Do teachers grade immigrants like they grade native children?

Lex Borghans* Lara Lebedinski†

Maastricht University IRES (UCLouvain)

Abstract

In Italy immigrant children have lower educational attainment and poorer academic performance than their Italian peers. An important policy question is whether these differences are reflected in the grading of teachers or that teachers reveal lower expectations for this group by giving them grades that are higher than justified based on performance. In this paper we analyse the difference in grading for math between immigrant children and their Italian counterparts, using PISA test scores as an anchor for the comparison. The main finding is teachers give higher grades in mathematics to immigrant children than native children with the same PISA score. Differences are especially large for children that do well in math.

*Email: lex.borghans@maastrichtuniversity.nl

†Email: lara.lebedinski@uclouvain.be

1 Introduction

Grades at school can have a motivational effect. If teachers have lower expectations for a specific group, they could under-value their knowledge in terms of grades. This can induce a response from the under-appreciated group through the so-called Pygmalion effect. Lower (higher) teacher expectations have been shown to induce students to perform worse (better)(Rosenthal and Jacobsen, 1968).Understanding what is happening at the level of the school and the causes of what is keeping immigrant children behind is a central question of the integration process.

The aim of this paper is to investigate whether teachers grade immigrant children differently compared to native children with the same knowledge? The analyses are based on a specific feature of the Program for International Student Assessment (PISA) data set that this data set contains objective test scores (PISA test scores) and subjective teacher assessment for each child. PISA tests are graded anonymously, whereas marks at school are, to a certain degree subjective, because they are given by a teacher who knows the pupil and its socio-economic background. The main finding is teachers give higher grades in mathematics to immigrant children than native children with the same PISA score. Differences are especially large for children that do well in math.

The performance of immigrant children on standardised tests as compared to their native peers differs largely among countries (Schnepf, 2007). While in some of the traditional immigration countries such as Australia and Canada, there are no significant differences between immigrant and native test scores, in others, such as US, the language barrier appears to be the largest impediment for foreign born pupils. However, in continental European countries (France, Netherlands, Switzerland) immigrant children face a disadvantage even after controlling for language skills and socio-economic background which are thought to be the main drivers of the native-immigrant educational gap.

Teacher bias in grading has been examined in several studies. Lavy (2008) challenges the widely held belief that teachers favour boys. By exploiting a natural experiment and by comparing blind and non-blind test scores, his findings on Israel suggest the opposite: teachers are actually over-valuing female abilities in both humanities and science. But biased grading does not have to originate necessarily from school teachers. Hanna and Linden (2012) demonstrate that even evaluators, who do not know the children, could be biased. In an experiment, evaluators were asked to grade tests with randomised children characteristics (age, gender and caste) (Hanna and Linden, 2012). Exams that were assigned to be from children from a lower-caste were graded significantly worse than exams assigned to be from higher-caste children. The behaviour of teachers was consistent with statistical discrimination. Graders discriminated children who were graded early in the examination process when the grading instrument or grade distribution was uncertain.

A related study by Kiss (2013) asks a similar question to ours, but focuses on second-generation immigrants in Germany.[1] While we appreciate the importance of educational outcomes of second generation immigrants, we hold that a successful integration of first generation immigrants is even more important and tackles the problem of integration earlier and at its core. What is surprising, we find that first generation immigrants actually get better grades conditional on PISA score while Kiss (2013) finds the opposite.

2 Data

2.1 PISA survey

PISA survey is a triennial survey which assesses the knowledge of 15 year old school pupils in mathematics, science and reading in a selected sample of countries. In each PISA round the test targets one specific subject. For this study we use 2003 when mathematics was the primary subject of examination.[2]

[1] Second generation immigrants are defined as immigrants born in the receiving country, but with at least one parent born abroad.

[2] In 2006 the main subject of examination was science, 2009 it was reading

National Programme Managers recruited qualified so-called coders to grade PISA exams. There was no explicit requirement for them to have tertiary degree, but they had to have a knowledge of mid-secondary level mathematics and science or the language of the test. Graders were expected to be familiar with ways secondary-school pupils express themselves. Suitable coders were considered teachers on leave, recently retired teachers and senior teacher trainees. Coders were expected to dedicate up to one month to grading. The front page of each exam contained the student identification number, but not the name of the student or any demographic characteristic.

2.2 Summary statistics

Summary statistics of the variables are provided in table 1. The average teacher mark is 6. Children in Italy have an average performance as compared to other participating countries in mathematics and reading, while they seem to be doing better in science. The average age is 15.71.

Table 2 shows the outcome and control variables for natives and immigrants. The average teacher grade does not differ between immigrants and natives, but PISA scores in all subjects are significantly different. Namely, native children perform much better. Among immigrants children there are fewer girls. In terms of education, parents of immigrant children are more educated, but their occupational status is lower.

Figure 1 shows for each grade the average PISA score. The relationship is fairly linear. Figure 2 shows the average PISA score for each grade for immigrants (red) and natives separately (blue). Given a school grade immigrants score worse on PISA than natives. Differences are especially large for children that do well in math.

3 Econometric Specification and preliminary results

Econometric model:

$$t\_assessment_{is} = \alpha_0 + \alpha_1 immig + \alpha_2 pisa\_score_{is} + \alpha_3 female_{is} + \beta X'_{is} + \gamma_s + \epsilon_{is} \quad (1)$$

where the outcome is teacher assessment ($t\_assessment_{is}$) of student i in school s. The set of controls includes PISA testscores ($pisa\_score_{is}$), a dummy for being a first generation immigrant ($immig$), a dummy for female ($female_{is}$), and a vector of additional controls such as age (age_{is}), education of parents and occupational status of their jobs (X'_{is}) and the school fixed effect, γ_s. Errors are clustered at the school level.

We assume that PISA and teacher assessments are evaluating the same skills. We also assume that class assignment is random. Ideally, we would like to estimate our models with teacher fixed effect because we are interested if the same teacher assigns different grades to pupils based on their background. PISA dataset does not contain a class or teacher identifier.

In Italy the highest grade is also the best grade.[3] Our variable of interest is the dummy immigrant, *immig*. A positive (negative) coefficient of the dummy immigrant means that cognitive test scores underestimate (overestimate) teacher assessment, conditional on the set of controls.

Table 3 summarises the baseline results. We find that immigrant children are graded significantly differently by teachers than by PISA officials. Let us briefly discuss the signs of the control variables. Mathematics has the expected sign and is statistically significant in all regressions. Girls are graded better by teachers than by PISA coders.[4]. Age has a positive coefficient. Only father's education has a positive effect on children's grades. One possible explanation why children coming from more educated families perform better on school tests than on PISA tests could be that more educated parents put more pressure on their children in the case of school tests because they carry more importance than PISA test scores.

[3]In some countries, such as Germany, the lowest number is the best grade.

[4]Lavy (2008) studies gender stereotypes using blind and non blind test scores in Israel and also finds a bias against boys in all subjects. The bias varies with teacher characteristics suggesting that it is a result of teacher behavior (not student behavior)

Age of arrival The age at which a child arrives in a new country can be very important factor for the success in education for several reasons. First, the language knowledge is affected by the age of arrival.The psychological literature postulates that there is a critical period for the language acquisition: children who arrive in a new country by the age of 9 become fluent in the language of that country while children who arrive later become less proficient. Second, the human capital acquired in another schooling system might not be fully transferable to the new educational system. Third, parents of recent immigrant might not be familiar with the educational system of the host country and might not find the right match in terms of school for their child. This can be a relevant factor for older children once the educational system becomes specialised.

We use this specification to capture the effect of age of arrival :

$$t\_assessment_{is} = \alpha_0 + \rho_1 age1 + \rho_2 age2 + \rho_3 age3 + \alpha_2 pisa\_score_{is} + \alpha_3 female_{is} + \alpha_4 age_{is} + \beta X'_{is} + \tag{2}$$

where everything is as in specification (3) with the exception of dummies $age1$, $age2$ and $age3$. Instead of the immigrant dummy ($immig$) we use 3 dummies to distinguish the age of arrival of immigrants. (1) immigrant children who arrived when they were 5 or younger, (2) immigrant children who arrived when they were 6 to 10 years old and (3) immigrant children who were older than 10 years old when they came to the new country.

Table 4 summarises the results. When we distinguish between the time of arrival of pupils in the host country, we find that the effect is mainly driven by the children who arrived after the age of 10.

Quantile regressions To understand if the effect is present for the whole distribution or is driven by a some part of the ability distribution, we divide children by PISA score in quartiles and re-estimate regression 3 for each quartile. Results are in table 5. The effect appears to be largest for the lowest quartile, but judging by the size of the coefficient for each quartile it doesn't appear that the effect is driven by a specific part of the distribution.

3.1 Possible explanations for bias

1. Language skills

   The PISA test in mathematics captures the language skills because children who do not have a good command of the test language can face difficulties in understanding test questions. On the other hand, it is possible that teachers are aware of these problems and that teachers take into account the language knowledge when evaluating children's mathematical skills.

   To test for this possibility we can include PISA reading test score in regressions (3). If the effect we find is a result of language deficiency, we would expect α_1 in regression (3) to become insignificant. We also use the language spoken at home to capture the knowledge of the host country language. Results are reported in table 6. Column (1) reports the results for the baseline model. In column (2), we include reading scores and in column (3) we add language spoken at home. PISA reading test scores have a predictive power for mathematics marks, the coefficent of *immig* remains significant. When language spoken at home is included, the coefficient of *immig* becomes insignificant, but it is still large. The coefficient *lang_home* is positive (though not significant) and it suggests that children using a different language at home than Italian are graded better by teachers. This could be an indication that teachers take into account the language limitations when evaluating immigrant children who do not speak Italian at home.

2. Do PISA and school grades measure different skills

   To examine if PISA assessment test different knowledge than the schools and teachers, we split the PISA results in 4 fields: (1) space and shape, (2) uncertainty, (3) change and relationship ability, (4) quantity ability. Suppose that the school mark gives more weight to topic X than to topic Y and immigrant and native students differ in their abilities in these two topics. Then the dummy immigrant should become insignificant once we regress marks on topic X. To test for this possibility, we regress marks on each topic separately. This is reported in table 7 in columns (3) to (6). Independently from the topic, immigrants are graded better by teachers conditional on PISA test scores.

School grades could also include behavioural problems like absenteeism or discipline in class which is not captured by PISA. To account for this possibility, we can include the answer to the following question "In the last two full weeks you were in school, how many times did you arrive late for school?" in the regressions. Table 7 in column (7) includes this self-reported measure of behaviour at school in the regression. The coefficient has the right sign and children who have been late more often are graded worse by their teachers conditional on PISA test scores.

3. Statistical discrimination

   If teachers were statistically discriminating, they would use observable characteristics to proxy for skills (Arrow, 1972; Phelps, 1972). How exactly teachers discriminate could be based on statistical actualities or teachers' expectations how pupils with certain characteristics should perform. The process of grading should limit statistical discrimination because teachers observe the skills of the children through written tests and oral participation.

4. Other explanations to be explored: (1) school tests could be culturally biased, (2) different timing of PISA and school grades, (3) different variance between natives and immigrants.

4 Tables

Table 1: Summary statistics outcome and control variables

	mean	p50	sd	min	max	count
Teacher mark mathematics	6.11	6.00	1.48	1.00	10.00	7655
PISA mathematics score	-1.59	-0.12	92.06	-362.56	305.14	7655
PISA reading score	1.66	9.69	94.33	-394.80	342.66	7655
PISA science score	16.97	23.63	101.95	-479.29	362.24	7655
Female (=1)	0.51	1.00	0.50	0.00	1.00	7655
Age	15.71	15.75	0.29	15.25	16.25	7655
Father's educ.	3.54	4.00	1.56	0.00	6.00	7470
Mother's educ.	3.48	4.00	1.54	0.00	6.00	7557
Occup. stat. father	43.58	43.00	16.02	16.00	90.00	7275
Occup. status mother	43.31	45.00	17.27	16.00	90.00	5849

Note: 500 has been substracted from PISA test scores.

Table 2: Difference between natives and immigrants

	(1) Natives	(2) Immigrants	(1)-(2) Difference
Teacher mark mathematics	6.11	6.05	0.63
			[0.10]
PISA mathematics score	-0.34	-33.15	32.81***
			[6.74]
PISA reading score	4.87	-58.88	63.76***
			[7.63]
PISA science score	19.20	-41.46	60.66***
			[8.22]
Female (=1)	0.52	0.44	0.08**
			[0.035]
Age	15.71	15.74	-0.03
			[0.02]
Father's educ.	3.52	4.24	-0.82***
			[0.12]
Mother's educ.	3.48	3.90	-0.43***
			[0.13]
Occup. stat. father	43.53	37.36	6.17***
			[1.14]
Occup. status mother	43.48	31.83	11.65***
			[1.43]

Table 3: Dependent variable: teacher mathematics assessment, main results, 2003, Italy

	(1) mark	(2) mark	(3) mark	(4) mark
1st gen. immigrant	0.155	0.220**	0.244**	0.226**
	(0.096)	(0.097)	(0.096)	(0.100)
Mathematics	0.006***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
Female (=1)			0.362***	0.372***
			(0.037)	(0.038)
Age				0.056
				(0.054)
Father's educ. med.				0.093**
				(0.039)
Father's educ. high				0.154***
				(0.048)
Mother's educ. med.				0.025
				(0.039)
Mother's educ. high				0.004
				(0.048)
_cons	6.113***	6.113***	5.928***	4.962***
	(0.016)	(0.015)	(0.024)	(0.850)
N	7655	7655	7655	7410
r2	0.163	0.257	0.267	0.271
School FEs	NO	YES	YES	YES

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Dependent variable: teacher mathematics assessment, split by age, 2003, Italy

	(1) mark	(2) mark
1st gen. immigrant	0.244**	
	(0.096)	
math	0.008***	0.008***
	(0.000)	(0.000)
Female (=1)	0.362***	0.359***
	(0.037)	(0.037)
arrival_age1		-0.164
		(0.179)
arrival_age2		0.160
		(0.177)
arrival_age3		0.549***
		(0.150)
_cons	5.928***	5.931***
	(0.024)	(0.024)
N	7655	7655
r2	0.267	0.268

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Dependent variable: teacher mathematics assessment, split by quantile (PISA score), 2003, Italy

	(1) baseline	(2) quartile 1	(3) quartile 2	(4) quartile 3	(5) quartile 4
1st gen. immigrant	0.244**	0.355*	0.099	0.212	0.253
	(0.096)	(0.192)	(0.195)	(0.228)	(0.223)
math	0.008***	0.006***	0.010***	0.009***	0.010***
	(0.000)	(0.001)	(0.002)	(0.002)	(0.001)
Female (=1)	0.362***	0.159*	0.438***	0.468***	0.404***
	(0.037)	(0.089)	(0.078)	(0.074)	(0.067)
_cons	5.928***	6.017***	5.944***	5.747***	5.640***
	(0.024)	(0.111)	(0.074)	(0.075)	(0.096)
N	7655	1915	1916	1915	1909
r2	0.267	0.206	0.228	0.228	0.254

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Dependent variable: teacher mathematics assessment, knowledge of language, 2003, Italy

	(1)	(2)	(3)
	mark	mark	mark
1st gen. immigrant	0.244**	0.281***	0.192
	(0.096)	(0.097)	(0.135)
math	0.008***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)
Female (=1)	0.362***	0.323***	0.339***
	(0.037)	(0.038)	(0.039)
Reading		0.001***	0.001***
		(0.000)	(0.000)
lang_home			0.236
			(0.145)
_cons	5.928***	5.945***	5.924***
	(0.024)	(0.024)	(0.025)
N	7655	7655	7182
r2	0.267	0.268	0.274

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

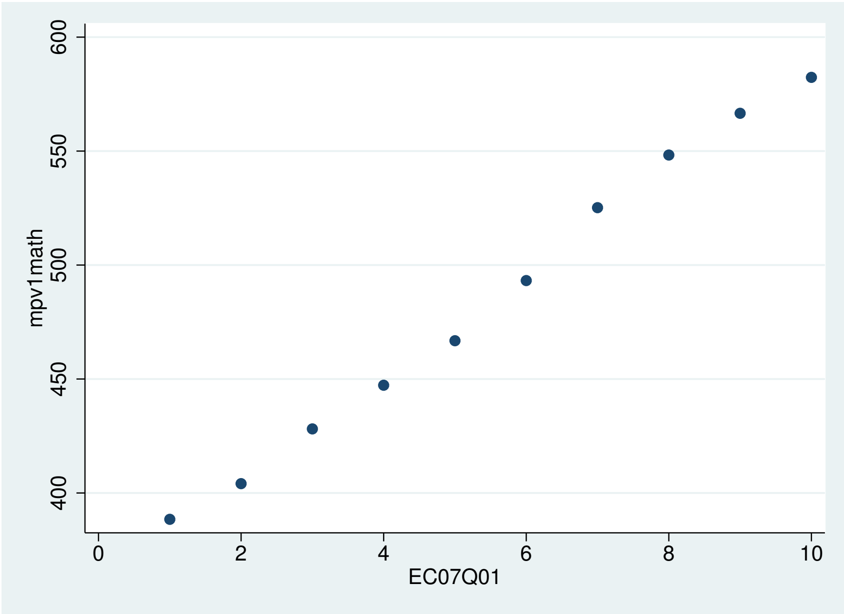
Table 7: Dependent variable: teacher mathematics assessment, split by mathematics area, 2003, Italy

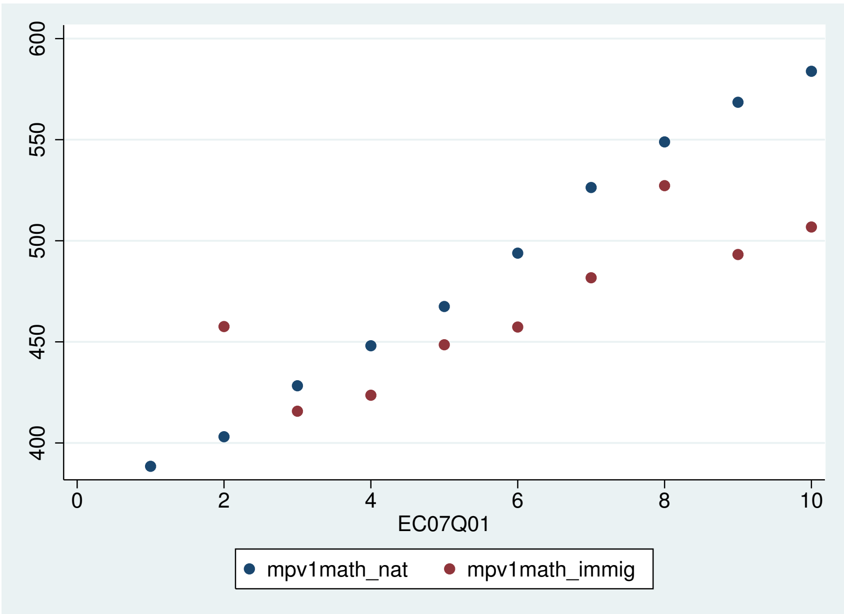
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	mark	mark	mark	mark	mark	mark	mark
1st gen. immig	0.244**	0.261***	0.206**	0.250**	0.293***	0.199**	0.258***
	(0.096)	(0.097)	(0.098)	(0.098)	(0.098)	(0.097)	(0.096)
math	0.008***						0.008***
	(0.000)						(0.000)
Female (=1)	0.362***	0.358***	0.325***	0.362***	0.364***	0.287***	0.345***
	(0.037)	(0.037)	(0.037)	(0.037)	(0.038)	(0.037)	(0.037)
Mathematics 1		0.002***	0.006***				
		(0.000)	(0.000)				
Mathematics 2		0.002***		0.007***			
		(0.000)		(0.000)			
Mathematics 3		0.001***			0.007***		
		(0.000)			(0.000)		
Mathematics 4		0.003***				0.006***	
		(0.000)				(0.000)	
late_for_sch							-0.105***
							(0.015)
_cons	5.928***	5.921***	5.912***	6.000***	5.945***	5.905***	6.108***
	(0.024)	(0.025)	(0.025)	(0.024)	(0.024)	(0.024)	(0.035)
N	7655	7655	7655	7655	7655	7655	7655
r2	0.267	0.266	0.238	0.249	0.236	0.251	0.272

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5 Figure


mpv1math
400
450
500
550
600
0
2
4
6
8
10
EC07Q01


400
450
500
550
600
0
2
4
6
8
10
EC07Q01
mpv1math_nat
mpv1math_immig

References

Arrow, K. (1972). Models of job discrimination. In A. H. Pascal (Ed.), *Racial Discrimination in Economic Life*, pp. 83–102. Lexington, MA: D. C. Health.

Becker, G. S. (1971). *The Economics of Discrimination.* University of Chicago Press.

Betts, J. R. and J. Grogger (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review 22*(4), 343–352.

Figlio, D. and M. Lucas (2004). Do High Grading Standards Affect Student Performance? *Journal of Public Economics 88*, 1815–1834.

Hanna, R. N. and L. L. Linden (2012). Discrimination in grading. *American Economic Journal: Economic Policy 4*(4), 146–168.

Kiss, D. (2013). Are immigrants and girls graded worse? Results of a matching approach. *Education Economics 21*(5), 447–463.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? evidence from a natural experiment. *Journal of Public Economics 92*(10-11), 2083–2105.

Phelps, E. (1972, 661). The statistical theory of racism and sexism. *American Economic Journal 62*(3), 659.

Rosenthal, R. and L. Jacobsen (1968). Pygmalion in the Classroom. *The Urban Review 3*(1), 16–20.

Schnepf, S. (2007). Immigrants' educational disadvantage: an examination across ten countries and three surveys. *Journal of Population Economics 20*(3), 527–545.

Spence, A. M. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355–74.