

"This is the *pre-peer reviewed version* of the following article: [Doubleday, A. F. and Lee, L. M.J. (2016), *Dissecting the voice: Health professions students' perceptions of instructor age and gender in an online environment and the impact on evaluations for faculty. Anat Sci Ed. doi: 10.1002/ase.1609*], which has been published in final form at doi: 10.1002/ase.1609]. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving."

TITLE: Dissecting the voice: health professions students' perceptions of instructor age and gender in an online environment and the impact on evaluations for faculty

AUTHORS:

Dr. Alison F. Doubleday, PhD (corresponding author)
Assistant Professor
Department of Oral Medicine and Diagnostic Sciences
University of Illinois at Chicago, College of Dentistry
527 DENT, MC 838
801 S. Paulina St.
Chicago, IL 60612
312-355-4010 (ph)
312-355-1222 (fx)
adouble@uic.edu

Dr. Lisa M.J. Lee, PhD
Associate Professor
Department of Cell and Developmental Biology
University of Colorado, School of Medicine
Mail Stop F435, Rm N5207C
13001 East 17th Place
Aurora, CO 80045
303-724-7460 (ph)
303-724-3420 (fx)
Lisa.L.Lee@ucdenver.edu

RUNNING TITLE: gender and age bias in online learning

Abstract

Gender and age biases are well documented in academia and many studies demonstrate bias in students' evaluation of instructors. An instructor's violation of gender-based expectations can translate to lower scores or negative comments on evaluations. To compound the matter, decreased contact hours and increased blended learning in health sciences curricula has elevated the importance of digital resources. The changing landscape of anatomical sciences education has come to rely on pre-recorded lectures or videos to support or replace face-to-face interactions. While there is evidence of bias in students' evaluations of online instructors, studies with health sciences students is lacking. First-year dental students at two institutions viewed a video on spinal cord anatomy narrated by one of four individuals of different gender and age. Students completed a prequiz and postquiz followed by a questionnaire evaluating the video and instructor based on instructor vocal characteristics. Results revealed gender and age bias in student evaluations of online instructors, with the older female voice receiving the lowest rankings. Students at Institution A scored the younger male and younger female voices highest. At Institution B the older male and younger female received the highest scores. The differences in results from the two samples suggest that the gender and age of in-class instructors has a significant impact on how students perceive members of these groups and on how online instructors are evaluated. This report has implications for the career trajectory of female faculty and perpetuation of bias against aging female faculty in the increasingly digital educational environment.

KEYWORDS: gender, bias, online learning, implicit bias, digital resources, student evaluations

Introduction

A multitude of studies exist outlining specific qualities of effective instructors and associated impact on student learning. While there is some consensus on “must-have” characteristics for effective teaching (Chickering and Gamson, 1987; Korthagen, 2004; Azer, 2005; McKinney, 2007; Vajoczki et al., 2011), answers to the question of how to assess or measure these qualities in instructors is less clear. Student evaluations remain the most common means of identifying qualities of a “good teacher” and for documenting effective teaching. Recent studies, however, have confirmed that student evaluations of instructors are subject to numerous influences, including gender and age biases. These findings call into question the validity and reliability of student evaluations of faculty (Arbuckle and Williams, 2003; Onwuegbuzie et al., 2007; MacNell et al., 2014) and the appropriateness of relying heavily on student evaluations for promotion and tenure decisions.

Compounding the issue is the rise in the use of computer-assisted learning resources over the last decade. For instructors moving to a blending learning format or to a flipped classroom, there can be a significant reliance on online resources. Additionally, student and instructor face-to-face time may be reduced and students can develop opinions about instructors, which are transferred to evaluation forms, based on characteristics exhibited through online interaction and digital resources, such as videos and narrated lectures.

The specific objectives of the study were to see if students’ perceptions of instructor characteristics (approachability, competence) are influenced by the perceived age and gender of the instructor, based on instructor vocal characteristics alone. While other studies have investigated gender bias in student evaluations, students in those studies typically experienced face-to-face interaction with the instructor being evaluated. We were specifically interested in simulating an online learning environment and in understanding student perceptions of an instructor without the benefit of synchronous interaction. We also wanted to know if the students’ perceptions of the instructor’s age and gender affected student performance on learning activities (quizzes) or on their desire and willingness to learn in an online format.

Gender and Age Biases in Evaluations of Faculty

Evaluation Processes

Gender and age biases in hiring, as well as in methods for evaluating faculty members as part of promotion and tenure guidelines have been documented previously. As early as 1976 Bourdieu acknowledged the gendered construction of evaluation criteria in academia and posited that claims of objectivity merely served to reinforce inequalities and support interests of those with power. Steinpreis et al. (1999) discuss gender bias in hiring and promotion credentials, specifically. The authors presented 238 academic psychologists with curriculum vitae for job applicants and for tenure applicants. Although job application materials were identical to each other, and the tenure materials were identical to each other, in every other respect, traditionally male and female names were put on the application materials before being distributed to participating subjects for review. The authors found that both men and women were more likely to hire a male job candidate than a female candidate with identical qualifications for an academic position. Both male and female academics were also more likely to rank the male applicant’s research, teaching, and service qualifications as adequate compared to the (identical) qualifications of the female job applicant. Although the authors did

not find any significant differences for the tenure candidates, male and female academics taking part in the study were more likely to request documentation or include cautionary comments for the female tenure candidate than for the male tenure candidate.

Van den Brink and Benschop (2011) argue that the concept of “academic excellence” is itself a gendered construct that privileges men and promotes a cycle of inequality and underrepresentation of women in academia. The authors examined academic appointment reports and interviews with academic search committee members in the Netherlands. Qualitative analyses revealed that committee members are more likely to give men the benefit of the doubt and more likely to request additional supporting documentation from women. The authors also note that age bias also occurs and frequently intersects with gender bias. For example, women who break from their career path for various reasons often meet criteria for excellence in publication at a later point than do men and that “there seems to be an unspoken age range in which professors should be appointed, and the system favours academics that have followed a traditional masculine career path” (Van den Brink and Benschop, 2011).

Letters of recommendation, also central to the hiring process have revealed some degree of gender bias, such as the use of more standout adjectives for male candidates (Schmader et al., 2007), and more communal adjectives (helpful, kind, sympathetic) for female candidates (Madera et al., 2009). Men also dominate publication output, globally and publications with women in the dominant author position receive fewer citations than publications with men in the dominant author position (Sugimoto, 2013). While this may be the result of underrepresentation of women in academia (in particular, in science fields) and not due to direct bias, the criteria and processes used for promotion and tenure surely have an impact on retention of female faculty. Van den Brink and Benschop (2012) point out that in their research this is particularly true for the field of medical sciences, in which a very small percentage of qualified women go on to apply for and hold academic appointments. These studies suggest that, although institutions attempt to use objective measures of evaluation for faculty, these processes are likely influenced by the same gender and age-based stereotypes found in the broader culture.

Student Perceptions

An important part of the evaluative process for faculty, student evaluations of faculty are influenced by gender and age biases held by students. It has been suggested that the roles faculty members assume in the classroom follow cultural gender norms and expectations and students use different criteria for evaluating male and female instructors (Sprague and Massoni, 2005; Korte et al., 2009). Korte and colleagues (2009) found that student evaluations for business school faculty depended on both the gender of the student and the gender of the instructor, with female students rating teachers higher, in general, and with male and female students emphasizing distinct characteristics for female and male business instructors. Female students rated female instructors higher than male instructors in class preparedness while male students rated female instructors higher than male instructors for content expertise, organized presentation, engaging speaker, industry experience. Both male and female students rated male faculty higher for relaxed demeanor, outgoing personality, sense of humor, and receptiveness to questions. Other studies with undergraduates have also shown that female students rate female instructors higher than do male students (Hancock, Shannon, and Trentham, 1992;

Feldman, 1993; Basow, 1995; Centra and Gaubatz, 2007). As student evaluations play an important role in promotion and tenure decisions for faculty, gender and age biases carry the potential to impact career trajectories of instructors, as well as those of students planning to enter academia (Dannels et al., 2009, McKay et al. 2012). Implicit biases held by students and administrators can impact retention of female faculty in particular (Goodwin and Stevens, 1993). The subsequent underrepresentation of particular age or gender groups within academia limits the opportunities for students to engage in positive interactions with these individuals and to build perceptions of people of different gender and age groups based on these experiences.

The Changing Classroom

There is evidence that students evaluate online courses and online faculty differently than they do in face-to-face courses, even if learning outcomes are positive and demonstrate no significant difference from those for students in face-to-face courses (Cheng, Lehman, & Armstrong, 1991; Davies & Mendenhall, 1998; Johnson et al., 2000). Studies with undergraduates show that students evaluate course instructors more harshly when they simply believe that the instructor is female. McNell and colleagues (2014) demonstrated that when instructors told students in their online course that they were female, without students seeing or hearing the instructor, students gave the instructor a significantly lower rating than when students were told that the instructor was male. This was true regardless of the actual gender of the course instructor. Arbuckle and Williams (2003) presented undergraduate students with a series of slides narrated by a gender-neutral voice and an “instructor” represented by an age and gender-neutral stick figure. Students were grouped into four subsets and provided textual information about the age and gender of the instructor (male under 35 years, female under 35 years, male over 55 years, female over 55 years). Students who were told the instructor was a young male consistently rated the instructor higher for measures of expressiveness, such as enthusiasm and using voice to convey meaningful information, than did students who were provided other demographic information for the instructor. In general, the instructors identified as “older” scored lower than the younger instructors.

Increased reliance on pre-recorded lectures and other digital materials for a flipped classroom environment means that hearing an instructor’s voice might be the most personal interaction many students have with their teacher. While students might feel a connection with the instructor after being exposed to his or her voice, studies in communication, advertising and business reveal that Americans also carry gendered perceptions of vocal characteristics. Pedelty and Kueker (2014) examined over 1000 television ads and concluded that over 75% of advertising voiceovers feature male voices as opposed to female voices. When female voiceovers are used, they are nearly always accompanied by an image of an attractive woman. The authors conclude that men are more often seen as the voice of authority or expertise, while women must be “seen in order to be heard” (Pedelty and Kueker, 2014).

When considering the recent increase in blended learning and online learning environments, settings in which students may have little to no direct contact with faculty, it is essential to consider how gender biases can influence student evaluations of instructors.

Implicit Bias and Health Professions Students

Although some studies have examined gender issues in traditionally male-dominated academic areas (Steel et al., 2002), the vast majority of studies have taken place in the context of undergraduate education or social science education and have rarely been extrapolated to professional and clinical education. As students in health care fields must go on to provide effective care and treatment for patients of all ages and gender, it is essential to understand the extent to which gender and age bias is present among these students and to develop strategies for helping students identify and challenge these biases.

The primary focus of existing studies in health professions fields appears to be the relatively recent increase of women entering traditionally male-dominated fields and the resulting questions surrounding gender differences in clinical practice and in the workplace (McKay et al., 2012, Nesbitt et al., 2003, Sinkford et al., 2003). Although some studies have investigated the role of gender in student perceptions of their learning environment and of faculty interactions (Prayoonwong and Nimnuan 2010), these studies have been relatively narrow in scope and may not be readily generalized to students and educators in other contexts. Most studies looking at implicit bias in health professions education have focused on gender bias in admissions policies, in evaluation of students and in representation of particular groups within health care fields. Axelson et al. (2010) identified gender bias in Medical Student Performance Evaluations, in which female medical students were rated higher than their male counterparts on criteria such as compassion, but significantly lower in categories such as “quick learner”.

In dental education some studies have reported incidences of gender bias within particular specialties (McKay et al., 2012; Rostami et al., 2010) and others suggest that gender differences may affect leadership status within the field (Yuan et al., 2012). Additionally, some studies have revealed that many of the same gender biases uncovered in the hiring and promotion processes at undergraduate academic institutions, are also present in medical schools. For example, Trix and Psenka (2003) demonstrated differences in letter length and in language used in medical school faculty position letters of recommendation for male and female applicants. It is likely that implicit biases reinforced during health professions education can impact the career trajectories of instructors and students, as well as interactions with colleagues, patients and subordinates. The lack of additional studies to replicate and support these observations, however, makes it difficult to characterize how common gender and age biases may be in health professions education, in which contexts they occur, and what interventions may be conducted if they do occur.

Aims and Goals of the Current Study

This study expands the current research on gender and age bias in student evaluations of faculty by specifically looking at instructor vocal characteristics, alone, and simulating an online or blended learning environment in order to determine whether gender and age biases are reflected in student evaluations of online lessons and instructors in this context. Although other studies have addressed these questions, they have primarily been investigated at undergraduate institutions and have not been extrapolated to health professions students and faculty. This study examines first-year dental student perceptions of faculty at two different institutions. Other researchers (Arreola, 2000) suggest that biases revealed by many studies are due to the specific courses/content that instructors are assigned to teach rather than to the

instructor's gender. Our study tests this assertion by looking at student perceptions of an online lesson covering the same discipline-based content across two distinct dental student populations.

Based on findings from the existing literature this study tested the following expectations:

Hypothesis 1: Students in a dental school anatomy class will rate a younger male higher than they would a younger female, older male, or older female, in areas of competence, authority, enthusiasm, and expertise, based on instructor vocal characteristics alone.

Hypothesis 2: Students in a dental school anatomy class will rate a female instructor higher than they would a male instructor in areas of conscientiousness, and caring, based on instructor vocal characteristics alone.

Hypothesis 3: As other studies have demonstrated little to no impact of instructor gender or age on learning outcomes for students, we hypothesize that gender and age of an instructor in a pre-recorded anatomy video will not impact whether or not a student can improve their knowledge of a subject after exposure to an online resource.

Methods:

Study Design

First year dental students at two Midwestern dental education institutions (referred to here as Institution A and Institution B) were recruited to complete a quiz about spinal cord anatomy (pre-quiz), watch one of four randomly-displayed spinal cord tutorial videos, take a different quiz (post-quiz) about spinal cord anatomy, and then complete a questionnaire on participant perceptions of the video and of the instructor.

Four narrated tutorial videos covering the basic anatomy of the spinal cord were developed for use in this study. The four videos are each 12 minutes long and are identical to each other in all respects except for the gender and age of the instructor reading the video script. For video production, instructors uninvolved with any aspect of dental school education, read the video script, verbatim, and followed similar pacing and vocal inflections. The pre-quiz and post-quiz were developed by content experts in anatomical sciences. The pre and post quizzes tested the same content but included different questions. The questionnaire evaluating student perceptions of the instructor was developed at Institution A. The questionnaire included many of the measures of expressiveness used in Arbuckle and Williams' 2003 study. Because we hypothesized that student perceptions of the instructor might also have an impact on how students feel about online learning and how they would choose to learn in the future, we also included items intended to measure student perceptions of the lesson and lesson format, in general. Because all of the items were literature based, content validity for the questionnaire was strong.

Participation in all aspects of the study took place entirely online. All participants completed the study anonymously using a device and browser of their choice and at a time and place convenient for their schedules. No IP addresses were collected. The study was conducted

first at Institution A in a mid-size college town. Approximately one year later the study was replicated at Institution B, in a large urban setting. Expanding the study to two institutions provided comparative results from two different institutional contexts.

Although the study design was consistent across both study locations, there were some key differences that reflect curricular differences between the two schools. The Institution A study was conducted in the fall semester, when students had not yet encountered much of the information in the video. At Institution B the study was conducted in the spring semester, well after students had already been exposed to some of the video content. At the time of the study, the Institution A anatomy program did not make regular use of narrated lectures in their curriculum. At Institution B, narrated lectures have been a standard component of the curriculum for many years, and students have access to numerous pre-recorded lectures. In addition, all in-person lectures are recorded for student review. Institution A has an integrated anatomical sciences curriculum (integrated histology, embryology, neuroanatomy, and gross anatomy). In contrast, Institution B has an integrated systems-based curriculum (no discipline based-courses) with minimal lecture time and an emphasis on learning in small group activities. The overall differences in curriculum are reflected in the way students experience anatomy content at each institution, as well. Anatomy at Institution A consists of both lecture and lab components, while at Institution B, anatomy content is sprinkled throughout the curriculum with a few key lectures incorporated during the year. Overall, the dissection laboratory is the primary vehicle for learning anatomy at Institution B. Lastly, dissemination of the study materials at Institution A was through the institution Learning Management System, Carmen, while materials were distributed to students at Institution B through Qualtrics (Qualtrics, Provo, UT). Similarities and differences between the two institutions are noted in Table 1.

Data Collection at Institution A:

In Fall 2012, first-year dental students at Institution A were recruited to take part in the study. Students received an invitation email containing a link to the learning module in the students' Learning Management System, Carmen. All participants took the pre-quiz and were then randomly shown one of the four spinal cord videos. After viewing the video, students were prompted to complete the post-quiz and questionnaire. Participation was anonymous and voluntary. All methods were approved by the Institutional Review Board (2012B0043).

Data Collection at Institution B:

In Spring 2013, first-year dental students at Institution B were recruited to participate in the study. Students were recruited via email and provided with a link to the learning module in Qualtrics. Within Qualtrics, participants were provided with the pre-quiz, one of the four spinal cord videos (randomly selected), the post-quiz and the questionnaire. Participation was anonymous and voluntary. All methods were approved by the Institutional Review Board (2012-1006).

All participant responses were analyzed in SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.). 67 students participated in the study at Institution A and 64 students participated at Institution B. ANOVA and Tukey's post hoc tests allowed for comparisons in responses among students who viewed the four different videos.

Results:*Institution A**Pre-quiz vs. Post-quiz results:*

Although most students at Institution A significantly improved on the posttest, compared to the pretest, comparison of pre-quiz and post-quiz scores among the video groups at institution A revealed no significant differences. The perceived gender and age of the instructor did not impact student achievement on these quizzes.

*Institution B**Pre-quiz vs. Post-quiz results:*

Similar to the results from Institution A, there were no significant differences in means or in degree of improvement on quiz scores among groups. Student performance on the pre-quiz and post-quiz was not found to be associated with the video (and associated instructor) viewed. In contrast to the results from Institution A, however, Institution B participants did not demonstrate a significant improvement from pre-quiz to post-quiz performance. All students at Institution B scored fairly high on both the pre-quiz and post-quiz. While there is slight improvement on the post-quiz, this result is negligible and not significant.

Survey results at Institution A:

Means for specific questions can be found in Table 2. Overall, students at Institution A gave higher scores to the younger male and younger female instructors for nearly every evaluative question. Overwhelmingly, the experienced female scored lowest for nearly every question. One-way between subjects ANOVA was conducted to compare the effect of the videos on student perceptions of the video instructor. Significant differences among the videos at the $p < .05$ level were found for the following course characteristics: learning module organization [$F(3,63) = 3.847, p = 0.014$]; whether the learning module was presented in a logical manner [$F(3,63) = 3.775, p = 0.015$]; whether the module used appropriate terminology [$F(3,63) = 3.619, p = .018$]. Post hoc comparisons using the Tukey's HSD test indicated that the mean score for module organization, as reported by the group viewing the older female's video differed significantly from the younger male ($M = -.797, p = .011$) and younger female ($M = -.695, p = .048$) videos. The mean score for the logical presentation of the older female video differed significantly from the younger male ($M = -.726, p = .023$) and younger female ($M = -.757, p = .025$) videos. The use of appropriate terminology for the older female video differed significantly from the younger male ($M = -.778, p = .029$) and younger female ($M = -.824, p = .028$) videos.

Significant differences among the videos at the $p < .05$ level were found for the following instructor characteristics: pleasantness of the instructor's voice [$F(3,63) = 4.692, p = 0.005$]. Post hoc comparisons using Tukey's HSD test revealed that the older male video differed significantly, with regard to mean ratings of the pleasantness of the instructor's voice, from the younger male ($M = -.579, p = .019$) and younger female ($M = -.554, p = .042$) videos. One-way MANCOVA revealed no significant differences in responses between male and female students.

To summarize the survey results at Institution A, the younger female and younger male instructors were rated highest on every questionnaire item. Students viewing the video narrated by the older female felt that the overall learning module was significantly less organized, logical and appropriate, in terms of terminology used, than did students viewing the videos narrated by the younger male and younger female. Additionally, the students viewing the older female video rated the pleasantness of the instructor voice significantly lower than did the students viewing the younger male and younger female videos.

Survey results at Institution B:

Means for specific questions can be found in Table 3. Overall, students at Institution B gave higher scores to the older male and younger female instructors for nearly every evaluative question. Similar to results from Institution A, the experienced female scored lower than any other instructor for nearly every question. One-way between subjects ANOVA was conducted to compare the effect of the videos on student perceptions of the video instructor. Significant differences among the videos at the $p < .05$ level were found for the following course characteristics: whether the student would choose to take a similar module again [$F(3,60) = 4.186, p = 0.009$]. Post hoc comparisons using the Tukey's HSD test indicated that the mean score for this item, as reported by the group viewing the older female's video, differed significantly from the mean score for the older male ($M = -1.267, p = 0.012$) and younger female ($M = -1.063, p = 0.042$) videos.

Significant differences among the videos at the $p < .05$ level were found for the following instructor characteristics: pleasantness of the instructor's voice [$F(3,60) = 3.761, p = 0.015$]; whether the instructor cares about student learning [$F(3,60) = 4.515, p = 0.006$]; whether the student would choose to take another course with this instructor [$F(3,60) = 3.440, p = 0.022$]; instructor enthusiasm [$F(3,60) = 3.683, p = 0.017$]; use of varied teaching methods [$F(3,60) = 3.461, p = 0.022$]; whether the instructor made the student feel accepted [$F(3,60) = 4.069, p = 0.011$]; use of proper and varied inflection [$F(3,60) = 3.277, p = 0.027$]; and whether the instructor was knowledgeable [$F(3,60) = 4.338, p = 0.008$]. Whether the instructor sounded relaxed or conscientious both approached significance at $p = 0.052$. Post hoc comparisons using Tukey's HSD test revealed that the older male video differed significantly, with regard to mean ratings of the pleasantness of the instructor's voice, from the younger female video ($M = -.893, p = 0.010$). With regard to whether the instructor cared about student learning, the mean score for the older male differed significantly from the videos for the younger male ($M = .946, p = .019$) and the older female ($M = 1.016, p = 0.009$).

The mean score for whether those viewing the older female video would opt to take another course from the instructor differed significantly from the mean score for those watching the younger female video ($M = -.882, p = .028$) and is close to significantly differing from the older male video ($M = -.816, p = 0.053$). The mean score for enthusiasm differed significantly between the older female and younger female videos ($M = -1.151, p = 0.017$). The mean score for use of novel and varied teaching methods differed significantly between the older female and younger female videos ($M = -.768, p = 0.031$). The mean score for whether the instructor made the student feel accepted differed significantly between the older female and younger female videos ($M = -.820, p = 0.037$). The mean score for use of proper and varied inflection significantly differed between the older male and older female videos ($M = 1.043, p =$

0.031). Finally, with regard to whether the instructor seemed knowledgeable, the mean score for the older female video differed significantly from the older male ($M = -7.29$, $p = 0.045$) and younger female ($M = -.904$, $p = 0.007$) videos. One-way MANCOVA revealed no significant differences in responses between male and female students

To summarize the survey results at Institution B, overall, the learning module was rated similarly across groups, although the younger female and older male received the highest scores for all questionnaire items. Students viewing the older female video indicated that they would not choose to learn using a similar module or take a course with the same instructor more often than did students viewing the older male and younger female videos. Students at Institution B who viewed the older female video rated the instructor significantly lower on a large host of characteristics (enthusiasm, care for student learning, making students feel accepted, being knowledgeable, and use of appropriate inflection) particularly in comparison to the groups who viewed the older male and younger female videos.

Discussion:

As we found no significant differences in improvement on quiz questions across the four exposure groups, our study is consistent with the existing literature. We accept hypothesis 3 and conclude that the perceived age and gender of an instructor in a pre-recorded lecture does not impact learning outcomes after exposure to an online resource. We did note that students at Institution A demonstrated a significant improvement between pre-quiz and post-quiz scores while students at Institution B did not. One hypothesis for differences in the results from the two study sites is that timing for distribution of the study materials likely played a role in students' ability to improve their quiz performances. Students at Institution A participated in the study during the fall semester, before students encountered much of the video content, while at Institution B, participants were provided with the videos and quizzes in the spring semester, after already obtaining a foundation with the content earlier in the curriculum. These differences highlight the need for instructors to carefully consider the goals of providing online resources to students and stress the importance of timing in allowing for these goals to be met. If the learning resource is being used as a vehicle for initial learning, presenting the resource earlier in the curriculum is likely to be better. If instructors provide the resource later, instructional goals might include review, formative assessment or preparing for application of the material.

Based on the existing literature, hypothesis 1 asserted that students viewing the younger male video would rate the instructor higher in areas of competence, authority, enthusiasm, and expertise and hypothesis 2 predicted that female voices would be rated higher than male voices in areas of conscientiousness and caring. Our findings, however, presented a much more complex picture. There are several differences in results across the two samples worth discussing as they may reflect the unique institutional context in which the data were collected.

First-year dental students at both institutions consistently rated the older female lower than the other three instructors for all questions. At institution A, the younger male and younger female instructors scored the highest rankings on the questionnaire items. The younger male instructor ranked highest for questions about course organization, use of appropriate scientific terminology, enthusiasm and pleasantness of voice while the younger

female scored highest for approachability and logical presentation of material. This trend at Institution A does seem to support our hypotheses but the differences in means are not statistically significant. Rather, the results initially suggest that age bias, rather than gender bias is influencing student evaluations. However, looking at Institution A results, the vast majority of differences seen across the groups are due to significantly lower scores for the older female instructor, suggesting that gender and age bias for this group of students, is interconnected.

Even more surprising, at institution B the *older* male and younger female scored higher than the younger male and older female for nearly every question. Once again, the older female receives the lowest scores across the board. To help explain this discrepancy between the two institutions and why the younger male and younger female ranked highest at Institution A while the older male and younger female ranked highest at Institution B, we examined the demographic profile of the anatomy faculty during the time in which students participated in the study. At institution A, the experts and authority figures for D1 (first-year dental school) anatomy were a man in his 40's and a woman in her 30's. At institution B, at the time of data collection, the experts and authority figures for D1 anatomy were a man in his 60's and a woman in her 30's. Although students likely interacted with many other faculty members at their institutions during this time, students spent a great deal of time each week with anatomy faculty in the dissection laboratory. These specific individuals were also regarded as the experts for the content covered in the spinal cord anatomy videos, even though their voices were not used in the videos. Results from these two samples suggest that we should reject hypotheses 1 and 2 and that student perceptions do reflect bias but they are also heavily influenced by the demographic profile of the faculty at their respective institutions. These results suggest that exposure to experts and authority figures of various age and gender groups may affect students' implicit bias towards or against these groups.

The idea that exposure to specific groups of people can impact one's perceptions of individuals belonging to those groups is not new. Project Implicit, a non-profit organization, investigates implicit social cognition through the Implicit Association Test (IAT; Greenwald, McGhee, and Schwartz, 1998), an instrument to measure the degree to which concepts (such as belonging to a particular group) is associated with attributes. The IAT has been used extensively to investigate implicit biases. Some studies demonstrate that, although implicit bias is pervasive, positive exposure to and positive interactions with members of individuals from different groups can alter perceptions (Dasgupta and Greenwold, 2001; Lowery, Hardin, & Sinclair, 2001; Wittenbrink, Judd, & Park, 2001; Sinclair et al., 2005). In their investigation on malleability of implicit bias, however, Joy-Gaba and Nosek (2010) conclude that exposure is a weak contributor to malleability and that prolonged exposure or persistent opportunities for exposure may be necessary to initiate change.

Future directions:

While this study furthers the discussion on gender and age bias in student evaluations of faculty, future work will be necessary to see whether dental students differ significantly from other student populations in their perceptions of instructors based on vocal characteristics. Recruiting students from additional courses in which the primary instructors vary in age and gender will allow for further comparisons to support the argument that positive exposure with faculty really does have an impact on student perceptions of age and gender.

There is some evidence that adult students and students in graduate programs assess instructors differently than do younger students and undergraduate students. Donaldson et al., (1993) found that adult graduate students include criteria for effective teaching, such as role-modeling, ability to motivate students, using a variety of teaching methods, and organizational skills, not mentioned by undergraduates. Because both samples used in this study come from first year dental student cohorts who may resemble graduate students more closely than undergraduates, it is essential to conduct additional studies investigating other student populations and their perception of instructors in an online environment. The fact that the students participating in this study were enrolled in a postgraduate program may also explain why our study found no significant differences in how male and female students rate male and female faculty, which conflicts with findings in much of the existing literature in undergraduate settings.

It will also be necessary to conduct studies that use different narrators/instructors for the videos used in this research, as it is possible that these particular voices elicit positive or negative reactions from viewers. Considering we obtained different results across two separate institutions, we feel this is not likely. Eventually, an experimental study will be needed in order to determine if exposure to instructors of a particular age and gender in a face-to-face environment can reduce implicit bias against these groups. Tied into this issue are questions of how much face-to-face exposure is necessary to alter implicit bias and how long lasting the effect is after exposure. Finally, we also need to consider the fact that gender and age were presented here as dichotomous variables when in reality each exists on a continuum. Further studies will be needed in order to address the complexity of these issues.

Conclusions:

Although the utility and reliability of student evaluations have recently been called into question ([Costin et al., 1971](#); [Onwuegbuzie et al., 2007](#); [Stark and Freishtat, 2014](#)), they continue to be an important part of tenure and promotion decisions at many institutions. The fact that the experienced female was evaluated so poorly, relative to the other instructors, in both samples, highlights some of the challenges women face in academia. Student evaluations, as part of the criteria used in promotion and tenure decisions, can have an effect on the career trajectory of female academics. The dearth of older females in leadership positions compounds the problem, as it is less likely that students will be exposed to this group often enough to alter perception and affect implicit bias against older women. In spite of this, the results presented here offer hope that efforts to increase student exposure to experts in underrepresented groups can lead to positive cultural change. This discussion also highlights the need to retain female faculty within academia and to encourage them to take on visible leadership positions within academic institutions. This is a particularly salient point considering that gender and age bias have the potential to negatively impact the number of instructors retained or promoted within an institution. As a result, gender and age bias can persist and be reinforced institutionally, as lack of retention of these individuals means that students are less likely to be exposed to members of these groups as they progress throughout their academic career.

Notes on Contributors:

Alison F. Doubleday, PhD is an Assistant Professor at University of Illinois at Chicago, College of Dentistry where she teaches gross anatomy and embryology to first-year dental students. Dr. Doubleday's primary research interest focuses on the use of technology to mediate collaborative learning.

Lisa MJ Lee, PhD is an Associate Professor at University of Colorado, School of Medicine. Dr. Lee focuses on developing innovative, cutting-edge virtual educational resources and effective pedagogies for the health science professionals and students.

REFERENCES CITED

- Arbuckle J, Williams BD. 2003. Students' perceptions of expressiveness: age and gender effects on teacher evaluations. Sex Roles 49: 507–16.
- Arreola RA. 2000. Developing a comprehensive faculty evaluation system. 2nd Edition. Bolton, MA: Anker Publishing. P. 79-93
- Axelson RD, Solow CM, Ferguson KJ, Cohen, MB. 2001. Assessing implicit gender bias in medical student performance evaluations. Eval. Health Prof 33:365–85.
- Azer SA. 2005. The qualities of a good teacher: how can they be acquired and sustained? J Royal Soc Med 98:67–69.
- Basow SA. 1995. Student evaluations of college professors: when gender matters. J Educ Psychol 87: 656–65.
- Bourdieu P. 1976. Le champ scientifique. Actes Rech Sci Soc 2: 88–104.
- Centra JA, and Gaubatz NB. 2000. Is there gender bias in student evaluations of teaching? J High Educ 71: 17–33.
- Cheng H-C, Lehman J, Armstrong P. 1991. Comparison of performance and attitude in traditional and computer conferencing classes. Am J Dist Educ 5:51–64.
- Chickering AW, Gamson ZF. 1989. Seven principles for good practice in undergraduate education. Biochem Educ 17: 140–41.
- Costin F, Greenough WT, Menges RJ. 1971. Student ratings of college teaching: reliability, validity, and usefulness. Rev Educ Res 41: 511–35.
- Dannels SA, McLaughlin JM, Gleason KA, Dolan TA, McDade SA, Richman RC, and Morahan PS. 2009. Dental school deans' perceptions of the organizational culture and impact of the ELAM program on the culture and advancement of women faculty. J Dent Educ 73: 676–88.
- Dasgupta N, Greenwald AG. 2001. On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. J Pers Soc Psychol 81: 800–814.
- Davies RS, Mendenhall R. 1998. Evaluation comparison of online and classroom instruction for HEPE 129--Fitness and Lifestyle Management Course. <http://eric.ed.gov/?id=ED427752>.
- Donaldson JF, Flannery DD, Ross-Gordon JM. 1993. A triangulated study comparing adult college students' perceptions of effective teaching with those of traditional students. Cont High Educ Rev 57: 147–65.

Feldman KA. 1993. College students' views of male and female college teachers: part II—evidence from students' evaluations of their classroom teachers. Res High Educ 34: 151–211.

Goodwin LD, Stevens EA. 1993. The influence of gender on university faculty members' perceptions of 'good' teaching. J High Educ 64:166–85.

Greenwald AG, McGhee DE, Schwartz JL. 1998. Measuring individual differences in implicit cognition: the implicit association test. J Pers Soc Psychol 74: 1464–80.

Hancock GR, Shannon DM, Trentham LL. 1993. Student and teacher gender in ratings of university faculty: results from five colleges of study. J Pers Eval Educ 6: 235–48.

Johnson SD, Aragon SR, Shaik N, Johnson SD, Aragon SR. 2000. Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. J Interact Learn 11: 29–49.

Joy-Gaba JA, Nosek BA. 2010. The surprisingly limited malleability of implicit racial evaluations. Soc Psychol 41: 137–46.

Korte L, Lavin A, Davies T. does gender impact business students' perceptions of teaching effectiveness? J Coll Teach Learn 10: 167–78.

Korthagen FAJ. 2004. In search of the essence of a good teacher: towards a more holistic approach in teacher education. Teach Teach Educ 20: 77–97.

Lowery BS, Hardin CD, Sinclair S. 2001. Social influence effects on automatic racial prejudice. J Pers Soc Psychol 81: 842–55.

MacNell L, Driscoll A, Hunt AN. 2014. What's in a name: exposing gender bias in student ratings of teaching. Inn High Educ December 5: 1–13.

Madera JM, Hebl MR, Martin RC. 2009. Gender and letters of recommendation for academia: agentic and communal differences. J Appl Psychol 94: 1591–99.

McKay, Julia C., and Carlos R. Quiñonez. 2012. The Feminization of Dentistry: Implications for the Profession. J Can Dent Assoc 78: c1.

McKinney K. 2007. On reflecting the reflectiveness of the teacher-scholar. TETYC 35: 20–29.

Nesbitt PE, Inglehart MR, Sinkford JC. 2003. Work environment perceptions of full-time dental educators: does gender matter? J Dent Educ 67: 916–24

- Onwuegbuzie AJ, Witcher AE, Collins KMT, Filer JD, Wiedmaier CD, Moore CW. 2007. Students' perceptions of characteristics of effective college teachers: a validity study of a teaching evaluation form using a mixed-methods analysis. *Am Educ Res J* 44: 113–60.
- [Pedelty M, Kuecker M. 2014. Seen to be heard? gender, voice, and body in television advertisements. *Comm Crit-Cult Stu* 11: 250–69.](#)
- [Prayoonwong T, Nimnuan C. 2010. Dental students' perceptions of learning environment. *SE Asian J Med Educ* 4: 49-54.](#)
- [Rostami F, Ahmed AE, Best AM, Laskin DM. 2010. The changing personal and professional characteristics of women in oral and maxillofacial surgery. *J Oral Maxillofac Surg* 68:381-5.](#)
- [Schmader T, Whitehead J, Wysocki VH. 2007. A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles* 57: 509–14.](#)
- [Sinclair S, Lowery BS, Hardin CD, Colangelo A. 2005. Social tuning of automatic racial attitudes: the role of affiliative motivation. *J Pers Soc Psychol* 89: 583–92.](#)
- [Sinkford JC, Valachovic RW, Harrison S. 2003. Advancement of women in dental education: trends and strategies. *J Dent Educ* 67:79-83.](#)
- [Sprague J, Massoni K. 2005. Student evaluations and gendered expectations: what we can't count can hurt us. *Sex Roles* 53: 779–93.](#)
- Stark PB, Freishtat R. 2014. An evaluation of course evaluations. Center for Teaching and Learning, University of California, Berkley. Retrieved from <http://www.stat.berkeley.edu/~stark/Preprints/evaluations14.pdf> (July 13, 2015)
- [Steele J, James JB, Barnett RC. 2002. Learning in a man's world: examining the perceptions of undergraduate women in male-dominated academic areas. *Psychol Women Quart* 26: 46–50.](#)
- [Steinpreis RE, Anders KA, Ritzke D. 1999. The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study. *Sex Roles* 41: 509–28.](#)
- Sugimoto, Cassidy R. 2013. Global gender disparities in science. *Nature* 504: 211-213.
- [Trix F, Psenka C. 2003. Exploring the color of glass: letters of recommendation for female and male medical faculty. *Discourse Soc* 14: 191–220.](#)
- Vajoczki S, Savage P, Martin L, Borin P, Kustra EDH. 2011. Good teachers, scholarly teachers and teachers engaged in scholarship of teaching and learning: a case study from McMaster University, Hamilton, Canada. *Can J SoTL* 2: 1-27.

Van den Brink M, Benschop Y. 2012. Gender practices in the construction of academic excellence: sheep with five legs. *Organization* 19: 507–24.

Wittenbrink B, Judd CM, Park B. 2001. Spontaneous prejudice in context: variability in automatically activated attitudes. *J Pers Soc Psychol* 81: 815–27.

Yuan JC, Lee DJ, Kongkiatkamon S, Ross S, Prasad S, Koerber A, Sukotjo, C. 2010. Gender trends in dental leadership and academics: a twenty-two-year observation. *J Dent Educ* 74:372-80.

Table 1

Differences between study design at Institution A and Institution B. These factors may have impacted some differences seen between the results for the two samples.

	Institution A	Institution B
Subject number	N = 67	N = 64
Male:Female Ratio	42:25	36:28
Average age	24 \pm 2	23 \pm 2.4
Semester	Fall semester	Spring semester
Exposure to online resources	No regular use of pre-recorded lectures	Pre-recorded lectures used throughout the curriculum
Curriculum	Anatomical sciences is integrated within a traditional curriculum	Entire curriculum integrated into organ systems
Learning anatomy	Face-to-face lectures and dissection laboratory	Pre-recorded lectures and dissection laboratory
Distribution of materials	Institution Learning Management System	Qualtrics

Item	Institution	Younger Female Mean (SD)	Younger Male Mean (SD)	Older Female Mean (SD)	Older Male Mean (SD)	Significance
1. Well-Organized	A	4.27* (.458)	4.37^ (.597)	3.57*^ (1.016)	4.16 (.688)	^0.011 *0.048
	B	4.13 (.957)	4.37 (1.025)	4.06 (.899)	4.53 (.516)	
2. Logical Presentation	A	4.40* (.507)	4.37^ (.597)	3.64*^ (1.082)	4.32 (.653)	^0.023 *0.025
	B	4.13 (1.025)	4.37 (1.025)	3.88 (.993)	4.67 (.488)	
3. Appropriate Terminology	A	4.47* (.516)	4.42^ (.607)	3.64*^ (1.082)	4.32 (.820)	^0.029 *0.028
	B	4.44 (.512)	4.00 (.894)	3.94 (.827)	4.47 (.516)	
4. Pleasant Voice	A	4.413^ (.352)	4.416* (.501)	3.64 (.745)	3.58*^ (.692) 3.93 (.704)	^0.042 *0.019 *0.010
	B	4.19* (.403)	3.81 (.911)	3.29* (.985)		
5. Teaching Methods	A	4.00* (.655)	4.00* (.667)	3.50* (.760)	3.74 (.562)	*0.020 *0.030
	B	4.06* (.443)	3.62 (1.025)	3.29* (.849)	4.0 (.655)	
6. Enthusiasm	A	3.53* (.743)	3.68* (.885)	2.93* (1.072)	3.16 (.834)	*0.015 *0.017
	B	3.56* (.814)	3.19 (1.223)	2.41* (1.176)	3.40 (1.056)	
7. Cared for my learning	A	3.67 (.617)	3.79 (1.032)	3.36 (.745)	3.63 (.761)	^0.009 *0.019
	B	3.63 (.719)	3.19* (.911)	3.12^ (1.111)	4.13*^ (.640)	
8. Made me feel accepted	A	3.93 (.799)	3.68 (.671)	3.36 (.745)	3.68 (.671)	^0.013 *0.037
	B	3.94* (.772)	3.56 (1.094)	3.12*^ (.781)	4.07^ (.704)	
9. Appropriate inflection	A	3.67 (.816)	3.47 (.964)	3.14 (.770)	3.74 (.653)	^0.031
	B	3.69 (.873)	3.25 (1.291)	2.82^ (1.131)	3.87^ (.743)	
10. Knowledgeable	A	4.00 (.926)	4.16 (.834)	3.86 (.535)	4.00 (.471)	^0.053
	B	4.38* (.500)	4.06 (.929)	3.47*^ (.874)	4.20^ (.676)	

Would take another course with this instructor	A	3.60 (.737)	3.74 (1.046)	3.36 (1.008)	3.58 (.838)	^0.053 *0.028
	B	4.00* (.632)	3.63 (1.025)	3.12*^ (1.054)	3.93^ (.704)	
Would choose to learn this way again	A	3.80 (1.424)	3.84 (.898)	3.50 (.855)	3.74 (1.098)	^0.012 *0.042
	B	4.06* (1.063)	4.00 (1.155)	3.00*^ (1.369)	4.27^ (.799)	

Table 2

One-way ANOVA, followed by Tukey's HSD test revealed significant differences in responses to items from the questionnaire for Institution A and Institution B. A ^ indicates that the significant difference in means is primarily due to gender. A * indicates that the significant difference in means is primarily due to age. For some items, significant differences were only noted at one institution