we obtain:

[25] $$V = \pi\bar{a} + \frac{\pi^2\sigma_a{}^2}{\gamma}\left[\frac{\left(\sigma_\varepsilon{}^2\sigma_a{}^2\frac{\pi^2}{\gamma}\right)^{\frac{1}{2}} - S^{\frac{1}{2}}\sigma_\varepsilon{}^2}{\left(\sigma_a{}^2\sigma_\varepsilon{}^2\frac{\pi^2}{\gamma}\right)^{\frac{1}{2}}\left(\sigma_a{}^2 + \sigma_\eta{}^2\right)}\right] - \frac{D}{S}$$

The optimal class size is obtained by deriving V with respect to S, which gives rise to the following First Order Condition (FOC):

[26] $$\frac{\partial V}{\partial S} = \frac{\pi^2}{2\gamma}\left[-\frac{\sigma_\varepsilon{}^2\sigma_a{}^2}{S^{\frac{1}{2}}\left(\frac{\sigma_a{}^2\sigma_\varepsilon{}^2\pi^2}{\gamma}\right)^{\frac{1}{2}}\left(\sigma_a{}^2 + \sigma_\eta{}^2\right)}\right] + \frac{D}{S^2} = 0$$

From [26] we obtain the optimal class size:

[27] $$S = \left[\frac{D}{\pi} \left(\frac{1}{\sigma_a{}^2 \sigma_\varepsilon{}^2 \gamma} \right)^{\frac{1}{2}} 2\left(\sigma_a{}^2 + \sigma_\eta{}^2\right) \right]^{\frac{2}{3}}$$

It is easy to see that when the variance σ_ε^2 increases the optimal class size reduces. On the other hand, when $\sigma_\eta{}^2$ increases it is optimal to define larger classes since the qualifications awarded by the school system are not a good signal of students abilities and as a result are scarcely rewarded on the labour market.

When the cost of education per student increases the optimal class size increases, while when the productivity of skills is higher it is optimal to reduce class size since this has a positive effect on the number of evaluations undertaken by students and hence on the precision of the evaluation system.

In this analysis, for the sake of simplicity, we have neglected the direct effect of class size on human capital accumulation process[4]. In a more general framework, considering this aspect, it is possible to show that in educational systems based on decentralized evaluations, the optimal class size is smaller compared to systems based on centralized evaluations. In fact, while in centralized evaluation systems the optimal class size only depends on the marginal benefit deriving from smaller classes in terms of student achievement and marginal costs related to higher expenditures for wages and rental capital, under decentralized evaluation systems class size also affects how informative evaluations are of individual skill. As shown in this section, larger classes may reduce the frequency of evaluations undertaken by teachers and worsen the informative value of evaluations. Instead, this effect does not play any role in the definition of optimal class size in centralized examination systems.

5. Concluding Remarks

In this paper we analyze a labour market with imperfect observability of workers' skills in which firms use grades (evaluations) obtained by students during the educational career in order to infer their productivity and, hence, determine their wages. Therefore, school grades are used by firms as a signal of abilities. We assume that firms form expectations on individual abilities solving a signal extraction problem.

In this framework, we study the effects of different evaluation systems – characterized by different measurement errors – on the reliability of the signal and, therefore, on the relationship

[4] See for example Lazear, 2001.

between wages and skills at individual level. We then analyze the effect that more precise evaluation systems produce on students effort in studying activities and on their welfare. We show that systems with lower measurement errors encourage students to provide a higher level of effort in studying activities. Employers, receiving a better signal of students' skills, are willing to pay a higher wage premium on the grade attained at school and as a result students are induced to exert more effort.

However, whereas high ability individuals strictly prefer more precise evaluation systems, low ability individuals may prefer less precise evaluations. We show that when labor productivity increases also individuals with relatively low abilities prefer more accurate school performance evaluation systems. This because less accurate evaluation systems on the one hand lead to a more egalitarian pay structure, improving welfare obtained by less able subject, but on the other hand, they reduce effort and, as a consequence, reduce the total output produced in the economy. When labour productivity is high, the wage reduction deriving from the lower level of effort tends to over compensate the positive effect that low ability individuals obtain from more egalitarian pay systems.

Stimulated by a growing theoretical and empirical literature which considers centralized examinations as an instrument to improve students performance and educational quality (Bishop and Wößmann, 2004; Wößmann, 2005 among others) we then use our framework to compare costs and benefits of centralized vs. decentralized evaluations. From the perspective followed in this paper, the advantage of a decentralized system is that it allows repeated evaluations of students performance and, hence, it is able to reduce measurement errors due to shocks hitting students. On the other hand, a centralized system, referring to a common standard, avoids errors deriving from different assessment standards that usually characterized locally graded exams. We show that the advantage of decentralized systems is higher when shocks affecting student performance have a higher variance.

In the final part of the paper we study the relationship between the number of evaluations and the class size under a decentralized system, showing that teachers (or schools), who care of students' welfare, reduce the number of evaluations when class size is higher but tend to increase evaluations if shocks hitting students are more important. A policy maker deciding class size reduces it when individual measurement errors are higher, whereas increases class size if teacher evaluations errors are the main source of errors.

Appendix A

A more general skill function assuming complementariety relationship between ability and effort

The aim of this appendix is to provide a framework in which effort and abilities are complementary with the aim to show that students with different abilities react different to an increase in the precision of the evaluation system. We assume a complementariety relationship between effort and abilities in defining individual skills. More precisely, we assume that skills acquired by students during their permanence at school are equal to $s = ae$.

Firms set wages inferring the real skills of workers on the basis of the evaluation they obtain at school. As a consequence, firms pay a wage $W = \pi E(e + a \mid v) = \pi(\beta_0 + \beta_1 v)$. Given the skill acquisition function, β_1 and β_0 are given by the following:

$$\beta_1 = \frac{Cov(s,v)}{Var(v)} = \frac{Cov(ea, ea+\varepsilon+\eta)}{Var(ea+\varepsilon+\eta)} = \frac{E\left((ea)^2 + ea\varepsilon + ea\eta\right) - E(ea)E(ea+\varepsilon+\eta)}{e^2 Var(a) + \sigma_\varepsilon + \sigma_\eta} = \frac{e^2 Var(a)}{e^2 Var(a) + \sigma_\varepsilon + \sigma_\eta}$$

$$\beta_0 = \left(\bar{e}^* \bar{a}\right)(1 - \beta_1)$$

Taking into account the wage firms are willing to pay for the qualification v_i, the student expected utility is given by:

[1A] $$E(U_i) = \pi[\beta_0 + \beta_1 E(v_i)] - \frac{\gamma e_i^2}{2} = \pi\left[\overline{e}\overline{a}\left(1 - \hat{\beta}_1\right) + \hat{\beta}_1(a_i e_i)\right] - \frac{\gamma e_i^2}{2}$$

Maximizing with respect to effort, we obtain the following first order condition:

[2A] $$\frac{\partial E(U_i)}{\partial e} = \pi\left[\hat{\beta}_1 a_i\right] - \gamma e_i = 0$$

Let [3A] denote an implicit function based on the previous FOC

[3A] $$F = \frac{\partial E(U_i)}{\partial e} = \pi\hat{\beta}_1(e) a_i - \gamma e_i = 0$$

We use the implicit-function theorem to analyse the effect produced by an increase in the variance of a, ε and η on the student effort. We obtain:

[4A] $$\frac{\partial e}{\partial \sigma_j} = \frac{\partial F / \partial \sigma_j}{-\partial F / \partial e} = \frac{\pi a_i \dfrac{\partial \beta_1}{\partial \sigma_j}}{-\partial F / \partial e} \qquad \text{where } j = a, \varepsilon, \eta$$

Since the denominator is always positive when SOC is respected, form [4A] it emerges that student effort reduces when σ_ε and σ_η increase. This result is in line with that discussed in section 3. However, form [4A], it is possible to see that more accurate evaluation systems exert heterogeneous effects on students' performance. For low ability students the incentive effect of more precise evaluation systems is relatively small. In fact, the effects of accurate evaluation systems increase with students' ability.

This result finds support in some empirical analysis showing that the effect of central examination on student performance are lower for low ability students relative to high-ability students (Wößmann, 2005)

References

Baker G., Gibbons R., Murphy K. (1994), "Subjective Performance Measures in Optimal Incentive Contracts", *Quarterly Journal of Economics*, vol. 109, n. 4, 1125-56.

Bishop J. and Wößmann L. (2004), "Institutional Effects in a Simple Model of Educational Production, *Education Economics*, 12 (1), pp. 17-38.

Bishop, J. (1997), "The Effect of National Standards and Curriculum Based Exams on Achievement", *American Economic Review*, 87(2), pp-260-64.

Bishop, J. (1996), "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning", *International Journal of Education Research,* 23 (8), pp. 653–752.

Bishop, J, (2006), "Drinking from the Fountain of Knowledge: Student Incentive to Study And Learn-Externalities, Information Problems and Peer Pressure", *Handbook of the Economics of Education*, North Holland.

Bishop, J. and Mane F. (2001), "The impacts of minimum competency exam graduation requirements on high school graduation, college attendance and early labor market success", *Labour Economics*, 8, pp. 203-222.

Dean L. and De Cicca P. (2001), "Higher standards, more dropouts? across time," *Economics of Education Review,* 20, pp. 459-473.

Dee T. and Jacob B. (2006), "Do High School Exit Exams Influence Educational Attainment or Labor Market Performance?", NBER Working Paper, N. W12199.

Glewwe, P. and Kremer M. (2006) "Schools, teachers, and educational outcomes in developing countries." In Handbook of the Economics of Education, edited by Eric A. Hanushek and Finis Welch. Amsterdam: North Holland:943-1017.

Lavy, V., (2004), "Performance Pay and Teachers' Effort, Productivity and Grading Ethics", *NBER Working Paper* No. 10622.

Jürges H, Büchel F. and Schneider K. (2005), "The Effect Of Central Exit Examinations On Student Achievement: Quasi-Experimental Evidence From TIMSS Germany", *Journal of the European Economic Association*, 3(5), pp. 1134-1155

Lazear E. (2001), "Educational Production", *Quarterly Journal of Economics*, pp.777-808.

Lazear, E., (2006), "Speeding, Terrorism, and Teaching to the Test", *Quarterly Journal of Economics*, 121, pp. 1029-1061.

Prendergast, C. (1999), "The Provision of Incentives in Firms", *Journal of Economic Literature*, vol. 37, n. 1, pp. 7-63.

Wößmann L. (2005), "The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA", *Education Economics*, 13(2), pp.143-169.