The B.E. Journal of Economic Analysis & Policy

Topics

Volume 9, *Issue* 1 2009 *Article* 16

Grading Across Schools

Valentino Dardanoni* Salvatore Modica†
Aline Pennsi‡

*Università di Palermo, vdardano@unipa.it
†Università di Palermo, modica@unipa.it
‡Ministero dello Sviluppo Economico, Italy, aline.pennisi@tesoro.it

Recommended Citation
Valentino Dardanoni, Salvatore Modica, and Aline Pennsi (2009) "Grading Across Schools," *The B.E. Journal of Economic Analysis & Policy*: Vol. 9: Iss. 1 (Topics), Article 16.
Available at: http://www.bepress.com/bejeap/vol9/iss1/art16

Grading Across Schools*

Valentino Dardanoni, Salvatore Modica, and Aline Pennsi

Abstract

This paper reports some facts about grading standards across a varied sample of 16 countries participating in the 2003 OCSE-PISA Survey. Our main finding is that in all countries except Ireland and the USA there is conspicuous heterogeneity in standards across schools (Table 3, Figures 1 & 2). In most of the countries where heterogeneity is present a grading-on-a-curve practice emerges, with grading standards increasing with average competence of the school's students (Table 4, Figures 3 & 4). Where this phenomenon is more pronounced, it may be related to existence of a tracking (as opposed to comprehensive) school system (Table 5, Figure 5).

KEYWORDS: grades vs. competence, grading heterogeneity

*We are indebted to Professor Fiona Scott Morton and the referees for their contribution to the improvement on a previous draft.

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 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 outside the classroom 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]

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 16 countries which report, in the students' questionnaire, information on school grades (precisely, whether the student got a pass grade in the last report in mathematics). Our main finding (last column of Table 3) is that grading policy is definitely not homogeneous across schools in 14 of these countries; this implies that, lacking specific information on each school, the information content of grades as signal of students' competence is probably quite low. In Ireland and the USA, on the other hand, heterogeneity is not found to be statistically significant.

In most of the countries where heterogeneity is present, a grading-on-a-curve pattern emerges whereby schools with weaker students tend to give higher grades for given level of competence (Table 4). Where this phenomenon is most pronounced, we speculate that such differences may be due to the existence of a tracking (as opposed to comprehensive) school system.

In the next section model and main results are presented. In section 3 we describe the common patterns found in the data. Section 4 contains some comments. Tables and figures are collected in section 5.

2 Heterogeneity of Grading Standards

For 20 countries the 2003 OECD-PISA Survey reports, for each student, data on competence in mathematics and whether she has obtained a pass grade

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

in her last school report.[2] This information yields a relationship between competence and grading, which we estimate to see whether and how it varies across schools in the different institutional contexts of the countries in the sample.

For student i in school s, competence in mathematics is denoted by x_{is}, and y_{is} denotes the binary variable taking value 1 if she had a pass grade in maths in her last report. Competence is measured by the average of the student's 5 plausible values in math scores reported in the PISA Survey.[3] As to grades we use the answers to question Q7 of Educational Career Questionnaire (variable EC07Q02), which asks whether grade in the student's last report was above the pass grade. Within the 20 countries which report this information, four (namely the Check and Slovackian Republics, Greece and Thailand) have an average pass grade greater than 95%; we decided to exclude these, so that our final sample is made of 16 countries. Finally, students' weights are not used.

For each country we fit the following logistic model with fixed effects at school level:

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

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.

Estimated β's are reported in Table 2. Estimates of the intercepts show substantial fixed effects in all countries, as can be seen from the distribution of the α_s's summarized in Table 3 and plotted in Figures 1-2.[4] A Wald test for the homogeneity of the fixed effects is strongly rejected for all countries, with the exception of Ireland and USA (cfr. last column of Table 3). To get a quantitative feeling of the impact of between school grading heterogeneity, recall from the logit link that for example a student with $x_{is} = 0$ (average

[2] PISA seeks to measure the ability of 15-year old to use their knowledge and skills to meet real-life challenges in a variety of situations. One of its key features –compared to other international assessments focusing on curricula– is to measure what students can do with rather than just what they have learned. Information on pass grades and, in some case on the exact grade score obtained in the classroom, is self-reported by students. We have not included cases which reported only on single grade scores because of the country differences in grade scales and because greater reliability is expected as regards pass mark.

[3] For details on the Survey we refer the reader to the OECD Publications (7; 8). In the Survey scores are scaled to an overall average of 500 and standard deviation of 100; we standardize them by subtracting 500 and dividing by 100.

[4] The number of schools in the various countries are lower in Table 3 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.

competence given centering) 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 $\alpha_s = 3$. Thus, a glance at the table reveals that in most countries grading is definitely heterogeneous, with wide school-level variations in grading practices. In the following section we investigate the school level correlation between grading standards and average competence in schools.

3 Heterogeneity of Standards and Average School Competence

3.1 Grading on a Curve

Grading on a curve means grading relatively to class mates, which implies that schools with better students set higher standards (i.e. give lower grades for given competence), or looking at the other tail, that weaker schools tend to inflate grades, perhaps not to fail too many students. In our sample a significant positive relation between school's median competence and grading standards occurs in 11 of the 14 countries where heterogeneity is present.

Recall that higher values of the intercept coefficient α_s of school s in the logit equation (1) reflect higher grades for given competence, i.e. low standards. Thus, defining $grading\_standards_s = -\alpha_s$, the positive relation in discourse is a positive relation between this variable and the school's students' median competence, which we denote by $median\_comp_s$. In the following regression we add the school's students' median socio-economic background (defined as the median value of the standardized PISA variable ESCS in the school, denoted $median\_bg_s$ for school s) to control for independent family pressure;[5] the estimated relation is

$$(2) \qquad grading\_standards_s = a + b \cdot median\_comp_s + c \cdot median\_bg_s + \epsilon_s \, .$$

Estimates of the slope coefficients in this equation are presented in Table 4 (we report also the cases of Ireland and USA for completeness). As anticipated, in all countries where grading is heterogeneous –with the exception of Island, Poland and Uruguay– median competence is significantly associated with grading standards; plots of grading standards against median competence are displayed in Figures 3-4. Concerning background, parents with

[5] We thank a referee for suggesting this to us. The PISA socio-economic and cultural background index (ESCS) combines information on the student's highest parental occupational status, highest parental education level, and home educational and cultural possessions (students are asked whether they have a room of their own, a desk to study, a computer at home, internet link, a calculator, classic literature, books of poetry, educational software, etc.)

higher socio-economic status might put pressure to raise grades, or on the contrary to raise standards; confirming this intuition, in our sample the effect of parental pressure is mixed; in most cases it is not statistically significant.

To evaluate the quantitative impact of the grading-on-a-curve phenomenon in the 11 countries where it is present we first compute from (2) the estimated grading standards of a low– and of a high–median-competence school, fixing schools' median social background at its zero mean. Low and high median-competence are taken one standard deviation apart from each other, midpoint being the mean of the distribution of school medians in the given country. 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.[6]

As seen in the first two columns of Table 5 the probability is generally higher in low-median schools. In PISA-Score scale, we may ask how much lower the competence level in a low-median school is allowed to be in order to get the probability of pass needed in a high-median one with a score of 420. The result is in the last column of Table 5. In Austria, Germany, Mexico and the Netherlands the difference is in the order of a full PISA-level. Notice that in the case of Germany and the Netherlands at least, the significant difference in standards is associated with a 'dual' system in which schools are in two performance-based clusters, which may be produced by the relatively diffuse tracking system.

The dual system referred to above is described in the upper panel of Figure 5, which depicts the case of The Netherlands. The distribution on the left is that of schools' median competence, whose pronounced bimodality reveals the two school clusters. The structure of these clusters is further clarified by the inverse relation on the right panel between median and standard deviation of competence at school level, which shows that competence variability in the better schools is lower, in other words that good students are concentrated in good schools. What the data show is a large grading difformity in the two clusters.

[6] In the Survey math competence is classified in 6 levels with lower boundaries spaced at 60 points from one another, the first and lowest starting at 360. A little more than 1/3 of the population is in the first level. Students with competence 420 are representative candidates for a failing grade.

3.2 Unstructured Heterogeneity

In 5 countries there is no significant relationship between grading standards and school performance (Table 4 and Figures 3-4).[7] The typical structure in these cases is that of the USA, displayed in the lower panel of Figure 5, with a bell-shaped distribution of school medians and within-school variance slightly increasing with school quality: in other words, there are relatively few "good" (i.e., high-competence) students in "weak" (i.e., low-performing in average) schools, but good and weak students alike populate the high performing ones.

To briefly comment on these cases we observe that in general 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). On the other hand, 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.[8]

For a country like the US, absence of significant school-level fixed effects (Table 3) and of correlation between standards and schools' median competence (Table 4) might be explained by the simultaneous influence of elements of centralization and low heterogeneity of schools.[9]

4 Final Comments

1. We have found that the informational content of grades as signal of competence is generally low. In principle, prospective employers may overcome this lack of information by combining students' grades with information on their school of origin (in terms of our model, on the school fixed effect). But such information is in general casual and locally circumscribed, hence limited in usefulness. In addition, significant heterogeneity in grading standards character-

[7]It would be interesting to see to what extent the results we have found are confirmed in other datasets. For the USA for example, as suggested by a referee one might use SAT scores and high school GPA's.

[8]The variance of competence between schools expressed as a percentage of the total variance within the country is on average equal to 35.7% in our sample. In the five countries where either there is no grading heterogeneity or grading standards do not depend on school's median competence, this percentage is equal to 15.9% (Ireland), 3.8% (Iceland), 12.6% (Poland), 43.8% (Uruguay), 15.9% (USA).

[9]In the USA mandatory exit exams are present in 22 States, see Kober et al. (6). Evidence on positive impact of CBEEE on competence is reported e.g. in Bishop (2) and Wößmann (12; 13). Bishop-Wößmann (3) also mention the link with the signalling of academic achievement.

izes a system at far distance from one with centralized exams, which according to recent research, see e.g. (11; 12), promotes higher students' achievements in terms of cognitive competence through more conducive incentives to teachers. We are therefore inclined to conclude that the dispersion in standards to the degree which data for most countries reveals is somehow excessive.

It must be observed that sending good signals about students' competence to the job market is not part of the teachers' job. Grades are fundamental multipurpose teaching instruments, and the heterogeneity we document demonstrates that they are indeed used with great elasticity. If one wishes to force them to also convey more circumscribed information about students' competence for use outside the school, the natural route is probably to increase the weight of external exams.

2. In talking of 'grading on a curve' we implicitly claim that students' competence in a school determines grading standards, but as confirmed e.g. by Figlio and Lucas (5) (on U.S. data) the reverse causal link may also be present, the idea being that setting tight standards stimulates learning. In the picture we have in mind, setting high standards may well be an effective instrument to motivate students and stimulate their progress, but it is one that teachers of schools with poor background students cannot afford to use.

5 Tables and Figures

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

Country	Mean	Std. Dev.	Min	Max;	# Schools;	% Pass
Aus	39.06	7.11	5	56	321	83.2
Aut	27.53	6.20	2	35	181	93.5
Ger	21.54	3.23	6	25	207	92.3
Hun	28.62	7.54	2	36	215	92.9
Idn	29.57	5.13	3	35	346	83.2
Irl	26.83	4.38	3	34	145	85.5
Isl	46.77	32.99	2	137	129	79.9
Ita	30.42	4.24	2	35	399	66.4
Lva	30.15	8.64	2	45	107	86.7
Mex	27.02	6.03	2	36	1107	77.8
Nld	25.49	3.34	12	30	150	71.8
Pol	27.41	4.40	2	35	165	75.8
Prt	31.00	5.94	2	40	153	56.6
Ury	25.62	7.10	2	35	241	64.7
USA	21.09	4.79	2	29	274	87.8
Yug	28.89	3.81	9	35	149	85.9

Table 2. Slope coefficients in equation (1)

Country	Coeff.	Std. Err.	t
Aus	0.8465	0.0343	24.66
Aut	1.1310	0.1198	9.44
Ger	1.4494	0.1134	12.78
Hun	1.1368	0.1201	9.47
Idn	0.7490	0.0616	12.15
Irl	0.8982	0.0733	12.26
Isl	1.7352	0.0832	20.86
Ita	1.2521	0.0406	30.82
Lva	1.9122	0.1200	15.93
Mex	0.8636	0.0307	28.15
Nld	1.3053	0.0837	15.60
Pol	2.0484	0.0781	26.21
Prt	1.2925	0.0593	21.80
Ury	0.6467	0.0472	13.71
USA	0.9361	0.0659	14.20
Yug	1.2283	0.0873	14.07

Table 3. Summary of Distribution of Intercepts

Country	Obs	Mean	Std. Dev.	Min	Max	Wald, p-value
Aus	315	1.65	0.65	-0.58	3.89	0.0000
Aut	113	2.41	1.08	0.39	4.49	0.0003
Ger	153	2.52	1.12	-0.20	4.85	0.0263
Hun	114	2.79	0.99	0.46	5.17	0.0000
Idn	330	2.75	0.83	0.10	4.77	0.0000
Irl	139	1.98	0.75	0.32	3.95	0.1566
Isl	110	1.46	0.99	-1.22	3.86	0.0000
Ita	388	0.94	0.91	-1.18	4.03	0.0000
Lva	91	2.82	0.99	0.33	5.29	0.0000
Mex	1010	2.23	1.08	-1.25	5.08	0.0000
Nld	149	0.54	1.22	-1.75	3.66	0.0000
Pol	154	1.93	0.84	0.06	4.62	0.0000
Prt	152	0.77	0.70	-1.03	3.24	0.0000
Ury	234	1.25	0.82	-1.30	3.75	0.0000
USA	221	2.29	0.72	0.55	3.89	0.7399
Yug	127	2.89	1.00	0.68	5.00	0.0000

Table 4. Slope coefficients in equation (2); dependent variable grading_standards

Country	median_competence coef	median_competence t-value	median_background coef	median_background t-value
Aus	.4688323	3.86	-.3463619	-2.74
Aut	.841302	5.70	.5337136	2.67
Ger	1.148178	7.54	-.0115183	-0.06
Hun	.587917	2.28	.1264128	0.39
Idn	.4905536	5.17	.0213621	0.27
Irl	-.204347	-0.88	.2972982	1.59
Isl	-.2306305	-0.79	-.371003	-1.78
Ita	.386022	4.37	-.2394218	-2.33
Lva	.6630175	2.63	.0749355	0.26
Mex	.9658953	11.67	.1542397	2.44
Nld	1.139498	7.06	.3481282	1.46
Pol	.081821	0.34	.5954776	2.74
Prt	.335571	2.38	-.0018985	-0.02
Ury	-.0736425	-0.61	.0119107	0.10
USA	.098694	0.76	-.1191441	-0.96
Yug	.5970497	2.56	.0640778	0.26

Table 5. Probability of Pass with PISA Score 420, in High- and Low-Performing School; Score fall allowed in latter to get pass probability of former

Country	High-Perf.Sch.	Low-Perf.Sch.	Score Fall
Aus	0.701	0.747	26.94
Aut	0.774	0.857	49.29
Ger	0.720	0.855	57.30
Hun	0.842	0.884	31.85
Idn	0.883	0.908	36.15
Ita	0.456	0.524	21.82
Lva	0.755	0.806	15.72
Mex	0.784	0.854	55.24
Nld	0.286	0.475	62.42
Prt	0.412	0.455	13.65
Yug	0.852	0.887	25.42

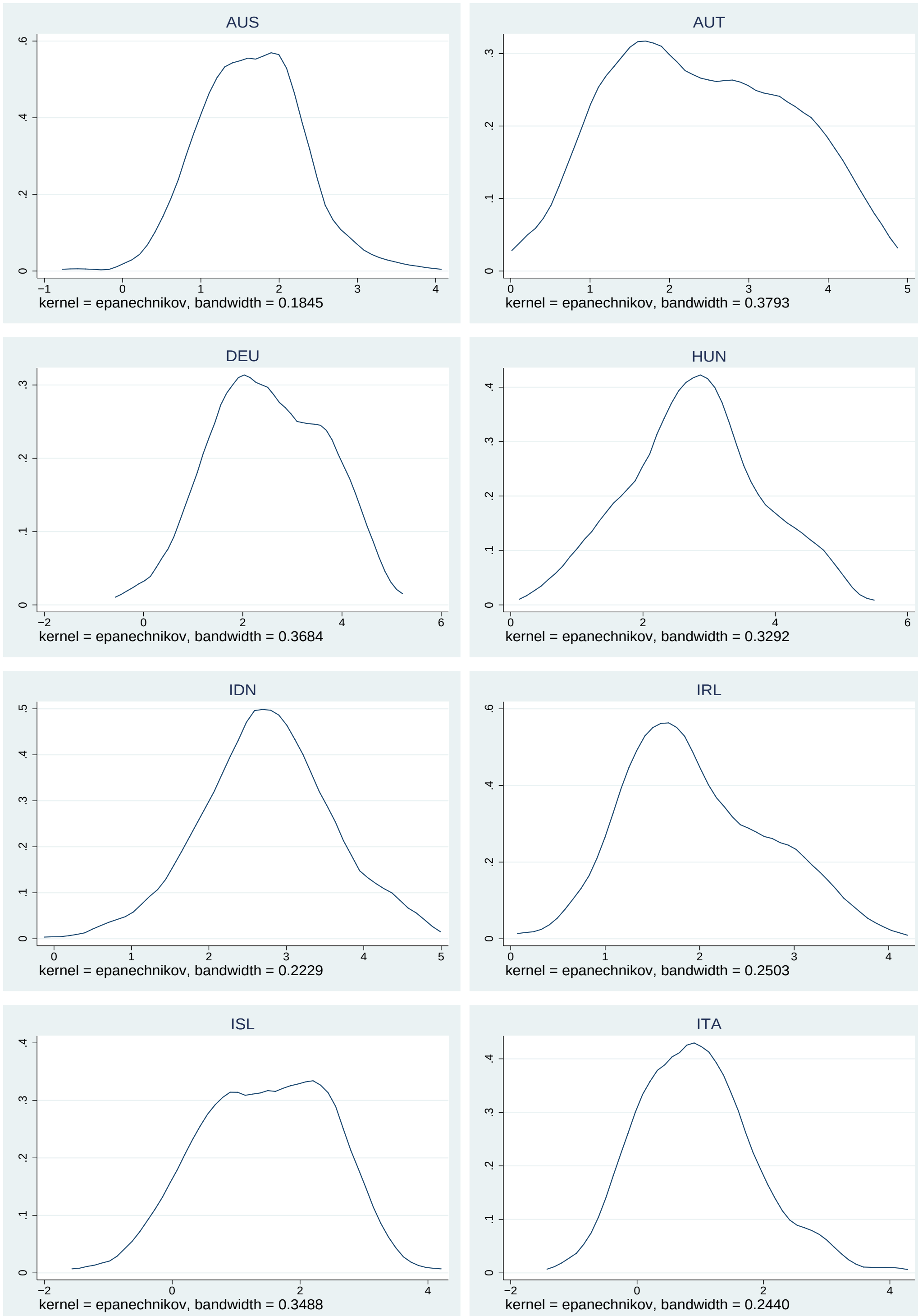

Figure 1: Density of Intercepts in Equation (1)

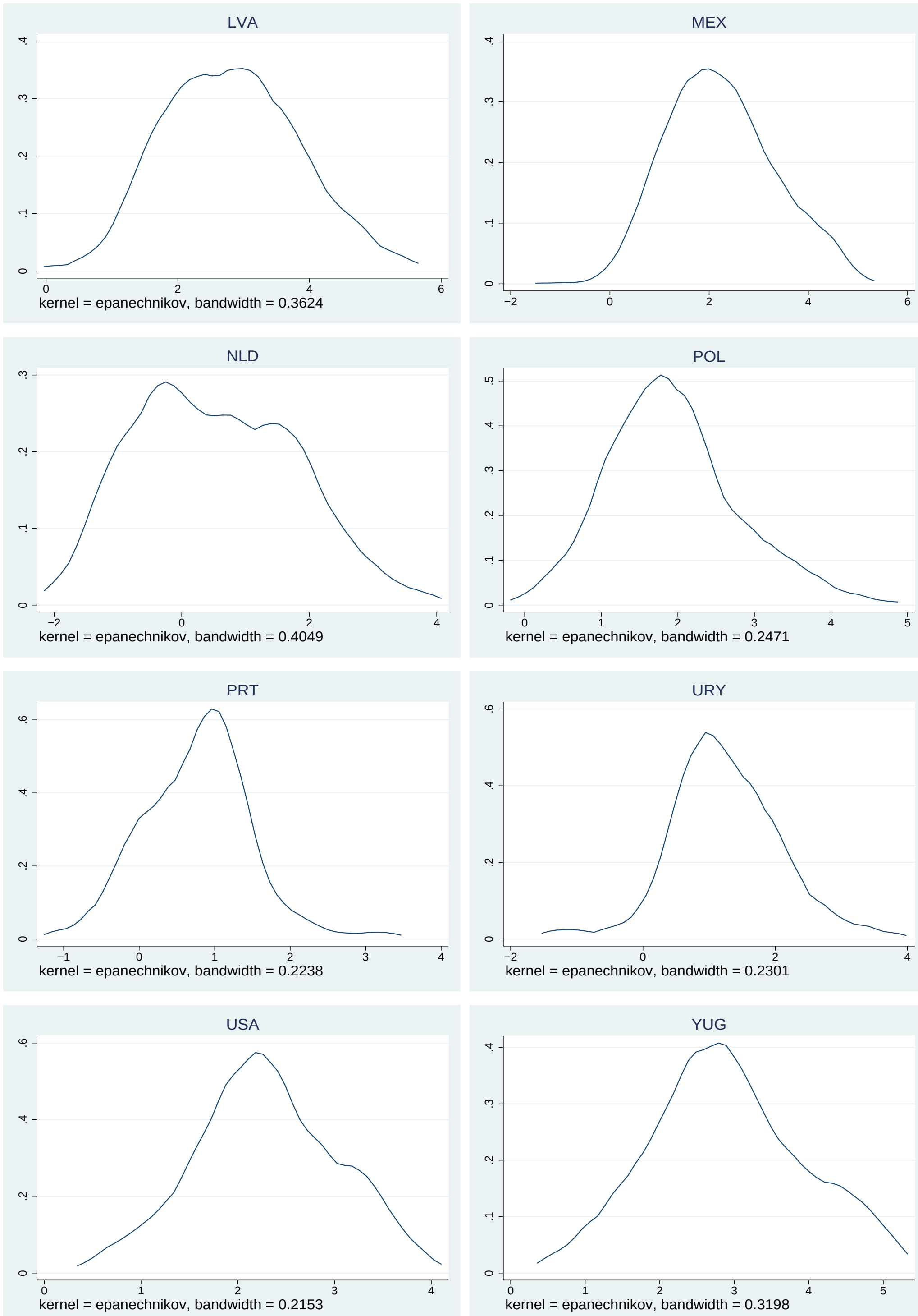

Figure 2: Density of Intercepts in Equation (1), Continued

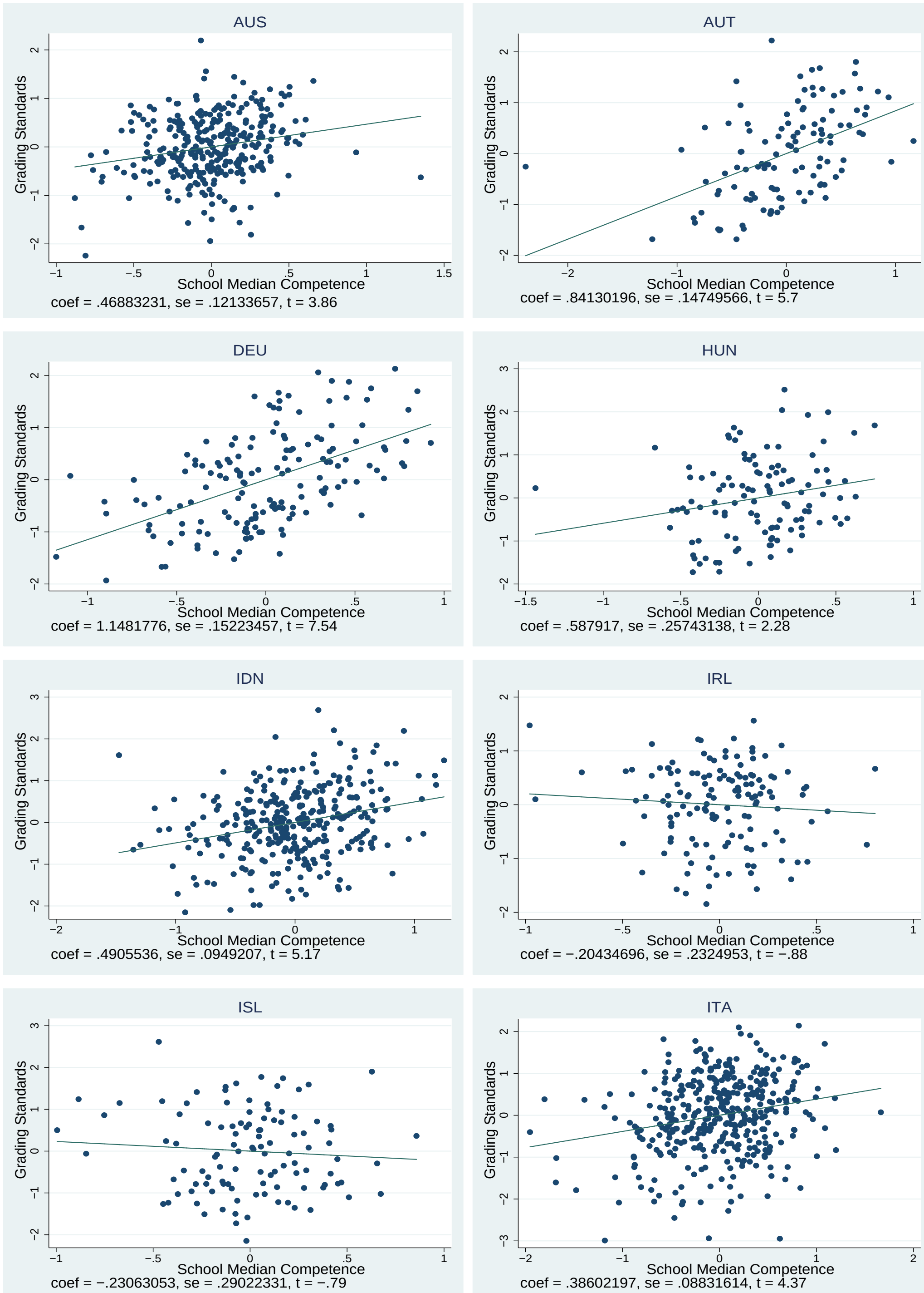

Figure 3: Grading Standards and Schools' Median Competence (both standardized in figure).

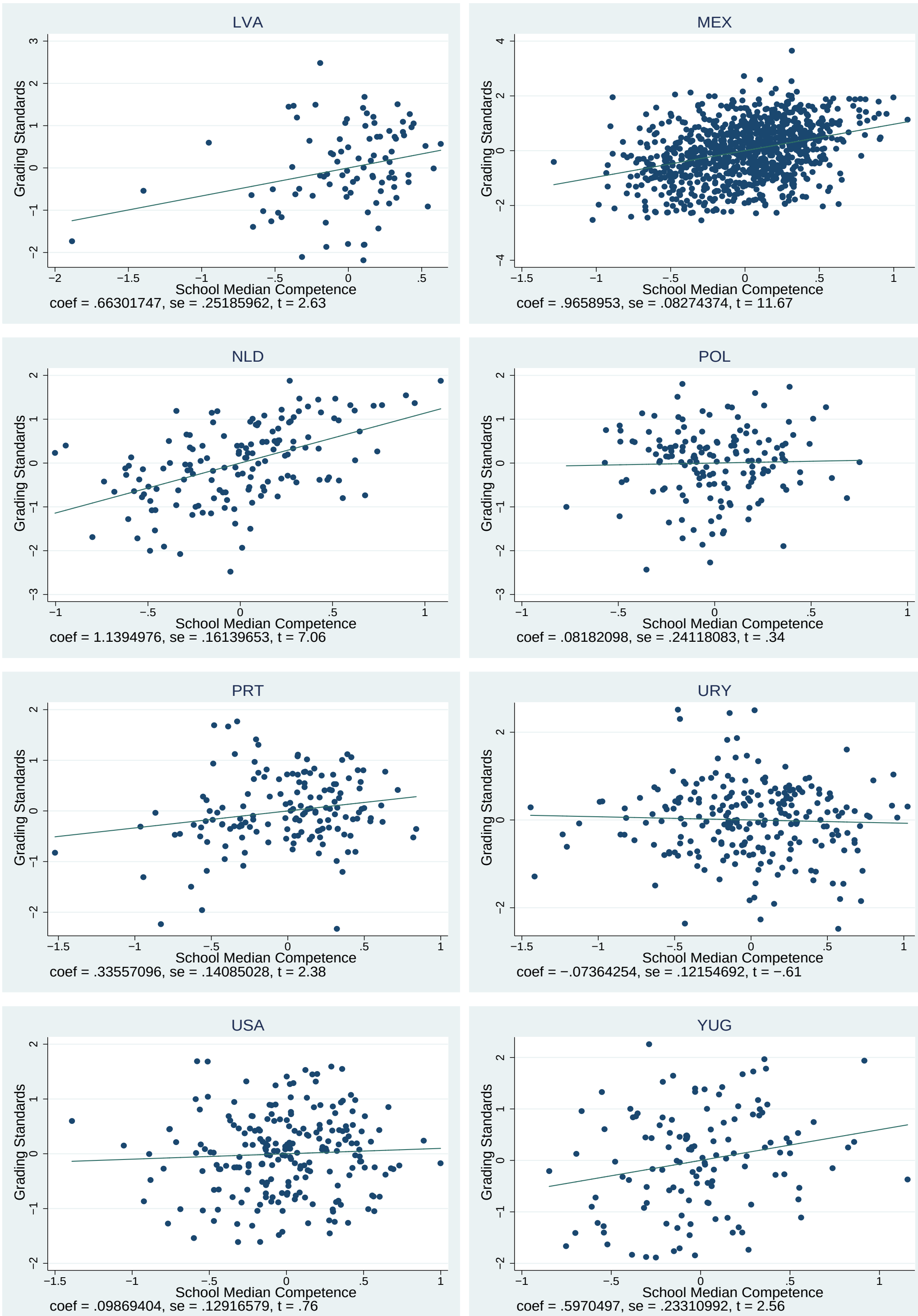

Figure 4: Grading Standards and Schools' Median Competence, Continued

Figure 5: Density of Schools Medians and Relation with Standard Deviation: The Netherland above, USA below

References

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

[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] Figlio, David N. and Maurice E. Lucas (2004): "Do high grading standards affect student performance?", *Journal of Public Economics* **88**, pp. 1815-1834

[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] 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 (2003): "Central Exit Exams and Student Achievement: International Evidence", in Paul E. Peterson, Martin R. West (eds.), *No Child Left Behind? The Politics and Practice of School Accountability*, Washington D.C., Brookings Institution Press.

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

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