The Good, the Bad, and the Ugly:

Teaching Evaluations, Beauty and Abilities

Michela Ponzo, Vincenzo Scoppa*

Using data from an Italian University we relate student evaluations of teaching quality to physical attractiveness of instructors, controlling for a number of teachers' and courses' characteristics. We first show that the beauty of teachers strongly affects teaching evaluations. To investigate whether the impact is due to productivity or discrimination, that is, if the better evaluations obtained by good-looking instructors are determined by their possess of greater abilities or by Becker-type customer discrimination, we propose a simple theoretical framework and build a measure of teachers' abilities that is used as control in the empirical model explaining teaching evaluations. We show that beauty affects teaching evaluations even controlling for ability, suggesting that customer discrimination is the key factor explaining the role of beauty.

JEL classification: I20; J24; J7.

Keywords: Beauty; Discrimination; Teaching Quality; Subjective Evaluations.

1. Introduction

Researchers have recently devoted an increasing attention to the effects of physical appearance on earnings and other labor market outcomes. The seminal paper by Hamermesh and Biddle (1994) finds evidence that beauty affects wages across occupations, both for males and females. Using data of U.S. and Canadian employees, they show that with respect to average looking people, plain individuals typically earn about 9 percent less while good looking people gain about 5 percent more. Harper (2000) finds similar evidence for the UK labor market: physical appearance has a substantial impact on earnings with a penalty for plainness exceeding the premium for attractiveness. Fletcher (2009) shows positive wage returns to attractiveness for US young high school graduates, controlling for a measure of abilities. All these findings provide strong evidence that attractiveness is positively associated with earnings.[1]

A few papers have analyzed if beauty also matters in particular high-skill occupations such as university teaching.[2] Hamermesh and Parker (2005) investigate the impact of professors' physical appearance on their courses' student evaluations of teaching.[3] They show that measures of perceived attractiveness have a substantial positive effect on instructional ratings by undergraduate students and the effects of teachers' looks on instructional ratings are larger for male instructors. In contrast, Süssmuth (2006) finds that perceived attractiveness of German university teachers is scarcely correlated with the instructional ratings: the statistical significance is weak and the magnitude of

* Department of Economics and Statistics, University of Calabria, Via Ponte Bucci, 87036 Arcavacata di Rende (CS), Italy. E-mail: michela.ponzo@unical.it; v.scoppa@unical.it. We thank the Dean of the Faculty of Economics at the University of Calabria for allowing us the use of student evaluations of teaching, Sara Laurita and Antonella Caparelli for helping in the use of the data and the students who kindly rated instructors' photographs. We are grateful to Federica Demaria, Maria De Paola, Sabrina Giordano, Daniel Hamermesh, Andrea Ottolina, Pierfrancesco Perri, Lorenzo Rocco, Concetta Scoppa for useful comments and suggestions.

[1] Further evidence finds associations between wages and height (Persico, Postlewaite and Silverman 2004) and between wages and obesity (Cawley, 2004).

[2] Physical attractiveness appears also to be a relevant factor in explaining the success of politicians, lawyers, economists, prostitutes. See the recent book of Hamermesh (2011) for a detailed account (and related references) of the evidence on the impact of beauty in a number of economic and social contexts.

[3] Teaching evaluations, in turn, tend to have an effect on instructors' wages, since university administrators often take into account these evaluations in setting salaries (Becker and Watts, 1999).

the effect much smaller. Sen, Voia and Wolley (2010) for 16 Canadian Universities find a positive effect of teachers' "hotness" on their earnings, on some measures of teaching quality and, at least in part, on research productivity. Unfortunately, in this paper the evaluation of instructors' beauty is made by the same students rating the quality of teaching, so there could be a reverse-causality problem.

The key question that remains open in this literature is whether employers or customers, having a preference for beauty, discriminate (in the Becker sense) in favor of attractive individuals or whether good looking people are paid more (or evaluated better) because they are more productive.

A realistic possibility – which complicates the picture – is that beauty is related to (or a cause of) other personality traits such as confidence, self-esteem, perseverance, and these non-cognitive skills may also have a direct impact on individual productivity. A number of recent studies find a relationship between beauty and self-confidence and other non-cognitive skills. Mobius and Rosenblat (2006), investigating the possible reasons for a beauty premium in a laboratory experimental labor market, find evidence that physically attractive workers are more confident and that their higher confidence increases the wages they receive. Persico, Postlewaite and Silverman (2004) show that taller people during adolescence acquire greater self-confidence and perseverance and develop more easily social interactions: these characteristics lead to higher achievements as adults. Cipriani and Zago (2011) find a positive relationship between students' attractiveness and their academic performance (only for males). They find that the students' look has a positive effect both on oral examinations (where physical aspect can be observed by evaluators) and on written examinations (where, the authors argue, beauty is typically not observable). The authors interpret these results as evidence against the discrimination explanation and in favor of productivity differences among good and bad looking individuals determined by confidence and self-esteem.

The aim of this paper is both to provide evidence of the impact of instructors' beauty on student evaluations of teaching for an Italian University and to contribute to verify if beauty is related to pedagogical productivity or if its impact on student evaluations is due to Becker-type discrimination.

We use data from student evaluations of teaching carried out at the University of Calabria. We relate student evaluations of teaching quality to physical attractiveness of instructors (as evaluated by external raters), controlling for a number of teachers' characteristics (gender, academic position, age) and courses' characteristics (number of attending students, Department offering the course, Degree Level, etc.). We first show that the beauty of teachers positively affects teaching evaluations, especially for female instructors and at the First Level Degree. To investigate whether the impact is due to discrimination or productivity, that is, if the better evaluations obtained by good-looking teachers are determined by their possess of greater abilities, we propose a simple theoretical framework exploiting the relationship between abilities, teaching and research productivity. We build a measure of teachers' abilities based on research productivity that it is used as a control variable in the empirical model explaining teaching evaluations. We show that beauty strongly affects teaching evaluations even controlling for individual ability (which, in turn, has a positive impact on teaching evaluations), suggesting that customers' discrimination is the key factor explaining the role of beauty.

The paper is organized as follows. Section 2 briefly describes the data and the way in which we build the main variables used in the analysis. Section 3 reports and discusses the results from several OLS regressions relating students' overall satisfaction with a course to a measure of the instructor's look. Section 4 proposes a framework to investigate the relationship between abilities and beauty. Section 5 offers some concluding remarks.

2. The Data

We use data on student evaluations of teaching from the Faculty of Economics at the University of Calabria. The University of Calabria is a middle-sized public University located in the South of Italy. It has currently about 35,000

students enrolled in different Degree Courses and at different levels of the Italian University system. Since the 2001 reform, the Italian University system is organized around three main levels: First Level Degrees (3 years of legal duration), Second Level Degrees (2 years more) and Ph.D. degrees. Students who have acquired a First Level Degree can undertake a Second Level Degree. After having accomplished their Second Level Degree, students can enroll in a Ph.D. degree.

Data on student evaluations of teaching come from all the courses taught over 7 academic years (from 2003–2004 to 2009–2010) in the Faculty of Economics which comprises 6 different Degree Courses (at the First and the Second Level): Business and Administration, Economics, Law, Statistics, Social and Economics Sciences, Tourism Economics. Each year these courses enroll in total about 1,500 freshmen. For the 7 years considered, we have observations on a total of 2,512 courses taught by 234 different teachers. However, we were able to find photographs (see below) for only 193 teachers, so we focus on the 2,338 courses they teach involving a total of 107,457 student evaluations. Mostly of the courses students have attended were compulsory. According to the University Regulations, attendance to classes is also compulsory, although checks are infrequent.

The Law no. 370 passed in 1999 has required Universities to evaluate the quality of teaching in each course through a survey among students. According to the Law, student evaluations of teaching should be an indicator taken into account by the Ministry of University to allocate public funds to each university and to stimulate – with additional funds – good teaching performance by instructors.[4]

Student evaluations of teaching are carried out between half and two thirds of the classes in each course, are anonymous and taken while the instructor is out of the classroom. Among the questions asked to students, we mainly use the question about the "Overall satisfaction for the course". The ratings students could give are: 1) "Very Positive"; 2) "Positive"; 3) "Negative"; 4) "Very Negative". However, in our dataset for each course we only observe the percentage of students giving a "positive" or a "very positive" evaluation (Answers 1 or 2) that we use as our dependent variable (called *Teaching Evaluations*).

Furthermore, students are also asked to evaluate: a) the clarity of the teacher in presenting the material (*Clarity*); b) whether the teacher stimulates interest in the subject (*Interest*). *Clarity* and *Interest* have been built in the same manner as *Teaching Evaluations*. The degree of correlation of the *Teaching Evaluations* with these two variables is very high (0.86 and 0.87, respectively). As a robustness check, we also use separately the variables *Clarity* and *Interest* as dependent variables and furthermore we aggregate the three evaluations using a principal component analysis (see Section 3.1).[5]

We asked 29 students to rate, independently, from 1 (lowest) to 10 (highest) the physical appearance of the instructors from facial photographs (taken on the instructors' departmental websites, personal web-pages or photos on the internet, or from non-digital photographs). The raters were students (16 females and 13 males – to reflect the fact that 55% of students attending the courses were females) enrolled in other Fields at the University of Calabria who did not know the evaluated instructors.[6]

In contrast to Hamermesh and Parker (2005) we did not ask our raters to evaluate beauty independently from age, as we judged rather hard to make an evaluation of this type.[7] However, to avoid spurious correlations, in our regressions we are able to control for the instructors' age and for the academic position of instructor (Full, Associate or Assistant Professor, Adjunct Instructor) that turns out to be highly correlated to age.

---

[4] On the difficulties regarding the effective implementation of these incentive mechanisms, see the detailed account in Perotti (2002).
[5] The rating forms also include questions on whether the instructor begins classes on time, he/she is available during office hours, students possess adequate knowledge allowing them to understand the subject, the rooms are satisfactory, the study load in the period was tolerable and so on.
[6] We asked these students whether they knew the instructors. Some of them declared to know 4 instructors (because they are or have been Dean or Department Chairmen). As a robustness check, we estimate excluding these 4 instructors but we obtain almost identical results.
[7] Hamermesh (2011) argues that people are not able to disentangle the effect of age from beauty, even if they are told to do so.

The average inter-item correlation is 0.49. The Cronbach's alpha statistic, a psycho-metric indicator measuring how closely related a set of items are as a group, is equal to 0.93 (on a scale between 0 and 1).

Following Mobius and Rosenblat (2006), from each evaluation r_{jk} made by rater *k* of the beauty of teacher *j*, we subtract the average value of evaluations given by each rater ($\bar{r}_k = \frac{1}{N}\sum_{j=1}^{N} r_{jk}$). This allows us to neutralize measurement errors due to different perceptions (or different standard) of beauty among raters. Then, for each instructor we take the average of all the raters' demeaned evaluations. Finally, we divide this variable by its standard deviation to obtain a standardized variable that we call *Beauty*.

We use administrative data on instructors regarding their gender, the courses taught (we classify courses according to the Department offering them: Business, Economics, Law, Mathematics, Statistics, and so on), academic position (Full, Associate or Assistant Professor, Adjunct Instructor). We gather information on the year of birth of teachers from CVs on-line. In the few cases in which we were not able to find the year of birth we have imputed it as the year of graduation minus 24.[8]

Table 1. Descriptive Statistics

Variables	Obs.	Mean	Std. Dev	Min.	Max.
Teaching Evaluations	2338	76.3204	19.8288	0	100
Clarity	2338	76.7655	20.7272	0	100
Interest	2338	73.6009	21.0490	0	100
Beauty	2338	0.0003	0.9933	-2.0890	2.6533
Female	2338	0.3995	0.4899	0	1
Full Professor	2338	0.2365	0.4250	0	1
Associate Professor	2338	0.2703	0.4442	0	1
Assistant Professor	2338	0.3529	0.4780	0	1
Adjunct Professor	2338	0.1403	0.3474	0	1
# Students in Class	2338	45.9611	44.1547	2	291
Second Level Degree	2338	0.3152	0.4647	0	1
Age	2313	44.4803	9.9446	24	73
Business	2338	0.2699	0.4440	0	1
Economics	2338	0.1848	0.3882	0	1
Law	2338	0.1681	0.3740	0	1
Computer Science	2338	0.0522	0.2224	0	1
Foreign Languages	2338	0.0492	0.2163	0	1
Mathematics	2338	0.0770	0.2666	0	1
Sociology	2338	0.0796	0.2707	0	1
Statistics	2338	0.0997	0.2996	0	1
History	2338	0.0197	0.1389	0	1

Notes: Observations at the course-level. Number of teachers: 193.

In Table 1 we show some descriptive statistics. Note that observations are at the course-level and some statistics could be slightly different if calculated at the teacher-level. *Teaching Evaluations* is on average equal to 76.3%, with a standard deviation of 19.83. *Beauty* has, by construction, a mean of zero and a standard deviation of 1. Female instructors make up 40% of the sample (110 males and 83 females), full professors are 24%, associate professors are 27% while assistant professors are 35%. The mean *Age* is 44. The average number of students attending the course is 46. About 32% of the courses are at the Second Level Degree. Courses are mainly in Business (27%), Economics (18%), and Law (17%).

[8] Nonetheless, age is missing for 3 instructors.

3. The Impact of Beauty on Student Evaluation of Teaching: Empirical Results

To evaluate the impact of beauty on students' teaching evaluations, we estimate by Ordinary Least Squares (OLS) the following equation:

[1] $$Teaching\ Evaluations_{ij} = \phi_0 + \phi_1 Beauty_j + \phi_2 X_j + \phi_3 W_{ij} + \varepsilon_{ij}$$

where $Teaching\ Evaluations_{ij}$ is the evaluation of course *i* taught by instructor *j*, $Beauty_j$ is a measure of physical attractiveness of instructor *j*, X_j is a vector of instructor's individual characteristics (gender, academic position, age), W_{ij} is a vector of course's characteristics (Degree Level, number of students attending the course, Department offering the course, academic year) and ε_{ij} is an error term capturing idiosyncratic shocks or unobserved course's or instructor's characteristics.

We weight observations by the number of students attending each course. Since typically the same instructor hold several courses,[9] we cluster standard errors at the instructor level. Moreover, standard errors are robust to heteroskedasticity.

Estimation results are reported in Table 2. In the first column we regress *Teaching Evaluations* only on *Beauty*. The coefficient is positive and highly statistically significant implying that an increase of one standard deviation in beauty raises the course ratings by 4.5 percentage points. Since *Beauty* could be correlated to other teacher's or course's characteristics, in column (2) we use as control variables dummies for gender and academic positions (the reference category is Assistant Professor), the number of students attending the course and a dummy variable taking into account the Degree Level (the reference category is First Level Degree). In column (3) to avoid possible biases due to heterogeneous evaluations among fields of study and to possible correlations between instructors' beauty and field of study, we include 8 dummies for the Departments offering the courses (leaving Business Economics as reference category).

Both in columns (2) and (3) the coefficient on *Beauty* is positive and statistically significant at the 1 percent level (*t*-stat=4.59 in column 3): ceteris paribus, better looking teachers receive higher instructional ratings. An increase of one standard deviation in beauty raises teaching evaluations by 5.58 points, which corresponds to about 0.28 standard deviations of *Teaching Evaluations*.

As regards the impact of other instructor's characteristics, we find that female teachers tend to attract worse evaluations than their male colleagues (although the effect is imprecisely estimated, *p*-value is around 0.13 in column 3). Full professors receive lower evaluation scores with respect to Assistant professors, maybe because the latter devote more time or put more enthusiasm on preparing classes, compensating for a lower teaching experience. The number of students in the class tend to reduce the evaluation of teaching quality, perhaps because learning is more difficult in a crowded and noisier environment. From column (3) it also emerges that courses in Economics, Foreign Languages, Statistics and Mathematics receive much worse instructional ratings.

[9] On average, each teacher in our sample taught 12 courses.

Table 2. Teaching Evaluations and Instructor's Beauty. OLS regressions

	(1) (All)	(2) (All)	(3) (All)	(4) (Males)	(5) (Females)	(6) (All)
Beauty	4.4989***	5.0885***	5.5809***	5.0554**	5.9330***	4.9393***
	(1.2534)	(1.1883)	(1.2147)	(1.9666)	(1.0685)	(1.8918)
Female		-4.4952**	-3.5956			-3.8854*
		(2.2704)	(2.3714)			(2.2877)
Full Professor		-5.1134	-4.8828*	-2.0440	-11.1471***	-4.8954*
		(3.5604)	(2.8987)	(3.4927)	(3.6661)	(2.8902)
Associate Professor		-3.4407	-1.2371	2.5709	-5.6901*	-1.3858
		(2.7881)	(2.5425)	(3.5255)	(3.0423)	(2.5162)
Adjunct Professor		-1.5127	-0.1352	5.3547	-4.5766*	-0.1591
		(2.1623)	(2.2822)	(3.6659)	(2.7230)	(2.2828)
# Students in Class		-0.0296*	-0.0297*	-0.0213	-0.0595***	-0.0297*
		(0.0175)	(0.0164)	(0.0198)	(0.0221)	(0.0164)
Second Level Degree		2.3869	1.7766	2.9010	0.3597	1.7585
		(1.8146)	(1.6890)	(2.1492)	(2.3335)	(1.6829)
Economics			-12.7018***	-18.6399**	-6.4096*	-13.0158***
			(4.3749)	(7.7969)	(3.3187)	(4.4936)
Law			-3.2840	-5.7267	-1.5181	-3.2640
			(3.4056)	(4.2121)	(4.2303)	(3.3425)
Computer Science			3.4169	1.8112	14.6214***	3.1869
			(3.0331)	(3.8981)	(3.3210)	(3.1345)
Foreign Languages			-17.1183***	-28.7038***	-15.3491***	-16.7722***
			(3.9836)	(3.7818)	(4.4091)	(4.0628)
Mathematics			-8.9339**	-11.4804**	-3.6991	-8.8919**
			(4.4365)	(5.1286)	(5.1340)	(4.4164)
Sociology			-6.6949**	-10.5509**	-3.1738	-6.3233*
			(3.3229)	(5.0921)	(4.6560)	(3.3443)
Statistics			-10.8101***	-11.9033***	-6.9950**	-10.5637***
			(2.7968)	(3.6815)	(2.9210)	(2.7429)
History			-3.6031	-1.3383	-2.1693	-3.1193
			(3.1914)	(4.9827)	(2.5714)	(3.1043)
Beauty*Female						1.3448
						(2.2406)
Observations	2338	2338	2338	1404	934	2338
Adjusted R-squared	0.058	0.099	0.186	0.195	0.229	0.187
Clusters	193	193	193	110	83	193

Notes: The Table reports Weighted Least Squares estimates (observations are weighted by the number of students attending each course). The dependent variable is *Teaching Evaluations.* In all the regressions we control for (7) year dummies. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the teacher level. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

To investigate if there exists a differential impact of physical attractiveness according to gender, we estimate the model (3) separately for male and female instructors (columns 4 and 5, respectively). Results show that *Beauty* has a strong significant impact both for males and females. From the magnitude of the coefficients, it seems that beauty is more relevant for females (5.93) than for males (5.05). However, in column (6) we estimate on the whole sample including an interaction term between *Beauty* and *Female* and the difference among gender in the impact of physical appearance is not statistically significant.

In column (1) of Table 3, we control for the instructor's *Age* which in our sample is correlated to *Beauty*[10] and could represent a relevant factor in determining teaching evaluations. Since we were not able to find age for 3 instructors, we lose 25 observations out of 2338. Estimates show that *Age* has no effect in explaining teaching evaluations, probably because the teacher's academic position is partially capturing the effect of *Age*.[11] More importantly, the coefficient on *Beauty* remains positive and almost unchanged in magnitude and statistical significance, implying that the effect we find is not related to the fact that older teachers turn out to be less attractive.

Since the Degree Level may generate different ratings depending on students' skills and on their maturity levels as well as on the difficulty of the subjects, we estimate our basic model (column 3 of Table 2) separately for

[10] We find a strong negative correlation between Age and Beauty (*t*-stat=-5. 36).
[11] Regressing Age on the dummies for academic positions, it emerges that all coefficients are highly statistically significant.

courses of First and Second Level Degree, reported, respectively, in columns (2) and (3) of Table 3. We find that *Beauty* is more important in the First Level Degree (the coefficient is 5.87), while it is less important, but nonetheless significant, in the Second Level Degree (4.48).[12] This suggests that higher abilities and more mature students are less focused on physical attractiveness of teachers and more on their efficacy in presenting the material.

Related to this issue, it could be that a factor as beauty becomes more important in influencing student evaluations of teaching in particular conditions, such as when classes are too crowded, or when the study load is excessive or when the prior knowledge of students is insufficient to understand the subject. In these contexts, students could give more weight to beauty in their evaluations rather than on the effective quality of teaching. To investigate this aspect, we use the information provided in the students' survey on *Crowded Classes*, *Excessive Study Load* and *Adequate Prior Knowledge*, interacting each of these variables, in turn, with *Beauty* (results are not reported). We do not find statistically significant differences in the impact of beauty according to the crowding of classes or to the study load. On the other hand, when we interact *Adequate Prior Knowledge* with *Beauty* we find a significant difference: beauty has a greater impact when prior knowledge is insufficient but the effect of beauty is positive and statistically significant even when prior knowledge is judged as adequate.

Table 3. Teaching Evaluations and Instructor's Beauty. Further specifications. OLS regressions

	(1) (All)	(2) (First Level)	(3) (Second Level)	(4) (All)	(5) (All)
Beauty	5.4699***	5.8678***	4.4779***	6.1365***	
	(1.2061)	(1.3513)	(1.3349)	(1.7544)	
Age	-0.0333				
	(0.1752)				
Female	-3.6233	-3.8908	-2.9353	-1.2231	-2.1824
	(2.4005)	(2.5623)	(3.0800)	(2.5016)	(2.4692)
Full Professor	-4.4163	-5.5366*	-2.2494	-2.2576	-4.3692
	(3.4046)	(3.0514)	(3.8500)	(2.7560)	(2.9024)
Associate Professor	-0.9174	-1.7215	0.5550	-0.9845	-1.6628
	(2.5846)	(2.7715)	(3.6621)	(2.9072)	(2.6435)
Adjunct Professor	0.1535	1.4518	-1.4001	0.5257	-0.2687
	(2.3722)	(2.5107)	(4.4786)	(2.5388)	(2.2705)
# Students in Class	-0.0293*	-0.0213	-0.0900***	-0.0057	-0.0223
	(0.0162)	(0.0172)	(0.0271)	(0.0201)	(0.0172)
Second Level Degree	1.7297				1.6681
	(1.6890)				(1.6814)
Very Attractive					6.3858***
					(2.4225)
Very Unattractive					-4.9786*
					(3.0143)
Observations	2313	1601	737	2337	2338
Adjusted R-squared	0.178	0.180	0.219	0.360	0.168
Clusters	190	181	154	193	193
Course Fixed Effects	NO	NO	NO	YES	NO

Notes: The Table reports Weighted Least Squares estimates (observations are weighted by the number of students attending each course). The dependent variable is *Teaching Evaluations*. In all the regressions we control for (9) Department dummies and (7) year dummies. In regression 4 we control for course fixed effects and we do not control for Department and year dummies. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the teacher level. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

In principle, although unlikely, the estimated effect could be biased if students with more favourable attitudes towards beauty are attracted to attend courses held by good-looking instructors or if some courses are assigned on the basis of the instructors' look, for example, if better looking teachers teach in easier courses. In order to take into account this possible bias, we exploit the fact that a number of courses have been taught by different instructors over the years or in the same academic year (in a small number of cases more classes for the same course

[12] Estimating a model on the whole sample including an interaction term between *Second Level Degree* and *Beauty* (not reported), we find that the difference is statistically significant.

were held due to the high number of enrolled students). A total of 581 different courses were held over the period considered. In column (4) of Table 3 we include course fixed effects but we do not control for Department and year dummies. The coefficient on *Beauty* will capture within-course differences in the impact of instructor's look on evaluation scores. The estimates with course fixed effects strongly confirm that more attractive instructors receive better teaching evaluations (the coefficient on *Beauty* is 6.14 and *t*-stat is 3.50).

Finally, to take into account possible non-linear effects of beauty, we build a dummy variable *Very Attractive* taking the value of one for instructors in the top 33% of the *Beauty* ranking (and 0 otherwise) and a dummy variable *Very Unattractive* taking the value of one for instructors in the bottom 33% of the *Beauty* ranking. We regress teaching evaluations on these two dummies leaving "Average Attractive Teachers" as reference category (column 5 of Table 3). We find that very attractive teachers obtain a better evaluation (+6.39 percentage points), while very unattractive teachers receive worse instructional ratings (–4.98 percentage points). In contrast to other studies examining the impact of beauty (see, e. g., Hamermesh and Biddle, 1994), penalties for unattractive teachers do not seem to be more relevant in magnitude than premiums for attractive ones.

3.1. Some Robustness Checks

To check the robustness of our empirical findings, instead of using *Teaching Evaluations* as dependent variable, we separately consider *Clarity* – measuring the clarity of the teacher in presenting the material – and *Interest* – measuring whether the teacher stimulates interest in the subject. Moreover, to obtain a comprehensive measure of satisfaction for the course, we undertake a principal component analysis summarizing the three previous indicators of instructional ratings (only the first component is considered), which we call *Overall Satisfaction*.

In columns (1)-(3) of Table 4 we replicate our benchmark estimates (column 3 of Table 2) using the three new measures as dependent variables (respectively, *Clarity, Interest*, and *Overall Satisfaction*). To save space, we do not report the coefficients on control variables. The impact of *Beauty* on these different indicators of teaching quality is stable across specifications, significant at the 1 percent level and similar in magnitude to the estimates obtained using *Teaching Evaluations*. Specifically, one additional standard deviation of perceived beauty raises the dependent variable by 7.3 (in the case of *Clarity*) or by 6.9 (when we consider *Interest*). The effect of better looking instructors on *Overall Satisfaction* is equal to about 0.55, which corresponds to 0.33 standard deviations of the dependent variable.

Another potential threat to the internal validity of our estimates comes from the fact the *Beauty* could be correlated to other characteristics or attitudes of instructors. If the instructors who are well organized in teaching or come on time to the class are also individuals who dress better or with good grooming – and thus their pictures receive higher ratings – we would attribute to beauty the effect of particular qualities of instructors that are independent of their looks. To account for this problem, we use the answers given by students to other two additional questions in the Survey of Student Evaluation of Teaching: *Classes Begin on Time* and *Instructor Available in Office*, measured again as the percentage of students giving a positive evaluation.

In columns (4) and (5) of Table 4 we consider as dependent variable *Teaching Evaluations* and we control in turn for *Classes Begin on Time* and *Available in Office*. From column (4), it emerges that teacher coming to class on time receive much higher teaching evaluations than others. Similarly, in column (5) teachers holding their announced office hours get better evaluation scores. Importantly, controlling for these two measures of instructor's aptitude, the coefficient on *Beauty* remains positive and highly statistically significant. However, the latter two regressions must be interpreted with caution since the student evaluations regarding *Classes Begin on Time* and *Available in Office* could be themselves affected by the instructor's beauty. If this is the case, they would be outcome variables that should not be used as control variables.

Moreover, following Hamermesh and Parker (2005), we build a dummy variable called *Formal Dress*, equal to one for male faculty members who are wearing neckties in their pictures and for female faculty who are wearing a jacket and blouse. We estimate our benchmark model including *Formal Dress* among control variables. Estimates in

column (6) of Table 4 show that instructors with formal dress in their picture do not differ in their instructional ratings from those wearing "casual dress" (the coefficient is 1.42 but the *t*-stat is 0.53), while the effect of *Beauty* on teaching evaluations remains quite large (+5.44) and highly statistically significant (*t*-stat=4.58).

Table 4. Robustness Checks. OLS regressions

	(1) Clarity	(2) Interest	(3) Overall Satisfaction	(4) Teaching Evaluation	(5) Teaching Evaluations	(6) Teaching Evaluations	(7) Teaching Evaluations
Beauty	7.3044***	6.8539***	0.5494***	5.2750***	3.0372***	5.4440***	6.0242***
	(1.4658)	(1.4035)	(0.1120)	(1.1326)	(0.8833)	(1.1896)	(1.3126)
Classes begin on time				0.5683***			
				(0.0683)			
Available in Office					1.0703***		
					(0.0429)		
Formal Dress						1.4168	
						(2.6595)	
Observations	2338	2338	2338	2338	2338	2338	2010
Adjusted R-squared	0.220	0.206	0.213	0.300	0.525	0.187	0.204
Clusters	193	193	193	193	193	193	147

Notes: The Table reports Weighted Least Squares estimates (observations are weighted by the number of students attending each course). The dependent variable are: *Clarity* (column 1), *Interest* (column 2), *Overall Satisfaction* (column 3) and *Teaching Evaluations* (columns 4, 5, 6 and 7). In all the regressions we control for teachers' and courses' characteristics as in Table 2, for (9) Department dummies and (7) year dummies. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the instructor level. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Sample selection bias caused by the fact that for some instructors we were not able to find the photographs should not be a major problem in our analysis: out of 2,512 courses taught in the sample period, we lose 174 observations, corresponding to less than 7 percent. With only a few exceptions, the instructors for whom we do not have photos are adjunct instructors that taught only for brief spells of time at the University of Calabria. To take into account possible sample selection bias, we estimate our preferred specification on a sample of only permanent instructors[13] (photos are missing for only 2 of them and we end up with 147 instructors). Estimates are reported in column (7) of Table 4: we find very similar results to those presented above: the coefficient on *Beauty* is equal to 6.02, significant at the 1 percent level.

Finally, as a further robustness check, to verify if the beauty ratings are consistent across raters, we also regress *Teaching Evaluations* separately on each rater's evaluation (r_{jk}). Confirming the high degree of agreement among raters, the coefficients of *Beauty* are always positive and, with only a few exceptions, statistically significant (results are not reported).

4. An Attempt to Deal with the Correlation between Beauty and Productivity

We have shown that instructor's physical attractiveness has a considerable impact on student evaluations of teaching. However, as explained above, it remains to be seen if good-looking teachers are effectively better teachers – a "productivity" effect is at work – or if their higher instructional ratings are simply the result of positive discrimination by students.

Ideally, to disentangle the impact of productivity from the extent of discrimination in favour of good-looking people (and against ugly people) we would have liked to estimate the following equation:

[2] $$Teaching\ Evaluations_{ij} = \phi_0 + \pi\left(Teaching\ Quality\right)_j + \phi_1 Beauty_j + \phi_2 X_j + \phi_3 W_{ij} + \nu_{ij}$$

[13] In the Italian academic system, all professors (Full, Associate and Assistant) have a permanent position.

where *Teaching Quality* is an objective (but unobservable) measure of teaching quality (which measures, for example, the knowledge effectively acquired by students) while the other terms are the same as in equation [1].[14]

If it were possible to estimate equation [2] , ϕ_1 would represent an accurate measure of discrimination since, if positive, it would imply a better evaluation for good-looking instructors in presence of equal teaching effectiveness.

In order to better understand how to interpret the estimated effect of *Beauty* when [2] cannot be estimated, let us suppose that teaching quality depends on teacher's comprehensive *Abilities* (which are also unobservable) and directly on *Beauty*, perhaps because, as argued by Hamermesh and Parker (2005), students pay more attention to, and learn more from, good-looking teachers:

[3] $$Teaching\ Quality_j = \gamma_0 + \gamma_1 Abilities_j + \gamma_2 Beauty_j$$

Therefore, γ_2 measures the direct impact of beauty on the quality of teaching, which may be or may not be equal to zero.

On the other hand, individual ability may depend, at least in part, on beauty. Several studies show that attractiveness has a direct effect on human capital formation: Mocan and Tekin (2010) point out that attractiveness for high school students has a direct impact on student achievement (measured by test scores), perhaps through better interactions with teachers and peers;[15] Persico, Postlewaite and Silverman (2004) argue that the height during adolescence has an impact on individual self-confidence; similarly, Mobius and Rosenblat (2006) find that good-looking individuals are more self-confident.[16] Therefore, we assume that:

[4] $$Abilities_j = \lambda_0 + \lambda_1 Beauty_j$$

where λ_1 should be positive.

If equations [3] and [4] adequately represent the determinants of *Teaching Quality*, then, *Beauty* and *Teaching Quality* would be correlated and the estimates of equation [1] shown in Section 3 – in which *Teaching Quality* is part of the error term – would lead to an upward biased estimate of our parameter of interest ϕ_1 (because of the positive correlation existing between *Beauty* and the error term). More precisely, the probability limit of the estimator of ϕ_1, $\hat{\phi}_1$, in equation [1] is: $p\lim\left(\hat{\phi}_1\right) = \phi_1 + \pi\gamma_1\lambda_1 + \pi\gamma_2$. Although this could be considered a reduced-form effect of beauty,[17] it does not allow us to distinguish between the effect occurring through productivity and the effect of discrimination.

To tackle this issue, considering that in the Italian academic system there is no distinction between teaching and research career and that all professors are supposed to be involved in teaching and research activities, let us assume that research productivity is also correlated to individual abilities:

[5] $$Research\ productivity_j = \beta_0 + \beta_1 Abilities_j$$

---

[14] Specifically, to go from equation [1] to equation [2] we assume that $\varepsilon_{ij} = \pi\left(Teaching\ Quality\right)_j + \nu_{ij}$.

[15] They also show that unattractive individuals are more likely to be involved in criminal activities as young adults.

[16] Some psychological studies (see, for example, Kanazawa 2011; Hatfield and Sprecher, 1986) find a strong correlation between intelligence and beauty.

[17] As Hamermesh and Parker (2005, p. 375) put it: "If there is a characteristic that is caused by a person's physical appearance and that also generates higher instructional ratings, then failing to measure it (and excluding it from the regressions) is correct. For example, if good-looking instructors are more self-confident because their beauty previously generated better treatment by other people, and if their self-confidence makes them more appealing instructors, it is their beauty that is the ultimate determinant of (part of) their teaching success".

Note that we are allowing that the impact of abilities (and, in turn, of beauty) is different between *Teaching Quality* and *Research Productivity*, since γ_1 may be different from β_1.

From equation [5], it is possible to write:[18] $Abilities_j = \frac{1}{\beta_1}\left(Research\ productivity_j\right) - \frac{\beta_0}{\beta_1}$

The latter expression for abilities can be substituted in the teaching quality equation [3]:

[6] $$Teaching\ Quality_j = \left(\gamma_0 - \gamma_1 \frac{\beta_0}{\beta_1}\right) + \frac{\gamma_1}{\beta_1}\left(Research\ productivity_j\right) + \gamma_2 Beauty_j$$

Finally, by substituting eq. [6] in eq. [2], we obtain:

[7]

$$Teaching\ Evaluations_{ij} = \phi_0 + \pi\left(\gamma_0 - \gamma_1 \frac{\beta_0}{\beta_1}\right) + \pi \frac{\gamma_1}{\beta_1}\left(Research\ productivity_j\right) + \pi\gamma_2 Beauty_j + \phi_1 Beauty_j + \phi_2 X_j + \phi_3 W_{ij} + v_{ij}$$

This equation shows that if we are able to control for a measure of individual research productivity, we could at least reduce the upward bias in the estimates of the effect of beauty on *Teaching Evaluations*. In particular, if there is no direct effect of beauty on teaching quality (that is, if $\gamma_2 = 0$), controlling for research productivity would allow us to interpret ϕ_1 as a measure of discrimination, disentangling in this way the impact of beauty on productivity.

There is a lack of evidence in the literature as regards the type of relationship existing between research productivity and teaching quality. One exception is De Paola (2009) who shows that professors with a greater number of scientific publications (or publications of higher quality) produce a positive impact on students' achievement, measured as the grades they obtain in subsequent courses in the same subject. Our finding (see below) that the coefficient on *Research Productivity* is positive and statistically significant reassures us that γ_1 is positive and that it is not much lower than β_1.

To the aim of taking into account individual productivity in our model, we have built a measure of scientific productivity for each instructor using the "Publish or Perish" software based on Google Scholar. More precisely, we have collected data on the number of publications, citations, *h* and *g* indexes, for each instructor. Using these data we have undertaken a principal component analysis to obtain a comprehensive measure of individual productivity (only the first component is considered), which we call *Research Productivity*.[19] Although in our sample there is a wide heterogeneity across Departments in *Research Productivity*,[20] this does not constitute a problem, since in our regressions we control for dummies of Department: in this way, we verify if research productivity has an impact on teaching evaluations within each field of study.

The OLS estimates of the main specifications in which we control for *Research Productivity* are reported in Table 5. Column (1) replicates our preferred specification adding *Research Productivity* as a control. We find that *Research Productivity* has a positive impact on *Teaching Evaluations*. More importantly, we show that controlling for *Research Productivity* the effect of *Beauty* is slightly reduced in comparison to the effects shown in Tables 2 and 3: one additional standard deviation of perceived attractiveness improves the teaching evaluations by about 4.84. This implies that in the previous estimates *Beauty* was in part capturing an effect due to its correlation with productivity. However, the statistical significance of *Beauty* remains high. To compare the effects, consider that one additional

[18] See Angrist and Pischke (2009) for a similar exercise.

[19] See De Paola and Scoppa (2011).

[20] For example, Law professors typically publish less in international journals and more in books (in Italian). It is more hard that this type of publications is present in the Google Scholar Archives.

standard deviation in *Beauty* has an effect on teaching evaluations corresponding to the effect of about 3 (=4.83/1.55) standard deviations in productivity.

In columns (2) and (3) we also report separately estimates for male and female instructors. There is still a significant beauty premium, relevant especially for females. The impact of *Beauty* on teaching evaluations remains almost the same even after controlling for *Age* (column 4) and for course fixed effects (column 5). In column 6 of Table 5 we use the dummies *Very Attractive* and *Very Unattractive*, and we show that also when controlling for research productivity it emerges a premium for being handsome and a penalty for being homely (in the latter case, very imprecisely estimated, *p*-value=0.23).

The effect of *Research Productivity* on teaching ratings in Table 5 is generally positive and statistically significant (although in some specifications the *p*-values are around 0.10 and in column 3 is not at all significant) suggesting that more productive teachers are probably better at presenting the material and obtain higher instructional ratings. If one is willing to believe that the direct effect of instructor's beauty on the attention paid by students is negligible (that is, $\gamma_2 = 0$), then the effect of *Beauty* in the estimates has been depurated by its impact through productivity and the estimated coefficient in Table 5 represents a measure of Becker-type discrimination in favor of good-looking instructors and against ugly ones.

Table 5. Dealing with the Correlation between Beauty and Ability. OLS regressions

	(1) All	(2) Males	(3) Females	(4) All	(5) All	(6)
Beauty	4.8386***	4.1097**	5.8428***	4.9479***	4.9039***	
	(1.0881)	(1.6952)	(1.1696)	(1.1327)	(1.2545)	
Female	-2.9720			-2.8308	-0.3400	-1.6936
	(2.1772)			(2.1791)	(2.3274)	(2.2303)
Full Professor	-5.6688*	-3.2575	-11.2355***	-6.1152*	-2.9083	-5.4252*
	(2.9020)	(3.5173)	(3.7228)	(3.5682)	(2.6744)	(2.8732)
Associate Professor	-2.0047	1.2929	-5.7390*	-2.2550	-1.6114	-2.5458
	(2.5874)	(3.5703)	(3.0879)	(2.6779)	(2.9682)	(2.6976)
Adjunct Professor	0.2323	4.5946	-4.4906	0.7041	1.5486	0.1834
	(2.3496)	(3.7563)	(2.7577)	(2.4067)	(2.7829)	(2.3422)
# Students in Class	-0.0299*	-0.0209	-0.0599***	-0.0302*	-0.0165	-0.0235
	(0.0167)	(0.0203)	(0.0224)	(0.0163)	(0.0207)	(0.0172)
Second Level Degree	2.0889	3.1490	0.4012	2.1264		2.0440
	(1.6626)	(2.0723)	(2.2490)	(1.6661)		(1.6462)
Research Productivity	1.5538*	2.0158*	0.1569	1.6141*	2.5274**	1.8598**
	(0.9240)	(1.2152)	(0.9746)	(0.9409)	(1.1641)	(0.8732)
Age				0.0475		
				(0.1679)		
Very Attractive						5.7922**
						(2.3117)
Very Unattractive						-3.7298
						(3.0955)
Observations	2338	1404	934	2313	2337	2338
Adjusted R-squared	0.197	0.216	0.228	0.189	0.377	0.184
Clusters	193	110	83	190	193	193
Course Fixed Effects	NO	NO	NO	NO	YES	NO

Notes: The Table reports Weighted Least Squares estimates (observations are weighted by the number of students attending each course). The dependent variable is *Teaching Evaluations*. In all the regressions we control for (9) Department dummies and (7) year dummies. In regression 5 we control for course fixed effects and we do not control for Department and year dummies. Standard errors (reported in parentheses) are corrected for heteroskedasticity and clustered at the teacher level. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

5. Concluding Remarks

Using a quite large dataset on student evaluation of teaching (with about 200 instructors and 2,300 courses), we have shown that beauty matters even in an intellectual profession like university teaching. Better looking instructors receive much higher instructional ratings. Controlling for a wide range of course's and instructor's characteristics, we find that an instructor with a level of beauty higher of one standard deviation receives an evaluation of his/her teaching quality of 5-6 percentage points higher, corresponding to about 0.30 standard deviations of *Teaching*

Evaluations. The effect is highly statistically significant in all specifications. Our estimates are in line with Hamermesh and Parker (2005) and much higher in magnitude than the effect estimated by Süssmuth (2006) for German students. In contrast to Hamermesh and Parker and in line with some old-age stereotypes, we find that beauty is more important for female than for male instructors.

We have tried to investigate whether the impact is due to discrimination or productivity, that is, if the better evaluations obtained by good-looking teachers are determined by their possess of greater abilities or simply by "customer" discrimination. In a simple theoretical framework we proxy instructors' abilities with their scientific productivity (exploiting the fact that in the Italian academic system all the professors are supposed to do teaching and research) and we use that measure in the empirical framework. Controlling for instructors' abilities, we show that while scientific productivity is significantly correlated to teaching evaluations, the impact of beauty on teaching evaluations remains positive and highly statistically significant (although slightly reduced), suggesting that the discrimination against ugly instructor is the key factor explaining the role of beauty.

References

Angrist, J., Pischke S., (2009), *Mostly Harmless Econometrics*, Princeton University Press, Princeton.

Becker, W. and Watts, M. (1999), "How departments of economics evaluate teaching", *American Economic Review. Papers and proceedings*, 89, 355-359.

Cawley, J. (2004), "The impact of obesity on wage. *Journal of Human Resources,* 39 (2), 451-474.

Cipriani, G. and Zago, A. (2011). Productivity or discrimination? Beauty and the exams", *Oxford Bulletin of Economics and Statistics*, 73 (3), 428-447.

De Paola, M., (2009) "Does Teacher Quality Affect Student Performance? Evidence from an Italian University", *Bulletin of Economic Research*, 61: 4, pp. 353-377.

De Paola, M., Scoppa, V., (2011) "Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy", Working Paper n. 6-2011, Dipartimento di Economia e Statistica, Università della Calabria.

Fletcher, J. (2009), "Beauty vs. brains: Early labor market outcomes of high school graduates", *Economics Letters*, 105, 321-325.

Hamermesh, D. (2011), *Beauty pays: Why attractive people are more successful*, Princeton University Press.

Hamermesh, D. and Biddle, J. (1994), "Beauty and the labor market", *American Economic Review,* 84 (5), 1174-1194.

Hamermesh, D. and Parker, A. (2005), "Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity", *Economics of Education Review*, 24, 369-376.

Harper, B. (2000), "Beauty, stature and the labour market: a British cohort study", *Oxford Bulletin of Economics and Statistics*, 62, 771-800.

Hatfield, E., and Sprecher, S. (1986), *"Mirror, mirror: The importance of looks in everyday life", Albany: SUNY Press.*

Kanazawa, S. (2011), "Intelligence and physical attractiveness", *Intelligence*, 39 (1), 7-14.

Mobius, M. and Rosenblat, T. (2006), "Why beauty matters", *American Economic Review,* 96 (1), 222-235.

Mocan, H. and Tekin, E. (2010), "Ugly criminals'", *Review of Economics and Statistics*, 92, 15-30.

Perotti, R., (2002), "The Italian University System: Rules vs. Incentives". Presented at the first conference on "Monitoring Italy", ISAE, Rome.

Persico, N., Postlewaite, A. and Silverman, D. (2004), "The effect of adolescent experience on labor market outcomes: the case of height", *Journal of Political Economy*, 112 (5), 1019-1053.

Sen, A., Voia, M., and Woolley, F. (2010), "Hot or not: How appearance affects earnings and productivity in academia", Mimeo, Carleton University.

Süssmuth, B. (2006), "Beauty in the classroom: Are German students less blinded? Putative pedagogical productivity due to professors' pulchritude: Peculiar or pervasive?", *Applied Economics, 38*, 231-38.