GENDER AND ETHNIC BIAS IN LETTERS OF RECOMMENDATION: CONSIDERATIONS FOR SCHOOL COUNSELORS

School counselors write letters of recommendation for students pursuing postsecondary education and help teachers and staff prepare for this task. Although letters of recommendation may impact admission and scholarship opportunities, research about equity and bias in letters is minimal as compared to standardized tests, teacher expectations, and grading practices. In this study, researchers analyzed letters of recommendation for evidence of gender and racial bias. Results demonstrate small but significant differences by gender and race in the average length of letters as well as the types of language used to describe students. This article discusses implications for school counselors.

Most universities require prospective students to submit one or more letters of recommendation when applying for undergraduate admission (The College Board, 2016). In many cases, the letters are used not only to determine admissibility, but also to determine eligibility for scholarships and honors invitations. Indeed, the National Association for College Admissions Counselors' (NACAC) *State of College Admissions Report* (2014) revealed that 70% of colleges and universities attribute considerable or moderate importance to letters of recommendation in their evaluation of candidates. Further, Kuncel, Kochevar, and Ones (2014) suggested this importance is evident for decades, where letters are the third most used predictor of college success after GPA and test scores.

One fundamental delivery mechanism within the ASCA National Model (American School Counselor Association [ASCA], 2012) is individual planning. In particular, high school counselors help students plan and prepare for postsecondary education. With respect to letters of recommendation, high school counselors play an especially large role. Counselors not only write their own letters of recommendation, but also provide guidance to teacher recommenders. The counselor competencies outlined by NACAC (2012) instruct counselors to "provide training, orientation, and consultation…to faculty, administrators, staff, and school officials to assist them in responding to the educational development and *precollege guidance*

Patrick Akos, Ph.D., is a professor at the University of North Carolina at Chapel Hill. Email: pta@unc.edu **Jennifer Kretchmar, Ph.D.**, is senior assistant director of admissions for research, also at the University of North Carolina at Chapel Hill.

and counseling needs of students" (p. 4). In general, counselors advise teachers to write letters demonstrating students' contributions in the classroom, while counselors themselves focus on students' interests outside the classroom and their contributions to the school and community (Von Bargen, 2014). Both teachers and counselors are charged to represent individual students and their institutions openly and honestly (NACAC, 2012). Finally, counselors also provide guidance to students about how to ask teachers for letters of recommendation.

Even within the 80% direct service recommended by ASCA (2012), the amount of time and resources counselors have at their disposal to devote to these specific tasks can vary significantly by school. According to NACAC's *State of College Admissions Report* (2014) and research by Clinedinst, Hurley, and Hawkins (2011), public high school counselors spend approximately 24% of their time (estimated as only 38 minutes per year with each student) on college access and admissions counseling compared to 52% for private high school counselors. And in 2013, 32% of public schools employed at least one counselor devoted exclusively to college counseling, compared to 71% of private schools (NACAC, 2014). Since individual planning is only one part of the school counselor role (e.g., responsive service, systems support), counselors are especially challenged by these labor-intensive tasks. In a recent *New York Times* article, for example, Harris (2014) reported that counselors with large caseloads often develop a stock recommendation letter that they tweak only slightly for each student. Counselors and teachers who are unable to get to know all of their students may be more susceptible to stereotypes and bias.

Because letters of recommendation have the potential to significantly impact educational opportunities for students, understanding the ways in which they might contain biases based on student characteristics like race and gender is important. Only one

study was located that explored bias in letters of recommendation written for undergraduate applicants. In 1989, LaCroix found that by relying too heavily on sex-stereotypical language, recommenders inadvertently shared both irrelevant and inaccurate information about students. We found no previous research, however, that investigated racial bias in letters of recommendation written for students applying to college, or examined the relationship between bias and outcomes like admission and scholarship selection. A great deal of attention has been given to other ways in which gender and racial bias operate as barriers to access and equity in

higher education—in standardized testing (e.g., Freedle, 2003), tracking of students toward or away from particular courses (e.g., Akos, Lambie, & Milsom, 2007), and grading practices (e.g., Azen, Bronner, & Gafni, 2002); this study extends our knowledge of factors impacting equity and access by investigating the presence of bias in letters of recommendation written for college admission.

BIAS IN LETTERS OF RECOMMENDATION

Gender
Although we could not locate research on bias in recommendation letters for undergraduate applicants, related studies documented gender bias in letters of recommendation for job applicants in line with Eagly's social role theory (Eagly, 1987; Eagly, Wood, & Deikman, 2000). This theory posits that gender beliefs arise from the specific social roles occupied by men and women, or more specifically, from the division of labor within society.

Because men and women are observed behaving differently, people develop different beliefs about what men and women can and should do. According to the theory, society perceives men to be agentic—dominant, assertive, competent, and competitive—while women are perceived to be communal—unselfish, friendly, emotional, and caring. Furthermore, because achievement is most closely associated with agency, and women are perceived to be communal, people find it more difficult to explain women's professional success (Valian, 1998). As a result, observers often attribute women's achievement to hard work rather than ability (Swim & Sanna, 1996; Valian, 1998).

> ## LETTERS ARE USED NOT ONLY TO DETERMINE ADMISSIBILITY, BUT ALSO TO DETERMINE ELIGIBILITY FOR SCHOLARSHIPS AND HONORS INVITATIONS.

When Trix and Psenka (2003) analyzed recommendation letters written for doctors applying to medical faculty positions, for example, they found that letters written for female applicants were shorter than letters written for male applicants, and that recommenders were more likely to describe female candidates in terms of their work ethic—or what they called "grindstone" adjectives (e.g., tireless, committed)—and male candidates in terms of talent and ability (e.g., genius, analytical). Schmader, Whitehead, and Wysocki (2007) replicated this study using letters of recommendation written for science faculty positions. Although they did not find significant differences in the length of letters written for men and women, or the use of ability and grindstone adjectives, their analysis did reveal that letters written for male candidates contained more standout adjectives—words like superb, wonderful, and magnificent—than letters written for female candidates. Furthermore, these standout adjectives positively correlated with ability words, and negatively correlated with grindstone adjectives,

suggesting that men were more likely to be described as having a superlative amount of natural ability. Similar bias and social role influences appear in the relationship between descriptors and hiring decisions. Madera, Hebl, and Martin (2009) found that women were more likely to be described in terms of communal adjectives (e.g., agreeable, sensitive, helpful), men in terms of agentic adjectives (e.g., directive, competent, independent), and that communal adjectives negatively correlated with hiring decisions.

EVEN AS CONSCIOUSLY HELD BELIEFS EVOLVE, RESEARCH SUGGESTS THAT MANY BIASES MAY BE IMPLICIT, EXISTING OUTSIDE OUR CONSCIOUS AWARENESS.

Much of the research on gender bias in letters of recommendation is grounded in Eagly's social role theory (Eagly, 1987; Eagly et al., 2000), described above. Beyond the general distinctions between agentic and communal, the overrepresentation of men in particular types of professions—mainly, math- and science-related disciplines—has led to more specific beliefs about men's and women's intellectual abilities in these domains (Wood & Eagly, 2010). In general, women are believed to be more creative and verbally skilled, and men more analytically and quantitatively skilled (Cejka & Eagly, 1999).

As social roles change, the theory suggests, so too will gender stereotypes. And indeed, social roles have changed a great deal in the past few decades. Women earned 58% of bachelor's degrees in 2006 compared to 43% in 1970 (National Center for Education Statistics [NCES], 2007), for example, and high school girls are now as likely as high school boys to take calculus (NCES, 2004). Similarly, women's participation in the workforce nearly doubled in the last half century (e.g., Wood & Eagly, 2010). Changes in gender stereotypes, however, may not be keeping

pace. Consistent with the notion of cultural lag—the idea that beliefs change more slowly than roles—much research indicates that gender stereotypes have remained stable over time (Bergen & Williams, 1991; Feingold, 1994; Lueptow, Garovich-Szabo, & Lueptow, 2001). Furthermore, some roles are not changing as fast as others; despite women's increased representation in higher education and the workforce, they are still less likely to pursue math and science majors in college, and they comprise just 15% of employed engineers and 25% of employed physicians (e.g., National Girls Collaborative Project, 2016).

Even as consciously held beliefs evolve, an emerging area of research suggests that many biases may be implicit, existing outside our conscious awareness. Using the Implicit Association Test, Rudman and Glick (1999) found an unconscious association between men and agency, and women and communal adjectives, regardless of consciously held beliefs. Lemm and Banaji (1999) referred to the gap between conscious and unconscious beliefs as "socially problematic" (p. 225), since people who do not explicitly endorse gender stereotypes might still be guided by implicitly held beliefs. Indeed, they later showed that an implicit men-agentic/women-communal stereotype predicted job discrimination against a female applicant, whereas a self-report measure did not (Rudman & Glick, 2001). Nosek, Banaji, and Greenwald (2002) found similar evidence for an implicit association between math/science and male and language/arts and female, which was predictive of both performance in math and science and preferences for studying in those fields.

Race

Racial bias in recommendation letters, by contrast, has been virtually ignored in the research literature. Although Bouton (1995) examined how recommenders from various ethnic and cultural backgrounds approach the task of letter writing differently, we could not locate empirical research that investigates whether or not the ethnicity or race of the student and/or job applicant biases the recommender. Research has demonstrated, however, that teachers are susceptible to negative stereotypes about ethnic groups (McKown & Weinstein, 2008; Reyna, 2008), and that teacher expectations may play a role in perpetuating the achievement gap between Caucasian and African American and Latino students (Ferguson 1998; McKown & Weinstein, 2008).

In research on racial stereotypes more generally, Devine and Elliot (1995) discovered that most White Americans have knowledge of the Black stereotype—one that is stable and highly negative—but that only a minority actually endorse the stereotypes. Among low-prejudiced individuals, for example, 88% selected low in intelligence as a trait of the black stereotype, though only 6% selected it as a personal belief. Among high-prejudiced individuals, on the other hand, 82% selected "lazy" as a trait of the Black stereotype, and 72% endorsed it as a personal belief. Nosek and colleagues (2007) documented a persistent implicit association between African American and "bad" and White American and "good," even among people who hold explicit egalitarian beliefs and attitudes.

Importantly, Devine and Elliot (1995) also discovered that high- and low-prejudiced individuals possess the same *knowledge* of the black stereotype; rejection of the stereotype, however, does not immediately eradicate it from one's mind. It is still a "well-organized, frequently activated knowledge structure" (p. 1140). In fact, it can be activated outside of one's conscious control, automatically, by the presence of certain stimuli resulting in prejudice-

like feelings and thoughts in low- and high-prejudiced individuals alike. Only when we have the time and cognitive resources can we avoid automatic activation of stereotypes.

To investigate the presence of gender and racial bias in letters of recommendation, we examined several primary research questions. First, we evaluated the relationship between the length of letters of recommendation written for students applying to college and applicant gender and race. Next, we focused on the descriptors used in the letters, examining the relationship between the types of adjectives used to describe applicants and applicant gender and race. We also included exploratory variables and interaction terms in our model to determine if the gender of the recommender or the subject taught by the recommender mediated our findings. Finally, we explored the relationship between word count, each of our descriptors, and educational outcomes, specifically, probability of admission and probability of selection for honors. Although we could have reasonably hypothesized a positive relationship between each variable and the outcomes, it is not entirely clear which characteristics universities most value.

METHOD

Participants
We analyzed letters of recommendation written by teachers for undergraduate applicants at a selective, Southeastern, public university with "very high research activity" (as indicated by the Carnegie Classification of Institutes of Higher Education). The original sample consisted of 5,255 letters of recommendation written for the same number of students applying for undergraduate admission. Nine percent ($n = 463$) of the letters of recommendation were excluded from the analysis due to missing data. The final sample consisted of 4,792 letters of recommendation.

Of the applicants in the final sample, 60% ($n = 2,859$) were female and

40% ($n = 1,933$) were male. In terms of ethnicity, 68% ($n = 3,278$) identified as White, 15% ($n = 713$) as Asian, 9% ($n = 436$) as Black or African American, 7% ($n = 316$) as Hispanic or Latino, 1% ($n = 47$) as Native American or Alaskan Native, and < 1% ($n = 2$) as Native Hawaiian or Other Pacific Islander. Students who identified with one of the minority groups that have been disproportionately underrepresented in the undergraduate student population—specifically, Black or African American, Hispanic or Latino, and Native American or Alaskan Native—were coded as *underrepresented* minority students; they constituted 17% ($n = 799$) of the sample. All other students (83%, $n = 3,993$) were coded as *nonunderrepresented* students.

The letters of recommendation were written by 4,126 recommenders, representing public and private high school teachers in a variety of disciplines. Thirteen percent ($n = 540$) submitted a letter of recommendation for more than one student in the sample. For the purpose of the analysis, recommenders were treated as if they were independent. The recommenders were 59% ($n = 2,440$) female and 41% ($n = 1,686$) male; 35% taught math/science-related disciplines, 65% taught subjects in the humanities. The race of the recommenders was unknown.

Procedure
After receiving institutional review board approval, we obtained letters of recommendation from the Office of Undergraduate Admission. Letters not submitted electronically as Microsoft Word documents were excluded from the study. The computer text analysis program called Linguistic Inquiry and Word Count, or LIWC (Pennebaker

et al., 2001) was used to analyze the letters. LIWC counts the total number of words in a document, as well as the percentage of words that fall into one of 74 predefined data dictionaries, or any number of user-defined data dictionaries. LIWC has been validated (Madera, Hebl, & Martin, 2009; Pennebaker et al., 2001) and is a widely used text analysis program.

Measures
This study borrowed the data dictionaries for grindstone, ability, and standout adjectives from Schmader et al. (2007) and the dictionaries for agentic and communal adjectives from Madera et al. (2007). Schmader and colleagues' dictionaries derived from the theoretical work of Trix

> WE EXPLORED THE RELATIONSHIP BETWEEN WORD COUNT, EACH OF OUR DESCRIPTORS, AND EDUCATIONAL OUTCOMES, SPECIFICALLY, PROBABILITY OF ADMISSION AND PROBABILITY OF SELECTION FOR HONORS.

and Psenka (2003), and Madera and colleagues' dictionaries derived from Eagly's social role theory (Eagly et al., 2000). To supplement their dictionaries, Madera et al. (2007) also used Wordnet (www.wordnet.princeton.edu), an online lexical reference tool that searches for synonyms, and the standout and grindstone dictionaries developed by Schmader et al. (2007). The complete list of words and word stems for each dictionary is provided in the Appendix. The present study also utilized LIWC's 186-word predefined dictionary for achievement.

With these five LIWC-generated measures, the study also included three control variables. These included the applicant's combined score on the Critical Reading/Verbal and Math (CR/V + M) portion of the SAT, a rating of the academic rigor of the applicant's high school curriculum, and a rating of the applicant's academic performance (i.e., grades). The ratings for program and performance were on a 1-10 Likert

scale (least rigorous to most rigorous), and were coded by application evaluators, with ratings relative to the applicant pool. A performance rating of 10 represented the highest possible score (i.e., best grades) while a performance rating of 1 represented the lowest possible score (i.e., poorest grades). For example, a student with a 10 in program and a 10 in performance might be a student pursuing the International Baccalaureate degree, earning straight A's. A student with a one in program and one in performance might be earning C's and D's in college preparatory classes.

students by including the ratings for high school program and performance, as well as the SAT (CR/V + M) score. In order to calculate effect sizes, we obtained modeled means for a *typical* applicant, defined as an applicant with a program rating of 6, performance rating of 7, and SAT (CR/V + M) score of 1200.

Our second research question was tested using a series of logistic regressions, modeling the proportion of words in each category (e.g. grindstone, ability, achievement, etc.) as a function of four main effects (gender of the applicant, gender of the recom-

between the gender of the applicant and the gender of the recommender. Analyses revealed that for male applicants, female recommenders wrote longer letters than did male recommenders, whereas for female applicants, the length of letters written by male and female recommenders was not significantly different. The length of letters did not significantly vary by ethnicity or gender of the applicant.

In order to calculate effect sizes, we obtained modeled means for a *typical* applicant, defined as an applicant with a program rating of 6, performance rating of 7, and SAT (CR/V + M) score of 1200. Under the model, letters written for the typical applicant by female recommenders were 14 words longer, on average, than letters written by male recommenders. When the applicant was male, however, letters written by female recommenders were 25 words longer, on average, than letters written by male recommenders; when the applicant was female, the letters were only 3 words longer. The gap in the average length of letters of recommendation written by humanities and math/science teachers was the largest; for the typical applicant, letters written by humanities teachers ($M = 415$) were 44 words longer than those written by math/science teachers ($M = 371$). Letters written for female applicants were 8 words longer on average than those written for male applicants, though this difference was not significant.

Our second research question was tested using a series of logistic regressions, modeling the proportion of words in each category (e.g., grindstone, ability, achievement, etc.) as a function of four main effects (gender of the applicant, gender of the recommender, ethnicity of the applicant, and course type) and three interactions terms (interaction between course type and the gender of the applicant, interaction between gender of the applicant and gender of the recommender, and interaction between the gender and ethnicity of the applicant). We again controlled for the academic credentials of the students by including the ratings

> LETTERS OF RECOMMENDATION FOR FEMALE APPLICANTS CONTAINED A HIGHER PROPORTION OF GRINDSTONE ADJECTIVES THAN LETTERS OF RECOMMENDATION WRITTEN FOR MALE APPLICANTS.

Two additional measures were included based on recommender characteristics: gender of the recommender and the type of course taught by the recommender. Gender was coded female or male and course type was coded math/science or humanities based on the course description provided by the applicant. Examples of math/science related disciplines included calculus, environmental science, biology, and chemistry; examples of humanities included English, Spanish, geography, history, and psychology.

Data Analysis
Using a 2 X 2 X 2 X 2 log linear model, we modeled word count as a function of four main effects (gender of the applicant, gender of the recommender, ethnicity of the applicant, and course type) and three interactions terms (interaction between course type and gender of the applicant, interaction between gender of the applicant and gender of the recommender, and interaction between the gender and race of the applicant). We controlled for the academic credentials of the

mender, ethnicity of the applicant, and course type) and three interactions terms (interaction between course type and the gender of the applicant, interaction between gender of the applicant and gender of the recommender, and interaction between the gender and ethnicity of the applicant). We again controlled for the academic credentials of the students by including the ratings for high school program and performance, as well as the SAT (CR/V + M) score.

RESULTS

Descriptive means and bivariate correlations for the variables are presented in Table 1. With regard to total word count, the length of letters significantly varied by gender of the recommender and course type; specifically, female recommenders wrote longer letters than male recommenders, and letters of recommendation written by humanities teachers were longer than those written by math/science teachers. There was also a significant interaction

for high school program and performance, as well as the SAT (CR/V + M) score. Results are detailed in Table 2.

Letters of recommendation for female applicants contained a higher proportion of grindstone adjectives than letters of recommendation written for male applicants. For our average female applicant, .98% of all words were grindstone adjectives. In a letter of recommendation 400 words in length, this represents 3.92 words. For the similarly credentialed male applicant, .90% of words were grindstone adjectives, or 3.60 words in a 400-word letter. The gender of the recommender was also significantly associated with the proportion of grindstone adjectives; for similarly credentialed applicants, female recommenders used more grindstone words than male recommenders. In a 400-word letter written by a female recommender, the average number of grindstone words was 4.00; in a similar length letter written by a male recommender, the average number of grindstone words was 3.56. In other words, female recommenders used 12% more grindstone adjectives than male recommenders.

The gender of the applicant did not significantly predict ability adjectives, communal adjectives, agentic adjectives, standout adjectives, or achievement words. There was, however, a significant interaction between gender of the applicant and course type on ability adjectives. Recommenders teaching science or math were less likely to use ability adjectives when describing female applicants than recommenders teaching the humanities; there was no significant difference in the proportion of ability adjectives used by math/science and humanities teachers when describing male applicants. The effect sizes were very small; the proportion of ability words used by math/science recommenders for female applicants was .62%, and by humanities recommenders .72%, a difference of one-half word in a 400-word document. The difference of one half word, however, from 2.48 words to 2.88 words, represents a 16% increase in

TABLE 1 MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF VARIABLES USED IN MODELS PREDICTING WORD COUNT AND ADJECTIVE USE

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Gender of applicant	--	--	--												
2. Gender of recommender	--	--	0.04**	--											
3. Ethnicity of applicant	--	--	-0.06**	-0.01	--										
4. % Agentic adjectives	0.007	0.006	-0.01	0.00	0.02	--									
5. % Communal adjectives	0.005	0.004	-0.01	-0.01	0.06*	-0.03*	--								
6. % Ability adjectives	0.008	0.006	0.00	0.02	-0.04**	0.09*	-0.02	--							
7. % Grindstone adjectives	0.010	0.007	-0.06**	-0.09**	-0.01	0.02	0.03*	-0.03	--						
8. % Standout adjectives	0.006	0.005	-0.04**	0.07**	-0.01	0.03*	-0.04*	0.08*	0.01	--					
9. % Achievement words	0.049	0.017	-0.03	-0.02	-0.05**	0.21**	-0.16**	0.10**	0.13**	0.06**	--				
10. Total word count in letter	419	192	-0.03	-0.03*	-0.07*	-0.01	-0.15*	-0.05*	-0.11*	-0.04*	0.77**	--			
11. SAT (CR/V + M)	1297	148	0.13**	0.02	-0.33*	-0.03	-0.08*	0.09*	-0.08*	0.05*	0.11**	0.17*	--		
12. Difficulty of high school program	6.12	2.66	0.01	0.00	-0.11*	-0.03	-0.06*	0.02	-0.04*	0.05*	0.10**	0.13*	0.43*	--	
13. High school grades	6.86	2.42	-0.08**	0.04**	-0.22*	0.02	-0.08*	0.01	-0.04*	0.09*	0.14**	0.14*	0.44*	0.33*	--

Note: $*p < .05, **p < .01$

TABLE 2 REGRESSION COEFFICIENTS FOR MODELS PREDICTING WORD COUNT, AND GRINDSTONE, ABILITY, AGENTIC, COMMUNAL, ACHIEVEMENT, AND STANDOUT ADJECTIVES

Dependent Variable	Word Count			Grindstone Words			Ability Words			Agentic Words			Communal Words			Standout Words			Achievement Words		
	β	SE	χ²	β	SE	χ²	β	SE	χ²	β	SE	χ²	β	SE	χ²	β	SE	χ²	β	SE	χ²
1. Gender of applicant	-0.01	0.01	1.05	-0.04	0.02	7.73	0.01	0.02	0.64	-0.01	0.02	0.16	-0.01	0.02	0.25	-0.01	0.02	0.36	-0.01	0.01	0.84
2. Gender of Recommender	-0.02	0.01	6.89	-0.06	0.01	32.30	0.01	0.01	0.35	0.01	0.01	0.62	-0.01	0.01	0.72	0.06	0.01	31.75	0.00	0.01	0.03
3. Interaction variables 1 & 2	-0.01	0.01	4.70	0.01	0.01	2.10	0.00	0.01	0.08	0.01	0.01	0.26	-0.01	0.01	0.35	0.01	0.01	1.80	0.01	0.01	2.53
4. Ethnicity of applicant	0.00	0.01	0.01	-0.05	0.02	9.58	0.00	0.02	0.06	0.01	0.02	0.18	0.05	0.02	9.40	0.01	0.02	0.43	-0.01	0.01	1.20
5. Course type	0.06	0.01	63.72	-0.06	0.01	35.69	0.05	0.01	17.86	0.04	0.01	10.93	0.00	0.01	0.09	0.01	0.01	1.10	-0.04	0.01	39.18
6. Interaction variables 1 & 4	0.01	0.01	1.06	0.00	0.01	0.06	0.02	0.02	0.93	-0.02	0.02	1.75	0.02	0.02	1.12	0.01	0.02	0.73	-0.02	0.01	4.25
7. Interaction variables 1 & 5	0.00	0.01	0.10	-0.02	0.01	2.51	-0.03	0.01	4.90	-0.01	0.01	0.31	0.00	0.01	0.03	0.00	0.01	0.01	-0.01	0.01	8.10

Note: Applicant gender was coded as female = -1 and male = 1. Ethnicity was coded as underrepresented minority = 1, non-underrepresented minority= -1. Course type was coded as science/math = -1 and humanities = 1. Results are after controlling for SAT (CR/V + M), the difficulty of a high school curriculum, and high school grades. Values in bold are statistically significant, $p < .05$.

the frequency of ability words used by humanities recommenders to describe female applicants compared to math/science recommenders describing female applicants.

Similarly, results showed a significant interaction between gender of the applicant and course type on achievement words. Recommenders teaching math or science were more likely to use achievement words when describing female applicants than recommenders teaching the humanities; the proportion of achievement words used by math/science recommenders for female applicants was 5.07% compared to 4.89% for humanities recommenders. In a 400-word letter, this represent 20.28 and 19.56 achievement words, respectively, or an increase of 3.7%. Recommenders teaching math or science were also more likely to use achievement words when describing male applicants compared to recommenders teaching the humanities; the proportion of achievement words used by math/science recommenders for male applicants was 5.15% compared to 4.69% for humanities recommenders. In a 400-word letter, this represents 20.6 and 18.76 achievement words respectively, or an increase of 9.8%.

Finally, the gender of the recommender significantly predicted standout adjectives, with male recommenders using more superlatives than female recommenders, but the gender of the recommender was not significantly associated with ability adjectives, agentic adjectives, communal adjectives, or achievement words. The proportion of standout adjectives used by male recommenders was .61%, and by female recommenders .54%, which represents a 13% increase from 2.16 to 2.44 words in a 400-word letter of recommendation.

Although the ethnicity of the applicant did not significantly predict standout adjectives, achievement words, or ability adjectives, the ethnicity of the applicant was significantly associated with grindstone adjectives and communal adjectives. Compared to letters of recommendation written for nonminority applicants, letters for

underrepresented minority applicants contained a lower proportion of grindstone adjectives (.90% vs. .99% for nonminority applicants) and a higher proportion of communal adjectives (.50% vs. .45% for nonminority applicants). This represents a 10% decrease in achievement-related words and an 11% increase in communal adjectives. The results also showed a significant interaction between the gender and the race of the applicant for achievement words. Fewer achievement words were used by recommenders to describe underrepresented male students relative to underrepresented female students, while there was no difference in the number of achievement words used by recommenders to describe nonminority male and nonminority female students. The effect sizes were very small; the proportion of achievement words used to describe underrepresented male students was 4.81% compared to 5.01% for underrepresented female applicants, a difference of one achievement-related word in a 400-word document, or a 4% decrease.

Finally, with regard to our exploratory inquiry, only total word count and agentic adjectives were significantly associated with probability of admission and honors selection (See Table 3). No other adjectives were significantly associated with either outcome. Under the model, typical applicants with letters of recommendation averaging 200 words in length had a 38% probability of admission; equally credentialed applicants with letters of recommendation averaging 600 words in length had a 48% probability of admission. To calculate effect sizes for honors selection, we obtained modeled means for a top applicant in the pool, defined as an applicant with a program rating of 8, performance rating of 8, and SAT (CR/V + M) score of 1400. Under the model, top applicants with letters of recommendation averaging 200 words in length had a 5.4% probability of being selected for honors; equally credentialed top applicants with letters of recommendation averaging 600 words in length had a 11.3% probability of being selected

TABLE 3 MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS FOR VARIABLES USED IN MODELS PREDICTING ADMISSION AND HONORS/MERIT SELECTION

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Admission probability	0.41	0.49	--													
2. Honors/merit probability	0.05	0.22	0.27*	--												
3. Ethnicity of applicant	--	--	0.04*	0.01	--											
4. Residency of applicant	--	--	-0.37*	-.09*	-0.02	--										
5. % Agentic adjectives	0.007	0.006	-0.01	0.02	0.02	-0.02	--									
6. % Communal adjectives	0.005	0.004	-0.09*	-0.03*	0.06*	0.08*	-0.03*	--								
7. % Ability adjectives	0.008	0.006	-0.03*	0.02	-0.04**	0.09*	0.09*	-0.02	--							
8. % Grindstone adjectives	0.010	0.007	-0.08*	-0.05*	-0.01	0.05*	0.02	0.03*	-0.03	--						
9. % Standout adjectives	0.006	0.005	0.03	0.02	-0.01	0.05*	0.03*	-0.04*	0.08*	0.01	--					
10. % Achievement words	0.049	0.017	0.12**	0.13**	-0.05**	-0.01	0.21**	-0.16**	0.10**	0.13**	0.06**	--				
11. Total word count in letter	419	192	0.18*	-0.18*	-0.07*	-0.06*	-0.01	-0.15*	-0.05*	-0.11*	-0.04*	0.77**	--			
12. SAT (CR/V + M)	1297	148	0.37*	0.27*	-0.33*	0.19*	-0.03	-0.08*	0.09*	-0.08*	0.05*	0.11**	0.17*	--		
13. Difficulty of high school program	6.12	2.66	0.45*	0.22*	-0.11*	-0.05*	-0.03	-0.06*	0.02	-0.04*	0.05*	0.10**	0.13*	0.43*	--	
14. High school grades	6.86	2.42	0.44*	0.22*	-0.22*	0.03*	0.02	-0.08*	0.01	-0.04*	0.09*	0.14**	0.14*	0.44*	0.33*	--

Note: *$p < .05$, **$p < .01$

for honors. Similarly, a top applicant whose 400-word letter of recommendation contained 12 agentic adjectives had a 10.3% probability of honors selection; equally credentialed top applicants whose 400 word letter of recommendation contained 4 agentic adjectives had a 5.2% probability of being selected for honors.

DISCUSSION

On the whole, and with respect to the particular variables of interest in this study, letters of recommendation written for male and female applicants were more similar than they were different. Letters of recommendation written for minority and nonminority applicants, with respect to the variables of interest in this study, were also more similar than they were different. As one of the first empirical tests of bias in letters of recommendation written for college applicants, these results are promising at first glance. But there were also some significant differences in the way male and female and minority and nonminority applicants were described—especially with respect to some of the exploratory variables included in the study—that should give pause and warrant further investigation.

Although length does not directly address content or bias, previous research found gender differences (Trix & Psenka, 2003) and it may serve as a proxy for bias. We found a positive relationship to letter length. Specifically, letters written for male applicants by female recommenders were longer than letters written for male applicants by male recommenders, whereas letters written for female applicants were the same length, regardless of the gender of the recommender. In addi-

tion, recommenders teaching humanities wrote longer letters, on average, than recommenders teaching math and science. Because we could not locate previous research that explored the interaction between gender of the recommender and applicant, we have no context into which to place these findings; it may be that the relationship between gender of recommender and gender of application is mediated by the subject taught by the recommender, suggesting a three-way interaction. Future research should explore the way multiple characteristics of recommenders and students might intersect to produce potential bias.

We also found a positive relationship between letter length and the recommender's evaluation; longer letters typically suggested stronger endorsements. Specifically, word count significantly and positively associated with both probability of admission and honors selection; for two students presenting equal academic credentials,

the student with the longer letter of recommendation had a greater chance of being admitted and selected for honors than the student with a shorter letter of recommendation. The specific combination of results—the significant association between length and educational outcomes paired with the lack of association between our adjectives and outcomes—might suggest that it is the content of longer letters, the inclusion of specific examples that illustrate characteristics of students, rather than just the descriptors themselves, that most persuade those reading the letters of recommendation. It also suggests that students should take care in choosing recommenders who know them well enough to provide specific examples of their abilities and achievement, rather than just describing them in more general terms.

Although the adjectives explored in this study were not significantly correlated with educational outcomes, some significant gender differences emerged in the frequency with which the descriptors were used. Recommenders used slightly more grindstone adjectives to describe female applicants than male applicants. Math and science recommenders used slightly fewer ability words to describe female applicants than humanities recommenders, but math/science and humanities

TABLE 4

REGRESSION COEFFICIENTS FOR MODELS PREDICTING ADMISSION AND HONORS/MERIT SELECTION

Dependent Variable	Admission			Honors/Merit		
	β	SE	χ²	β	SE	χ²
1. Total word count	0.001	0.000	16.38	0.00	0.00	36.74
2. Agentic adjectives	-1.01	9.24	0.01	37.31	17.46	4.56
3. Communal adjectives	-6.04	12.12	0.25	25.16	24.38	1.06
4. Ability adjectives	-9.69	8.27	1.37	21.94	14.46	2.30
5. Grindstone adjectives	-5.58	8.01	0.49	-0.16	15.17	0.00
6. Standout adjectives	-2.89	10.53	0.08	-14.03	19.84	0.50
7. Achievement words	-2.650	3.50	0.25	-7.84	6.70	1.37

Note: Applicant gender was coded as female = -1 and male = 1. Ethnicity was coded as underrepresented minority = 1, non-underrepresented minority = -1. Course type was coded as science/math = -1 and humanities = 1. Results are after controlling for SAT (CR/V + M), the difficulty of a high school curriculum, and high school grades. Values in bold are statistically significant, $p < .05$.

FEWER ACHIEVEMENT WORDS WERE USED BY RECOMMENDERS TO DESCRIBE UNDERREPRESENTED MALE STUDENTS RELATIVE TO UNDERREPRESENTED FEMALE STUDENTS.

recommenders used ability words with the same frequency to describe male applicants. In some ways, these results partially support social role theory and findings by Trix and Psenka (2003) who found that recommenders were more likely to describe female candidates in terms of their work ethic—or what they called "grindstone adjectives"—and male candidates in terms of talent and ability. The female candidates in their study were applying for jobs in the sciences, fields still traditionally dominated by men (Schmader, Whitehead, & Wysocki, 2007). If gender stereotypes are grounded in social roles, as social role theory suggests (Eagly et al., 2000), we might expect to find more bias in letters of recommendation when the position being recommended for is perceived as a gender stereotypical role. The role of student, on the other hand, is arguably more gender neutral. The role of student did not serve as a gender cue until a female student was paired with a recommender writing about the student's achievement and/or ability in math and science.

In addition to gender, some significant differences emerged in the frequency with which particular descriptors were used by recommenders to describe underrepresented applicants. For example, recommenders used slightly fewer grindstone adjectives (e.g., thorough) to describe underrepresented applicants. Additionally, achievement words were also significantly associated with the ethnicity of the applicant, but a significant interaction with applicant gender also emerged. Recommenders used fewer descriptors of prior accomplishment for male underrepresented candidates; results showed no differences in the proportion of achievement descriptors for female underrepresented candidates. While these data are not conclusive, racial bias could impact male students disproportionately. One of the more surprising outcomes of the study was the lack of association between the descriptors and probability of admission and selection for honors; in some cases, the associa-

tions were even negative, though not significant. Why might adjectives that describe positive traits of applicants like work ethic, talent, and achievement show little relationship to these outcomes? One possible explanation is letter inflation. Kuncel et al. (2014) suggested letters are often superficially positive in tone, making it more

difficult to make distinctions among candidates. Perhaps letters that are effusive in praise—especially when they do not contain specific examples to back up their claim—might actually work against a student (Knouse, 1983). A second possible explanation is the nature of the applicant pool. That is, letters of recommendation may influence decisions differently for people applying for a single open position like a job, rather than applicants applying for one of thousands of open seats in a first-year class. Understandably, readers may be able to pay more attention to the details of letters of recommendation when sifting through only a handful than when presented with thousands. Finally, perhaps the impact of letters of recommendation is greater for students on the margins; for students who already present a flawless profile elsewhere, the specifics of his or her letters might weigh less in the final decision. For students whose grades are spotty, or who do not elaborate on their activities outside the classroom, perhaps a teacher's or counselor's recommendation could tip a decision one way or the other.

Implications for School Counselors
It is important for school counselors and teachers to understand that bias is not inevitable. Even unconscious stereotypes can be deactivated by using social cues and individuating information, and by limiting time constraints and other cognitive demands (Devine,

Plant, & Buswell, 2000). In other words, people can inhibit stereotyping when they are motivated to do so and when they have sufficient cognitive resources (Kite, Deaux, & Haines, 2007). We believe school counselors are in an ideal position to confront their own bias, and help teachers confront their bias as well.

> ## LETTERS OF RECOMMENDATION WRITTEN FOR MALE AND FEMALE APPLICANTS WERE MORE SIMILAR THAN THEY WERE DIFFERENT.

Writing a letter for a student they do not know well enough may only harm the student in the long run. These data suggest students should ask for letters from teachers with whom they have developed a relationship. If a counselor or teacher has to say yes, they should ask students to provide more specific behavioral information about themselves when needed (e.g., via student interest forms, etc.) and gather information about the student from other teachers and counselors.

Specific to content of letters, a focus on motivation, drive and persistence— or conscientiousness—should be primary because cognitive characteristics are represented in other measures in admissions. These characteristics should be familiar to school counselors as they are prominent in the ASCA mindsets and behaviors in the foundation of the ASCA National Model. For example, self-confidence (mindset) and creativity, self-discipline, and adaptability (behaviors) provide context not available in traditional measures. When describing students, school counselors and teachers should use concrete examples and not just adjectives. Previous research hints to the benefits of letter length; longer letters may be more effective because they provide specific examples. Simply describing a student as "intelligent" is not as effective as providing a more three-dimensional account—e.g., "Jane not only has an unmatched ability to recall specific facts, but she

makes connections between ideas and events in a way I have rarely witnessed in a student of her age."

To further consider language, when providing examples that describe a student in terms of their work ethic and natural ability, teachers or counselors should pause and ask themselves how much they know about the student in terms of each of these characteristics. While ASCA mindsets and behaviors can be powerful descriptors, teachers do not always have evidence to attribute classroom success to one or the other—they might not know how many hours a student studies at night vs. how naturally talented they are in math, for example. We all tend to explain success differently for different groups of people; recommenders should make sure that examples, and the balance between the examples they provide, do not reflect these biases (e.g., to communicate a student's natural affinity for physics, a letter should not primarily describe how hard he works). Additional sources (e.g., Jones, 1990) for guidance on writing letters of recommendations expand on these and guide larger policy and procedures.

Limitations and Future Research

Like most research, conclusions are specific to or limited by sample and methodology choices. Our findings relate to bias as measured by letter length and adjective use. Neither can capture bias in a comprehensive way. Further, our sample of letters was for applicants at one public, research-intensive university in the Southeast. Future research should utilize broader samples or letters of recommendation and consider varied methods (e.g., content analyses, direct survey, or interview of writers of letters or campus admission committees) to explore bias. In particular, bias around STEM areas for female students is ripe for inquiry. This study is only one possible line of research on the important school counselor task of college access/admissions counseling and individual planning. Futher research can explore guidance on personal statements and

the most effective ways to deliver services amidst challenging school counselor role demands. Recognition of the social justice aim of school counseling is especially important in high-poverty schools. The gender and racial composition of a school may mediate bias and best practice.

Data from this study suggest that bias in letters of recommendation may not only be influenced by social roles and stereotypes, but also may play a role in their continued existence. School-based counseling makes a distinct and substantial contribution to the college enrollment and destinations of low SES populations (Belasco, 2013). It is incumbent upon school counselors to take action to increase the cultural competence in a school and prevent bias in this important role in college access. There is no school professional more important, especially in disadvantaged populations, than the high school counselor (Belasco, 2013). ■

REFERENCES

Akos, P., Lambie, G., & Milsom, A. (2007). Early adolescents' aspirations and academic tracking: An exploratory investigation. *Professional School Counseling, 11*, 57-64. doi:10.5330/PSC.n.2010-11.57

American School Counselor Association. (2012). *ASCA National Model: A framework for school counseling programs*. Alexandria, VA: Author.

Azen, R., Bronner, S., & Gafni, N. (2002). Examination of gender bias in university admissions. *Applied Measurement in Education, 15*(1), 75-94. doi:10.1207/S15324818AME1501_05

Belasco, A. (2013). Creating college opportunity: School counselors and their influence on postsecondary enrollment. *Research in Higher Education, 54*, 781-804. doi:10.10007/s111620013-9297-4.

Bergen, D. J., & Williams, J.E. (1991). Sex stereotypes in the United States revisited: 1972-1988. *Sex Roles, 24*, 413-423.

Bouton, L. F. (1995). A cross-cultural analysis of the structure and content of letters of reference. *Studies in Second Language Acquisition, 17*, 211-244. doi:10.1017/S0272263100014169

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin, 25*, 413-423. doi:10.1177/0146167299025004002

Clinedinst, M., Hurley, S., & Hawkins, D. (2011). *State of college admission 2011*. Washington, DC: National Association for College Admissions.

The College Board. (2016). How to get a great letter of recommendation. Retrieved from https://bigfuture.collegeboard.org/get-in/your-high-school-record/how-to-get-a-great-letter-of-recommendation.

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Society for Personality and Social Psychology, 21*(11), 1139-1150. doi:10.1177/01461672952111002

Devine, P. G., Plant, E. A., & Buswell, B. N. (2000). Breaking the prejudice habit: progress and obstacles. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 185-208). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.

Eagly, A. H., Wood, W., & Diekman, A. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123-174). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*, 429-456. doi:10.1037/0033-2909.116.3.429

Ferguson, R. F. (1998). Teachers' perceptions and expectations and the black-white test score gap. In C. Jencks and M. Phillips (Eds.), *The black-white test score gap* (pp. 273-317). Washington, DC: Brookings Institution Press.

Freedle, R. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review, 73*(1), 1-44. doi:10.17763/haer.73.1.8465k88616hn4757

Harris, E. A. (2014, December 26). Little guidance: 500 students per counselor. *New York Times*, pp. A1-A3.

Jones, D. (1990). College letters of recommendation: A staff development opportunity. *School Counselor, 90*, 153-156.

Kite, M. E., Deaux, K., & Haines, E. L. (2007). Gender stereotypes. In F. L. Denmark & M. A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories* (pp. 205-236). Westport, CT: Praeger Publishers.

Knouse, S. B. (1983). The letter of recommendation: Specificity and favorability of information. *Personnel Psychology*, *36*, 331-341. doi:10.1111/j.17446570.1983.tb01441.x

Kuncel, N., Kochevar, R., & Ones., D. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment, 22*(1), 101-108. doi:10.1111/ijsa.12060

LaCroix, P. (1985). Sex in recs: Gender bias in recommendation writing. *Journal of College Admissions*, *108*, 24-26.

Lemm, K. & Banaji, M. R. (1999). Unconscious attitudes and beliefs about women and men. In U. Pasero & F. Braun (Eds.), *Perceiving and performing gender* (pp. 215-233). Opladen, Germany: Westdutscher Verlag.

Lueptow, L. B., Garovich-Szabo, L., & Lueptow, M. B. (2001). Social change and the persistence of sex typing: 1974-1997. *Social Forces*, *80*, 1-36. doi:10.1353/sof.2001.0077

Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology*, *94,* 1591-1599. doi:10.1037/a0016539

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap*. Journal of School Psychology*, *46*, 235-261. doi:10.1016/j.jsp.2007.05.001

National Association for College Admissions Counselors. (2012). *Fundamentals of college admissions counseling* (3rd ed.). Arlington, VA: Author.

National Association for College Admissions Counselors. (2014). *State of college admissions report*. Arlington, VA: Author.

National Center for Education Statistics. (2004). Trends in educational equity of girls and women. Retrieved from http://nces.ed.gov/pubs2005/equity/Section4.asp doi:10.1177/0957926503014002277

National Center for Education Statistics. (2007). Historical summary of faculty, students, degrees, and finances in degree-granting institutions: Selected years, 1869-70 through 2005-06*.* Retrieved from https://nces.ed.gov/programs/digest/d07/tables/dt07_178.asp

National Girls Collaborative Project. (2016). *The state of girls and women in STEM*. Retrieved from https://ngcproject.org/sites/default/files/ngcp_the_state_of_girls_and_women_in_stem_2016_final.pdf

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = Male, Me = Female, Therefore Math <> Me. *Journal of Personality and Social Psychology*, *83*, 44-59.

Nosek, B .A., Symth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., … Banaji, M.R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 1-53. doi:10.1080/10463280701489053

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC 2001): A computerized text analysis program.* Mahwah, NJ: Erlbaum.

Reyna, C. (2008). Ian is intelligent but Leshaun is lazy: Antecedents and consequences of attributional stereotypes in the classroom. *European Journal of Psychology of Education*, XXIII, 439-458. doi:10.1007/BF03172752

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle-managers. *Journal of Personality and Social Psychology, 77*, 1004-1010.

Rudman, L.A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*, 743-762.

Schmader, T., Whitehead, J., & Wysocki, V. H. (2007). A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles*, *57*, 509-514. doi:10.1007/s11199-007-9291-4

Swim, J. K., & Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin*, *22*, 507-519.

Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse and Society*, *14*, 191-220.

Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: MIT Press.

Von Bargen, D. (2014). Tips for writing letters of recommendation. *ASCA School Counselor*, D5-D7.

Wood, W., & Eagly, A. H. (2010). Gender. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 629-667). New York, NY: Wiley.