

GRADING ACROSS SCHOOLS

VALENTINO DARDANONI, SALVATORE MODICA, AND ALINE PENNISI

ABSTRACT. This paper studies the relationship between students' cognitive ability and their school grades, and in particular whether this relationship changes across schools in a given country at a given time. We use data for Australia, Germany, The Netherlands and USA from the OECD-PISA 2003 Survey. We find, in all cases, a substantial heterogeneity in grading practice across schools. Also, with the exception of the USA, a common pattern emerges: weaker schools give higher grades for any level of competence. Possible explanations are discussed.

JEL Classification Number I21

Keywords Grades-Vs-Competence, Grading Heterogeneity.

INTRODUCTION

Evaluating students' cognitive achievements is key to support decisions of future employers, parents, school and college boards and policy makers. The measurement of achievements by means of cognitive tests raises thorny problems, since no test is perfect, and repeated tests carry the risk that "only what gets measured gets done". School grades, on the other hand, are costless, abundant, frequent, and population-wide; 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.¹

Date: July 21, 2008.

¹The classic works are those by Arrow [1], Spence [9] and Stiglitz [10]. More recently, Costrell [4] argues that detailed information on students' competence increases welfare.

The present paper looks at the information value of grades by exploring whether, in a given country at a given time, grading policy varies across schools. We estimate a multi-level logistic model with fixed effects at school level using data from the OECD-PISA 2003 Survey for four countries which chose to report, in the students' questionnaire, some information on school grades (precisely, whether the student got a pass grade in the last report in mathematics): Australia, Germany, The Netherlands and USA. Our main finding is that grading policy is definitely not homogeneous across schools within each of these countries. Moreover, in all countries except the USA a definite pattern emerges, whereby schools with weaker students tend to give higher grades for given level of competence. This dependence is much stronger in the European countries than in Australia. We speculate that such differences may be due to whether higher performing students are grouped together and separated from lower performing students, to the existence of a vocational versus comprehensive tracking systems, and to the strength of elements of centralization in the exit exams.

RESULTS

For some countries, notably Australia, Germany, The Netherlands and the USA, the 2003 OECD-PISA Survey reports, for each student i in school s , data on competence in mathematics and whether she has obtained a pass grade.² This information gives the rather unique possibility of analyzing the relationship between true competence and grading, and to enquire whether this relationship varies at the school level in different institutional contexts.

²Question Q7 of Student Questionnaire, variable EC07Q02. To save space, for any information and detail on the PISA 2003 Survey we refer the reader to the OECD Publications [7, 8].

Grading Heterogeneity. Student competence in mathematics as measured by PISA score is denoted by x_{is} , and y_{is} denotes the binary variable taking value 1 if she had a pass grade in math.³ For each country we fit the following logistic model with fixed effects at school level:

$$\Pr(y_{is} = 1 \mid x_{is}) = \Lambda(\alpha_s + \beta x_{is}), \quad (1)$$

Λ being the logit link $\Lambda(t) = \exp(t)/(1 + \exp(t))$. A summary of the number of students per school, number of schools and percentage of students with pass grade in each country is in Table 1 below.⁴

Table 1. Summary of Students per School; Number of Schools; % of Pass

Country	Mean	Std. Dev.	Min	Max;	# Schools;	% Pass
Aus	39.30	7.08	5	56	321	83.2
Ger	21.73	3.17	6	25	207	92.3
Nld	25.66	3.19	12	30	150	71.8
USA	21.17	4.77	2	29	274	87.9

The estimates of regression (1) show surprisingly strong fixed effects in all countries, as can be seen from the distribution of the α_s 's as summarized in Table 2 and plotted in Figure 1.⁵ To get a feeling on the quantitative impact of grading heterogeneity within each country, recall for example that a student with $x_{is} = 0$

³PISA Scores, scaled countrywise to an average of 500 and standard deviation of 100, have been re-scaled in (1) below to $(\text{Score}-500)/100$.

⁴From the last column of Table 1 one sees that in The Netherlands the percentage of passes is much lower than in the other countries; this may perhaps be due to the different 'message space' of grades, for in NLD the grade scale is between 1 and 10, with 1-5 being below pass, while the others have a grading scale made of 5 or 6 different grades with typically one grade being below pass. Since the PISA Survey is conducted in April-May, teachers with a greater choice of below-pass grades more easily send work-stimulating messages.

⁵The number of schools in the various countries are lower in Table 2 than in Table 1; the reason is that some schools are not included in the computation of the estimates owing to constancy of the dependent variable.

has a probability pass of 0.5 in a school with $\alpha_s = 0$, 0.73 in a school with $\alpha_s = 1$, 0.88 in a school with $\alpha_s = 2$, and 0.95 in a school with $\alpha_s = 3$. A glance at Table 2 reveals that grading heterogeneity is much more substantial in European countries, with, for example, the standard deviation of the fixed effect being almost twice as much in The Netherlands than in Australia.

Table 2. Summary of Distribution of Intercepts

Country	Obs	Mean	Std. Dev.	Min	Max
Aus	315	1.64	0.64	-0.59	3.60
Ger	154	2.53	1.12	-0.19	4.89
Nld	149	0.53	1.22	-1.75	3.66
USA	221	2.29	0.72	0.55	4.00

Grading on a Curve? If different schools are characterized by different distributions of competences, the desire not to fail too many students in weaker schools and the general tendency of 'grading on a curve' (that is, using a relative valuation system) may imply that weaker schools will tend to inflate grades and pass more students for given level of competence than better ones. In fact, in the case of Germany and The Netherlands (and in Australia, but to a negligible degree) the fixed effects identified above have a distinct common pattern: in weaker schools, for any given competence, a student is more likely to get a pass grade. This fact emerges from regressing, for each country, the intercept coefficient α_s of school s on the median competence z_s of the students in that school:

$$\alpha_s = a + b z_s + \epsilon_s. \quad (2)$$

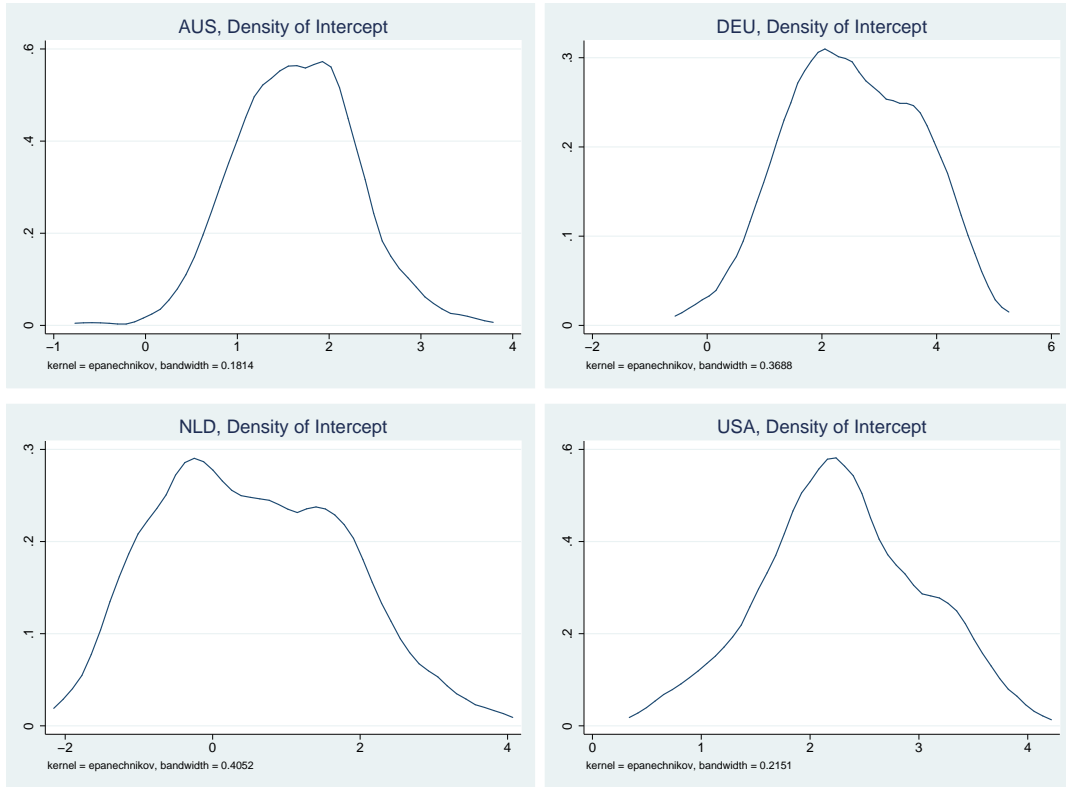


FIGURE 1. Density of Intercepts in Equation (1)

A summary of the slope coefficients b in this equation are presented in Table 3 below. As one can see, in the USA the slope is effectively zero while in the other countries it is negative, in Germany and The Netherlands being more than five times as high as in Australia. We discuss this difference in the next section. A plot of the α_s against z_s and the fitted lines are shown in Figure 2.

Table 3. Slope coefficients in equation (2)

Country	Coef.	Std. Err.	t	P> t
Aus	-0.204	0.0728	-2.80	0.005
Ger	-1.144	0.0844	-13.55	0.000
Nld	-1.341	0.0879	-15.26	0.000
USA	-0.026	0.0953	-0.16	0.784

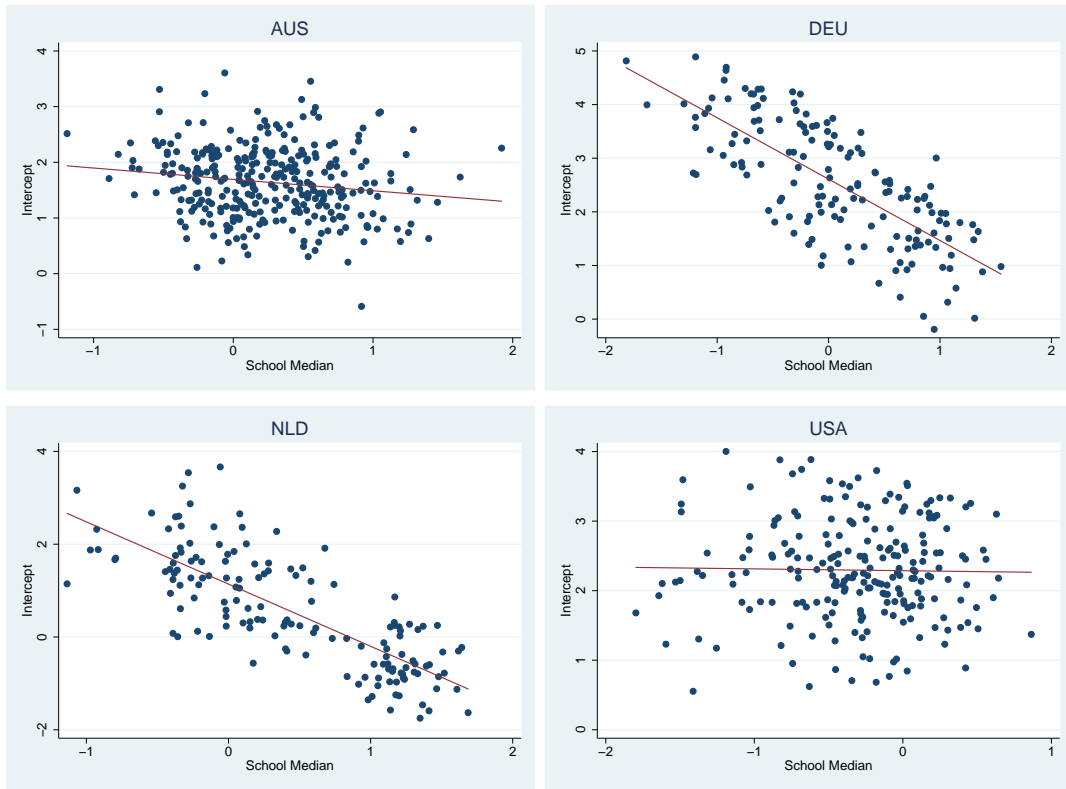


FIGURE 2. Intercepts in Equation (1) against School Medians

To evaluate the quantitative impact of this distortion we first compute from (2) the estimated α_s of a ‘weak’ (low-median) and of a ‘good’ (high-median) school, where weak and good are taken one standard deviation apart from each other in the distribution of school medians in the given country (midpoint being the mean of the distribution); and then, from (1), we compute the probability of pass for a student with competence at the boundary between the 1st and 2nd PISA level (score 420) in the two types of schools. As seen in Table 4, the probability is higher in weak schools. In PISA-Score scale, we may ask how much lower the competence level in a weak school is allowed to be in order to get the probability of pass needed in a good one with a score of 420. The result is in the last column

of Table 4. Recall that a difference of around 60 covers a full PISA-level; on the other hand, a difference of around 10 is negligible.

Table 4. Probability of Pass with PISA Score 420, in Good and Weak School; Score fall allowed in weak school to get pass prob. of good one

Country	Good School	Weak School	Score Fall
Aus	0.713	0.733	11.7
Ger	0.725	0.858	57.8
Nld	0.271	0.493	73.5
USA	0.823	0.825	1.4

DISCUSSION

There are two institutional elements linked to grading, namely centralization and schools' heterogeneity. Heterogeneous grading across schools cannot occur if grades are centralized at country level (it is much limited even if only exit exams are centralized). And if all schools have a similar pool of students, that is, if between-schools variance of competence is low, then heterogeneous grading cannot be due to schools grading on a curve.

In Table 5, along with country-level average score, we report the competence variance between schools as fraction of total (between- plus within-school) variance. Between-school variance is a measure of how much the higher performing students are grouped together in the same schools and separated from the lower performing students. Low between-school relative variance carries with it low heterogeneity.

The relatively weak school-level fixed effects (Table 2) and the lack of correlation between grading practices and the distribution of competences within school (Figure 2) found in the data for the US (and to a large extent also for Australia)

Table 5. PISA Scores

Country	Average Score	Between-School Var / Total Var
Aus	524	0.21
Ger	503	0.52
Nld	538	0.58
USA	483	0.26

might be explained by the simultaneous influence of elements of centralization and low heterogeneity of schools.⁶

Figure 3 below compares distributions of school medians and their relation with within-school variability. In the upper panel, the US presents a ‘normal’ bell-shaped distribution of school-means with within-school variance slightly 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 case of Germany and The Netherlands on the other hand (the latter depicted in lower panel), the pronounced 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. The data show a large grading difformity in the two clusters. In turn, the clusters may be produced by the relatively diffuse tracking system, which may then be at the root of grading heterogeneity in Germany and The Netherlands.⁷

⁶In the USA mandatory exit exams are present in 22 States, see Kober et al. [5], and the tradition of external tests is long (the Scholastic Achievement Test was introduced in 1901). In Australia the situation is similar to the USA, with exit exams standardized at state level in various degrees in the nine States (Masters et al. [6]). Evidence on positive impact of CBEEE on competence is reported e.g. in Bishop [2] and Wößmann [11, 12]. Bishop-Wößmann [3] also mention the link with the signalling of academic achievement.

⁷Actually, the case of Germany is complicated by the presence of regional differences regarding centralization of exams. As reported in [12], 7 of the 16 Landers have external exit exams at the end of secondary school, and with one exception these outperform those where exams are

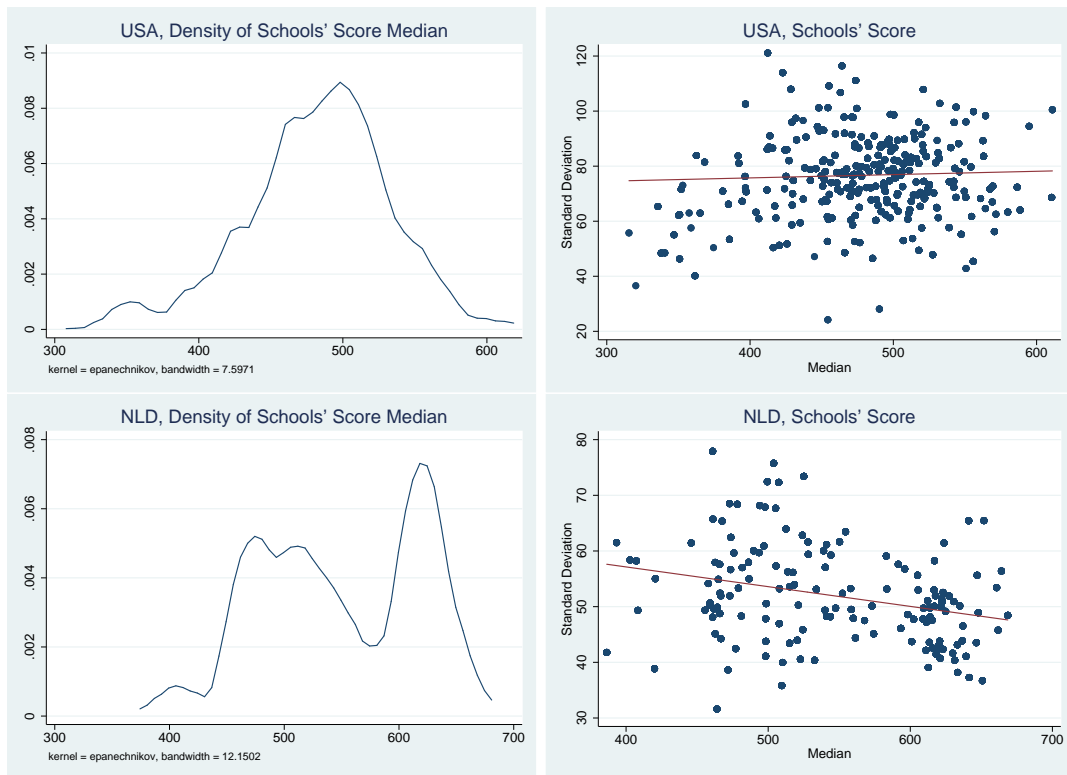


FIGURE 3. Density of Schools Medians and Relation with Standard Deviation

Summing up, it seems that grading distortions in the two pair of countries are different, in Germany and The Netherlands being more pronounced but more predictable than in Australia and the USA. Evaluation of the relative quality of the resulting signals falls outside the scope of this paper.

REFERENCES

- [1] Arrow, Kenneth J. (1973): “Higher education as a filter”, *Journal of Public Economics* **2**, pp. 193-216

graded on a local basis; thus grading heterogeneity could also, at least partially, originate in locally based exams in weak environments.

- [2] 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
- [3] Bishop, John H. and Ludger Wößmann (2004): "Institutional Effects in a Simple Model of Educational Production", *Education Economics* **12** no.1, pp. 17-38
- [4] Costrell, Robert M. (1994): "A Simple Model of Educational Standards", *American Economic Review* **84**, pp. 956-971
- [5] Kober, Nancy, Dalia Zabala, Naomi Chudowsky, Victor Chudowsky, Keith Gayler, and Jennifer McMurrer (2006): "State High School Exit Exams: A Challenging Year", Center on Education Policy, Washington, D.C.
- [6] Masters, Geoff and Gabrielle Matters (2007): "Year 12 Curriculum Content and Achievement Standards", Australian Council for Educational Research
- [7] OECD (2004): "Learning for Tomorrow's World –First results from PISA 2003", OECD, Paris
- [8] OECD (2005): PISA 2003 Technical Report, OECD, Paris
- [9] Spence, Michael (1973): "Job market signalling", *Quarterly Journal of Economics* **87**, pp. 355-74
- [10] Stiglitz, Joseph E. (1975): "The theory of screening, education, and the distribution of income", *American Economic Review* **65**, pp. 283-300
- [11] Wößmann, Ludger (2005): "The effect heterogeneity of central exams: Evidence from TIMSS, TIMSS-Repeat and PISA", *Education Economics* **13** no.2, pp. 143-169
- [12] 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`