An Evaluation of the Relationship between Instructor Appearance and College Student Evaluations of Teaching

Nathan E. Gonyea, nathan.gonyea@esc.edu, SUNY Empire State College
Marie Osick, mosick92@hotmail.com, Johnson City Central School District
Elizabeth G. Bradley, Elizabeth.Bradley@esc.edu, SUNY Empire State College

Despite the media attention given to research purporting to show a relationship between faculty member appearance and college student evaluations of teaching (SETs), there are significant methodological issues with much of the research on this topic. Specifically, most of this research has used data from the RateMyProfessors.com website, which is of questionable quality. This research also measures college students' perceptions of faculty members' appearance and SETs at the same point in time, which makes it difficult to make conclusions concerning the direction of the relationship between the two variables. The present study sought to address these methodological issues by measuring college student perceptions of faculty members' appearance and SETs at various points over the course of one semester, using a sample of 408 college students enrolled in 12 courses at a comprehensive regional institution in the Northeastern United States. Specifically, the study investigated whether students' pre-course perceptions of a faculty member's appearance and end-of-course perceptions of a faculty member's appearance were correlated with end-of-course SETs. Results indicated that college students' pre-course perceptions of appearance were not related to SETs; however, end-of-course perceptions were positively correlated with SET ratings. A subsequent regression analysis found that end-of-course student perceptions of a faculty member's appearance were significantly predicted by

students' overall perception of the course at midterm and the perception of the course's workload when controlling for students' initial perception of the faculty member's appearance and other factors, suggesting that end-of-course perceptions of appearance are impacted by students' experience in the course.

Keywords: c*ollege student evaluations of teaching, attractiveness, appearance*

There has been a considerable amount of media attention (e.g., Shea, 2012; Wilson, 2010) given to studies that have found a relationship between college students' ratings of professors' appearance, specifically attractiveness, and college student evaluations of teaching (SETs; i.e. surveys of students' perceptions of a course or instructor administered at the end of a course). However, this research suffers from a number of significant methodological limitations, one being that most of this research uses data from one website, RateMyProfessors.com (RMP) (Boehmer & Wood, 2017; Coladarci & Kornfield, 2007; Davison & Price, 2009; Felton, Mitchell, & Stinson, 2004; Felton, Koper, Mitchell, & Stinson, 2006; Felton, Koper, Mitchell, & Stinson, 2008; Freng & Webber, 2009; Kindred & Mohammed, 2005; Mangan & Fleck, 2011; Otto, Sanford, & Ross, 2008; Riniolo, Johnson, Sherman, & Misso, 2006; Rosen, 2017). Research investigating the use of the RMP dataset to evaluate the relationship between attractiveness and SETs has brought the validity of this research into question (Gonyea & Young, 2012). Specifically, rating attractiveness and teacher quality at the same point in time opens up the possibility that the former is impacted by the latter. Additional research has discovered that RMP evaluations are negatively biased and do not predict college student scores on multi-item measures of SET (Murray & Zdravkowiv, 2016). RMP is also biased by a halo effect and creates more of a likeability scale, rather than being a valid measure of teaching effectiveness (Clayson, 2014).

Research that has avoided the first limitation by collecting data from official SETs or researcher administered SETs has suffered from another limitation: the use of small numbers of college students who have not taken courses from the professor and instead evaluate their attractiveness based on a photograph (Buck & Tiene, 1989; Campbell, Gerdes, & Steiner, 2005; Hamermesh & Parker, 2003; Ponzo & Scoppa, 2012; Wolbring & Riordan, 2016). The small n of

the sample who rated the faculty member brings into question the generalizability of those ratings to a larger sample of college students who provided the SET ratings. Another issue threatening the generalizability of the attractiveness ratings is the use of photographs. Photographs of a faculty member's face do not provide the full information concerning attractiveness available to college students enrolled in a class. Therefore, even if the small number of college students' views of attractiveness could be generalized to the larger group of college students given the SETs, that small number of college students is not rating overall attractiveness, but rather facial attractiveness. The purpose of the present study is to address the limitations of previous research by investigating the underlying assumption of past research that the direction of the relationship between appearance and SETs is that a faculty member's appearance influences a student's evaluation (as opposed to a faculty member's teaching influencing a student's perception of the faculty member's attractiveness). It also expands the definition of appearance to include both attractiveness (as in previous research) and overall appearance. Specifically, this study evaluates the relationship between college students' ratings of faculty members' appearance (both attractiveness and overall appearance) measured at the beginning and end of the course with end-of-course college student evaluations of teaching ratings. A second analysis investigates the assumed direction of the relationship between ratings of appearance and SETs by evaluating whether end-of-course ratings of appearance can be predicated by SETs when controlling for beginning-of-course ratings of appearance.

Review of Literature

The research on appearance and college student evaluations of teaching (SETs) falls into one of three categories – correlational studies using RateMyProfessors.com (RMP), correlational studies using "independent" ratings of attractiveness, and experimental studies using documents

(i.e. descriptions of fictitious faculty and their teaching). The largest body of research on appearance involves analyzing data from RMP for correlations between attractiveness ('hotness') and other variables (e.g., easiness and overall quality; Felton et al., 2004; Felton et al., 2006; Felton et al., 2008; Gonyea & Young, 2012; Kindred & Mohammed, 2005; Otto et al., 2008; Riniolo et al., 2006). All of these studies have found a significant relationship between the RMP hotness ratings and other course ratings found on RMP (Felton et al., 2004; Felton et al., 2006; Felton et al., 2008; Gonyea & Young, 2012; Kindred & Mohammed, 2005; Otto et al., 2008; Riniolo et al., 2006). However, the limitations of RMP as a source of data put these results into question. While RMP provides large amounts of information, the quality of that information is questionable. There is no way to validate who made the ratings or that the raters were actually enrolled in the class. Further, all of the ratings (e.g., 'hotness,' easiness, and clarity) are made at the same point in time, which makes it impossible to know if perceptions of attractiveness impact perceptions of overall quality, vice versa, or if a third variable impacts both. Finally, there is evidence that self-selection bias results in a different population of student raters on RMP than what would be representative of students in the classroom, resulting in lower evaluations on RMP (Legg & Wilson, 2012). Gonyea and Young (2012) evaluated the efficacy of using RMP in faculty attractiveness research by comparing the relationship between RMP ratings of attractiveness and RMP ratings of overall quality with the relationship between an independent measure of attractiveness (i.e., one given based on a photograph by college students who have never met the instructor) and RMP ratings of overall quality. They found that there was a significant correlation between the two RMP ratings, but not between the independent attractiveness rating and RMP overall quality. This finding suggests that RMP attractiveness ratings are not independent of faculty SET ratings and suggests that the direction of the

relationship hypothesized in previous studies (i.e., attractiveness impacting SET ratings) may not be correct. While this study identified a weakness in the research using RMP, it does not provide definitive information concerning the relationship between attractiveness and overall quality ratings.

A similar line of research uses data from campus databases or researcher-administered SETs (Buck & Tiene, 1989; Campbell et al., 2005; Hamermesh & Parker, 2003; Ponzo & Scoppa, 2012). While this strategy corrects the problems of not being able to verify whether or not raters were college students enrolled in a course with a faculty member and the timing of the attractiveness and overall quality ratings, it has the same limitation concerning "objective" ratings as Gonyea and Young's (2012) study. Additionally, all of these studies use small numbers of college students to rate professor attractiveness (i.e., less than 30; Campbell et al., 2005; Hamermesh & Parker, 2003; Ponzo & Scoppa, 2012). With such small samples, it is feasible that the beliefs concerning what makes a professor attractive may be idiosyncratic and therefore not generalizable to the significantly larger group of college students who rated overall quality.

The limitations of this independent attractiveness rater approach may explain the inconsistent findings of studies using this method. Hamermesh and Parker (2003) and Ponzo and Scoppa (2012) both found significant results using independent attractiveness ratings (6 undergraduates and 29 undergraduates respectively) and large databases of college student evaluations (both studies used over 16,000 evaluations). Conversely, Campbell et al. (2005) did not find significant relationships between attractiveness and overall quality ratings. They used 11 undergraduates to rate the photographs of 53 faculty members who they had never met. The researchers then regressed the attractiveness ratings and a series of control variables on faculty

members' SET ratings from 70 classes. After controlling for the other variables (e.g., rigor of course, workload expectation, and professor gender), attractiveness was not a significant predictor of SET ratings.

The least utilized method of investigating the relationship between attractiveness and SETs is an experimental approach. Buck and Tiene (1989) attempted to experimentally isolate the effect of faculty member attractiveness on ratings. They began by identifying photographs of highly attractive and highly unattractive faculty members (based on the ratings of 23 undergraduates) and pairing each photograph with two different teaching philosophies. A group of education college students then reviewed the photographs and philosophies and rated the quality of the instructor based on this information. They reported no main effect for attractiveness. The major limitation of this study is that the raters never actually saw the faculty members teach. Therefore, these findings may not be applicable to ratings of actual teaching.

While researchers have taken numerous approaches to evaluating the relationship between faculty member appearance (almost exclusively attractiveness) and SETs, the limitations of the existing research make it difficult to provide definitive information to college faculty and administrators concerning how they should interpret the SETs of faculty members within the context of faculty members' physical appearances. In order to address this issue, the present study expands the scope of appearance to include both attractiveness (all of the studies reviewed only defined appearance as attractiveness) and overall appearance (as defined by attractiveness, professionalism of dress, and stylishness of dress) and had the same participants rate both faculty appearance and teaching quality. Further, the present study had the participants rate faculty member appearance twice, once at the beginning and once at the end of the semester.

Hypotheses

It is hypothesized that pre-class ratings of faculty members' attractiveness and overall appearance (as defined by attractiveness, professionalism, and stylishness) will not be significantly correlated with end-of-class college student evaluations of teaching; however, it is hypothesized that end-of-class ratings of faculty members' attractiveness and overall appearance will be significantly positively correlated with end-of-class college student evaluations of teaching. Further, it is hypothesized that factors related to college students' general attitude toward a class and instruction (i.e., change in anticipated grade, difference between expected workload and actual workload, and midterm college student evaluation of teaching score) will be significant predictors of college students' final ratings of faculty members' overall appearance when controlling for the sex of both the college student and the professor, and pre-class appearance ratings.

Method

Participants

Participants included 408 undergraduate (44% lower division and 49% percent upper division) and graduate (6%) college students enrolled in education, social science, and natural science courses at a public masters-level college in the Northeastern U.S., and 12 faculty members. The most commonly reported majors were in the fields of education (53%), social sciences (19%), and STEM (16%). Participants were predominately female (78%) and Caucasian (90%), with a median age of 20 years and a median GPA of 3.3. Seventy-one percent of the college students reported being enrolled in the course because it was required for their major or minor, and 67% chose their particular instructor for the course because only one section of the course was offered or only one section of the course fit into their schedule. College students who

had previously taken a course with the same instructor were excluded from study participation. The college's overall population of students was over 80% Caucasian and over 60% female. Education majors, which made up 53% of the sample, were more Caucasian and more female than the college population as a whole. All of the faculty members were Caucasian and 9 of the 12 were female. At the time of data collection, approximately 90% of the institution's faculty were Caucasian.

Instruments

Data for this study came from a larger study of factors related to college student evaluations of teaching. Data were collected from pre-, midterm-, and post-class surveys. The pre-class survey consisted of demographic questions and questions concerning college students' initial impressions of the course and instructor. The midterm survey consisted of questions pertaining to college students' perception of the course and instructor after midterm grades were released. The post-class survey consisted of questions pertaining to the college students' final perceptions of the course and instructor (see Appendix for List of Student Perception of Instruction (SPI) Questions).

College students' sex, perception of course workload, and anticipated grade were evaluated by single items on both the pre- and post-class surveys. The workload item was rated on a 5-point scale from 0 (Very Little) to 4 (Unreasonably High). The difference in expected and actual workload used in the regression analysis was calculated by subtracting the workload rating from the pre-class survey from the workload rating from the post-class survey. Therefore, a positive change in workload score means that college students rated the actual workload of the class on the final survey higher than their initial expectation of the workload of the class from the pre-survey. The change in grade variable was similarly calculated by subtracting a college

student's initial grade expectation (on an E [0] to A [4] scale) from their end-of-course expected grade, which means a positive score reflects an increase in a college student's expected grade from pre-class to post-class survey.

Both the midterm and final perceptions of the course were evaluated using the campus instrument called the Student Perception of Instruction (SPI) survey. The survey consists of ten items scored on 4-point scale (Strongly Disagree to Strongly Agree) and one overall evaluation item scored on a 5-point scale. For the purpose of this study, the ten 4-point scale items were summed to create a cumulative SPI score. See the Appendix for a list of the 10 SPI questions used in this study.

College students' perceptions of their instructor's appearance were measured in two ways. The first was similar to the single-item method used in most other studies. Specifically, participants were asked to rate the physical attractiveness of their instructor on a 6-point scale from 0 (Very Unattractive) to 5 (Very Attractive). The second method was on a 3-item appearance scale that combined the attractiveness item with items about the professionalism (Very Unprofessional to Very Professional) and stylishness (Very Unstylish to Very Stylish) of their instructor's appearance (See Table 1).

Table 1.

Descriptive Statistics

Variable	Mean	SD	Range	N	Alpha
Attractiveness Pre-class	2.83	1.09	5	408	
Attractiveness Post-class	2.82	.96	5	408	
Appearance Pre-class	9.68	2.03	13	408	.52
Appearance Post-class	9.38	2.20	12	408	.60
Sex of College student*	.78	.42	1	408	
Sex of Instructor*	.81	.39	1	408	
Expected Grade Change	-.10	.62	6	405	
Change in Workload	-.09	.79	7	408	
Midterm SET Score	32.28	4.98	28	209	.91
Final SET Score	33.91	5.73	30	408	.93

* 0 (Male) to 1 (Female)

Design

The present study used a correlational design consisting of three surveys administered to college students enrolled in the courses of participating instructors. The study used both bivariate correlations and a stepwise linear regression to test the hypotheses.

Procedures

Faculty were recruited to participate in the study via email and a campus message board posting. College students enrolled in sections of classes taught by participating faculty were asked to complete three surveys. The first was conducted during the last 15 minutes of the first

class of the semester, the second was administered online after midterm grades were released, and the final survey was administered on the last day of class prior to final exams. The response rate on the midterm survey was approximately half of that in the face-to-face administrations, which is reflected in the number of participants included in the regression analysis that includes the midterm SET rating from this survey. There were no significant differences in the means on the final SET rating or final instructor appearance ratings between the midterm survey completers and non-completers. All three surveys included questions designed to create an anonymous ID (e.g., "What is the last letter of your last name?" and "What are the last two digits of your cellphone number?") that could be used to match the surveys.

Results

The hypotheses were evaluated using four binary correlations and one stepwise linear regression. The hypothesis that pre-course ratings of faculty members' attractiveness and overall appearance would not be significantly correlated with end-of-class college student evaluations of teaching was supported, $r(408) = -.01$, $p = .86$ and $r(408) = .04$, $p = .42$ respectively. The hypothesis that end-of-class ratings of faculty members' attractiveness and overall appearance would be significantly positively correlated with end-of-class college student evaluations of teaching was supported, $r(408) = .14$, $p = .004$ and $r(408) = .39$, $p < .001$ respectively (see Table 1 for descriptive statistics).

The final hypothesis that change in anticipated grade, difference between expected workload and actual workload, and midterm college student evaluation of teaching score would predict college students' final ratings of faculty members' overall appearance when controlling for the sex of both the college student and professor and pre-class appearance ratings was evaluated using a stepwise linear regression. In the first step, appearance was entered to identify

the extent to which college students' initial evaluation of their instructor's appearance predicted their final evaluation of their instructors' appearance. The next step introduced the two other control variables: sex of college student and instructor. The final step introduced the predictor variables: change in anticipated grade, difference between expected workload and actual workload, and midterm college student evaluation of teaching score. The hypothesis was partially supported. Change in workload and midterm SET were both significant predictors of the final appearance rating when controlling for the other variables, while change in anticipated grade was not, $R^2 = .30$, $F(6, 200) = 14.35$, $p < .001$. The final model explained 30% of the variance, an increase of 12 percentage points from the second step that included only the three control variables, $R^2 = .18$, $F(3, 203) = 15.06$, $p < .001$. Of note is that the pre-class rating of a faculty member's appearance accounted for only 16% of the post-class rating of a faculty member's appearance (see Table 2 for stepwise linear regression results).

Table 2.

Stepwise Linear Regression

Variable	B	SE	ß	F	df	R^2
Step 1				40.29***	1, 205	.16
Appearance Pre-Class	.47	.07	.41***			
Step 2				15.06***	3, 203	.18
Appearance Pre-Class	.49	.08	.42***			
Sex of College student	.75	.41	.12			
Sex of Instructor	-.43	.42	-.07			
Step 3				14.36***	6, 200	.30
Appearance Pre-Class	.45	.07	.39***			
Sex of College student	1.08	.39	.17**			
Sex of Instructor	-.37	.40	-.06			
Expected Grade Change	-.16	.21	-.05			
Change in Workload	.60	.18	.20***			
Midterm SET Rating	.14	.03	.29***			

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Discussion

The results of the first two hypotheses suggest that a faculty member's attractiveness or overall appearance, as measured by an initial rating without prior knowledge of the faculty member's teaching abilities, teaching style, or personal characteristics, is not related to college student evaluations of teaching. This is consistent with the findings of Gonyea and Young (2012). When combined with the findings concerning the correlations between final

attractiveness and appearance scores and final SET ratings, these findings bring into question the conclusions drawn by previous research that used attractiveness and SET ratings taken at a single point in time (e.g., Felton et al., 2004; Felton et al., 2006; Felton et al., 2008; Kindred & Mohammed, 2005; Otto et al., 2008; Riniolo et al., 2006). Specifically, the finding that correlations between end-of-class ratings of attractiveness and overall appearance are significant (though weak) and the finding that correlations for the pre-course ratings and final SETs were zero supports the conclusion that students' initial impression of a faculty member's attractiveness does not impact the student's SET for that faculty member. Therefore, the present findings suggest that previous studies' findings were potentially confounded by the impact of course-specific factors that influenced college students' perceptions of a faculty member's attractiveness at the end of a course. Given the weakness of the present study's correlations for end-of-class attractiveness and overall appearance ratings, and end-of-class SETs, there may be little meaningful relationship between faculty member appearance and SETs.

The impact of course-related events and factors on final appearance ratings was evaluated in the final analysis. The finding that both the expected versus actual workload and midterm SET rating significantly predicted final appearance ratings further supports the conclusion that the differences in the correlational results for pre-class and post-class appearance ratings and final SET rating were due to the impact of college students' experiences in the course. Thus, the final appearance ratings appear to be acting, in part, as a measure of college students' general attitudes toward a course. This supports previous researchers' hypotheses concerning a halo effect (e.g., Clayson, 2014), only in the opposite direction those researchers assumed.

While this study addressed many of the limitations in past research on attractiveness and SETs, like all research, it has limitations. Specifically, the present study was conducted on a

single campus with that campus' SET form. While this instrument is an improvement on the single-item scales used in much of the previous research, it is not necessarily representative of all SET instruments currently in use. Similarly, the appearance measure, while an improvement over the single item scales of past studies (e.g., Felton et al., 2008; Kindred & Mohammed, 2005; Riniolo et al., 2006) for which reliability cannot be calculated, had relatively low internal consistency. Future research should focus on developing a more robust measure of instructor appearance with greater internal consistency.

The sample was another area of limitation. It consisted of education, social science, and STEM courses taught by 12 faculty members that enrolled predominately female and Caucasian education, social science, and STEM majors.. Future research should address these limitations by using a variety of SET instruments at campuses throughout the world, with classes, college students, and faculty from a broader array of disciplines and demographic backgrounds.

In conclusion, results of this study suggest that faculty members' appearance (attractiveness or overall appearance), as measured by an initial rating without prior knowledge of a faculty member's teaching abilities, teaching style, or personal characteristics, is not related to college student evaluations of teaching. While more research is needed to confirm the present study's findings, its results combined with the methodological weaknesses of much of the past research in this area highlighted by Gonyea and Young (2012), shed serious doubt on the much publicized and touted relationship between a faculty member's attractiveness and their SET ratings.

Practical implications of this research relate to how college faculty and administrators interpret SET results. The current findings suggest that faculty attractiveness does not impact SET results; thus, the attractiveness of a faculty member should not impact SET results or their

interpretation by promotion and tenure committees or administrators using these results to inform a decision concerning a faculty member's instructional effectiveness. Future research should investigate the relationship between faculty attractiveness and SET ratings in a more diverse and representative sample of students and faculty in order to evaluate whether the present study's findings are generalizable across demographic groups.

References

Boehmer, D. M., & Wood, W. C. (2017). Student vs. faculty perspectives on quality instruction: Gender bias, "hotness," and "easiness" in evaluating teaching. *Journal of Education for Business, 92*(4), 173-178. doi: 10.1080/08832323.2017.1313189

Buck, S., & Tiene, D. (1989). The impact of physical attractiveness, gender, and teaching philosophy on teacher evaluations. *Journal of Educational Research, 82*(3), 172-177.

Campbell, H. E., Gerdes, K., & Steiner, S. (2005). What's looks got to do with it? Instructor appearance and college student evaluations of teaching. *Journal of Policy Analysis and Management, 24*(3), 611-620. doi: 10.1002/pam.20122

Clayson, D. (2014). What does RateMyProfessors.com actually rate? *Assessment & Evaluation in Higher Education, 39*(6), 678-698. doi:10.1080/02602938.2013.861384

Coladarci, T., & Kornfield, I. (2007). RateMyProfessors.com versus formal in-class college student evaluations of teaching. *Practical Assessment, Research & Evaluation, 12*(6). Retrieved from http://pareonline.net/pdf/v12n6.pdf

Davison, E., & Price, J. (2009). How do we rate? An evaluation of online college student evaluations. *Assessment & Evaluation in Higher Education, 34*(1), 51-65. doi: 10.1080/02602930801895695

Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2004). Web-based college student evaluations of professors: The relations between perceived quality, easiness and sexiness. *Assessment & Evaluation in Higher Education, 29*(1), 91-108. doi: 10.1080/0260293032000158180

Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2006). Attractiveness, easiness, and other issues: College student evaluations of professors on RateMyProfessors.com. *The Social Science Research Network Website*. doi: 10.2139/ssrn.918283

Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2008). Attractiveness, easiness, and other issues: College student evaluations of professors on RateMyProfessors.com. *Assessment & Evaluation in Higher Education, 33*(1), 45-61. doi: 10.1080/02602930601122803

Freng, S., & Webber, D. (2009). Turning up the heat on online teaching evaluations: Does "hotness" matter? *Teaching of Psychology, 36*(3), 189-193. doi: 10.1080/00986280902959739

Gonyea, N. E., & Young, J. (2012). *'Hot' professors are good professors or good professors are 'hot' professors: Evaluating the relationship between professor attractiveness and college students' evaluations of teaching using data from RateMyProfessors.com.* Poster presented at the annual meeting of the American Educational Research Association Meeting, Vancouver, BC.

Hamermesh, D. S., & Parker, A. M. (2003). Beauty in the classroom: Professors' pulchritude and putative pedagogical productivity. *Economics of Education Review, 24*(4), 369-376. doi: 10.1016/j.econedurev.2004.07.013

Kindred, J., & Mohammed, S. N. (2005). "He will crush you like an academic ninja!" Exploring teacher ratings on RateMyProfessors.com. *Journal of Computer-Mediated Communication, 10*(3). Retrieved from http://jcmc.indiana.edu/vol110/issue3/kindred.html. doi: 10.1111/j.1083-6101.2005.tb00257.x

Legg, A. M. & Wilson, J. H. (2012). RateMyProfessors.com offers biased evaluations. *Assessment & Evaluation in Higher Education, 37*(1), 89-97.

Mangan, M. A., & Fleck, B. (2011). Online college student evaluation of teaching: Will professor "hot and easy" win the day? *Journal on Excellence in College Teaching, 22*(1), 59-84.

Murray, K. B. & Zdravkovic, S. (2016). Does MTV really do a good job of evaluating professors? An empirical test of the internet site RateMyProfessors.com. *Journal of Education for Business*, *91*(3), 138-147. doi: 10.1080/08832323.2016.1140115

Otto, J., Sandford, D. A., Jr., & Ross, D. N. (2008). Does RateMyProfessor.com really rate my professor? *Assessment & Evaluation in Higher Education, 33*(4), 355-368. doi: 10.1080/02602930701293405

Ponzo, M., & Scoppa, V. (2012). The good, the bad, and the ugly: Teaching evaluations, beauty and abilities (Working Paper No. 04-2012). Retrieved from http://www.ecostat.unical.it/RePEc/WorkingPapers/WP04_2012.pdf

Riniolo, T. C, Johnson, K. C., Sherman, T. R., & Misso, J. A. (2006). Hot or not: Do professors perceived as physically attractive receive higher college student evaluations? *The Journal of General Psychology 133*(1), 19-35. doi: 10.3200/GENP.133.1.19-35

Rosen, A.S. (2017). Correlations, trends and potential biases among publicly accessible web-based student evaluations of teaching: A large-scale study of RateMyProfessors.com data. *Assessment & Evaluation in Higher Education, 43*(1), 31-44.

Shea, C. (2012, April 10). Shocking: Attractive professors get better college student

evaluations [Blog post]. Retrieved from http://blogs.wsj.com/ideas-

market/2012/04/10/shocking-attractive-professors-get-better-college student-

evaluations/

Wilson, R. (2010). Professors: Hot at their own risk. *The Chronicle of Higher Education*.

Retrieved from: http://chronicle.com/article/Professors-Hot-at-Their-Own/123822/

Wolbring, T. & Riordan, P. (2016). How beauty works: Theoretical mechanisms and two

empirical applications on college students' evaluation of teaching. *Social Science*

Research, 57, 253-272. doi: 10.1016/j.ssresearch.2015.12.009

Appendix

List of Student Perception of Instruction (SPI) Questions

1. Written course requirements and policies were clearly stated and distributed at the beginning of the semester.
2. The instructor was well prepared for class.
3. The instructor's presentation of course material was well organized.
4. The instructor encouraged college students to ask questions and make comments relevant to course content.
5. The instructor treated college students with respect.
6. The instructor presented and explained ideas effectively.
7. Methods of assessment fairly reflected course content.
8. Methods of assessment were graded fairly.
9. Classes met regularly as scheduled.
10. Overall, this class provided a valuable learning experience.