NBER WORKING PAPER SERIES

BEAUTY IN THE CLASSROOM:
PROFESSORS' PULCHRITUDE AND
PUTATIVE PEDAGOGICAL PRODUCTIVITY

Daniel S. Hamermesh
Amy M. Parker

Beauty in the Classroom: Professors' Pulchritude and Putative Pedagogical Productivity
Daniel S. Hamermesh and Amy M. Parker
NBER Working Paper No. 9853
July 2003
JEL No. J7, I2

ABSTRACT

Adjusted for many other determinants, beauty affects earnings; but does it lead directly to the differences in productivity that we believe generate earnings differences? We take a large sample of student instructional ratings for a group of university professors, acquire six independent measures of their beauty and a number of other descriptors of them and their classes. Instructors who are viewed as better looking receive higher instructional ratings, with the impact of a move from the 10th to the 90th percentile of beauty being substantial. This impact exists within university departments and even within particular courses, and is larger for male than for female instructors. Disentangling whether this outcome represents productivity or discrimination is, as with the issue generally, probably impossible.

Daniel S. Hamermesh
Department of Economics
University of Texas
Austin, TX 78712-1173
and NBER
hamermes@eco.utexas.edu

Amy M. Parker
Department of Economics
University of Texas
Austin, TX 78712-1173
amparker@mail.utexas.edu

It was God who made me so beautiful. If I weren't, then I'd be a teacher.
                    [Supermodel Linda Evangelista]

I. Introduction

An immense literature in social psychology (summarized by Hatfield and Sprecher, 1986) has examined the impact of human beauty on a variety of noneconomic outcomes. Recently economists have considered how beauty affects labor market outcomes, particularly earnings, and have attempted to infer the sources of its effects from the behavior of different economic agents (Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998). The impacts on these monetary outcomes are implicitly the end results of the effects of beauty on productivity; but there seems to be no direct evidence of the impacts of beauty on productivity in a context in which we can be fairly sure that productivity generates economic rewards.

A substantial amount of research has indicated that academic administrators pay attention to teaching quality in setting salaries (Becker and Watts, 1999). A number of studies (e.g., Katz, 1973; Siegfried and White, 1973; Kaun, 1984; Moore *et al*, 1998) have demonstrated that teaching quality generates *ceteris paribus* increases in salary (but see DeLorme *et al*, 1979). The question is what generates the measured productivity for which the economic rewards are being offered. One possibility is simply that ascriptive characteristics, such as beauty, trigger positive responses by students and lead them to evaluate some teachers more favorably, so that their beauty earns them higher economic returns.

In this study we examine the productivity effects of beauty in the context of undergraduate education.[1] In particular, we consider the impact of professors' looks on their instructional ratings in the courses that they teach. In Section II we describe a data set that we

---

[1]Linking professors' looks to their pedagogical productivity does not appear to have been done previously, but Goebel and Cashen (1979) and Buck and Tiene (1989) did ask students in various grades to comment on the teaching ability that they would expect from individuals of varying levels of beauty based on a set of photographs. Ambady and Rosenthal (1993), the only study to look at actual teaching evaluations (of 13 TAs in a single course) focused on their nonverbal behavior but did touch on the effects of their attractiveness.

have created to analyze the impact of beauty on this indicator of professors' productivity. In Section III we discuss and interpret the results of studying these impacts. Section IV presents the implications of the analysis for interpreting the impact of an ascriptive characteristic on economic outcomes as stemming from productivity effects or discrimination.

II. Measuring Teaching Productivity and Its Determinants

The University of Texas at Austin, like most other institutions of higher learning in the United States, requires its faculty to be evaluated by their students in every class. A student administers the evaluation instrument while the professor is absent from the classroom. The rating forms include: "Overall, this instructor was very unsatisfactory (1); unsatisfactory (2); satisfactory (3); very good (4); excellent (5);" and "Overall, this course was very unsatisfactory, unsatisfactory…." In the analysis we concentrate on responses to the second question, both because it seems more germane to inferring the instructor's educational productivity, and because, in any event, the results for the two questions are nearly identical.

We chose professors at all levels of the academic hierarchy, obtaining professorial staffs from a number of departments that had posted all faculty members' pictures on their departmental websites. An additional ten faculty members' pictures were obtained from miscellaneous departments around the University. The average evaluation score for each undergraduate course that the faculty member taught during the academic years 2000-2002 is included. This sample selection criterion resulted in 463 courses, with the number of courses taught by the sample members ranging from 1 to 13. The classes ranged in size from 8 to 581 students, while the number of students completing the instructional ratings ranged from 5 to 380. Underlying the 463 sample observations are 16,957 completed evaluations from 25,547 registered students.

We also obtained information on each faculty member's sex, whether on the tenure track or not, minority status and whether he/she was not educated in an English-speaking country.[2] Table 1 presents the statistics describing these variables and the information about the classes. These descriptive statistics are generally unsurprising: 1) The average course rating is below that for the professor him/herself; 2) The average rating is around 4.0 (on the 5 to 1 scale), with a standard deviation of about 0.5; 3) Non-tenure track faculty are disproportionately assigned to lower-division courses.

One might be surprised that the course and professor ratings are actually slightly (but insignificantly) lower in the upper-division courses, which contain mostly majors who should be favorably disposed to the instructor and the material. The cause of this apparent anomaly is that higher-quality teachers are matched to the lower-division courses that typically contain more students.[3] Indeed, in a regression relating course ratings to class size and level, including fixed effects for each instructor, class size has a substantial negative impact on instructional ratings.[4]

Each of the professors' pictures was rated by each of six undergraduate students: Three women and three men, with one of each gender being a lower-division, two upper-division students (to accord with the distribution of classes across the two levels). The raters were told to use a 10 (highest) to 1 rating scale, to concentrate on the physiognomy of the professor in the picture, to make their ratings independent of age, and to keep 5 in mind as an average. In the

---

[2]This last variable is designed to account for the possibility of lower productivity of foreign teachers (see Borjas, 2000, but also Fleisher *et al*, 2002) that might also be correlated with perceptions of their looks. In fact, in our sample this correlation is only -0.02.

[3]This near invariance of ratings to class size may result from the maximizing behavior by administrators, who assign faculty to classes so as to equalize their marginal products, as implied by Lazear (2001).

[4]Included in the regression were variables measuring the course level, whether the course was for only one credit, and a quadratic in class size. The coefficients on the quadratic in class size were -.00493 (s.e. = .00147) and .00000713 (s.e. = .0000053). (The pair of terms in class size was highly significantly different from zero.) Implicit in these estimates is a decline in the instructor's evaluation until class size reaches 345 students, which in our sample includes all but 5 of the 463 classes.

analyses we unit normalized each rating. To reduce measurement error the six normalized ratings were summed to create a composite standardized beauty rating for each instructor.

Table 2 presents statistics describing the ratings of the professors' beauty. The students clearly had some difficulty holding to the instruction that they strive for an average rating of 5, as the averages of three of the six raw ratings were significantly below that, and none was significantly above (perhaps reflecting the students' inability to judge these older people, perhaps reflecting the choices implied in the epigraph). Moreover, the standardized ratings show that five of the six sets of ratings were skewed to the right. There was some concern, based on observations in earlier research, that the distribution of ratings of female faculty might have higher variance than that of males. While the variance was slightly higher, the Kolmogorov-Smirnov statistic testing equality of the two distributions had a p-value of 0.077.

Despite these minor difficulties, a central concern—that the assessments of beauty be consistent across raters—was achieved remarkably well. The fifteen pairwise correlation coefficients of the standardized beauty ratings range from 0.54 to 0.72, with an average correlation coefficient of 0.62. These indicate substantial agreement among the raters about the looks of the 94 faculty members.

III. The Impact of Beauty on Teaching Ratings

A. Basic Results

The basic model specifies a faculty member's teaching ratings as determined by a vector of his/her characteristics, X, and by a vector of the course's characteristics, Z. Included in X are whether the professor is female, whether he/she is a minority, whether not a native English speaker, and whether on the tenure track. The central variable in X is our composite measure of standardized beauty. Z includes whether the course is upper- or lower-division, and whether it is for one credit.[5] (Twenty-seven of the classes were one-credit labs, physical education or other

---

[5]Age and a quadratic in age were included in other versions of the basic equation. These terms were never significantly nonzero as a pair or individually, and they had essentially no impact on the coefficients of the

low-intensity activities that students tend to view differently from regular classes.) Where sample sizes permit we examine the determinants of course evaluations in lower- and upper-division courses separately, since the students in the former may be more focused on the instructor him/herself and less on the degree to which the instructor can exposit the course material.

Table 3 presents weighted least squares estimates of the equations describing the average course evaluations. As weights we use the number of students completing the evaluation forms in each class, because the error variances in the average teaching ratings are larger the fewer students completing the instructional evaluations. We present robust standard errors that account for the clustering of the observations (because we observe multiple classes for the overwhelming majority of instructors) for each of the parameter estimates.

The striking fact from the estimates in the first column is the statistical significance of the composite standardized beauty measure. The effects of differences in beauty on the average course rating are not small: Moving from one standard deviation below the mean to one standard deviation above leads to an increase in the average class rating of 0.46, close to a one-standard deviation increase in the average class rating.[6] A complete picture of the importance of beauty in affecting instructors' ratings is presented in Figure 1. For instructors at each percentile of the distribution of beauty, the Figure shows the rating that the instructor would obtain if he/she had

---

other terms in X and Z. Similarly unimportant was an indicator of whether the faculty member was tenured. If the one-credit classes are excluded the coefficient on standardized beauty rises to 0. 283.

[6]This impact is at the intensive margin—among students who showed up in class on the day the course evaluations were completed. If we examine the extensive margin—the impact on the fraction of students attending class on that day—we also find a positive and nearly statistically significant effect of composite standardized beauty.

other characteristics in X and Z at the sample means. The instructional rating varies by nearly two standard deviations between the worst- and best-looking instructors in the sample.

That inferring the impact of professors' looks on measures of their instructional productivity requires evaluations of their looks by several raters is demonstrated by sequential reestimates of the basic equation that include each of the six raters' evaluations individually. While the class ratings are significantly related to each rater's views of the instructors, the estimated impacts range only from 0.12 to 0.23, i.e., below the estimates based on the composite standardized measure. There is substantial measurement error in the individual beauty ratings.

Minority faculty members receive lower teaching evaluations than do majority professors, and non-native English speakers receive substantially lower ratings than do natives. Lower-division courses are rated slightly lower than upper-division courses. Non-tenure-track instructors receive course ratings that are surprisingly almost significantly higher than those of tenure-track faculty. This may arise because they are chiefly people who specialize in teaching rather than combining teaching and research, or perhaps from the incentives (in terms of reappointment and salary) that they face to please their students.

Perhaps the most interesting result among the other variables in the vectors X and Z is the significantly lower rating received by female instructors, an effect that implies reductions in average class ratings of nearly one-half standard deviation. This disparity departs from the consensus in the literature that there is no relationship between instructor's gender and instructional ratings (Alexander, 1993).

To explore this sex difference further we estimate the basic model separately for classes taught by male and female instructors. The results are shown in Columns 2 and 3 of Table 3. At the means of the variables the predicted instructional rating is lower for female instructors—the negative coefficient on the indicator in Column 1 is not an artifact of a correlation of perceived beauty and gender. The reestimates show, however, that the impact of beauty on professors' course ratings is much lower for female than for male faculty. Good looks generate more of a

premium, bad looks more of a penalty for male instructors, just as was demonstrated (Hamermesh and Biddle, 1994) for the effects of beauty in wage determination.

Columns 4 and 5 show the results of estimating the equation separately for lower- and upper-division classes. The impact of beauty on instructional ratings, while statistically significant in both equations, is over twice as large in lower-division classes. Indeed, the same much bigger effects are found for two of the other variables that affected instructional ratings in the sample as a whole, whether the instructor is on the tenure track or is female. We might be tempted to conclude that class ratings by more mature students, and students who are learning beyond the introductory level in a subject, are less affected by factors such as beauty that are probably unrelated to the instructor's knowledge of the subject. Yet the impacts of being a minority faculty member or a non-native English speaker are just as large in the estimates for upper-division courses as in those for lower-division courses. It is unclear why the impacts of these variables among those in X are not attenuated in the more advanced courses. These estimates may imply the existence of discrimination by students in their evaluations, or it may result from shortfalls in the ability of those instructors to transmit knowledge.

B. Robustness Tests

One might be concerned that a host of statistical problems plagues the estimates shown in Table 3 and means that our results are spurious. One difficulty is a potential measurement error: Raters may be unable to distinguish physical attractiveness from good grooming and dress. Were this merely classical measurement error, we would have no difficulties. A subtle problem arises, however, if those who dress better, and whose photographs may thus be rated higher, are the same people who take care to be organized in class, to come to class on time, to hold their announced office hours, etc. What if our measure of beauty is merely a proxy for the general quality of the faculty member independent of his/her looks?

To account for this possibility we created an indicator equaling one for male faculty members who are wearing neckties in their pictures and for female faculty who are wearing a

jacket and blouse. Formal pictures are on the websites of one-sixth of the faculty (weighted by numbers of students), and this indicator is added to a respecified version of the basic equation for which the results were shown in Column 1 of Table 3. The estimated impacts of this indicator and of composite standardized beauty are presented in the first row of Table 4. While instructors who present a formal picture do receive higher class ratings, the inclusion of this additional measure reduces the estimated impact of beauty only slightly. The effect of composite standardized beauty remains quite large and highly significant statistically. We may conclude that the potential positive correlation of measurement error in the beauty ratings with unobservable determinants of teaching success does not generate serious biases in our estimates.

Perhaps the most serious potential problem may result from a type of sample selectivity. Consider the following possibility: Among a group of people (a department) those who place their photographs on their websites will, until equilibrium in the game is reached, be better-looking than those who do not present their photographs. They may also be people who are "go-getters" in other aspects of their lives, including their classroom teaching. If that is true, those instructors who are among the few in a department whose picture is available will be better looking and be better instructors, while those from departments with all pictures available will on average be average-looking and average instructors.

To examine this potential problem we reestimate the basic equation on the subsample of 84 faculty members, teaching 414 classes, in which an entire department's faculty's pictures are available. The results of estimating the basic equation over this slightly reduced sample are shown in the second row of Table 4. Compared to the basic estimate (0.275), accounting for this potential problem reduces the estimated impact of composite standardized beauty slightly and implies that a two-standard deviation change in beauty raises the course rating by 0.39 (three-fourths of a standard deviation in course ratings). Apparently this kind of selectivity matters a bit, but it does not vitiate the basic result.

The next possibility does not represent a potential bias in the basic results, but rather supposes that they may be masking some additional sample information. There is some indication (Hamermesh and Biddle, 1994; Hamermesh *et al*, 2002) that the effect of beauty on earnings is asymmetric, with greater effects of bad than of good looks. Does this asymmetry carry over into its effects on productivity in college teaching? To examine this possibility we decompose the composite standardized beauty measure into positive and negative values and reestimate the basic equation allowing for asymmetry. The results are shown in the third row of Table 4. The effect on course ratings of looking better than average is slightly below and opposite in sign to the effect of looking worse than average.[7] There is only slight evidence of asymmetry in the impact of instructors' beauty on their course ratings.

Another potential issue is that courses may attract students with different attitudes toward beauty. These may be correlated with the instructional ratings that the students give and may also induce departmental administrators to assign courses to instructors based on their looks. Some courses may also generate different ratings depending on their difficulty, their level, and other differences, and these may be correlated with the instructor's looks. The gender mix of students may differ among courses, and this too may affect the estimated effects of beauty. To examine

---

[7]The t-statistic on the hypothesis that they are equal and opposite sign is 0.41. This may not contradict results indicating asymmetric effects of beauty on earnings. Many more individuals are rated above average in looks than are considered below average, so that the asymmetry might not exist if the beauty measure itself were symmetric, as it is by construction here.

these possibilities we take advantage of the fact that 157 of the 463 classes in our sample are instructed by more than one faculty member over the two years of observation. These courses involve 54 different instructors (of the 94 in the sample). We reestimate the basic equation on this subsample adding course fixed effects. Thus any estimated effect of beauty will reflect within-course differences in the impact of looks on instructional ratings.

The results are presented in the final row of Table 4. The estimated impact of composite standardized beauty on class evaluations is somewhat smaller than in the other estimates, but still substantial. This is mostly due to sampling variability: Reestimating the basic equation of Table 3 over this reduced sample of 157 classes yields an impact of composite standardized beauty on instructional ratings of 0.190 (s.e.=0.079).[8]

IV. Conclusion and Interpretations

The estimates leave little doubt that measures of perceived beauty have a substantial independent positive impact on instructional ratings by undergraduate students. We have accounted for a variety of possibly related correlates, and have shown that the estimated impacts are robust to potential problems of selectivity, correlated measurement error and other difficulties. The question is whether these findings really mean that beauty itself makes professors more productive in the classroom, or whether students are merely reacting to an irrelevant characteristic that differs among instructors.

The first issue is that our measure of beauty may merely be a proxy for a variety of related unmeasured characteristics that might positively affect instructional ratings. To the extent that these are positively correlated with beauty but not caused by it, our results overstate the impact of beauty. That we have held constant for as many course and instructor characteristics

---

[8]If we include a vector of indicators for departments in the basic equation in Table 3, we find a somewhat larger effect than here, although one that is still smaller than that in the basic equation.

as we have should mitigate some concerns about this potential problem. 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 professors 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 their teaching success.

A second, and more important issue is whether higher instructional ratings mean that the faculty member is a better teacher—is more productive in stimulating students' learning. The instructional ratings may putatively reflect productivity, but do they really do so? Discussions of this question among administrators and faculty members have proceeded since instructional evaluation was introduced, and we do not wish to add to the noise. Regardless of the evidence and of beliefs about this issue, however, instructional ratings are part of what universities use in their evaluations of faculty performance—in setting salaries, in determining promotion, and in awarding special recognition, such as teaching awards. Thus even if instructional ratings have little or nothing to do with actual teaching productivity, university administrators behave as if they believe that they do, and they link economic rewards to them. Thus the ratings are at least one of the proximately affected outcomes of beauty that in turn feed into labor-market outcomes.

The most important issue is what our results tell us about whether students are discriminating against ugly professors or whether students really do learn less (assuming that instructional ratings reflect learning). For example, what if students simply pay more attention to good-looking professors and learn more? We would argue that this is a productivity effect—we would claim that the instructors are better teachers. Others might (we think incorrectly) claim that the higher productivity arises from students' (society's) treating them differently from their worse looking colleagues and is evidence of discrimination. Disentangling the effects of differential outcomes resulting from productivity differences and those resulting from

11

discrimination is extremely difficult in all cases, as we believe this unusual illustration of the impact of beauty on a physical measure that is related to earnings illustrates.

The epigraph to this study may be correct—someone who does not qualify to be a supermodel might well go into teaching. Even in college teaching, however, our evidence demonstrates that a measure that is viewed as reflecting teaching productivity, whether it really does so or not, is also one that is enhanced by the instructor's pulchritude.

REFERENCES

Ambady, Nalini and Robert Rosenthal. 1993. "Half a Minute: Predicting Teacher Evaluations from Thin Slices of Nonverbal Behavior and Physical Attractiveness," *Journal of Personality and Social Psychology*, 64 (March): 431-41.

Becker, William, and Michael Watts. 1999. "How Departments of Economics Evaluate Teaching," American Economic Association, *Papers and Proceedings*, 90 (May): 355-59.

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

Borjas, George. 2000. "Foreign-born Teaching Assistants and the Academic Performance of Undergraduates," American Economic Association, *Papers and Proceedings*, 89 (May): 344-49.

Buck, Stephen, and Drew Tiene, "The Impact of Physical Attractiveness, Gender, and Teaching Philosophy on Teacher Evaluations," *Journal of Educational Research*, 82 (January/February, 1989): 172-7.

DeLorme, Charles; R. Carter Hill, and Norman Wood, 1979. "Analysis of a Quantitative Method of Determining Faculty Salaries," *Journal of Economic Education*, 11 (Fall): 20-5.

Feldman, Kenneth A. 1993. "College Students' Views of Male and Female College Teachers: Part II. Evidence from Students' Evaluations of their Classroom Teachers," *Research in Higher Education,* 34 (April): 151-211.

Fleisher, Belton; Masanori Hashimoto and Bruce Weinberg, "Foreign GTAs Can Be Effective Teachers of Economics," *Journal of Economic Education*, 33 (Fall 2002): 299-326.

Goebel, Barbara, and Valjean Cashen, 1979. "Age, Sex and Attractiveness as Factors in Student Ratings of Teachers: A Developmental Study," *Journal of Educational Psychology*, 71 (October): 646-53.

Hamermesh, Daniel, and Jeff Biddle, 1994. "Beauty and the Labor Market," *American Economic Review*, 84 (December): 1174-94.

------------------------; Xin Meng, and Junsen Zhang. 2002. "Dress for Success: Does Primping Pay?" *Labour Economics*, 9 (October): 361-73.

Hatfield, Elaine, and Susan Sprecher, 1986. *Mirror, Mirror....* Albany: State University of New York Press.

Katz, David, 1973. "Faculty Salaries, Promotions, and Productivity at a Large University," American Economic Association, *Papers and Proceedings*, 63 (May): 469-77.

Kaun, David, 1984. "Faculty Advancement in a Nontraditional University Environment," *Industrial and Labor Relations Review*, 37 (July): 592-606.

Lazear, Edward, 2001. "Educational Production," *Quarterly Journal of Economics*, 116 (August): 777-803.

Moore, William J.; Robert Newman, and Geoffrey Turnbull**,** 1998. "Do Academic Salaries Decline with Seniority?" *Journal of Labor Economics*, 16 (April): 352-66.

Siegfried, John, and Kenneth White, 1973. "Financial Rewards to Research and Teaching: A Case Study of Academic Economists," American Economic Association, *Papers and Proceedings*, 63 (May): 309-15.

Table 1. Descriptive Statistics, Courses, Instructors and Evaluations[a]

Variable	All	Lower Division	Upper Division
Course evaluation	4.022 (0.525)	4.060 (0.563)	3.993 (0.493)
Instructor evaluation	4.217 (0.540)	4.243 (0.609)	4.196 (0.481)
Number of students	55.18 (75.07)	76.50 (109.29)	44.24 (45.54)
Percent evaluating	74.43	73.52	74.89
Female	0.359	0.300	0.405
Minority	0.099	0.110	0.090
Non-native English	0.037	0.007	0.060
Tenure track	0.851	0.828	0.869
Lower division	0.339	--------	--------
Number of courses	463	157	306
Number of faculty	94	42	79

[a]Means with standard deviations in parentheses. All statistics except for those describing the number of students, the percent evaluating the instructor and the lower-upper division distinction are weighted by the number of students completing the course evaluation forms.

Table 2. Beauty Evaluations, Individual and Composite

	Average	Std. Dev.	Standardized: Minimum	Maximum
Individual Ratings:				
Male, Upper Division	4.43	2.18	-1.57	2.10
Male, Upper Division	4.87	1.65	-2.34	2.50
Female, Upper Division	5.18	2.05	-2.03	1.84
Female, Upper Division	5.39	2.10	-2.10	2.20
Male, Lower Division	3.53	1.70	-1.49	2.04
Female, Lower Division	4.14	1.88	-1.67	2.05
Composite Standardized Rating:				
	0	0.83	-1.54	1.88

Table 3. Weighted Least-Squares Estimates of the Determinants of Class Ratings[a]

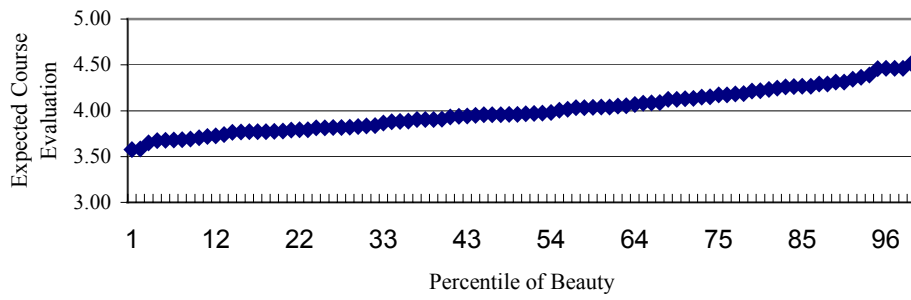| Variable	All	Males	Females	Lower Division	Upper Division
Composite stdzd. beauty	0.275 (0.059)	0.384 (0.076)	0.128 (0.064)	0.359 (0.092)	0.166 (0.061)
Female	-0.239 (0.085)	--------	--------	-0.345 (0.133)	-0.093 (0.104)
Minority	-0.249 (0.112)	0.060 (0.101)	-0.260 (0.139)	-0.288 (0.156)	-0.231 (0.107)
Non-native English	-0.253 (0.134)	-0.427 (0.143)	-0.262 (0.151)	-0.374 (0.141)	-0.286 (0.131)
Tenure track	-0.136 (0.094)	-0.056 (0.089)	-0.041 (0.133)	-0.187 (0.141)	0.005 (0.119)
Lower division	-0.046 (0.101)	0.005 (0.111)	-0.228 (0.129)	--------	---------
R^2	.279	.359	.162	.510	.126
N courses	463	268	195	157	306
N faculty	94	54	40	42	79

[a]Robust standard errors in parentheses here and in Table 4. All the estimating equations also include an indicator equaling one if the course is a one-credit offering.

Table 4. Alternative Estimates of the Relation Between Beauty and Class Ratings (lower- and upper-division classes, N=463 unless otherwise noted)

			Variable	
	Composite stdzd. beauty	Formal picture	Composite stdzd. beauty: Above mean	Below mean
1. Photo bias (individual)	0.229 (0.047)	0.243 (0.088)		
2. Photo bias (department) (N = 414)	0.236 (0.049)			
3. Asymmetric beauty effect			0.237 (0.096)	- 0.318 (0.133)
4. Course fixed effects (N = 157)	0.177 (0.107)			

[a]The equations reported in Rows 1-3 also include all the variables included in the basic equation in Column 1 of Table 3. The equation reported in Row 4 excludes variables in the vector Z.

Figure 1. Beauty and Course Evaluations

Expected Course Evaluation (y-axis): 3.00, 3.50, 4.00, 4.50, 5.00

Percentile of Beauty (x-axis): 1, 12, 22, 33, 43, 54, 64, 75, 85, 96

Legend: ◆ Expected Course Evaluation