Student Appearance and Academic Performance

Rey Hernández-Julián

Christina Peters*

Metropolitan State University of Denver

December 2016

Abstract

Studies have shown that attractive people have higher earnings. In this paper, we test the hypothesis that physical attractiveness proxies for unobserved productivity. We compare the impact of attractiveness on grades in college courses where instructors directly observe the student's appearance to courses where they do not. We find that in traditional classrooms, appearance matters: both below- and above-average appearance female students earn lower grades. In regressions including student fixed effects, we find that students of above-average appearance earn significantly lower grades in online course environments compared to traditional courses, a finding driven mainly by courses taught by male instructors. Our empirical evidence provides little support for the hypothesis that appearance is a proxy for productive traits, but instead suggests that the return to appearance is due to discrimination.

Keywords: appearance, discrimination, student performance

JEL Codes: I21, J71

---

*Corresponding author is Hernández-Julián: rherna42@msudenver.edu, CB 77 POBox 173362, Denver, CO 80217-3362. Peters: cpeter80@msudenver.edu, CB 77 POBox 173362, Denver, CO 80217-3362. We would like to thank Geoffrey Lehr, Steven Matthias, all the raters that participated in the process, Hani Mansour, Daniel I. Rees, Angela K. Dills, and anonymous referees. We are also grateful to session participants at the 2016 American Economic Association annual meetings for helpful comments. This project has been approved by the MSU Denver IRB.

1. Introduction

Studies have shown that there are significant rewards to being more attractive in both labor and dating markets (Hamermesh and Biddle 1994, Biddle and Hamermesh 1998, Hamermesh 2011). Among men, the homeliest earn nine percent less than the average looking, while the best looking earn five percent more. In women, the least attractive earn four percent less than the average, while the most beautiful earn five percent more (Hamermesh and Biddle 1994). Despite these substantial returns to appearance, no study has used real-world outcomes to test whether better-looking individuals are actually more productive.

If appearance is correlated with productive traits, then it should have a return even when the individual cannot be seen. Unfortunately, it is difficult to find a setting in the labor market that enables the researcher to compare the productivity of a worker when he is both observed and not observed. In this paper, we instead exploit a unique source of variation in student academic outcomes. Specifically, we compare how well appearance predicts academic performance in courses where the student is seen (i.e. traditional lecture courses) to ones where the student is not (i.e. online courses).

We use data from student records and ID-card photographs at a large, public, open-admission institution located in Denver, Colorado, United States. In regressions without student fixed effects, we find that female student grades in traditional environments are lower for those with below-average appearance and, surprisingly, also for those females with above-average appearance. There appears to be no such difference for male students according to appearance type. In online courses, the return to appearance is not significant for any students, though the coefficients on appearance from online and traditional courses are not significantly different from each other.

Because the results described above do not include student fixed effects, they are likely biased. Student-level unobservable characteristics, such as income, drive, or academic interest, may be correlated with appearance, selection into course type, and grades. Thus, a better identification strategy uses only information from students who are observed in both traditional and online courses, subsequently estimating the return to appearance based on the difference between in-class and online performance of those students. In these regressions, we find that both male and female students with above-average appearance perform significantly worse in online courses than they do in traditional ones. This penalty for attractive students in online courses relative to their performance in traditional classrooms appears to be driven by courses taught by male professors. The main empirical result of our paper provides evidence against the hypothesis that appearance serves as a proxy for productivity, suggesting that the return to beauty is better explained by discrimination.

2. Background

To date, a large literature has established a significant return to appearance across several areas, most notably in labor markets (Hamermesh and Biddle 1994, Biddle and Hamermesh 1998, Hamermesh 2011). The literature also indicates that appearance is related to student academic outcomes, particularly grades. French et al. (2009) use data from the National Longitudinal Study of Adolescent to Adult Health to show that better looking, better-groomed high-school students are more likely to get higher grades.

Numerous studies have also established the strong presence of discrimination based on other individual physical characteristics such as race, gender, and class (Goldin and Rouse 2000; Bernard and Mullanaithan 2004; Becker 2010; Hanna and Linden 2012; Kuhn and Shen 2013). In terms of discrimination by physical appearance, the primary channel proposed by previous studies operates through Becker-type discrimination on the part of either employers or customers. In this setting, customers are more likely to purchase goods from more attractive workers, which increases the employer's willingness to pay for their labor. Attractive individuals tend to sort into fields where a return to their appearance is more likely to be present, but the estimated returns persist even after controlling for such sorting (Hamermesh and Biddle 1994, Hamermesh 2011).

Even if appearance itself is not preferred for discriminatory reasons, the employer may prefer the job applicant with the higher appearance rating because he associates that appearance with productivity. Indeed, there is some evidence that appearance may be correlated with higher productivity. Kanazawa and Kovar (2004) and Kanazawa (2011) argue that assortative mating leads the fittest, most intelligent males to mate with the most beautiful women, thereby producing offspring that are both physically attractive and intelligent. Jackson et al. (1995) document that attractiveness is related not only to perceived competence but also to actual competence in children, though not in adults. Consistent with this hypothesis, Mobius and Rosenblat (2006) use a simulated labor market experiment to document that the beauty premium is transmitted through higher confidence, stronger oral skills, and through being *incorrectly* perceived as more productive. Without more detailed information on productivity, employers may use appearance as a signal of otherwise unobservable productive traits. It could then be the case that in a non-experimental environment decision makers are not necessarily mistaken when they associate appearance with productivity.

If there is a productivity component associated with appearance, those with higher appearance ratings should attain better outcomes even in situations when they are not being observed. Not many such scenarios exist in typical labor markets; one example might be a sales worker who handles transaction both by phone/online and in person. An ideal data set would allow the researcher to examine sales outcomes using those two methods for more attractive people compared to less attractive ones. If a worker with a higher appearance rating has higher measured outcomes not just in the context where he or she is seen, but also in the context where she remains unseen, the overall return to appearance exists in both settings. Unfortunately, a data set that collects this information is not readily available.[2]

Rather than examining labor market data, this paper estimates returns to appearance in academic outcomes. In particular, we examine whether appearance has any predictive power on student outcomes in two contexts. In the first, student appearance and their academic work are both easily observable by the evaluating professor. In the second context, however, only their academic work is observed. If our results find a similar return to appearance even when the students are not being seen, we interpret this as evidence that appearance is correlated with otherwise unobserved productivity. If,

[2] Mobius and Rosenblat (2006) overcome this issue by creating an experiment and using data collected from a simulated labor market.

instead, appearance is less predictive of success when the student is not observed, the evidence supports the hypothesis that the return to appearance is due to discrimination.

Cipriani and Zago (2011) and Deryugina and Shurchkov (2015) are the closest studies that we have found which use student performance to examine the impact of appearance on college grades. Using a sample of Italian students, Cipriani and Zago (2011) compare their performance on oral exams, where they are seen by the evaluators, to their performance on written exams where they are not seen. In this environment, where the exams are optional and not required, the better-looking students are more likely to take exams, more likely to opt for oral rather than written exams, and earn higher scores on both types of exams. Their findings are driven by results on male students. Deryugina and Shurchkov (2015) examine data from a women's college in the United States, and they find that controlling for pre-college test scores, better-looking students sort into more beauty-based areas of study but earn only slightly higher grades.

We expand on these studies by using data from an institution in the United States with a larger and more representative student body. Our empirical results contrast with the findings of Cipriani and Zago (2011) and Deryugina and Shurchkov (2015), perhaps due to the different institutional environment of our study. However, we also benefit from having many more observations than Cipriani and Zago (2011) and Deryugina and Shurchkov (2015). In addition, Cipriani and Zago's (2011) analysis was complicated by the fact that in their university, each student can choose both when to take an exam and whether to take it at all. Their estimation thus required a two-step strategy, while we are able to instead take advantage of an estimation strategy common to other studies of college-student performance (Dills and Hernández-Julián 2008, Carrell et al. 2011, and Lindo et al. 2014).

3. Data and Methodology

3.1 Data

Our sample consists of students from the Metropolitan State University of Denver, an open-enrollment public institution located in the Western United States. The institution serves approximately 22,000 students, many of whom are older than typical first-time freshmen. For this study, we have access to student academic records, including course enrollment and subsequent grades earned between Spring 2006 and Fall 2011. We exclude from the sample any students who are under 18 years old, observations of grades other than A, B, C, D, or F, and students for whom there is missing information.[3]

For a subset of the students, we obtained the photographs taken for their student identification cards. Students need these cards to get a bus pass, enter the fitness center, and access library materials. Our set of student images does not match exactly with our set of student grade information; some photographs belong to students who had no earned grades during the sample period, and in a few cases, students never got an ID card. We thus limit the sample of images to those for whom we have student records, and then we subsequently sort the images randomly before constructing ratings of appearance. After receiving approval from both the institution and the Institutional Review Board, we recruited individuals who were over 18 years of age and neither students nor faculty at

[3] Specifically, we drop grade observations of NC for No Credit, I for Incomplete, P for Pass, or AW for Administrative Withdrawal.

the institution in order to rate the images. These raters consisted of both males and females of many ages and races, and each worked independently and anonymously. We recruited a total of 28 raters, each rating an average of about 400 images. The IRB approval did not allow us to track information on the raters, though we attempted to maintain an even balance between male and female raters and also to recruit raters from diverse age, educational, racial, and ethnic backgrounds. The raters were shown the photograph from a student ID and asked to rate that image on a scale of 1-10. The raters were also always given the option to not rate an image if it made them uncomfortable or if the image was of someone they knew. Thus, a few of the pictures were not rated because either the individual had a visible disability, the rater determined the picture was unclear, the rater knew the subject, or another unknown reason.

The first 50 images shown to each rater were identical. This strategy allowed us to establish a baseline and to prime all raters from the same set of images. However, we exclude these 50 individuals from our sample.[4] For all the other ratings, we create a gender-specific normalized appearance rating for individual *i* rated by rater *r* by subtracting the rater- and gender-specific mean value from each rating, and then dividing it by the rater- and gender-specific standard deviation.[5]

Normalizing enables the ratings of different raters to be more easily comparable, as some rate higher than others, and some rate more widely. In order to check for inter-rater reliability, we calculate Cronbach's alpha (Cronbach 1951) using the first 50 images

[4] Since these 50 images are rated many times, it is unclear what the value of their rating should be: it could be the mean of all the ratings, or the median or mode value. If we take the mean value of all the ratings, there is no obvious way to normalize the ratings to make them comparable to the others. Many students do not include ACT or other academic information because they are admitted under the university's age-based open admission policy. The estimation sample is significantly younger and has higher grades than the full student body.

[5] Because all ratings are demeaned by rater, including rater fixed effects in our specification generates small, insignificant estimates that have no impact on the other estimates.

(the ones seen by all raters) and find a coefficient of 0.945. This coefficient meets the standards for clinical application as defined by Bland and Altman (1997), which gives us confidence that our sample has a high level of inter-rater reliability.

We follow the existing literature and separate students into three groups: those with below-average appearance, those with average appearance, and those with above-average appearance. We define an individual as below average if his normalized appearance rating is less than -1, average if the value falls between -1 and 1, and above average if the value is greater than 1. After merging these normalized ratings with our sample of grade observations, we have 167,554 observations. Because some student traits are not available for some students (student ACT exam scores are included as a control variable in our regressions, and this statistic is only reported for 37 percent of students in the full sample), our estimation sample becomes further limited to 90,090 grade observations earned by 4,543 students. Table 1 presents descriptive statistics for this preferred estimation sample.

Normalized appearance ratings for students in online courses are, on average, not significantly different from those in traditional ones. The mean online appearance rating is 0.127. Below-average appearance ratings are given to 16.2 percent of the sample, while above-average appearance ratings are given to 18.5 percent of the sample. Below-average appearance ratings are significantly more likely to be given to students in online courses compared to traditional ones.

The mean student age is 26.5 years old (the median age is 26), 42 percent of the grade observations are earned by male students, and approximately 65 percent of the grades are earned by students who identify as white. The mean ACT score is 20.6, which places them in the 49th percentile of test takers nationally. Institutional data reports a freshman retention rate of 65 percent, a transfer-student retention rate of 70 percent, and a six-year graduation rate of about 25 percent (MSU Denver IR Data Book n.d.). These numbers align closely to the traits of non-selective two- or four-year institutions, the type of institution that the typical college student in the US attends (Freedman 2013, Department of Education 2013).

Figure 1 shows the normalized appearance distributions for students in online and traditional courses. Both distributions have a long right tail and some bumpiness where the in-rater normalization has resulted in several students earning the same post-normalization rating at particular values. A Kolmogorov-Smirnov test for equality of the distribution functions fails to reject the null hypothesis that the distributions between online and traditional courses are the same. Consistent with Table 1, however, the distribution does show that students in online courses are more likely to be below-average appearance.

3.2 Methodology

We first estimate a regression of the following form:

$$(1) \quad grade_{ijkt}=\alpha_1 \, \textit{above-average appearance}_i + \alpha_2 \, \textit{below-average appearance}_i \; \beta X_{it}+\gamma Y_{jkt}+s_k+\tau_t+u_{ijkt}$$

where $grade_{ijkt}$ is the grade of student i taking course-section j of subject-level k in term t. Grades are measured as 4, 3, 2, 1, 0, representing A, B, C, D, and F.[6] *Above-average* and *below-average appearance* are time-invariant measures of the student's appearance (with

[6] During the time period of the sample, instructors could not assign letter grades with plus-minus designations.

average appearance being the excluded category), and X_{it} is a vector of student traits (sex, age, age-squared, and race), some of which vary over time. Y_{jkt} measures sex of the professor. s_k are subject-level specific fixed effects, one for all 100-level Economics courses, one for all 100-level English courses, another for all 200-level Economics courses, etc. τ_t are semester-year fixed effects to capture any grade inflation.[7] To test our hypothesis that the return to appearance varies between those environments where the student is seen and where the student is not seen, we further include interactions between the appearance measures, whether the course meets online or traditionally, and the sex of the student.[8] We also estimate regressions where we split the regression sample by the gender of the professor.

Our estimation strategy is similar to that used in other research predicting college grades (Dills and Hernández-Julián 2008, Carrell et al. 2011, and Lindo et al. 2014). Although previous work examining student grades as a function of appearance has relied on student GPA (French et al. 2009), we believe that the use of individual course grades is a better strategy, because it allows for course-specific controls such as the subject and level of the course.

One concern with our specification is that students with higher appearance ratings may have some other advantages that help them earn higher grades. For instance, richer students may be able to afford more purchases that improve their appearance, and they may also be able to afford to hire tutors to help their grades. Such advantages would

[7] Standard errors are clustered by student in all specifications.

[8] We also split the sample by generating a variable, which we call 'objective course', that limits the sample to courses in mathematics and hard sciences. We expect appearance to matter less in these courses, since grading is likely to less subjective. Here we find estimates of appearance that are much smaller in magnitude and less significant compared to our full specifications. However, even in this sample of only 'objective' courses, we find that female students of above-average appearance still earn lower grades in online courses when the courses are taught by male instructors. Unfortunately, though, this limited sample leaves us with too few observations to include student fixed effects.

create an upward bias on the estimated correlation between appearance and grades. Although this initial equation cannot include student fixed effects to absorb this kind of variation (because our ratings of student appearance do not vary over time), a comparison of traditional vs. online courses enables us to include student fixed effects θ_i that control for all observed and unobserved time-invariant traits of the student. Thus, equation (2) includes these student fixed effects as well as a binary indicator for whether the course is online and a dummy variable for the interaction between the student's appearance rating and whether the course is online. That equation becomes:

$$(2) \qquad grade_{ijkt} = \eta\ online_{jkt} + \varphi_1\ online_{jkt}*above\text{-}average\ appearance_i$$
$$+ \varphi_2\ online_{jkt}*above\text{-}average\ appearance_i + \theta_i + \gamma Y_{jkt} + s_k + \tau_t + u_{ijkt.}$$

In this specification, it is impossible to include any unvarying student traits (such as gender or appearance dummies) or interactions of these traits (such as the interaction of gender with the appearance dummy) because they do not vary across observations for the same student, making them perfectly collinear with the fixed effects. It is still possible to include traits and interactions that vary across courses, such as the interaction between online courses and appearance, and the interaction between online courses, appearance, and student gender. In equation (2), the primary coefficients of interest are φ_1 and φ_2, which will indicate whether any estimated return to appearance is different between online and in-class courses, even after controlling for observed and unobserved student traits through the student fixed effects. This variable is interacted with student gender and also examined separately in equations where we split the sample by the gender of the professor.

4. Results

4.1 Appearance and Course Choices

One concern with a comparison of students in online vs. traditional courses is that students may systematically self-select into course types. If more attractive students see a return to their appearance in traditional courses, but not in online ones, perhaps they respond by enrolling in those courses at higher rates. The regressions in Table 2 examine this question across all teaching environments. If better-looking students expect to earn higher grades in traditional classrooms, where they would be seen, then they will be more likely to select into these courses. Further, if the gender of the professor of the size of the course matter differently to students based on their appearance, we would expect them to systematically sort into these courses as well.

Table 2 presents estimation results using estimation equation (2), where course traits instead of grades are the dependent variable. We estimate whether a student prefers male professors, online courses, larger courses, or objective courses as a function of his or her appearance. We include both a measure of appearance as well as appearance interacted with student sex in case appearance predicts selection differently for male or female students. Male students appear more likely to enroll in courses with male professors and less likely to take online courses. Student appearance, though, is unrelated to professor gender, enrollment in online courses, and enrollment in objective courses for both male and female students. Female students with above-average appearance ratings are significantly 1.6 percentage points more likely to enroll in a class with over 30 students, but this effect does not hold for male students.

Overall, then, the regressions in Table 2 show no clear pattern of appearance having a large impact on student course choices, even though Table 1 suggests that below-average appearance individuals are unconditionally more likely to take online courses. The lack of evidence for sorting in these regressions implies that favoritism towards better-looking students does not inform student course selections.

4.2 Appearance and Grades

Table 3 presents the results of estimating student grades as a function of student appearance and other student and course-specific traits. Column (1) shows a regression of student grades on the appearance measures, whether the course is taken online, and ACT score. We find that students with above-average appearance earn lower grades, as do male students and students in online courses. The finding that male students earn lower grades is not unusual and is consistent with previous research (Lindo et al. 2012, Dills and Hernández-Julián 2008). In addition, online courses have been previously found to have lower grades, partly due to a greater prevalence of Fs (approximately twenty percent of online grades in our sample are F, compared with eleven percent in courses not online). Grades of C, D, and F are all more common in online courses than in classes that meet in person. Online, they comprise almost 52 percent of the grades, while in traditional courses they make up fewer than 32 percent of grades. This difference in grade distribution is partly due to weaker performance (Wachenheim 2009) and partly due to lower completion rates (Carr 2000, Moody 2004).

Column (2) includes interaction terms for male with below-average appearance and male with above-average appearance. The results here show a statistically significant penalty to below-average appearance and above-average appearance for female students, but not for male students. Specifically, below-average appearance females earn grades that are 0.078 points lower on a 4-point scale, which is a quarter of the distance between an A- and a B+. Female students of above average appearance also earn grades that are lower by a similar magnitude of 0.0757 grade points. For male below-average appearance students, the sum of the coefficients on *below-average appearance* and *male X below-average appearance* is not significant; thus, we fail to reject the null hypothesis that below-average appearance males earn similar grades as average appearance males. We also test whether the sum of the coefficients on above-average appearance males (as measured by *above-average appearance* plus *male X above-average appearance*) is statistically significant, obtaining a p-value of 0.25. Consequently, it appears that males of both below- and above-average appearance earn grades similar to those of average males.

Our overall initial finding is that the return to appearance in college grades depends on the sex of the individual: we estimate a significant penalty to females who depart from average appearance but insignificant estimates for males. These results set us up to address the primary question of this paper: do better-looking students still see a return to appearance when they cannot be seen? Columns (3) and (4) of Table 3 split the sample by whether the course is taken in a traditional classroom or online. In online courses, professors and students communicate mostly via email and exchange course content through the computer. Although some software packages include images of the students automatically, the ones used in our sample do not. Some professors might request that the students upload images or video, but this is not typical, particularly in the earlier years of our sample.[9] Online courses are also often first-year or introductory courses, making it even less likely that the professors are familiar with the students' names and faces. Since professors in online courses typically do not have access to the students' appearance, we can test whether appearance pays off in environments where the students are not being seen.

We find that all coefficient estimates on appearance are insignificant for online courses: in settings where students are not observed, their appearance has no predictive power on grades. In contrast, column (4), the sample of traditional courses, shows strong predictive power of appearance in that environment, with magnitudes similar to those in column (2).

Column (5) explicitly tests whether the coefficients in columns (3) and (4) are significantly different from each other by including interaction terms for online courses in the full sample. We find that coefficients are, once again, similar to those in column (2), as the large majority of observations come from traditional environments. None of the interactions with online are statistically significant at conventional levels.

The next two columns split the sample by the gender of the professor. We find that female professors assert stronger penalties for below-average appearance female students in traditional course environments, with no grade difference between average and above-average female students. Female student appearance has no predictive power in online courses taught by female professors. Male student appearance does not significantly predict grades in either environment when the professor is female.

[9] Even if professors in online courses do use photographs and become aware of their students' appearance, this merely exposes students in the online courses to the 'appearance' treatment, biasing our results towards zero.

When the professor is male, we find that above-average appearance female students earn lower grades, and that this penalty is larger in online courses. Above-average appearance male students, like females, receive lower grades in online courses than they do in traditional classrooms.

Throughout, the estimates presented in Table 3 suggest some better-looking students perform worse in online courses relative to traditional ones. However, these results may be biased because they do not account for student-level unobservables that may be correlated with both appearance and the choice of class. These traits, such as income, drive, or academic interest, may be correlated with student appearance, selection into course type, and grades. For example, a more dedicated individual may be more willing to travel to campus for courses, invest in personal grooming, or do the work necessary to earn higher grades. As a consequence, we prefer the estimation strategy described in equation (2) above, which uses information from students who are observed in both traditional and online courses and estimates the return to appearance based on the difference between in-class and online performance of those students.

Table 4 presents results from this specification. Because the student fixed effects are collinear with both the student-unvarying traits and some interactions, the list of coefficients in these regressions is shorter. These collinear variables include *male*, *below-average appearance*, *above-average appearance*, *male X below-average appearance*, and *male X above-average appearance*. Table 4 also estimates these results both jointly and separately by professor gender. The sample sizes reported in Table 5 are slightly larger than in Table 4, because the student fixed effects capture traits that had previously been included as explicit control variables (thus limiting our sample size).

Column (1) displays the estimated return for in-class courses with all students and professors. Once again, students earn lower grades overall in online courses, but this difference is 0.233 grade points larger for female students with high appearance ratings (about two-thirds of the difference between an A- and a B+). This penalty does not exist for below-average appearance females nor is it significant for male students. Splitting the sample by professor gender shows that the result in the full sample is primarily driven by the male professors. Among female professors, none of the appearance coefficients, interactions, or the sums of the interactions generate significant estimates. Among male professors, though, above-average appearance students of both genders are penalized relative to their performance in traditional classrooms. Table 5 provides the most convincing test of our hypothesis, showing that the return to appearance does in fact diminish in those environments where an individual cannot be seen. Further, that penalty on above-average appearance appears to be concentrated among male professors.

4.3 Large vs. Small Classes

Splitting the sample by class size provides a way to revisit whether there is still a return to appearance in classes where the students are less likely to be observed by their professor. Larger classes will have lower levels of interaction with the professor, so professors may be less likely to notice or remember a given student's appearance. Our regressions separate small and large classes at 30 students. The average class size for our sample is 31, so choosing 30 as the cutoff allows us to separate the sample so that about 60 percent of the grade observations are from small classes and 40 percent from large ones. Here we find much weaker results than the regressions that compare online with traditional courses. The one significant coefficient is a penalty for above-average appearance females, relative to males, in larger classrooms when the professor is male. However, since the largest classes in our sample rarely reach beyond 50 students, our analysis of large vs. small courses is comparing courses of 20-30 students to courses of 40-50 students. We believe this finding is consistent with the previous results; large class can be seen as an intermediate point between a high-visibility small class and a no-visibility online class. As a result, the coefficient estimates are much smaller and less significant than those in online courses, with the remaining significant result for above-average females with male professors. Where it is more difficult to gain a return to appearance, the estimated return to appearance diminishes.

5. Implications and Conclusion

Consistent with the previous literature, our student fixed effects estimations show that appearance matters: more attractive students earn higher grades when they are seen relative to when they are not seen. The fact that there is a difference between the return to appearance in online and traditional courses supports discrimination as a mechanism behind the rewards to appearance.

It is important to recognize that our estimated return to appearance could still be productive. For instance, Mobius and Rosenblat (2006) find that more attractive workers possess stronger oral communication and social skills which lead to better interactions with their employers. A similar skill difference may be driving the results in our setting as well: it is possible that instructors are more likely to incorporate in-class discussions or oral presentations as part of a student's grade in traditional courses compared to online courses. If attractive students have better oral and social skills, then they would be expected to score higher on such assignments and thus obtain higher grades in in-class courses. Mobius and Rosenblat (2006) further observe that more attractive workers appear to have more confidence, and this increased confidence directly increases wages. Perhaps in our setting, attractive students have increased confidence only when being directly observed, and this confidence thus works to increase their grades only in a traditional course setting where instructors see them.

In addition, it is also possible that throughout the course of a semester, professors pay less attention and offer less support to less attractive students. As a result, these students learn less, accumulate less human capital, and perform worse in the evaluation of the course. The more attractive students do earn higher grades, but these higher grades are actually a result of higher learning. However, the reason they are learning more is because of their appearance. In this case, appearance *does* produce more learning.

We remain unable to separate the paths through which discrimination may penalize those who are less attractive: either through harder grading, or through less learning. Further research should therefore focus on disentangling these mechanisms. If professors pay more attention to attractive students or preparing assignments that work in their favor, there are clear implications to the return to appearance in labor markets. The higher earnings of more attractive individuals may not entirely be due to discrimination on the part of employers, but also at least partly due to the higher productivity gained as a result of preferential treatment by professors.

6. References

Becker, Gary. 2010. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Bernard, M. and Mullanaithan, S. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *The American Economic Review* 94(4): 991-1013.

Biddle, J. E. and Hamermesh, D. S. 1998. "Beauty, productivity, and discrimination: lawyers' looks and lucre." *Journal of Labour Economics*, 16: 172–201.

Bland, J.M., and Altman, D. G. 1997. "Statistics notes: Cronbach's alpha." *BMJ* 314: 572.

Carr, S. 2000. "As distance education comes of age, the challenge is keeping the students." *The Chronicle of Higher Education* 46(23): A39-A41.

Carrell, S. E., Maghakian, T., and West, J. E. 2011. "A's from Zzzz's? The Causal Effect of School Start Time on the Academic Achievement of Adolescents." *American Economic Journal: Economic Policy* 3(3): 62-81.

Cronbach, L.J. 1951. "Coefficient alpha and the internal structure of tests." *Psychometrika* 16(3): 297–333.

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

Deryugina, T. and Shurchkov O. 2015. "Does Beauty Matter in Undergraduate Education?" *Economic Inquiry* 53(2): 940-961.

Dills, A. K. and Hernández-Julián, R. 2008. "Course Scheduling and Academic Performance." *Economics of Education Review* 27(6): 646-654.

Freedman, J. 2013, September 20. "The Typical College Student Is Not A 'Typical' College Student (And Other Fun College Demographics Data)." *Forbes*. Retrieved from http://www.forbes.com/sites/joshfreedman/2013/09/20/the-typical-college-student-is-not-a-typical-college-student-and-other-fun-college-demographics-data/.

French, M. T., Robins, P. K., Homer, J. F., and Tapsell, L. M. 2009. "Effects of physical attractiveness, personality, and grooming on academic performance in high school." *Labour Economics*, 16(4): 373-382.

Goldin, C. and C. Rouse. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." *American Economic Review* 90(4): 715-741

Hamermesh, D. S. 2011. *Beauty Pays*. Princeton, NJ: Princeton University Press.

Hamermesh, D. S. and Biddle, J. E. 1994. "Beauty and the labour market." *American Economic Review*, 84(5): 1174–1194.

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

Jackson, L. A., Hunter, J. E., and Hodge, C. N. 1995. "Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review." *Social Psychology Quarterly* 58(2): 108-122.

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

Kanazawa, S. and Kovar, J. L. 2004. "Why Beautiful people are more intelligent." *Intelligence* 32(3): 227-243.

Kuhn, P. and Shen, K. 2013. "Gender Discrimination in Job Ads: Evidence from China." *Quarterly Journal of Economics* 128(1): 287-336.

Lindo, J. M., Swensen, I. D., and Waddell, G. R. 2012. "Are Big-time Sports a Threat to Student Achievement?" *American Economic Journal: Applied Economics* 4(4): 254-274.

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

Moody, J. 2004. "Distance education: Why are the attrition rates so high?" *The Quarterly Review of Distance Education* 5(3): 205-210.

MSU Denver Office of Institutional Research. n.d. *Institutional Research Data Book.* Retrieved from http://www.msudenver.edu/oir/institutionalresearchdatabook/.

United States Department of Education. 2013. "Total fall enrollment in degree-granting postsecondary institutions, by control and level of institution: 1970 through 2012." Retrieved from http://nces.ed.gov/programs/digest/d13/tables/dt13_303.25.asp

Wachenheim, Cheryl J. 2009. "Final Exam Scores in Introductory Economics Courses: Effect of Course Delivery Method and Proctoring." *Review of Agricultural Economics* 31(3): 640-652.

Figure 1. Appearance Distribution by Course Type

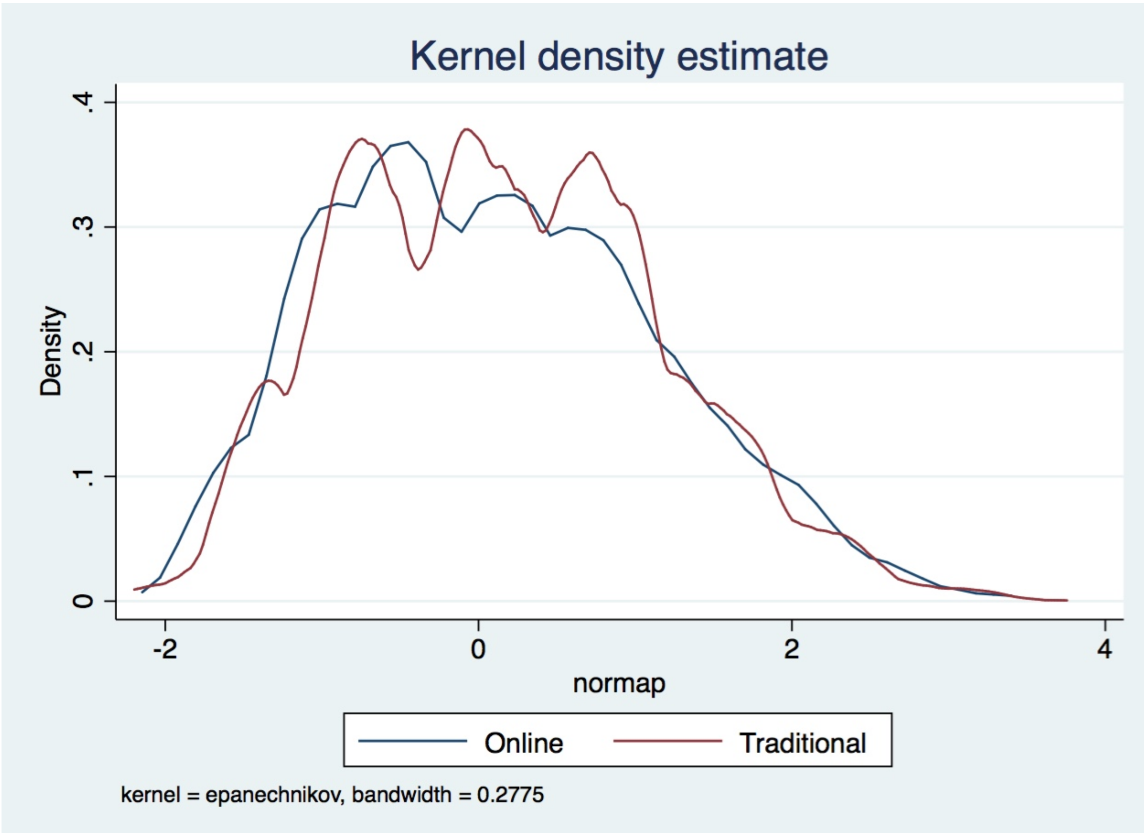

Table 1. Summary Statistics

	all observations			online			traditional		
	N	*mean*	*std. dev.*	*N*	*mean*	*std. dev.*	*N*	*mean*	*std. dev.*
grade	90,090	2.777	1.225	778	1.960	0.049	89,312	2.785***	1.222
normalized appearance	90,090	0.127	0.996	778	0.070	1.06	89,312	0.127	0.995
below average appearance	90,090	0.162	0.368	778	0.194	0.395	89,312	0.161**	0.368
above average appearance	90,090	0.185	0.206	778	0.206	0.404	89,312	0.185	0.388
male	90,090	0.427	0.495	778	0.463	0.499	89,312	0.427**	0.495
ACT score	90,090	20.594	3.653	778	20.852	3.713	89,312	20.591**	3.652
white	90,090	0.650	0.002	778	0.649	0.476	89,312	0.650	0.478
age	90,090	26.502	2.402	778	27.781	3.411	89,312	26.491***	2.388
male professor	90,090	0.518	0.499	778	0.632	0.482	89,312	0.517***	0.499

*** p<0.01, ** p<0.05, * p<0.1 for the tests of significant differences between online and traditional course means.

Table 2: Course Choices and Student Appearance

	(1) Male Professor	(2) Online	(3) Class Size Over 30	(4) Objective
Below Average	-0.0078	0.0001	0.0070	-0.0253
Appearance	(0.007)	(0.002)	(0.006)	(0.029)
Above Average	-0.0071	0.0014	0.0166***	0.0113
Appearance	(0.006)	(0.002)	(0.005)	(0.025)
Male	0.0277***	-0.0029**	0.0129***	0.0027
	(0.005)	(0.001)	(0.004)	(0.020)
Male × *Below Avg*	0.0144	0.0007	-0.0106	0.0292
Appearance	(0.010)	(0.003)	(0.009)	(0.044)
Male ×*Above Avg*	0.0055	-0.0018	-0.0173**	0.0381
Appearance	(0.009)	(0.003)	(0.008)	(0.041)
Online	-0.0637***		-0.164***	0.606***
	(0.018)		(0.017)	(0.016)
ACT	0.001**	0.0005***	-0.0014***	0.0037*
	(0.001)	(0.000)	(0.000)	(0.002)
Observations	97,165	101,591	101,591	42,797
R-squared	0.194	0.126	0.478	0.045

All regressions include controls for the student's age, age squared, whether the student is white, semester fixed effects, and subject-level fixed effects. Standard errors clustered at the student level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Appearance and Grades

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				grade			
			online	traditional	all observations	female professor	male professor
Below Average	-0.0195	-0.0780*	0.175	-0.0795*	-0.0780*	-0.105**	-0.0474
Appearance	(0.032)	(0.046)	(0.261)	(0.045)	(0.046)	(0.050)	(0.049)
Above Average	-0.0684**	-0.0757**	-0.199	-0.0742**	-0.0729**	-0.0583	-0.0876**
Appearance	(0.030)	(0.036)	(0.212)	(0.036)	(0.036)	(0.039)	(0.040)
Male	-0.132***	-0.154***	0.221	-0.158***	-0.155***	-0.194***	-0.123***
	(0.024)	(0.029)	(0.187)	(0.029)	(0.029)	(0.033)	(0.032)
Male ✕ Below Avg		0.113*	-0.202	0.114*	0.111*	0.148**	0.0734
Appearance		(0.065)	(0.372)	(0.064)	(0.064)	(0.072)	(0.069)
Male ✕Above Avg		0.0175	-0.429	0.0214	0.0179	0.0147	0.024
Appearance		(0.063)	(0.368)	(0.062)	(0.063)	(0.072)	(0.066)
Online	-0.491***	-0.492***			-0.451***	-0.596***	-0.355***
	(0.069)	(0.069)			(0.090)	(0.127)	(0.103)
Below Average Appearance					-0.00602	0.3790	-0.2560
✕ Online					(0.248)	(0.384)	(0.261)
Above Average Appearance					-0.306	0.0216	-0.468*
✕ Online					(0.207)	-0.31	(0.258)
Below Average Appearance					0.22	0.101	0.336
✕ Online ✕ Male					(0.325)	-0.46	(0.365)
Above Average Appearance					-0.0319	-0.231	0.0417
✕ Online ✕ Male					(0.322)	(0.419)	(0.426)
ACT	0.0368***	0.0369***	0.0492**	0.0370***	0.0370***	0.0365***	0.0382***
	(0.003)	(0.003)	(0.021)	(0.003)	(0.003)	(0.004)	(0.004)
Observations	90,090	90,090	778	89,312	90,090	43,445.00	46,645.00
R-squared	0.119	0.119	0.148	0.116	0.119	0.123	0.131

All regressions include controls for the student's age, the professor's gender, age squared, whether the student is white, semester fixed effects, and subject-level fixed effects. Standard errors clustered at the student level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Student Fixed Effects Estimates by Course Type

	(1)	(2)	(3)
		grade	
Professor Gender	All	Female	Male
Online	-0.610***	-0.649***	-0.584***
	(0.042)	(0.068)	(0.104)
Below Average Appearance	-0.0563	0.115	-0.117
× Online	(0.123)	(0.162)	(0.147)
Above Average Appearance	-0.233**	-0.171	-0.265*
× Online	(0.109)	(0.167)	(0.150)
Below Average Appearance	0.0963	-0.0256	0.147
× Online × Male	(0.146)	(0.257)	(0.148)
Above Average Appearance	0.0747	0.215	-0.0196
× Online × Male	(0.157)	(0.279)	(0.140)
Observations	167,554	80,394	87,160
R-squared	0.378	0.428	0.402

All regressions include controls for the student's age, the professor's gender, age squared, whether the student is white, semester fixed effects, and subject-level fixed effects. Standard errors clustered at the student level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Student Fixed Effects Estimates by Class Size

	(1)	(2)	(3)
		grade	
Professor Gender	All	Female	Male
Student Gender	All	All	All
Class Size over 30	-0.0631***	-0.0548***	-0.0656***
	(0.009)	(0.014)	(0.013)
Below Average Appearance ✕ Class Size over 30	0.0173	0.0438	-0.00774
	(0.023)	(0.033)	(0.031)
Above Average Appearance ✕ Class Size over 30	-0.0153	0.0097	-0.0604*
	(0.024)	(0.033)	(0.035)
Below Average Appearance ✕ Class Size Over 30 ✕ Male	-0.0232	-0.0523	0.00606
	(0.031)	(0.045)	(0.042)
Above Average Appearance ✕ Class Size Over 30 ✕ Male	0.00871	-0.00655	0.0445
	(0.034)	(0.050)	(0.045)
Observations	167,554	80,394	87,160
R-squared	0.431	0.477	0.458

All regressions include controls for the student's age, the professor's gender, age squared, whether the student is white, semester fixed effects, and subject-level fixed effects. Standard errors clustered at the student level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1