SCHOOL GRADING AND INSTITUTIONAL CONTEXTS

VALENTINO DARDANONI, SALVATORE MODICA, AND ALINE PENNISI

Abstract. We study how the relationship between students' cognitive ability and their school grades depend on institutional context. In a simple abstract model we show that unless competence standards are set at above-school level or variation of competence across schools is low, students' competence valuation will be heterogeneous, with weaker schools inflating grades or flattening their dependence on competence, therefore reducing the information content and comparability of school grades.

Using data from the OECD-PISA 2003 Survey the model is applied to a sample of 4 countries, namely Australia, Germany, Italy and The Netherlands. We find that in Australia schools heterogeneity does not affect grading practices; in the other countries grades are inflated in weaker schools, uniformly in Germany and The Netherlands, to a larger extent for weaker students in Italy.

JEL Classification Number I21

Keywords Grades-vs-Competence, Schools' Heterogeneity, External Exams, PISA 2003.

1. Introduction

Evaluation of students' cognitive achievements supports decisions of future employers, parents, school and college boards and policy makers. The measurement of achievements by cognitive tests raises thorny problems, since their validity is controversial, and concern arises that "only what gets measured gets done". School grades, on the other hand, are costless, abundant, frequent, and population-wide;

Date: January 14, 2010.

but to be useful they should accurately reflect underlying competence, since the lower their information content, the higher the signaling noise generated by the sender and the de-codification costs incurred by the receiver. The weaker the signal of student competence, the worse the information that others (e.g. universities and employers) must rely upon in order to make admissions or remedial decisions, and the less grades can be used by schools themselves as a tool to motivate students.

The present paper studies how a country's educational system affects the way grading policy varies across schools. We present a simple theoretical model which shows how a school's grading policy may depend on the distribution of competence of its own students, as it happens for example when teachers 'grade on a curve', in which case weaker schools tend to grant higher grades for given level of achievement. The model investigates the relationship between schools' evaluation and actual competence as a function of some characteristics of the educational system prevailing in the country. More specifically, the model identifies four classes of institutional settings, each one implying a well defined relationship between grades and competence across schools. The broad conclusion that theory points to is that the information on competence which grades contain is determined by the competence distribution across schools and on the institutional setting. In the common case of grading on a curve for example, to acquire information on a student's competence one needs to know not only her grades, but also additional information on the average students' competence in her school, so the link bewteen grades and competence is rather weak.

An econometric model is derived from theory and estimated for a sample of five countries, namely Australia, Germany, Italy and The Netherlands, using the

OECD-PISA 2003 test scores and the information reported in the students' questionnaire on school grades. Using a probit model we obtain, for each country, an estimate of the relationship between the teachers' evaluation of students' competence and its level as measured by PISA test scores, and how this relationship depends on the mean and variance of students' competence within each school. We find that each country corresponds rather closely to one of the institutional settings that the theoretical model identifies. In particular, in the Australian educational system grading policies do not seem to vary significantly at the school level, while in Germany, The Netherlands and Italy there seem to be a rather substantial dependence, for a given level of students' competence, between school grades and school characteristics.

To the best of our knowledge, investigating how the relationship between grading and actual competence varies at the school level, and how it depends on above-schools institutional settings, is rather novel, and we believe that the model proposed here and the empirical results which support it may help to shed light on the connections between school heterogeneity, institutional settings and the informational content of school grades.

In the next section we present the theoretical model. Section 3 contains the empirical findings, and section 4 concludes with some policy reflections.

2. Theory

As mentioned above, our goal is to study how the relationship between cognitive competence and grades varies across schools, in different institutional contexts. To this end we must model the fact that within each country, teachers' evaluations may differ across schools. This we do next.

In a given institutional context $c = 1, \ldots, C$ (we will use 'institutional context' and 'country' interchangeably) there are S_c schools, and in each school $s_c = 1, \ldots, S_c$ there are $n(s_c)$ students with competence levels $x_1, \ldots, x_{n(s_c)}$, which we assume to be independent real random variables extracted from some school-dependent cdf F_{s_c}. Teachers in school s must choose a (possibly s-dependent) valuation $y_i \in \mathbb{R}$ to students with competence x_i, where of course better students should get higher valuations. Thus in each given country, the teachers' task in school s is to choose an increasing valuation map $y = v_s(x)$ which is to be used in their school.

Remarks. (i) Valuations y_i's are ultimately mapped into an actual grade which, depending on the country in which the school operates, typically belongs to a set of ordered categories.

(ii) In this model the distribution of competence across schools is exogenous, with endogenous schools' grading policy. One may object that the causal link may go the other way around: for example, families may know that a certain school has low standards and enrol their children in that school to grant them high grades. While this is logically correct, we neglect such strategic behaviour in the belief that competence within schools is mostly determined by socio-economic factors.

2.1. **Heterogeneity within Countries.** We start by considering how grading policies vary between schools operating in the same institutional context. The subscript c from s_c will be omitted in the discussion which follows.

To model the fact that usually the main issue in grading within schools is what to do with the weak and the strong students, we assume the existence of external constraints which take the form of two reference competence levels $x_c^-(s) < x_c^+(s)$, a low and a high one —which may depend both on the given school and on the country where it operates— for which grades must be fixed at $y_c^- < y_c^+$.

Assumption 1. *School-specific valuations v_s must satisfy the following constraints:*

$$v_s(x_c^-(s)) = y_c^- \ , \quad v_s(x_c^+(s)) = y_c^+ \ . \tag{1}$$

For example, y^- may denote the minimum valuation required for the pass grade, while y^+ may be the minimum valuation required for some higher grade. The two reference competence levels $x_c^-(s), x_c^+(s)$ may be for example quantiles (like in "$n\%$ of the students must pass"), or may be fixed independently of school parameters (like in "To get an A the student must know this and that").

We normalize students' competence level so that in each country it has zero mean and unit standard deviation, and assume that in each school s the low reference competence is below the national average (i.e. $x_c^-(s) < 0$) and the high reference competence is above it ($x_c^+(s) > 0$).[1]

Regarding the choice of v_s we wish to formalize the idea that teachers, when choosing v_s, are constrained by students' perception of unfairness on their part, so that students' relative evaluations must be related to their relative competence. This can be modeled as the requirement that given any two students with competence levels x and x', the difference in their valuations $v_s(x) - v_s(x')$ must be nondecreasing in $x - x'$:

Assumption 2. *$v_s(x) - v_s(x')$ is nondecreasing in $x - x'$ for all $x, x' \in \Re$.*

Our first result is the following:

[1] That the lower and upper tails of the competence distribution have independent non-negligible effects on economic growth has been recently discussed by Hanushek and Wößmann [5].

Proposition 1. *Under Assumptions 1 and 2, there are school-dependent intercept $\alpha(s)$ and slope $\beta(s) > 0$ such that*

$$v_s(x) = \alpha(s) + \beta(s)\, x \tag{2}$$

where $\alpha(s)$ and $\beta(s)$ are given by

$$\alpha(s) = \frac{x_c^+(s) y_c^- - x_c^-(s) y_c^+}{x_c^+(s) - x_c^-(s)}, \;\; \beta(s) = \frac{y_c^+ - y_c^-}{x_c^+(s) - x_c^-(s)}\,. \tag{3}$$

Proof. For arbitrary $x, x' \in \Re$, since $x - x' = (x - x') - 0$ the assumption implies

$$v_s(x - x') - v_s(0) = v_s(x) - v_s(x')\,.$$

Let now $z = -x'$ and use the above equation twice to obtain $v_s(x + z) - v_s(0) = v_s(x) - v_s(0 - z) = v_s(x) + v_s(z) - 2v_s(0)$, that is

$$v_s(x + z) + v_s(0) = v_s(x) + v_s(z)\,.$$

Letting $f(x) = v_s(x) - v_s(0)$, one then has

$$f(x + y) = f(x) + f(y)\,.$$

This is a Cauchy equation, whose only increasing solution is $f(x) = cx$ for some $c > 0$ (Aczel [1], Theorem 1 page 34). Hence $v_s(x)$ is linear as claimed. The constraints in equation (1) can now be used to give a system of two linear equations into two unknowns which can be solved as claimed. □

Linearity of the valuation function in students' competence levels implies that, within each country, schools' heterogeneity affects the valuation process only through the intercept and slope parameters $\alpha(s)$ and $\beta(s)$. The latter depend on c via x_c^-

and x_c^+, so the next step is to investigate how valuation depends on s in different types of institutional contexts.

2.2. **Heterogeneity of Countries.** An institutional context is characterized by the constraints which determine the two reference points $(x_c^-(s), y_c^-)$ and $(x_c^+(s), y_c^+)$ for each school. Recall that the values y_c^- and y_c^+ have been assumed school-independent; on the other hand, even within the same country, each school may be characterized by quite different distribution of students' competence, so $x^-(s)$ and $x^+(s)$ may vary across schools. Thus, in our model there are four possibilities which describe different institutional scenarios: i) $x^-(s)$ and $x^+(s)$ are both s-independent; ii) $x^-(s)$ and $x^+(s)$ are both s-dependent; iii) $x^+(s)$ is s-dependent; iv) $x^-(s)$ is s-dependent. In detail, the four institutional settings can be described as follows:

[A] *Absolute Valuation.* Grades follow common procedures at above-school level. In this case, constraints on school-level grading amount to setting a common scale, that is, $x_c^-(s)$ and $x_c^+(s)$ are fixed independently of school, at $x_c^- < 0$ and $x_c^+ > 0$.

[R] *Relative Valuation.* The proportion of students below y^- and above y^+ –determined by probability levels p^-, p^+ respectively– is fixed above the school level. In terms of the constraints (1), this amounts to having $x_c^-(s)$ and $x_c^+(s)$ determined as the p^--th and p^+-th quantiles. This is equivalent to scenario A if competence distribution is invariant across schools; if on the other hand school populations are heterogeneous, quantiles will generally be lower the weaker the school population.

[AL] *Absolute Lower Bound.* In this case there is a minimum absolute acceptable level of competence required for the valuation y^-; on the other hand, the upper tail (valuations above y^+) is determined in relative terms within each school by

p^+. The formal translation of this case implies x_c^- being school-independent, and $x_c^+(s)$ as being the p^+-th quantile in school s.

[RL] *Relative Lower Bound.* In this case in each school s there is a maximum acceptable fraction of failed students, but the high competence level is fixed in absolute terms. This implies that $x_c^-(s)$ is the p^--th quantile in school s while x_c^+ is fixed.

These specifications need not be determined by written rules; as we shall see, they may be inferred implicitly from analysis of teachers' behavior.

We come to the main purpose of this section, that is to study how the valuation function v_s varies across schools when c belongs to one of these institutional contexts. Given linearity this amounts to studying how, in each different setting, the intercept $\alpha(s)$ and slope $\beta(s)$ vary depending on the distribution of competence levels F_s in the school.

We concentrate on the mean μ_s and standard deviation σ_s of F_s, assuming that higher moments have a negligible effect on the valuation function.[2] It is then convenient to simplify notation further: given identification of s with F_s and the latter with its first two moments (μ_s, σ_s), a school is effectively identified with a pair (μ, σ). In the sequel we shall then write $s = (\mu, \sigma)$.

Using now subscripts for partial derivatives we proceed under the following

Assumption 3. *Let $q^-(\mu, \sigma) < 0 < q^+(\mu, \sigma)$ be the p^--th and p^+-th quantiles of $F_{(\mu,\sigma)}$. Then*

$$(i)\ q_\mu^+ = q_\mu^- > 0\,, \quad (ii)\ q_\sigma^- \leq 0,\ q_\sigma^+ \geq 0\,.$$

[2]In fact, in our application even the second moment is usually not significant.

Recall that $q^-(s) < 0 < q^+(s)$ follows from our assumption that, in all schools, low and high reference competence levels are not above/below the national average, which has been normalized to zero. Assumption 3 says that an increase in average competence in a given school implies a uniform upward shift of the two reference quantiles; and the low (high) reference quantile does not increase (decrease) when the dispersion of competence levels in the school increases. This assumption holds for example when F_s belongs to a family of location-scale distributions with μ and σ as location and scale parameters.[3] The implications of Assumption 3 are described in the next proposition.

Proposition 2. *Under assumption 3, in the four scenarios* [A], [R], [RL], [AL] *the coefficients* α, β *defined in Proposition 1 satisfy:*

[A] $\alpha_\mu = \beta_\mu = \alpha_\sigma = \beta_\sigma = 0$	[R] $\alpha_\mu < 0,\ \beta_\mu = 0,\ \beta_\sigma \leq 0$
[AL] $\alpha_\mu < 0,\ \beta_\mu < 0,\ \alpha_\sigma \leq 0,\ \beta_\sigma \leq 0$	[RL] $\alpha_\mu < 0,\ \beta_\mu > 0,\ \alpha_\sigma \geq 0,\ \beta_\sigma \leq 0\,.$

Proof. We omit the c subscript, and write for example x^+_σ for $\partial x^+_c(\mu, \sigma)/\partial\sigma$. In case [A], α, β are independent of (μ, σ) since x^-, x^+ are. In case [R], x^- and x^+ are the p^--th and p^+-th quantiles of F_s, so $(x^+ - x^-)_\mu = 0$ by assumption 3(i), whence $\beta_\mu = 0$; and $\alpha = y^+ - \beta x^+$, whence $\alpha_\mu = -\beta x^+_\mu < 0$. Finally, $\beta_\sigma = -\beta(x^+ - x^-)_\sigma/(x^+ - x^-) \leq 0$ by assumption 3(ii). Notice that the sign of α_σ is not determined in this case. In case [AL], x^- is fixed and x^+ is the p^+-th quantile, so one easily checks that $\beta_\mu < 0$; and $\alpha_\mu = -\beta_\mu x^- < 0$ from $x^- < 0$.

[3] Indeed, when F_s belongs to a family of location-scale distributions there is a fixed c.d.f. H such that the competence variable, for any school, is distributed as $H((x - \mu)/\sigma)$, so that if z is the p-th quantile, then $z = \mu + H^{-1}(p)\,\sigma$; since $q^-(s) < 0 < q^+(s)$ one has $H^{-1}(p^-) < 0 < H^{-1}(p^+)$, and the claim follows.

Also, $\beta_\sigma = -\beta x_\sigma^+/(x^+ - x^-) \leq 0$ from $x_\sigma^+ \geq 0$; and then $\alpha_\sigma = -\beta_\sigma x^- \leq 0$ from $x^- < 0$. For case [RL] the argument is analogous to the one just given. □

The proposition implies that in institutional settings where [A] holds there is a homogeneous valuation across different school types. This is the benchmark, undistorted system. Its simplest implementation is identifiable with country-level curriculum-based external exit examinations, but we shall see this is not strictly necessary for nationwide standards to emerge. In the other cases, if variation in competence across schools is non-negligible, departure from absolute valuations implies that, within the same country, both the intercept and the slope of the valuation function may become school-specific; in these cases not only valuations in some schools may be uniformly inflated (intercept effect), but also in some schools the less capable students are over-evaluated and the strong ones penalized (slope effect). If these effects are substantial, the grading signal may become much less informative of the students' underlying ability. Quantitative estimates are given in section 3.3 (figure 2 page 17 illustrates).

3. Application to the OECD-PISA 2003 Survey

3.1. **An Estimable Equation.** The theory developed in the previous section provides a framework which can be used for empirical estimation. Proposition 1 shows that, under our assumptions, evaluation v_{is} of student i in school s is a linear function of her competence x_{is}, with school-specific slope and intercept. Taking a first order approximation of α and β with respect to school's mean and standard deviation μ_s and σ_s, and ignoring higher order moments, using again subscripts for partial derivatives for α and β, in each country the valuation of student i in school s in can be written as

$$v_{is} = a + bx_{is} + \alpha_\mu \mu_s + \alpha_\sigma \sigma_s + \beta_\mu \mu_s \, x_{is} + \beta_\sigma \sigma_s \, x_{is} \tag{4}$$

with $b > 0$, while the signs of the other coefficients depend on the institutional setting where schools operate as spelled out in Proposition 2.

Students evaluations may also depend on a vector of student-specific covariates, some of which may be observable by the econometrician while some other may represent residual unobservable heterogeneity. We then augment equation (4) as:

$$v_{is} = a + bx_{is} + \alpha_\mu \mu_s + \alpha_\sigma \sigma_s + \beta_\mu \mu_s \, x_{is} + \beta_\sigma \sigma_s \, x_{is} + \boldsymbol{\gamma}' \boldsymbol{z}_{is} + \epsilon_{is} \tag{5}$$

where $\boldsymbol{z}_i$ is a vector of observable covariates and ϵ_{is} denotes an idiosyncratic error term. In practice students' evaluations v_{is} are not observed, but can be considered as a continuous latent representation of the observed binary variable p_{is}, which takes value 1 if the student obtains a pass grade (that is, $v_{is} > \bar{v}$, where $\bar{v}$ denotes the school-independent valuation level necessary to pass), and 0 otherwise. Under the assumption that ϵ_{is} has a standard normal distribution, it follows that the following probit equation holds

$$\Pr(p_{is} = 1 \mid x_{is}, \mu_s, \sigma_s, \boldsymbol{z}_{is}) = \Phi(a_0 + bx_{is} + \alpha_\mu \mu_s + \alpha_\sigma \sigma_s + \beta_\mu \mu_s \, x_{is} + \beta_\sigma \sigma_s \, x_{is} + \boldsymbol{\gamma}' \boldsymbol{z}_{is}), \tag{6}$$

where Φ denotes the standard normal link.[4] In our application, whose data are described in detail below, p denotes whether the student has obtained a pass grade in mathematics, x is the PISA measured mathematical competence score, μ and σ respectively measure the mean and standard deviation of PISA mathematical competence in each school, and $\boldsymbol{z}$ contains two covariates, namely student's gender

[4]Note that the intercept a in equation (5) is not identifiable since $\bar{v}$ is not observed.

and his/her socio-economic family background. With these specifications, the probit equation (6) is what we estimate. As mentioned above, for estimation purposes the competence variable x is standardized in each country.

3.2. **The Data.** We use data from the 2003 OECD-PISA Survey, which focused on mathematics, cfr. [8]. Competence is measured by a rescaling of the PISA scores (for details on score assignment cfr. [9]). Data on grading for the dependent variable are taken from the (optional) student's educational career questionnaire, which the five countries we consider chose to administer. [5] The PISA socio-economic and cultural background index (SE) combines information on the occupational, educational and cultural environment of students' household.

Of interest for our study is how total variance in competence decomposes in variability between schools (between μ_s's in ou model) and within schools. A large relative weight of between-school variance indicates that the higher performing students are grouped together in the same schools and separated from the lower performing students. In such a situation, if valuation is Relative, the difference in competence between students with same grades coming from different schools may be considerable. On the other hand, when differences between schools are small the Relative Valuation case is closer to Absolute; as remarked on page 7, if in the limit between schools variance is zero, models A, R and RL coincide. In the four countries we are considering there are large differences, with between-school variance being around 20-25% in Australia, while in Germany, Italy and The Netherlands being over 50%. This may be due to existence of early tracking and the implied differentiation between vocational and more comprehensive schools.

[5] Question Q7, variable EC07Q02: "In last school report, how did your mark in mathematics compare with the pass mark?"

TABLE 1. Between-School Variance; % of Pass

Country	Between-School Var., % of Total	% Students Above Pass
AUS	21,1	83,2 (0,7)
DEU	51,7	92,3 (0,6)
ITA	52,2	62,0 (1,0)
NLD	58	72,2 (1,2)

It is instructive to look at the whole distribution of school mean μ besides its variance, and at its relation with within-school variability σ. Indeed, the upper and lower panels of Figure 1, referring to Italy and the Netherlands respectively, reveal two different pictures. The first presents a 'normal' bell-shaped distribution of μ with within-school variance increasing with school quality: there are relatively few good students in weak schools, but good and poor students alike populate the high performing ones. In the lower panel on the other hand, the bimodality of the mean distribution describes a system partitioned in two performance-based school clusters, a story reinforced by the fact that competence variability in the better schools is lower. Germany is similar to The Netherlands, and as we shall see in these two countries the difference between strong and weak schools in terms of grades are the most pronounced; this fact may have its roots in this 'duality' of their school systems. Finally, in Australia the histogram is bell-shaped and the regression line is slightly downward sloping.[6]

The second column of table 1 reports percentages of students with grades above pass, and also in this case there are non-trivial differences: in AUS and DEU

[6]The slope coefficients of the linear regressions of school standard deviation on school mean shown in the figure, and the others referring to Australia and Germany are all significantly different from zero.

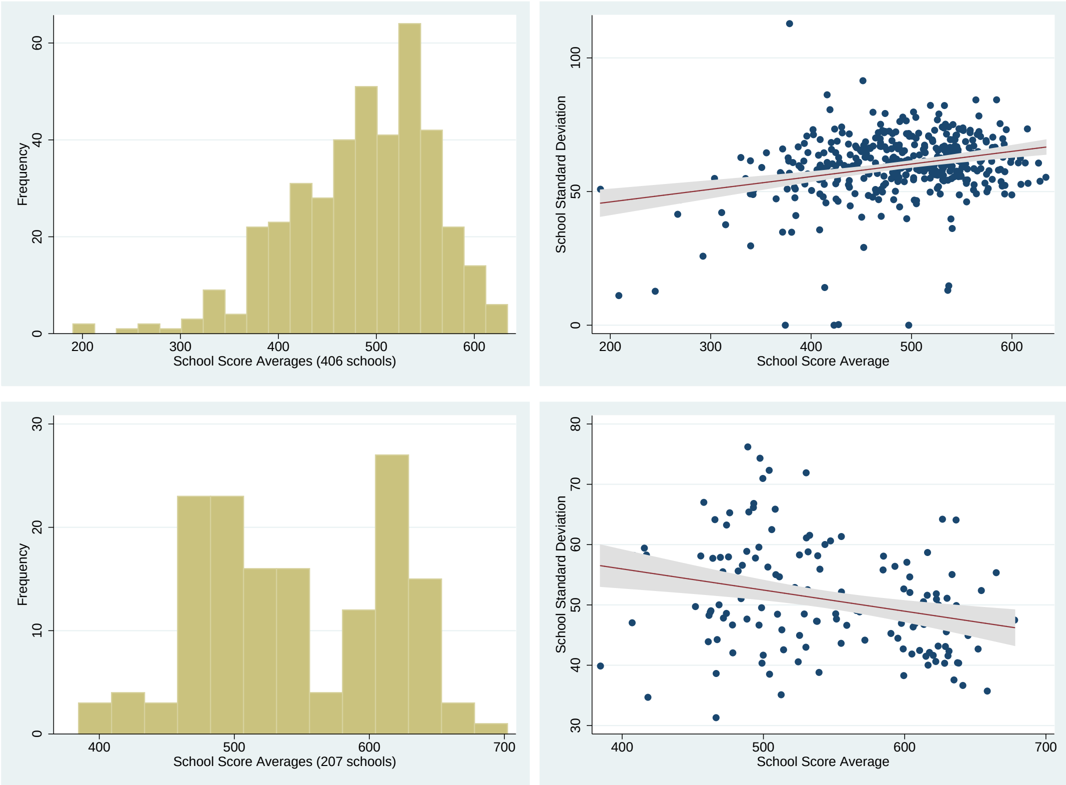

Figure 1. Schools Means and Std: Italy above, The Netherlands below

around 90% of the students are above the pass grade, while in ITA and NLD around 65%. This difference may be due to the different "grades' message space" as we may call it: in ITA and NLD the grade scale is between 1 and 10, with 1-5 being below pass; the others have a grading scale typically made of 5 or 6 different grades with typically the first two grades being below pass. Since the survey is conducted in April-May in all countries so that the 'last mark' is before the final quarter, teachers with a greater choice of below pass grades can send richer warning, work-stimulating messages.

3.3. **Estimation Results.** Our estimation of equation (6), whose results are contained in Table 2, is carried out using the sample weights information given in the OECD-PISA study, and adjusting the standard errors of the estimates to take into account the cluster structure induced by the school level sampling. The reported estimates are obtained using STATA's survey probit weighted ML routine with robust linearized SE.

TABLE 2. Dependent Variable: Probability of Pass

		AUS		**DEU**		**NLD**		**ITA**	
Variable	Coef.	Value	t	Value	t	Value	t	Value	t
x		.597	4.61	.589	2.85	.997	4.18	.565	3.08
μ	α_μ	-.137	-2.22	-.517	-5.91	-.642	-8.72	-.455	-10.23
$\mu\, x$	β_μ	-.037	-0.74	.013	0.27	-.016	-0.32	.0730	2.56
σ	α_σ	-.394	-1.64	-.178	-0.46	-.308	-0.73	1.000	3.95
$\sigma\, x$	β_σ	-.197	-1.39	.011	0.04	-.634	-1.69	.147	0.55
male		-.033	-0.99	.010	0.17	.149	2.90	-.422	-10.85
backgr.		.058	2.13	.070	2.04	-.018	-0.58	.108	5.22
const.		1.389	6.53	1.639	6.24	.739	2.79	-.165	-0.94

On the basis of these estimates we assessed correspondence between model and facts. We first carried out a Wald test, in each country, for the hypothesis that $\alpha_\mu = \beta_\mu = \alpha_\sigma = \beta_\sigma = 0$, i.e. that grading conforms to the hypothesis [A]. This hypothesis is not rejected for Australia, with p-value equal to 0.195. For the other countries, a glance at the table above reveals that in DEU and NLD, among school-specific competence parameters only α_μ has a significant (negative) sign, suggesting that in below-average-competence schools grades may be uniformly inflated. Therefore, grading practices in DEU and NLD are compatible with hypothesis [R]. On the other hand, in Italy there are strongly significant intercept and slope effects as a function of schools' competence heterogeneity, with signs (when significant) compatible with hypothesis [RL].

To evaluate the quantitative impact of these distortions we compare, for each country, the difference in valuation between a high-performing –'good' to abbreviate– and a low-performing –'weak'– school, at a low and at a high level of student performance. Thus we have to specify in terms of μ and σ typical good and weak schools, and two appropriate levels of competence which may be considered representative of good and poorly-performing students. Given benchmark school chosen with μ_s and σ_s equal to their country average, good and weak schools are taken with μ at the 75th and 25th percentiles respectively, with corresponding standard deviations adjusted along the regression line of σ on μ (cfr. figure 1). As to performance, there are six PISA levels in Mathematics, as described in [8] (p.48); we have taken the threshold between the first and the second level —score 420— as low performance, and that between the fifth and sixth —score 670— as high.

The question is then the following: given the estimated valuation of a student scoring 420 in an average school, what is the competence score of a student who receives the same valuation in a good [resp. weak] school? The same question is then repeated for the 670 score. Intuitively, if schools evaluation are relative, in a good school it should take a higher score for any given valuation students have higher competence, so the good school line lies below the other. The results are in figure 2 (the average school lines are not shown), where the lines are drawn on the basis of the coefficients of the probit regression presented in table 2.

A glance at figure 2 reveals that in Germany and The Netherlands the difference in school grading is substantial; given the same teachers' evaluation, there is a full PISA-level difference in competence between good and bad schools, both at the low and high end of the spectrum. Again, the large difference is partially due to the

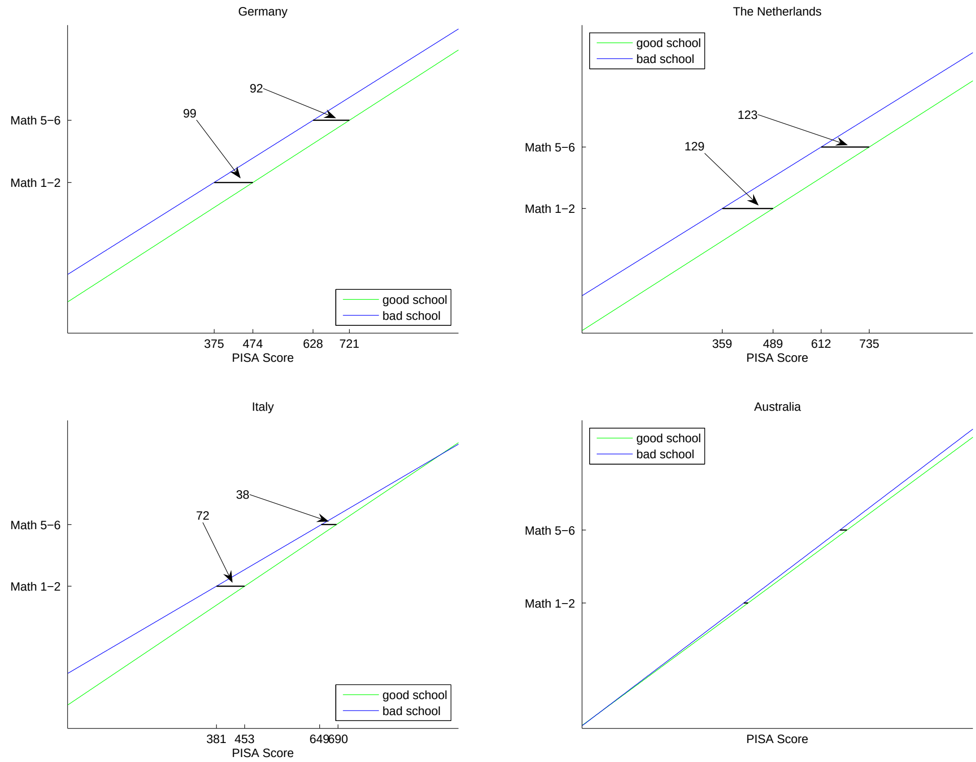

FIGURE 2. PISA Scores needed for valuation corresponding to thresholds between Math Levels 1-2 and 5-6 (420 and 670 in average school)

.

tracking system present in the two countries, and this may alleviate the signaling problem if it is common knowledge that vocational schools follow different grading practices. In the case of Australia the two lines essentially coincide. Italy is the only country in our sample which falls in [RL], the difference being more marked for low than for high levels. In Italy exams are effectively decentralized, and weaker schools are located especially in the South. Thus the 'political' need not too fail too high a fraction of students from poorer areas may determine a school-dependent $x^{-}(s)$, producing higher grades at the bottom end of the distribution. On the other hand there is a strong national cultural tradition, which apparently induces teachers to require high standards from the best students throughout the country.

The detrimental consequence is that the strong students from poorer areas are in the worst position to differentiate themselves from the others through grades.

We close this section by mentioning the gender and socio-economic background effects. In the PISA 2003 survey males tend to perform better than females in mathematics. Somewhat surprisingly, the results in table 2 say that given performance, in some countries male seem to be penalized in terms of grades, and sometimes substantially so. On the other hand, except in The Netherlands, students coming form higher socio-economic background apparently tend to receive higher grades for given level of competence.

4. Conclusions

This paper studies the informational value of school grades as a signal of underlying competence, in different institutional contexts. We spell out in a simple theoretical model four classes of systems which may produce distortions at the school level (such as when weaker schools grant higher grades at given skill levels). In the benchmark case, with competence standards fixed at system level, school grades reflect competence independently of school type. With different patterns of system behavior (e.g. not failing more than a given percentage of students), grades are usually inflated in weaker schools, uniformly or to a larger extent for weaker students.

The theoretical model is applied to data from the OECD-PISA 2003 survey in a sample of 5 countries, namely Australia, Germany, The Netherlands and Italy. According to our estimates, in Australia heterogeneity does not affect grading practices; in the other countries grades are inflated in weaker schools, uniformly in Germany and The Netherlands, to a larger extent for weaker students in Italy.

Implementing system-wide curriculum-based external exit examinations is of course a sufficient condition for system-wide competence standards.[7] According to our empirical estimates it may not be necessary. In the case of Australia for example, competence standards appear to be fixed at system (country) level, but external exams are held sub-system (state) level. In the other cases, the extent of distortion appears to depend on the variance of school quality and possibly on other characteristics of its distribution.

References

[1] Aczel, J. (1966): *Lectures on functional equations and their applications*, Academic Press

[2] Bishop, John H. (1997): "The Effect of National Standards and Curriculum-Based Examinations on Achievement", *American Economic Review* **87** no.2, pp. 260-264

[3] Bishop, John H. (2006): "Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn-Externalities, Information Problems and Peer Pressure", in Eric A. Hanushek, Finis Welch (eds.), *Handbook of the Economics of Education*, Vol. 2, Amsterdam, North-Holland, pp. 909-944

[4] Bishop, John H. and Ludger Wößmann (2004): "Institutional Effects in a Simple Model of Educational Production", *Education Economics* **12** no.1, pp. 17-38

[5] Hanushek, Eric A. and Ludger Wößmann (2007): "The Role of Educational Quality in Economic Development", mimeo April

[6] Kober, Nancy, Dalia Zabala, Naomi Chudowsky, Victor Chudowdsky, Keith Gayler, and Jennifer McMurrer (2006): " State High School Exit Exams: A Challenging Year", Center on Education Policy, Washington, D.C.

[7] Koretz, Daniel and Mark Berends (2001): "Changes in HIgh School Grading Standards in Mathematics, 1982–1992", *RAND Education*, Prepared for the College Entrance Examination Board

[7] Evidence on positive impact of CBEEE on competence is reported in Bishop [2, 3] and Wößmann [10, 11]. Bishop-Wößmann [4] also mention the link with the signalling of academic achievement. Kober et al. [6] warn of the possibility that nationwide standards be too high and raise the drop-out rate.

[8] OECD (2004): “Learning for Tomorrow’s World –First results from PISA 2003”, OECD, Paris

[9] OECD (2005): PISA 2003 Technical Report, OECD, Paris

[10] Wößmann, Ludger (2005): ”The effect heterogeneity of central exams: Evidence from TIMSS, TIMSS-Repeat and PISA”, *Education Economics* **13** no.2, pp. 143-169

[11] Wößmann, Ludger (2007): “Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries”, PEPG 07-02 Harvard University

Facoltà di Economia, Università di Palermo

E-mail address: vdardano@unipa.it

Facoltà di Economia, Università di Palermo

E-mail address: modica@unipa.it

Ministero dello Sviluppo Economico, Roma

E-mail address: aline.pennisi@tesoro.it