Does Beauty Matter in Undergraduate Education?[*]

Tatyana Deryugina, University of Illinois[+]

Olga Shurchkov, Wellesley College[++]

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE OR CIRCULATE.

Abstract:

The evidence for the "beauty premium" is widespread: from politics to professional sports to the labor market, more attractive individuals achieve greater success in terms of earnings and status, on average. However, previous literature has failed to uncover the mechanism behind this correlation, because most studies lack measures of ability. Unlike previous studies, we find that our ability measures at the undergraduate level (such as SAT scores and quantitative reasoning scores) are *negatively* correlated with attractiveness. We find no evidence of an additional bias in favor of more attractive students in our data. This suggests that the beauty premium we observe later in a person's life is driven by either selecting into industries where the person has a comparative advantage or by discrimination.

I.   Introduction

In many settings, discrimination based on characteristics such as gender, age, race, and national origin is illegal. Appearance-based discrimination, while not currently unlawful, has been the subject of several lawsuits in recent years.[1] In parallel, the academic literature has documented a positive correlation between earnings and perceived attractiveness for both men and women (Hamermesh and Biddle, 1994; Hamermesh and Biddle, 1998). However, the causes of this correlation, dubbed the "beauty premium," remain unclear.[2] In particular, in the absence of relevant ability measures, most existing studies are unable to test whether attractiveness and ability are correlated and if attractiveness still matters once ability is controlled for.

In this paper, we test for the existence of a beauty premium among female college students. Our unique dataset allows us to control for several proxies of ability and isolate the relationship between course grades and attractiveness, net of ability. Moreover, we can directly test whether our measures of ability, which include SAT scores, freshman quantitative reasoning (QR) scores, and admissions ratings assigned to applicants, are correlated with attractiveness. While all these evaluations are assigned without observing the applicant's appearance, course grades are not. Thus, the former provide measures of ability that are independent of any beauty bias, while the latter reflect both ability and potential bias.

We find that attractiveness is negatively correlated with both SAT and QR scores. Specifically, a one standard deviation increase in attractiveness is associated with scoring 7.8 points lower on the math SAT, 11.6 points lower on the verbal SAT, 6.3 points lower on the writing SAT, and 0.28 points lower on the QR test (out of 18). To our knowledge, our study is the first to find that attractiveness is *negatively* correlated with ability, as measured by these tests. However, once we control for the math and verbal SAT scores, we find no relationship between admissions scores and attractiveness. This suggests that other applicant characteristics observed by the admissions committee, such as recommendation letters and personal essays, which we do not observe, are not correlated with attractiveness on net.

Our preliminary findings also indicate that, once we control for SAT scores or one of the other ability measures, there is no significant relationship between course grades and attractiveness. The same is true of overall GPA, freshman GPA, and GPA in humanities, social sciences, or sciences. We also find that there is no beauty premium in courses taught by either male or female professors. However, we find some heterogeneity by course size. Specifically, there is a substantial beauty premium in smaller courses: a one standard deviation increase in attractiveness is associated with a 0.026 grade increase (on a 4-point scale) in courses with the lowest quartile of students (13 or fewer). Our analysis is robust to including extensive controls, such as course level (introductory, intermediate, or advanced), the student's race, year of enrollment, semester-by-course type (humanities, social sciences, or sciences), and department (e.g., English, Economics, Math) fixed effects.

Surprisingly, there are very few papers that consider the relationship between attractiveness and ability. Prokosch et al. (2005) look at the relationship between body symmetry

---

[1] See for example MSNBC "Librarian sues Harvard of discrimination," March 21, 2005.

[2] The importance of beauty has been studied in contexts other than the labor market. See Ravina (2009) for the beauty premium in credit markets; Andreoni and Petrie (2008) for the beauty premium in public goods games; Wilson and Eckel (2006) for the beauty premium in trust games; Berggren et al. (2007) for the beauty premium in electoral outcomes.

and cognitive test results in 78 young males, finding that body symmetry is positively correlated with cognitive performance. Satoshi (2011) shows that there is a positive association between IQ test results and physical attractiveness in British and American children of both sexes. In a laboratory experiment, Mobius and Rosenblat (2006) show that more attractive subjects do not perform better in a maze solving task. However, it is not clear whether their results generalize to broader measures of ability.

The question of whether attractiveness and ability are correlated is very important for interpreting the results of the previous literature, which considers the relationship between attractiveness and earnings but lacks measures of ability beyond educational attainment (Hamermesh and Biddle, 1994; Hamermesh and Biddle, 1998). A key exception is Berri et al. (2011), who consider the relationship between earnings and attractiveness of NFL quarterbacks (as measured by facial symmetry), controlling for performance. Averett and Korenman (1996) also control for test scores. However, they consider the relationship between obesity and earnings, rather than attractiveness and earnings.

In an experimental setting, Deryugina and Shurchkov (2012) use realistic tasks to test for both the existence of a beauty premium and performance differentials between less and more attractive subjects. They find that there is no performance differential by attractiveness in two of the tasks (bargaining and data analysis) and a *negative* performance differential in the data entry task.

The paper most closely related to ours is von Bose (2012), who looks at the wages and educational attainment of more versus less attractive youths from the National Labor Survey of Youths (NLSY). In contrast to our findings, she finds that more attractive youths, especially males, have higher high school GPAs than less attractive ones. However, unlike our study, this paper does not shed light on the mechanism behind the observed premium and does not have disaggregated data at the course level.

In addition to being relevant to the labor market, our study provides important insights into appearance-based discrimination in undergraduate education. Several larger academic institutions in the US (many law schools, for instance) have begun to move toward a single- or double-blind exam grading policy. Since the administrative costs of blind exam grading are significantly higher than those of the non-blind status quo, many academic institutions have elected not to enforce such policies. However, our study is the first to test for the existence of appearance-based biases that would justify the need for a blind grading system, although some have tested for familiarity bias more generally. Our results do not offer support for blind grading on the basis of eliminating a bias in favor of more attractive students, since we do not find a beauty premium in our data.

The rest of the paper is organized as follows. Section II describes our data. Section III outlines the conceptual framework and the empirical strategy. Section IV discusses the findings. Section V concludes.

II. Data

Our dataset consists of 794 female alumni who graduated from an unnamed college between the years 2002 and 2011.[3] To measure attractiveness, we used the pictures taken when the alumni were freshmen. A key advantage of our data is that the pictures are not chosen by the student: All are photographed for their student ID cards by campus officials. The pictures were subsequently rated by current male and female students from a college in another state. Each picture was rated by at least 25 male and 25 female raters. We then demean and average the ratings to obtain the mean attractiveness rating of each alumna. For more details about the rating procedure, see Appendix A.

The mean attractiveness rating is then matched to the alumna's academic record, which includes SAT scores, course-level grades, demographic information, financial aid, international status, and scores from a quantitative reasoning (QR) test that all freshmen are required to take. Like the SAT, the QR test is scored blindly. Moreover, we observe each student's admissions ratings, as assigned by three or more application reviewers. With few exceptions, application reviewers likewise do not observe the student's appearance.[4]

We also have detailed information on course characteristics, including department, course level (introductory, intermediate, or advanced), total enrollment, and the gender of the instructor. This is important, as admissions ratings, SAT scores, and QR scores may be noisy measures of ability. For example, attractiveness may be correlated with communication skills, which are not well measured by these tests. The richness of our data allows us to test whether the beauty premium varies with course-specific characteristics, such as course size and subject matter. Thus, we should be able to rule out many competing explanations for any residual correlation between attractiveness and grades.

Our data are not without limitations. First, we do not observe parental income, a potentially important control. However, we do observe the amount of need-based and non-need-based loans and grants that a student receives, which we use as a proxy for income. We also do not observe post-college outcomes. Thus, we cannot test whether a higher GPA is predictive of a higher salary, for example.

Table 1 presents the summary statistics for our sample. The mean attractiveness rating is 0 by construction and ranges from -2.7 to 2.4. The admissions rating ranges from -2 to -1, with higher ratings corresponding to a higher chance of admission. Average GPA in the sample is fairly high, ranging from 3.23 in the sciences to 3.50 in humanities. On average, 94% of the students pass the quantitative reasoning test. The average grant amount is nearly $50,000. Need-based and other loans are substantially smaller, averaging $1,626 and $696, respectively.

---

[3] This study was conducted with IRB approval. Individuals had to consent to have their photographs included in the study. About 5,000 alumni were contacted for consent. Informed consent form and description of the project sent to the alumni are available upon request.

[4] Some international applicants have TOEFL scores that are accompanied by a picture. In some cases, applicants are interviewed by a member of the admissions staff or by an alumna. However, the application reviewers only have access to the interviewer's comments which do not contain information about the applicant's appearance.

Table 1: Summary Statistics

	Mean	Std. dev.	Min	Max	Obs
Attractiveness rating	0	0.86	-2.69	2.42	794
Admissions rating	-1.35	0.13	-2	-1	794
Total GPA	3.47	0.31	0	4	794
Freshman GPA	3.40	0.35	2.04	4	793
Humanities GPA	3.50	0.27	2.04	4	794
Social Science GPA	3.43	0.39	1.84	4	738
Science GPA	3.23	0.50	1.40	4	790
Math SAT score	684	60	490	800	765
Verbal SAT score	704	61	450	800	765
Writing SAT score	704	67	490	800	551
QR test score	13.25	2.60	2	18	794
Passed QR test	0.94	0.23	0	1	794
Grants	49,363	56,228	0	202,198	794
Need-based loans	1,626	3,265	0	17,675	794
Other loans	696	2,762	0	24,700	794

One concern is that the sample of women who gave consent to participate in our study may not be representative of all students at the college.[5] To test for this, we compare the mean test scores, admissions ratings, and year of enrollment for the entire population of alumni who graduated between 2002 and 2011 and for the consenting group. The results are shown in Table A1. Overall, the group of consenting students has significantly higher test scores and admissions ratings. They also enrolled in the college about half a year later than the general population of women, on average. As long as there is no systematic difference in the relationship between attractiveness and test scores/GPA between the consenting group and the general student population, our analysis is valid despite the baseline differences.

III. Conceptual framework

The ideal experiment to test whether there is a beauty bias would involve all students taking an identical test and subsequently randomizing whether the tests are graded blindly. We are not aware of a situation where such an experiment arises naturally, but we believe that our data allow us to approximate such a scenario. Of course, as with any non-controlled experiment, there are identification concerns that we do our best to address.

---

[5] Selection effects could not be avoided in this study because of the informed consent process in compliance with the IRB guidelines.

Using data from a natural setting rather than an experimental one has two key advantages. First, our estimates are much more likely to have external validity. An experiment would not allow us to simulate the extensive interactions that may occur in a classroom setting. In other words, results from the ideal experiment may not generalize to a situation where individuals interact with each other repeatedly. Second, we have many more students in the sample than would be feasible in an experimental setting, allowing us to have more precise estimates.

The key to our analysis is that we have two types of ability measures: one that is measured without observing the student's appearance (SAT scores, QR scores, and admissions ratings) and one where appearance is observable (grades). Our identifying assumption is that there are no unobserved skills that are (a) uncorrelated with the skills measured by the SAT tests, QR tests, or admissions ratings and (b) correlated with attractiveness. If this assumption holds, then we can attribute any excess correlation between grades and attractiveness to beauty-based discrimination.

To test for the correlation between attractiveness and ability, we estimate the effect of beauty rating on several ability measures:

$$Ability_i = \alpha_1 Rating_i + X_i'\gamma + \varepsilon_i \qquad (1)$$

where i represents the individual alumna, $Ability$ refers to scores on SAT (Verbal, Math or Writing) and QR tests, or admission rating; $Rating$ refers to the demeaned beauty rating; and X represents the vector of student characteristics, including race fixed effects, student loans, and year of enrollment fixed effects.

Our preferred specification in estimating the effect of attractiveness on college GPA includes measures of ability as controls:

$$Grade_{ijt} = \beta_1 Rating_i + Ability_i'\delta + Z_i'\rho + \theta_j + \mu_{jt} + \varepsilon_{ijt} \qquad (2)$$

where i represents the individual alumna, $Grade$ refers to course grade; $Rating$ refers to the demeaned beauty rating; $Ability$ measures ability according to SAT scores (Verbal and Math), or QR test scores, or admission ratings; Z represents the vector of student characteristics, including race fixed effects, student loans, and year of enrollment fixed effects; θ_j refer to the course-level fixed effects, including department fixed effects, and μ_{jt} denotes course-by-semester fixed effects.

A number of scenarios could violate our identification assumption. For example, more attractive women could also be more confident, all else equal. This might lead them to do better in courses where participation is an important part of the grade while not helping them do better on the SATs. Although we cannot fully rule out this possibility, the richness of our dataset allows us to see if the beauty premium varies in a way that would be predicted by alternative explanations, such as another unobservable skill.

IV.   Results and discussion

Our key contribution is to test whether attractiveness is correlated with ability (as measured by test scores) and if there is a correlation between attractiveness and academic performance once we control for ability. We begin our analysis by estimating the relationship between QR/SAT scores and attractiveness. Because all of these tests are scored blindly, there is no concern that examiners are discriminating against or for more attractive people. Any correlation between test scores and attractiveness will thus reflect skill differences rather than bias.

The results are shown in Table 2. Our preferred specification includes controls for the year of enrollment, the student's race, and logs of financial aid amounts by category (need-based loans, other loans, and grants). A one standard deviation increase in attractiveness is associated with a 7.8 point decrease in the individual's math SAT score (about 1.1% of the mean), a 11.6 point decrease in the verbal score (about 1.6% of the mean), and a 6.3 point decrease in the writing score (about 0.9% of the mean). Finally, more attractive people also score about 0.28 points lower on the first-year QR score (about 2.1% of the mean). These results are very robust to varying the controls.

Table 2: Relationship between attractiveness and test scores

	Math SAT			Verbal SAT		
Attractiveness rating	-7.953**	-7.784**	-7.794***	-12.560***	-12.321***	-11.594***
	(3.090)	(3.040)	(2.783)	(2.892)	(2.895)	(2.764)
Year of enrollment f.e.	N	Y	Y	N	Y	Y
Financial aid controls	N	N	Y	N	N	Y
Race f.e.	N	N	Y	N	N	Y
Observations	765	764	764	765	764	764
R-squared	0.01	0.06	0.22	0.02	0.04	0.11
	Writing SAT			QR test		
Attractiveness rating	-7.318*	-7.308*	-6.263*	-0.254**	-0.252**	-0.276**
	(3.797)	(3.779)	(3.677)	(0.127)	(0.127)	(0.121)
Year of enrollment f.e.	N	Y	Y	N	Y	Y
Financial aid controls	N	N	Y	N	N	Y
Race f.e.	N	N	Y	N	N	Y
Observations	551	551	551	794	793	793
R-squared	0.01	0.03	0.12	0.01	0.01	0.12

Note: Robust standard errors clustered at the student level in parentheses. Significance levels: *10 percent, ** 5 percent, *** 1 percent.

To our knowledge, the finding that more attractive people perform worse on standard ability tests is new and goes against the hypothesis that more attractive individuals earn more because of some unobservable skill, including intelligence. Although SAT scores do not fully capture intelligence, in our sample they are predictive of higher course grades, as we show later. This suggests that, absent bias or another unobservable skill that is uncorrelated with doing well on the SATs/QR, more attractive individuals should not receive higher grades.

Of course, SAT and QR scores provide only a limited measure of ability. It is possible that more attractive individuals possess some other skill that we cannot observe and that is correlated with GPA. We cannot fully rule out its existence, but we can improve on the previous analysis by looking at the relationship between the admissions rating and attractiveness, controlling for SAT scores. Although the admissions committee does not observe everything about the applicant, they have a much richer information set than we do, including extracurricular activities, recommendation letters, and personal essays. Through these, it is possible that the admissions committee receives signals about other skills that predict college success and that may be correlated with attractiveness. However, with very few exceptions, the admissions committee does not observe the applicant's appearance. Thus, we do not have to worry about beauty-based discrimination affecting the scores they assign to the applicant.

The results of this analysis are shown in Table 3. Although the admissions ratings of more attractive alumni are worse on average, we find that this is entirely driven by their lower SAT scores. Once we control for math and verbal SAT scores, there is no relationship between the admissions rating and attractiveness. This rules out the possibility that attractiveness is correlated (on net) with characteristics that admissions officers can observe but we cannot. While the admissions rating result does not fully rule out an unobservable characteristic correlated with both appearance and GPA, it diminishes the likelihood of its existence.

Table 3: Relationship between attractiveness and admissions ratings

	(1)	(2)	(3)	(4)	(5)	(6)
Attractiveness rating	-0.018***	-0.018***	-0.018***	-0.005	-0.005	-0.006
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Math SAT score				0.008***	0.008***	0.007***
				(0.001)	(0.001)	(0.001)
Verbal SAT score				0.007***	0.007***	0.007***
				(0.001)	(0.001)	(0.001)
Year of enrollment f.e.	N	Y	Y	N	Y	Y
Financial aid controls	N	N	Y	N	N	Y
Race f.e.	N	N	Y	N	N	Y
Observations	794	793	793	765	764	764
R-squared	0.01	0.04	0.13	0.32	0.32	0.35

Note: Robust standard errors clustered at the student level in parentheses. Significance levels: *10 percent, ** 5 percent, *** 1 percent.

We proceed to look at the relationship between course-level grades and attractiveness. The results are shown in Table 4. Because of the detailed nature of the course-level data, we are able to include many controls that may be correlated with attractiveness, such as course type (humanities, sciences, social sciences) and department (e.g., math, French, economics). We also include course-type-by-semester fixed effects, year of enrollment fixed effects, race fixed effects, and financial aid amounts as controls. Standard errors are clustered at the student level.

Our results indicate that there is no significant correlation between attractiveness and course-level grades, although the point estimates are positive. The inclusion of SAT scores increases the magnitude of our point estimates by quite a bit, but they remain statistically insignificant. The results are robust to excluding the controls mentioned above and to using QR test scores or admissions ratings as ability controls.[6]

When we look at cumulative GPA at the student level (Appendix Table A1), we find that the beauty rating is marginally significant in some cases. However, because we cannot control for students selecting into courses, we view the course-level results as more reliable.

Table 4: Relationship between attractiveness and course grades

Attractiveness rating	0.001	0.015	0.005	0.019	0.018	0.018
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Math SAT score		0.010***		0.010***	0.009***	0.010***
		(0.002)		(0.002)	(0.002)	(0.002)
Verbal SAT score		0.006***		0.006***	0.005***	0.005***
		(0.002)		(0.002)	(0.002)	(0.002)
Course type f.e.	Y	Y				
Course type by semester f.e.			Y	Y	Y	Y
Course level f.e.			Y	Y	Y	Y
Race f.e.					Y	Y
Loans/grants					Y	Y
Subject f.e.						Y
Observations	20,608	19,916	20,608	19,916	19,916	19,916
R-squared	0.039	0.061	0.072	0.094	0.106	0.128

Note: Standard errors clustered at the student level in parentheses. Significance levels: *10 percent, ** 5 percent, *** 1 percent. All regressions include year of enrollment fixed effects. Course level is either beginning, intermediate, or advanced. Course type is either humanities, social sciences, or sciences. Subject fixed effects represent a specific subject code, such as English, Economics, Physics, etc.

---

[6] For space reasons, we do not show the specifications using QR test scores or admissions ratings as controls. Results are available upon request.

Finally, we test for heterogeneity in the beauty premium between small and large courses and between male and female instructors. We expect the beauty premium, if it exists, to be stronger in small courses, where the professor has a greater probability of knowing what the students look like. We use two measures of course size – an indicator for below and above median enrollment (18 or fewer v. 19 or more students) and indicators for enrollment quartiles.

The results are shown in Table 5. Overall, it appears that the beauty premium does not exist in courses taught by either male or female instructors, and point estimates are actually higher in courses taught by females. Once we include our preferred controls, there is also no evidence of a beauty premium in courses with below median or above median enrollment. However, there is a significant beauty premium in the smallest courses (13 students or less), with more attractive students receiving grades that are 0.026 points higher.[7] This effect on grades is equivalent to the average woman in our sample scoring 26 points (3.8%) higher on the math SAT or scoring 52 points (7.4%) higher on the verbal SAT. In turn, a 26-point increase in math SAT is correlated with an admissions rating that is 0.18 higher, while the 52-point increase in the verbal SAT is correlated with an increase in the admissions rating of 0.36 points. This is a substantial increase over the mean of 1.35 and can significantly affect an applicant's chance for admission.

Our results are subject to several limitations. First, our sample is conditional on being admitted to college. This is not likely to bias our results, as we find that admissions ratings are uncorrelated with attractiveness, once SAT scores are controlled for. Second, we do not observe any post-graduation outcomes, such as earnings. Thus, we cannot claim that the absence of a beauty premium in the college setting translates into an absence of a beauty premium in the workplace. The behavior of college professors may generally differ from the behavior of managers in the workplace. Professors may exert more effort to correct any bias they may have in order to be fair to the students. Finally, the college from which we have data may differ from the average undergraduate institution. However, we believe that it is fairly representative of an elite college or university.

---

[7] The smallest courses tend to be upper-level courses in humanities and languages. Because we include course level and department fixed effects in our preferred specification, our results are not driven by this fact.

Table 5: Course-level heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Attractiveness rating x female professor	0.020	0.019	0.019						
	(0.013)	(0.013)	(0.013)						
Attractiveness rating x male professor	0.012	0.011	0.011						
	(0.012)	(0.012)	(0.013)						
Attractiveness rating x below median size				0.030*	0.029*	0.026			
				(0.017)	(0.016)	(0.016)			
Attractiveness rating x above median size				0.021*	0.019*	0.019			
				(0.012)	(0.012)	(0.012)			
Attractiveness rating x bottom quartile in size							0.029**	0.029**	0.026**
							(0.014)	(0.013)	(0.013)
Attractiveness rating x second quartile in size							0.012	0.01	0.011
							(0.013)	(0.013)	(0.013)
Attractiveness rating x third quartile in size							0.018	0.016	0.018
							(0.015)	(0.016)	(0.016)
Attractiveness rating x top quartile in size							0.006	0.005	0.006
							(0.015)	(0.015)	(0.015)
Loans/grants	N	Y	Y	N	Y	Y	N	Y	Y
Race f.e.	N	Y	Y	N	Y	Y	N	Y	Y
Subject f.e.	N	N	Y	N	N	Y	N	N	Y
Observations	19,027	19,027	19,027	19,545	19,545	19,545	19,545	19,545	19,545
R-squared	0.09	0.11	0.13	0.09	0.11	0.13	0.09	0.11	0.13

Note: Standard errors clustered by student in parentheses. Significance levels: *10 percent, ** 5 percent, *** 1 percent. All regressions include controls for SAT math and verbal scores, course levels (beginning, intermediate, advanced), year enrolled fixed effects, and year by semester by course type fixed effects (humanities, social sciences, sciences). In addition, main effects for male professor, above median course size, and course size quartiles are included in the respective regressions.

V.   Conclusion

The issue of beauty-based discrimination has gained increasing attention in recent years. Previous literature has found that more attractive people earn more on average. However, the source of the beauty premium has remained elusive. In particular, the most prominent studies lack measures of ability. As a result, they are unable to test whether the beauty premium is driven by higher unobserved ability of more attractive people or by discrimination.

We contribute to the previous literature both by looking at whether the beauty premium exists earlier in life, in the form of college grades, and by controlling for several measures of ability. Unlike previous studies, we find that our ability measures (test scores) are *negatively* correlated with attractiveness. This suggests that the beauty premium we observe later in a

person's life is driven by either selection into industries where the person has a comparative advantage or by discrimination.

Moreover, we find that there is no relationship between a student's course grades and her attractiveness, whether or not ability measures are controlled for. However, we do find a beauty premium in the smallest courses. Thus, on average, our results do not offer support for blind grading on the basis of eliminating a bias in favor of more attractive students. Overall, our findings also suggest that the beauty premium in wages does not stem from a beauty premium in college.

The results suggest several directions for future research. First, reproducing the analysis with a mixed-gender group of college graduates would enhance our understanding of gender differences in the role of appearance in undergraduate education. Second, tracking the post-graduation outcomes, such as labor force status, earnings, and history of promotions of the alumni in our sample would shed light on whether a beauty premium for these individuals develops later, after they have entered the labor market.

References

Andreoni, James and Ragan Petrie. 2008. "Beauty, gender and stereotypes: Evidence from laboratory experiments." *Journal of Economic Psychology*, 29(1): 73–93.

Averett, Susan and Sanders Korenman. 1996. "The Economic Reality of the Beauty Myth." *Journal of Human Resources*, 31(2): 304-330.

Berggren N., H. Jordahl, and P. Poutvaara. 2007. "The Looks of a Winner: Beauty, Gender, and Electoral Success," CESifo Working Paper Series 2002, CESifo Group Munich.

Berri, David J., Rob Simmons, Jennifer Van Gilder, and Lisle O'Neill. 2011. "What does it mean to find the face of the franchise? Physical attractiveness and the evaluation of athletic performance." *Economics Letters*, 111(3): 200–202.

von Bose, Caroline. 2012. "Child Stars vs. Ugly Ducklings: Does adolescent attractiveness contribute to the beauty premium?" Working paper.

Deryugina, T. and O. Shurchkov. 2012. "Are Appearances Deceiving? The Nature and Evolution of the Beauty Premium." Working paper.

Hamermesh, D. and J. Biddle. 1994. "Beauty and the Labor Market." *American Economics Review,* 84(5): 1174-1194.

Hamermesh, D. and J. Biddle. 1998. "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre." *Journal of Labor Economics,* 16(1): 172-201.

Hamermesh, D. 2011. "Beauty Pays: Why Attractive People Are More Successful." Princeton University Press, Princeton, NJ, USA.

Mobius, Markus M., and Tanya S. Rosenblat 2006. "Why Beauty Matters." *American Economic Review*, 96(1): 222–235.

Ravina, E. 2009. "Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets", under revision.

Wilson, R. K., and C. C. Eckel. 2006. "Judging a Book by its Cover: Beauty and Expectations in a Trust Game," *Political Research Quarterly*, 59(2): 189-202.

Appendix A. Rating procedure and instructions

All pictures were rated by at least 25 female and 25 male raters. Raters were students at a college in a different state and were pre-screened to ensure that they were not familiar with students from the college of interest.

Raters were shown pictures of each student and asked to rate her physical appearance on a 1-10 point scale. Five of the numbers had descriptions describing the level of attractiveness corresponding to that number (see experimental instructions on the next page). Raters were instructed to choose the numbers without descriptions if they felt the student's appearance fell in between the two descriptions.

Each rater was shown four sets of about 100 photos. The order of the photos within each set was randomized for each rater. In early stages of the experiment, we compared the mean and standard deviation of ratings across different sets to see if having subjects rate 400 pictures led to fatigue. There was no significant difference in either the mean or standard deviation of ratings for earlier and later sets, which led us to conclude that 400 pictures was not an excessive amount. We did not use data from three raters who chose 1's 40% or more percent of the time. The "1" option was the closest to the "Next" button. Thus, these subjects were most likely trying to get through the experiment as quickly as possible.

Instructions for the experiment

You are about to participate in an experiment involving the perception of appearance. Once the experiment begins, you will see a photograph of an individual along with the following prompt:

Rate this person's physical appearance using the following scale:

10 strikingly handsome or beautiful
9
8 good-looking (above average for age and sex)
7
6 average looks for age and sex
5
4 quite plain (below average for age and sex)
3
2 homely
1

Choose the number that best corresponds to your evaluation. Choose the numbers without descriptive text (1, 3, 5, 7, and 9) if you feel the person's appearance falls in between the descriptions found in the adjacent numbers.

After you have chosen a number, click "Next". You will then see another photograph and be asked to repeat the procedure. Continue selecting the number you feel best reflects your assessment of the individual's appearance until you are told to stop.

Appendix B. Student-level Analysis

In Table A2, we show the relationship between GPA and attractiveness, with and without controlling for SAT scores, QR scores, and admissions rating. In Panel A, we use freshman GPA because we may expect the correlation between freshman year attractiveness and grades to diminish over time. In Panel B, we use cumulative GPA. We find limited evidence that GPA is affected by attractiveness, and the results are sensitive to which controls are included. Once we control for test scores, race, and grants/loans (our preferred specification), the correlation between cumulative GPA and attractiveness is positive and marginally significant. Specifically, a one standard deviation increase in attractiveness is associated with a 0.026 increase in total GPA.

Again, the main concern with this analysis is that SAT scores, admissions ratings, and QR scores measure ability differently than the average college course. Although we cannot address this concern perfectly, we can separate college courses into those that require different skills, namely humanities, social sciences, and sciences. We then compute students' GPA in each area and replicate the analysis in Table A2. The results are shown in Table A3. For space reasons, we only show the specifications with a full set of controls – the results are not significantly affected by their inclusion. Overall, there is no evidence of a heterogeneous relationship between attractiveness and the different GPA measures.

Appendix Tables

Table A1: difference between the general student population and those giving consent

	Panel A: Test scores and enrollment date					
	Math SAT	Verbal SAT	Writing SAT	QR test score	Admissions rating	Year enrolled
All students	674	684	683	12.94	-1.38	2002
Consenting students	684	704	704	13.25	-1.35	2003
Consenting - all	9.13***	20.08***	21.36***	0.31***	0.03***	0.53***
Observations	5,894	5,894	4,544	6,158	6,157	6,160

Note: Significance levels: *10 percent, ** 5 percent, *** 1 percent.

Table A2: Relationship between attractiveness and GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: freshman GPA								
Attractiveness rating	-0.003	-0.004	0.014	0.011	0.014	0.011	-0.042	-0.031
	(0.017)	(0.016)	(0.017)	(0.016)	(0.016)	(0.016)	(0.065)	(0.056)
Passed QR x attractiveness rating							0.05	0.038
							(0.067)	(0.059)
Ability controls	None	None	SAT	SAT	Admissions rating	Admissions rating	QR	QR
Race f.e.		Y		Y		Y		Y
Loans/grants		Y		Y		Y		Y
Observations	793	793	764	764	793	793	793	793
R-squared	0.05	0.12	0.12	0.15	0.18	0.21	0.14	0.17
Panel B: overall GPA								
Attractiveness rating	0.015	0.015	0.027*	0.025*	0.028*	0.026*	0.007	0.013
	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)	(0.071)	(0.063)
Passed QR x attractiveness rating							0.016	0.010
							(0.073)	(0.065)
Ability controls	None	None	SAT	SAT	Admissions rating	Admissions rating	QR	QR
Race f.e.		Y		Y		Y		Y
Loans/grants		Y		Y		Y		Y
Observations	793	793	764	764	793	793	793	793
R-squared	0.03	0.09	0.07	0.10	0.12	0.15	0.09	0.13

Note: Robust standard errors in parentheses. Significance levels: *10 percent, ** 5 percent, *** 1 percent. All regressions include year of enrollment fixed effects.

Table A3: Varying the GPA measure

	Humanities GPA			Social sciences GPA			Sciences GPA		
Attractiveness rating	0.015	0.019	-0.001	0.008	0.002	-0.064	0.033	0.036	0.135
	(0.012)	(0.011)	(0.060)	(0.018)	(0.018)	(0.089)	(0.023)	(0.023)	(0.112)
Passed QR x attractiveness rating			0.014			0.061			-0.107
			(0.062)			(0.091)			(0.115)
Math SAT score	0.004**			0.010***			0.022***		
	(0.002)			(0.003)			(0.003)		
Verbal SAT score	0.004***			0.006**			0.002		
	(0.002)			(0.003)			(0.003)		
Admissions rating		0.578***			0.794***			1.011***	
		(0.075)			(0.108)			(0.143)	
QR test score			0.016***			0.031***			0.062***
			(0.004)			(0.007)			(0.008)
Passed QR			-0.025			-0.12			-0.089
			(0.057)			(0.078)			(0.102)
Observations	764	793	793	711	737	737	761	789	789
R-squared	0.17	0.23	0.17	0.15	0.17	0.14	0.13	0.13	0.15

Note: Robust standard errors in parentheses. Significance levels: *10 percent, ** 5 percent, *** 1 percent. All specifications include year and race fixed effects, as well as financial aid controls.