STUDENT APPEARANCE AND ACADEMIC PERFORMANCE

Rey Hernández-Julián

Christina Peters*

Metropolitan State University of Denver

December 2015

Abstract

Studies have shown that attractive people have higher earnings. In this paper, we test the hypothesis that physical attractiveness may be a proxy for unobserved productivity. We compare the impact of attractiveness on grades in college courses where instructors can directly observe the student's appearance to courses where they do not. We confirm that appearance matters: attractive female students earn higher grades than unattractive ones. Moreover, we provide evidence that this return to appearance is significantly smaller for both male and female students in online course environments. Thus, 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. We would like to thank Steven Matthias, all the raters that participated in the process, Hani Mansour, Daniel I. Rees, and Angela K. Dills. This project has been approved by the MSU Denver IRB.

1. Introduction

Studies have shown that there are significant rewards in labor and dating markets to being more attractive (Hamermesh and Biddle 1994, Biddle and Hamermesh 1998, Hamermesh 2011, Hamermesh 2012). 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, none of the studies have identified a clear mechanism for this return, in part because it is difficult to separate a return to appearance that is a result of discrimination from one that is due to higher productivity.

If appearance is correlated with unobserved 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 seen and not seen. In this paper, we instead exploit a unique source of variation in student academic outcomes. Specifically, we compare how well appearance can predict 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 the Metropolitan State University of Denver, a large, public, open-admission institution in Denver. Our first set of results indicates that female students with below-average ratings of appearance have significantly worse grade outcomes. Male students, on the other hand, see little return to their appearance. We then include student fixed effects and estimate whether the return to appearance varies between in-class and online environments. Here we find that an improvement in appearance has a smaller impact in online classes than in traditional courses. We interpret this result as evidence that the return to appearance is more likely a result of discrimination than a reflection of otherwise unobserved productivity.

Our paper lies at the intersection of the literatures on physical appearance, discrimination, and determinants of student performance. We provide an important contribution to these literatures by empirically disentangling a mechanism for the return to beauty. Overall, our empirical results provide evidence against the hypothesis that appearance serves as a proxy for productivity and suggest 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). Of particular interest to academics is the finding that better-looking professors get higher ratings of instruction from students (Hamermesh and Parker 2005). Instruction ratings increase linearly by about 2 standard deviations as one moves from the least attractive to the most attractive. A separate study by Bokek-Cohen and Davidowitz (2008) shows that returns to appearance are only present among male professors when evaluated by female students. In both scenarios, however, the authors are unable to disentangle the source of the return: do students favor otherwise identical professors and give them higher ratings because of their beauty? Could it be that looks are directly productive because they encourage students to attend class and pay attention? Or alternatively, in the context of the hypothesis that this paper tests, do more attractive professors receive higher ratings both because they pay more attention to their looks and they *also* pay more attention to their students and their work? In other words, is attractiveness merely a proxy for other unobservable traits of the professor?

The literature also indicates that appearance is related to students' academic outcomes, particularly grades. French et al. (2009) use data from the National Longitudinal Study of Adolescent Adult Health to show that better looking, better-groomed high-school students are more likely to get higher grades. French et al.

(2014) also document that these better-looking individuals are more likely to marry. In fact, the wage returns to beauty are large when compared to the returns to ability (Fletcher 2009). As with the other studies, however, they are unable to identify a clear mechanism for the return to appearance.

There is a basis in the prior literature for supporting the potential mechanism of pure discrimination by appearance. Numerous studies have 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; Fryer and Levitt 2004; Becker 2010; Hanna and Linden 2012; Kuhn and Shen 2013). Although other papers on physical appearance have been unable to clearly identify discrimination as the definitive mechanism, the literature does suggest a few potential channels for this return. The primary channel proposed by previous studies operates through Becker-type discrimination on the part of either employers or customers. In this setting, people want to be around better-looking individuals, so employers will be willing to pay them more. In addition, customers are also more likely to purchase goods from more attractive workers, which further increases the employer's willingness to pay for their labor. Thus, in this last case, worker beauty is in fact productive to the employer, even though it is not *socially* productive. (Hamermesh discusses this idea at length in *Beauty Pays* (2011)). Although attractive individuals also sort into those fields where a return to their appearance is more likely to be present, the estimated returns persist even after controlling for such sorting (Hamermesh and Biddle 1994, Hamermesh 2011).

An alternative explanation for the return to appearance proposes that beauty is correlated with unobservable traits that make an individual more productive. For instance, individuals who are especially detail-oriented may be more likely to have a flattering haircut or better grooming and sense of style. Attention to detail itself may also be directly productive in many forms of employment. An employer, then, may prefer the more attractive applicant not because of the attractiveness itself, but because it conveys information about other productive traits, such as attention to detail and dedication to the task at hand. Gehrsitz (2014) showed that more attractive individuals not only have higher hourly wages, but they also work more hours. If that is the case, then the estimated return to beauty potentially captures these productivity traits and is biased upward by that omitted variable. [1]

More generally, if individuals are given an endowment of appearance but have the ability to enhance this endowment through some trait that is correlated to productivity, employers will see two parts to appearance. The first component is an individual's endowment, while the second component is that same individual's effort-induced improvement. This effort-induced improvement could come through better grooming, exercise, investing time in more flattering clothes or makeup, or even through plastic surgery. Overall, although appearance will be a noisy measure, and employers may be unable to differentiate between its two potential sources on the margin, they may still prefer a candidate with a higher appearance rating for a reason that has nothing to do with discrimination. Even if appearance itself is not preferred for the Becker-style reasons above, the employer may prefer the job applicant with the higher appearance rating because he associates that appearance with productivity.

If there is a productivity component associated with appearance, then 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 in those two methods for more attractive people compared to less attractive ones. The unobserved productivity hypothesis predicts that those

[1] An additional mechanism could be that earnings and appearance both predict having a better family background or higher social status. In our context, social status is likely similar among all the students as they chose to attend the same institution of higher education. An additional path may be that those with higher earnings make more money, so they use that money to improve their appearance. We ignore that path in this context, as it is unlikely to be the case that having higher grades allows students to somehow purchase appearance improvements. Students also take their pictures upon enrollment, so the picture is taken *before* any of the grades are earned.

with higher appearance ratings would have higher measured outcomes not just in the context where he or she is seen, but also in the context where he remains unseen. In that case, the overall return to appearance will not be significantly different between the two settings. Unfortunately, a data set that collects all this information is not readily publicly available.

Our study instead exploits student data to empirically distinguish between the hypotheses that the return to appearance stems from discrimination versus unobserved productivity. Rather than examining labor market data, we estimate 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 finding as evidence that appearance is in fact correlated with otherwise unobserved productivity. If, instead, appearance is less predictive of success when the student is not observed, the evidence supports the understanding that the bulk of the return to appearance is due to discrimination.

Cipriani and Zago (2011) is the only other work that we have found which uses student performance to answer this question. Using a sample of Italian students, the study compares their performance on oral exams, where they are seen by the evaluators, to their performance on written exams where they are not seen. The students have an option of when and whether to take these exams at all, and the better-looking students are more likely to take exams. Better-looking students are also more likely to opt for oral rather than written exams and more likely to pass, but notably, they earn higher scores on both types of exams. Thus, their results suggest that appearance may reflect student productivity. These findings are driven by results on male students, which further suggests that the value of appearance as a proxy of otherwise unobservable productive traits may vary by sex.

We expand on their work by using a larger sample of student data from the United States. Our empirical results contrast with their findings, perhaps due to the different institutional environment of our study. However, we also benefit from having many more observations—thousands, compared to approximately 300 used in Cipriani and Zago (2011). In addition, their study 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, Maghakian, and West 2011, and Lindo, Swensen, and Waddell 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 Denver, Colorado. The institution has approximately 22,000 students, many of whom are older than typical first-time freshmen. For this study, we have access to students' academic records, including their enrollment in courses and subsequent grades earned between Spring 2006 and Fall 2011. We exclude from the sample any students who are under 18 years old, and we also drop observations of grades other than A, B, C, D, or F.[2] The resulting dataset consists of 1,139,772 observations earned by 77,067 students (see Table 1). The mean student age is 30 years old (the median age is 28), 46 percent of the students are male, and approximately 60 percent of the students identify as white. Because the university is required by law to admit any students over age 20 who have a high-school diploma or GED certificate, not all students have their high-school GPA or standardized test scores attached to their student records. Of the 37 percent of students who do have ACT scores, the mean score is 20.5, which places them in the 49th percentile of test takers. Institutional data reports a freshman

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

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, which is the type of institution that the typical college student in the US attends (Freedman 2013, Department of Education 2013).

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, access the fitness center and recreation center, and use the library. 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 MSU-Denver in order to rate the images. These raters consisted of both males and females of many ages and races, and they each worked anonymously. They were shown the image and then asked to rate that image along 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 the rater did not rate the image for another reason.

3.2 Methodology

The first 50 images shown to each rater were the same. 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.[3] For all the other ratings, we

[3] 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.

create a normalized appearance rating for individual i rated by rater r by subtracting the rater-specific mean value from each rating, and then dividing it by the rater-specific standard deviation.

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 Cochran's alpha (Cochran 1951) using the first 50 images (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.

After merging these ratings with the student records, our final sample consists of 6,777 individuals and 168,092 grade observations. Table 1 presents descriptive statistics for this sample. For traits that vary at the student level, each observation is a student; for those that vary with each observation, we include each grade earned.

We first use this sample to estimate a regression of the following form:

$$(1) \qquad grade_{ijkt} = \alpha\, appearance_i + \beta X_{it} + \gamma Y_{jkt} + s_k + \tau_t + u_{ijkt}$$

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. *appearance* is a time-invariant measure of the student's appearance, and X_{it} is a vector of student traits (sex, age, and race), some of which vary over time. Y_{jkt} includes characteristics of the class such as the 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 fixed effects to capture any grade inflation.[4] We also split the sample by

[4] Standard errors are clustered by student.

traits of the class, such as its size, whether it meets online, the sex of the professor, and the sex of the student. Splitting the sample by the sex of the professor and the student may be particularly informative, as it would reveal if the responsiveness to student appearance varies by professor or student sex.

Our estimation strategy is similar to that used in other research predicting college grades (Dills and Hernández-Julián 2008, Carrell, Maghakian, and West 2011, and Lindo, Swensen, and Waddell 2014). Although previous work examining student grades as a function of appearance has relied on student term 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 subject and level of the course.

Most of the literature on appearance to date uses a 1 to 5 scale of attractiveness (Hamermesh and Biddle 1994, French et al. 2009). Our study purposefully departs from this literature in favor of a 1 to 10 scale because we are interested in an individual's potential ability to improve on their appearance within the range that it may be malleable. Given a standard 1 to 5 rating, it is likely impossible for an individual endowed at birth with a 3 to improve him or herself to a 4, or for a 2 to become a 3 through effort. On a 10-point scale, though, it is more feasible for an individual to go from a natural 5 to a 6 based on a better haircut, better grooming, getting braces, washing their face, or other margins on which, through effort, the individual can become more attractive. In other words, the marginal effort that raises an individual's appearance even slightly can be more easily captured in a 1 to 10 rating than in a 1 to 5 rating. This type of increase is key to our analysis: our goal is to test whether the individuals who go through this kind of effort also have a higher level of otherwise unobservable productivity that makes them more willing to exert higher levels of effort in other environments as well.

We then vary the specification to include a binary indicator for whether the course is online as well as a dummy variable for the interaction between the student's appearance rating and whether the course is online. Although our previous equation could not include fixed effects (because our ratings of student appearance do not vary over time), a comparison of traditional vs. online courses also enables us to include student fixed effects that control for all observed and unobserved time-invariant traits of the student, because we can interact our appearance measure with an indicator for course type as follows:

$$(2) \qquad grade_{ijkt} = \eta\ online_{jkt} + \varphi\ online_{jkt} * appearance_i + \theta_i + \gamma Y_{jkt} + s_k + \tau_t + u_{ijkt}$$

Here the coefficient of interest is φ, 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 θ_i. Thus, if we find a return to appearance in the first specification and a negative sign here, then within a student, the return to appearance for a given student is higher when he is seen than when he cannot be seen. This result would be considered evidence of appearance-based discrimination. However, if the coefficient $\varphi$is instead zero or small, we will interpret that result as evidence that the return to appearance in the literature is due to traits of the student that are present even in settings where their appearance is not directly observed.

4. Results

4.1 Appearance and Grades

Table 2 presents the results of estimating student grades as a function of student appearance and other student and course-specific traits. When student appearance is measured as a within-rater normalized value, it does not appear to be a significant predictor of student performance. In general, male students earn lower grades, grades are lower in online courses, and male professors seem to be easier graders. The specific penalty on males in this regression is -0.117 grade points, which is almost half the difference between an A- and a B+ (see Column 1).

One reason that online courses have lower grades stems from a greater prevalence of F's. Approximately 20 percent of grades earned in online courses are F, compared with 11 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 make up almost 52 percent of the grades, while in in-class 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).

In results not reported, we also find that white students earn grades that are approximately 0.2 grade points higher than non-white students. Moreover, grades appear to increase with age, but at a decreasing rate. This result is consistent with the literature, which has found that individuals become more confident about their appearance as they age (McCarthy 2014), and one important mechanism through which appearance can improve outcomes is through increasing self-confidence (Judge, Hurst, and Simon 2009).

The second column of Table 2 incorporates student ACT scores as an additional control in the regression from the first column. Including this variable cuts the sample size in half because many students are admitted without requiring this score. Although this sub-sample is self selected, we find coefficient estimates on the indicators for male, male professor, and online course that are similar to those in column 1.

The next columns include an interaction with the student's sex, allowing the impact of appearance on grades to vary between males and females. Once we include these interactions, the results change substantially: for female students, the impact of appearance as measured by the within-rater normalized value becomes significant and positive: a one standard deviation increase in appearance increases a grade by 0.024 grade points on a 4-point scale. For males, however, the estimate (measured by the sum of the coefficients on *normalized appearance* and *normalized* *appearance × male)* is not significant at conventional levels. When ACT score is included, the female return becomes smaller in magnitude and is no longer significant.[5]

One reason that there may be little significance in the normalized measure is that the impact of appearance is not continuous but categorical. In other words, observers may not respond to small changes in appearance but instead use perceived appearance to sort others into types. Once observers sort individuals into a category, they treat others differently depending on their category. As a consequence, we follow the previous literature and separate students into 3 groups: those with below-average appearance, those with average appearance, and those with above-average appearance. We present the results for this estimation in Table 3. Using the appearance rating defined above based on within-rater normalized values, we then 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. Students with average appearance ratings are the excluded group. Here we find an impact of -0.067 grade points for having below-average appearance (Column 1). This estimate becomes smaller and insignificant when we add ACT score (Column 2).

The next columns divide students into the three categories as before but add an interaction between *male* and the appearance categories. Column (3) shows a significant penalty to below average appearance for female students. Such students earn grades that are 0.141 lower on a 4-point scale, almost half the distance between an A- and a B+, while female students of above average appearance earn grades that are 0.005 higher than average students (although this difference is not statistically significant). The coefficient estimates for the impact of appearance on

[5] A separate interesting question is whether ACT score is a predictor of appearance. In a regression available upon request, we find that, controlling for sex, age, and race, ACT score negatively predicts appearance (significant at the 10 percent level). The ACT score could reduce the significance of the return to appearance if it captures historical discrimination that the student has received in their education; for instance, if their high-school instructors discriminated in favor of the more attractive.

male grades are not significant at conventional levels. In the regression in Column (4), we include ACT scores as an additional control and obtain a result similar to those in previous regressions. The return to female appearance remains significant, while the return to male appearance remains insignificant at conventional levels.

Thus, our overall initial finding is that the return to appearance in college grades depends on the sex of the individual: we estimate a positive and significant return for females but a mostly insignificant return for males. This pattern of results is still observed for all students when we estimate the impact of appearance using categories. 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?

4.2 Online vs. Traditional Courses

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 by the institution in our sample do not. Some professors might request that the students upload images, or in some courses video, but this is not typical, particularly in the earlier part of our sample. 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. 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.

Table 4 presents results using the specification from equation (2) estimated both jointly and separately for professor and student sexes. This strategy allows male and female professors to respond independently to the appearance of male and female students. The sample sizes reported in Table 4 are slightly larger than in Tables 2 and 3, because the student fixed effects capture student 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. Students earn lower grades overall in online courses, but as a student's appearance rating increases, the predicted difference in grades between online and traditional courses increases as well. In particular, a one standard deviation increase in normalized appearance decreases predicted online grades by -0.0695 grade points. Compared to the estimates from Table 2, this difference is larger than the return to appearance in a traditional course. Thus, as attractiveness increases, so does the difference in performance between online courses and traditional courses. Columns 2 and 3 indicate that this effect is largely concentrated among female students.

If we examine the relative impact of the higher appearance ratings for female professors only, we find estimates that are larger for both the full sample and female students, yet again insignificant for male students. When the professor is male, we find that better appearance ratings lower grades in online courses for all students, but this estimate is not significant for the split sample. Throughout, the estimates presented in Table 4 suggest that the impact of appearance in online courses is large compared to the overall estimates of the return to appearance given by Table 2. Thus, it appears that not controlling for student fixed effects produces estimates that are biased downward.

Table 5 estimates the same regressions as Table 4 but with appearance now split into three categories (as in Table 3), with average-appearance students being the omitted group. Here we find that above-average appearance is negatively associated with grades in online courses for all students, with the largest effect appearing for male students. The magnitude of each of these estimates is larger than the returns to appearance estimated in Table 3, again suggesting that earlier estimates without student fixed effects were biased. In addition, when this sample is decomposed by professor sex, it appears that this effect is now concentrated among male professors, with female professors instead rewarding appearance identically in both in-class and online courses. The results on the effect of professor sex are therefore different when we use a categorical measure of appearance versus a continuous measure of appearance.

Our overall results suggest a significant penalty to appearance in online courses compared to in-class courses, and this penalty exceeds the estimated return to appearance for grades in Tables 2 and 3. A student who has a higher rating of appearance is predicted to perform more weakly in online courses, relative to his other courses, than a student with a lower appearance rating. Such a result is inconsistent with a hypothesis that physical appearance is simply a proxy for unobserved productivity, but instead suggests the presence of discrimination.

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 cases where the students are seen less 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. To the extent that the return from appearance is due to discrimination, it will be harder to get that return in larger courses where the professor sees individual students less.[6] On the other hand, the students with appearances most likely to sway a grade are also the ones most likely to be noticed by a professor. A smaller estimated return to appearance in larger courses might therefore be interpreted as evidence in favor of pure discrimination.

Our regressions separate small and large classes at 30 students. At MSU Denver the average class size is 31, and 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, regardless of whether appearance is measured as a continuous (Table 6) or categorical (Table 7) variable, we find much weaker results than the regressions that compare online with traditional courses. When appearance is measured categorically in Table 7, male students do see a

[6] Using a continuous measure of class size interacted with the student's sex and appearance rating generates similar results; we present these instead for ease of interpretation.

penalty to appearance in large courses, as do any male students who deviate from the average in courses with female professors. Beyond that result, however, we find no evidence that appearance has a differential correlation with grades in small courses compared to large ones. Thus, these estimates suggest that having a larger class does not limit the professor's ability to identify and respond to student appearance, as the return to appearance is typically identical in both environments.

4.4 Course Choices

If more attractive students see a return to their appearance in traditional courses, but not in online ones, do they respond by enrolling in those courses at higher rates? The regressions in Table 8 examine this question across all course types from the previous regressions. Better-looking female students appear less likely to enroll in courses with male professors, but this finding is sensitive to the inclusion of ACT score. Female students with above-average appearance are, controlling for ACT score, more likely to enroll in larger classes. However, the magnitude of these results is extremely small.

The regressions in Table 8 show no clear pattern of appearance having a significant impact on student course choices. More attractive and less attractive students are equally likely to have male professors, enroll in online courses, and take large classes. These results hold for both male and female students. The lack of evidence for sorting in these regressions implies that favoritism towards better-looking students is either not well known or too small in size to inform student decisions.

5. Implications and Conclusion

Consistent with the previous literature, we find that appearance matters: attractive female students receive higher grades in college courses compared to their unattractive peers. Furthermore, we provide evidence that in environments where students cannot be seen, more attractive students perform relatively worse than in traditional environments. The fact that there is a difference between the return to appearance in online and traditional courses supports discrimination as the mechanism behind rewards to appearance.

It is important to recognize that our estimated return to appearance could still be productive. Rather than assuming that the entire difference we document is due to discrimination against the less attractive[7], a second mechanism may be at work instead: throughout the course of a semester, professors may 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 two 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 two mechanisms. If professors are paying more attention to attractive students, helping them earn higher levels of human capital, 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 discrimination by professors.

[7] For instance, two students produce identical work, but the more attractive student receives a higher grade than the less attractive one.

6. References

Bokek-Cohen, Y. and Davidowitz, N. 2008. “Beauty in the Classroom: Are Female Students Influenced by the Physical Appearance of their Male Professors?” *Journal of Education and Human Development* 2(1): 14-16.

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.

Chronback, 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.

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/.

Fletcher, J. 2009. “Beauty vs. Brains: Early Labor Market Outcomes of High School Graduates.” *Economics Letters* 105(3): 321-325.

French, M. T., Popovici, I., Robins, P.K., and Homer, J.F. 2014. “Personal traits, cohabitation, and marriage.” *Social Science Research* 45: 184-199.

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.

Fryer R, and Levitt S. 2004. “The Causes and Consequences of Distinctively Black Names.” *Quarterly Journal of Economics* 119(3):767-805.

Gehrsitz, M. 2014. “Looks and Labor: Do Attractive People Work More?” *Labour* 28(3): 269-287.

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. 2012. “Tall or Taller, Pretty or Prettier? Is Discrimination Absolute or Relative?” *IZA Journal of Labor Economics* 1:2.

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

Hamermesh, D. S. and Parker, A. 2005. “Beauty in the Classroom: Instructors’ Pulchritude and Putative Pedagogical Productivity.” *Economics of Education Review* 24(4): 369-376.

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

Judge, T. A., Hurst, C. and Simon, L. S. 2009. “Does It Pay to Be Smart, Attractive, or Confident (or All Three)? Relationships Among General Mental Ability, Physical Attractiveness Core Self-Evaluations, and Income.” *Journal of Applied Psychology* 94(3): 742-755.

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.

McCarthy, J. 2014. “Older Americans Feel Best About Their Physical Appearance.” *Gallup Well Being*, Accessed online Oct. 14, 2014 at http://www.gallup.com/poll/172361/older-americans-feel-best-physical-appearance.aspx.

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.

Table 1. Summary Statistics

	all students			male students			female students		
	N	*mean*	*std. dev.*	*N*	*mean*	*std. dev.*	*N*	*mean*	*std. dev.*
grade	168,092	2.848	1.239	76,620	2.765	1.269	91,472	2.918	1.210
normalized appearance	6,777	0.00	1.00	3,133	-0.220	0.940	3,644	0.199	1.010
below average appearance	6,777	0.154	0.361	3,133	0.204	0.403	3,644	0.111	0.314
above average appearance	6,777	0.167	0.373	3,133	0.108	0.311	3,644	0.218	0.413
male	6,777	0.462	0.499	3,133	1.000	0.000	3,644	0.000	0.000
ACT score	3,113	20.689	3.688	1,377	20.95	3.833	1,736	20.4820	3.556
online	6,777	0.015	0.12	3,133	0.015	0.122	3,644	0.014	0.119
white	6,777	0.601	0.490	3,133	0.609	0.488	3,644	0.594	0.491
age	6,777	30.373	8.048	3,133	30.191	7.076	3,644	30.530	8.224
male professor	167,554	0.520	0.499	76,342	0.588	0.492	91,212	0.463	0.499

Table 2. Appearance and Grades

	(1)	(2)	(3)	(4)
		grade		
normalized appearance	0.0062	0.000675	0.0239**	0.0103
	(0.01)	(0.01)	(0.01)	(0.02)
male	-0.117***	-0.131***	-0.119***	-0.131***
	(0.02)	(0.03)	(0.02)	(0.03)
male x			-0.0414**	-0.0242
normalized appearance			(0.02)	(0.02)
ACT score		0.0369***		0.0369***
		(0.00)		(0.00)
online	-0.385***	-0.492***	-0.385***	-0.493***
	(0.05)	(0.07)	(0.05)	(0.07)
male professor	0.0349***	0.0319***	0.0351***	0.0319***
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	167,554	90,090	167,554	90,090
R-squared	0.099	0.118	0.099	0.118

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 Categories and Grades

	(1)	(2)	(3)	(4)
		grade		
below-average appearance	-0.0668***	-0.0393	-0.141***	-0.130**
	(0.02)	(0.04)	(0.04)	(0.06)
above-average appearance	-0.023	-0.0244	0.00522	-0.0127
	(0.02)	(0.03)	(0.03)	(0.03)
male	-0.117***	-0.131***	-0.119***	-0.139***
	(0.02)	(0.02)	(0.02)	(0.03)
male x			0.120**	0.141*
below-average appearance			(0.05)	(0.07)
male x			-0.0965**	-0.057
above-average appearance			(0.05)	(0.06)
ACT score		0.0367***		0.0368***
		(0.00)		(0.00)
online	-0.384***	-0.492***	-0.385***	-0.495***
	(0.05)	(0.07)	(0.05)	(0.07)
male professor	0.0349***	0.0318***	0.0350***	0.0318***
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	167554	90090	167554	90090
R-squared	0.099	0.119	0.1	0.119

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 4: Return to Normalized Appearance Online v. Traditional Classes

	(1)	(2)	(3)
		grade	
	all students	male students	female students
ALL PROFESSORS			
online	-0.412***	-0.371***	-0.451***
	(0.03)	(0.05)	(0.00)
normalized appearance	-0.0695**	-0.0776	-0.0485**
x online	(0.03)	(0.05)	(0.00)
Observations	168,092	76,620	91,472
R-squared	0.431	0.441	0.423
FEMALE PROFESSORS			
online	-0.467***	-0.466***	-0.467***
	(0.06)	(0.10)	(0.07)
normalized appearance	-0.0964*	-0.109	-0.0956*
x online	(0.05)	(0.09)	(0.06)
Observations	80,394	31,457	48,937
R-squared	0.477	0.502	0.457
MALE PROFESSORS			
online	-0.396***	-0.356***	-0.439***
	(0.04)	(0.06)	(0.06)
normalized appearance	-0.0760**	-0.0795	-0.0586
x online	(0.04)	(0.06)	(0.05)
Observations	87,160	44,885	42,275
R-squared	0.458	0.462	0.461

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 5: Return to Categorical Appearance Online v. Traditional Classes

	(1)	(2)	(3)
		grade	
	all students	male students	female students
ALL PROFESSORS			
online	-0.379***	-0.329***	-0.421***
	(0.04)	(0.06)	(0.00)
below-average appearance	-0.02	0.02	-0.11
x online	(0.09)	(0.11)	(0.00)
above-average appearance	-0.159*	-0.246**	-0.103**
x online	(0.08)	(0.12)	(0.00)
Observations	168,092	76,620	91,472
R-squared	0.431	0.441	0.423
FEMALE PROFESSORS			
online	-0.481***	-0.458***	-0.499***
	(0.07)	(0.11)	(0.09)
below-average appearance	0.16	0.18	0.17
x online	(0.15)	(0.23)	(0.17)
above-average appearance	-0.0321	-0.152	0.0136
x online	(0.14)	(0.24)	(0.17)
Observations	80,394	31,457	48,937
R-squared	0.477	0.502	0.457
MALE PROFESSORS			
online	-0.334***	-0.294***	-0.372***
	(0.05)	(0.08)	(0.07)
below-average appearance	-0.0767	-0.0251	-0.20
x online	(0.10)	(0.13)	(0.18)
above-average appearance	-0.266***	-0.341**	-0.221*
x online	(0.10)	(0.16)	(0.13)
Observations	87,160	44,885	42,275
R-squared	0.458	0.462	0.461

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 6: Return to Normalized Appearance Large v. Small Classes

	(1)	(2)	(3)
		grade	
	all students	male students	female students
ALL PROFESSORS			
large class	-0.0644***	-0.0693***	-0.0624***
	(0.01)	(0.01)	(0.00)
normalized appearance	-0.00301	-0.00059	-0.000917
x large class	(0.01)	(0.01)	(0.00)
Observations	168,092	76,620	91,472
R-squared	0.43	0.441	0.422
FEMALE PROFESSORS			
large class	-0.0497***	-0.0497***	-0.0421***
	(0.01)	(0.01)	(0.02)
normalized appearance	-0.00258	-0.00258	-0.0023
x large class	(0.01)	(0.01)	(0.01)
Observations	80,394	80,394	48,937
R-squared	0.476	0.476	0.456
MALE PROFESSORS			
large class	-0.0728***	-0.0680***	-0.0813***
	(0.01)	(0.02)	(0.02)
normalized appearance	-0.00422	-0.00168	0.000701
x large class	(0.01)	(0.01)	(0.01)
Observations	87,160	44,885	42,275
R-squared	0.458	0.461	0.46

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 7: Return to Categorical Appearance Large v. Small Classes

	(1)	(2)	(3)
		grade	
	all students	male students	female students
ALL PROFESSORS			
large class	-0.0600***	-0.0578***	-0.0638***
	(0.01)	(0.01)	(0.00)
below-average appearance	-0.01	-0.02	0.01
x large class	(0.02)	(0.03)	(0.00)
above-average appearance	-0.0206	-0.0560*	0.000167
x large class	(0.02)	(0.03)	(0.00)
Observations	168,092	76,620	91,472
R-squared	0.43	0.441	0.422
FEMALE PROFESSORS			
large class	-0.0463***	-0.0431*	-0.0536***
	(0.01)	(0.02)	(0.02)
below-average appearance	(0.01)	-0.0773**	0.05
x large class	(0.03)	(0.04)	(0.04)
above-average appearance	-0.0125	-0.106**	0.0233
x large class	(0.03)	(0.05)	(0.03)
Observations	80,394	31,457	48,937
R-squared	0.476	0.502	0.456
MALE PROFESSORS			
large class	-0.0686***	-0.0675***	-0.0729***
	(0.01)	(0.02)	(0.02)
below-average appearance	0.00319	0.00838	-0.02
x large class	(0.03)	(0.03)	(0.04)
above-average appearance	-0.0268	-0.0148	-0.0276
x large class	(0.02)	(0.04)	(0.03)
Observations	87,160	44,885	42,275
R-squared	0.458	0.461	0.46

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 8: Appearance and Course Choice

	(1)	(2)	(3)	(4)	(5)	(6)
	male professor		online		class size over 30	
below average	0.0048	-0.0015	-0.000826	-0.00259	0.00125	0.00768
appearance	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
above average	-0.00987**	-0.00816	0.00065	0.0000351	0.00441	0.00892**
appearance	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
male	0.0284***	0.0292***	-0.00424***	-0.00331***	0.00748***	0.0109***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
male × below	-0.00517	0.000739	0.00199	0.00469	0.00112	-0.0119
average appearance	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
male × above	0.0084	0.00717	-0.000737	-0.00317	-0.0011	-0.00364
average appearance	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
online	-0.0781***	-0.0637***			-0.153***	-0.164***
	(0.01)	(0.02)			(0.01)	(0.02)
ACT score		0.000994*		0.000501***		-0.00142***
		(0.00)		(0.00)		(0.00)
	0.527***	0.594***	-0.0446***	-0.0125	0.671***	0.795***
	(0.02)	(0.13)	(0.01)	(0.07)	(0.02)	(0.10)
Observations	182,331	97,166	191,166	101,592	191,166	101,592
R-squared	0.207	0.194	0.16	0.126	0.465	0.478

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