

Examining Student Evaluations of Black College Faculty: Does Race Matter?

Bettye P. Smith **University of Georgia**
Billy Hawkins **University of Georgia**

The purpose of this study was twofold. First, to describe the undergraduate student ratings of teaching effectiveness based on the traditional 36-item end-of-course evaluation form used in the College of Education (COE) at a southeastern Research Extensive predominantly White institution. Second, using critical race theory (CRT) to compare the teaching effectiveness for the tenure-track faculty in this study based on race (White, Black, and Other racial groups including Asians, Latinos, and Native Americans). Three academic years of undergraduate level courses were used to analyze student ratings for 28 items (26 multidimensional, which address specific topics or a single aspect about instruction and 2 global/overall, which address value of course and teaching ability) on the end-of-course evaluation form. Eight of the 36 items request demographic information from the student. The findings showed that of the three faculty racial groups, Black faculty mean scores were the lowest on the 26 multidimensional items. On the two global items, which are used in making personnel decisions, Black faculty mean scores were also the lowest of the faculty groups analyzed.

Keywords: *student evaluations, race, faculty, Black faculty, predominantly White institutions*

INTRODUCTION

In the past twenty years, significant shifts in racial groups have been the norm for many states within the United States. Based on the 2006 Census data, several states have witnessed an increase in the presence of non-Whites including California (27%), New York (21%), New Jersey (19.5%), Florida (18.5%), and Nevada (17.4%) (Advertising Age's: American Demographics, 2006). In Georgia, 40% of the state's total population is comprised of non-Whites, and of this percentage, 30% are Black (U. S. Census Bureau, 2006). According to these statistics, the United States is becoming more racially and culturally diverse. Nevertheless, in higher education, there is a resistance to multicultural curricula (Chizhik & Chizhik, 2005; Grant, 1994) and racially diverse departments across the United States (Bromberg, 1993). A more racially diverse society indicates that colleges and universities must produce students who are prepared to navigate a racially and culturally diverse environment.

Many institutions of higher education have sought to increase the cultural and racial diversity among faculty members by recruiting and hiring faculty members from under-represented groups (Orlans, 1992). This practice of recruiting and hiring conversely has created adverse situations at Research Extensive institutions where the mission is teaching, research, and service. The most noted and unexplored adverse situations are problems associated with teaching evaluations and diverse faculty, particularly Black faculty. A review of existing literature reveals a plethora of research and professional literature on course evaluations and teaching effectiveness. Furthermore, there is an existing body of research and professional literature on Black faculty and teaching experiences in higher education generally (Cooper, Massey, & Graham, 2006; Li & Beckett, 2006; Smith & Womble, 2000; Steward & Phelps, 2000) and in predominantly White institutions (PWIs) specifically (Allison, 2008; Hendrix, 2007; Stanley, 2006; Steward & Phelps, 2000; Vargas, 2002). However, a paucity of that literature includes quantitative course evaluations for Black faculty. In fact, Huston (2005) ascertained that there was relatively little in the way of empirical, quantitative research on the intersection of race and

gender regarding course evaluations. Therefore, this study is designed to fill the void in the literature concerning quantitative student evaluations of teaching effectiveness for Black faculty.

Student evaluations are the most common form of evaluating teacher effectiveness among American universities (Cashin, 1988; Hobson & Talbot, 2001; Richardson, 2005; Wagenaar, 1995). However, the criticisms of these evaluations are varied ranging from the belief that they are an unpredictable form of measurement (DeFina, 1996) to the idea that they often over-interpret the effectiveness of the instructor (Cashin, 1989) and there is no relationship between teaching and learning (Clayson, 2009). Nevertheless, the College of Education (COE) at a southeastern Research Extensive institution is no exception in using student evaluations to measure teaching effectiveness. The COE adapted a 36-item student evaluation form that is used in many of the college's eight departments that have undergraduate programs. Some departments in the college have created and standardized a student evaluation form that more adequately reflects their disciplines. In other cases, faculty has elected to use an online version of the student evaluation form. However, the faculty in the COE use some form of standardized student evaluation. Of the 36 items on the college's adapted evaluation form, the COE uses the last two items (#35—*overall value of course* and #36—*overall teaching ability*) to make personnel decisions regarding promotion and tenure, annual merit pay increases, and merit awards (college, university, national, or international). Although much focus is on knowledge production at a Research Extensive institution, there is still an expectation for faculty to maintain good teaching ratings especially on the two global/overall items previously mentioned. This practice of giving attention to the two global items is troubling to Black faculty since it has the potential to negatively affect their careers.

Black faculty in the COE contends that their student evaluation ratings do not accurately reflect their teaching effectiveness. Black faculty also perceived that their ratings are numerically lower on items that are considered in personnel decisions (i.e., global items 35 and 36), but higher on those items that are not strongly considered. In turn, Black faculty's prospects for promotion, tenure, merit increases, awards (including college, university, national or international), and even leadership roles are negatively affected by these student ratings. This phenomenon of lower teaching evaluations from students creates a problematic situation for the college and its personnel decisions. However, in the absence of empirical research, Black faculty's assertions are neither supported nor refuted; rather they are simply anecdotal. This lack of empirical research on student evaluations of teaching related to race coupled with the changing demographics in higher education departments, and the controversy surrounding student ratings and their subsequent negative affect on faculty, formed the impetus for this study in the COE.

CONCEPTUAL FRAMEWORK

Since the focus of this study is teaching effectiveness of Black faculty, a race-based theory was sought as a foundation. After examination of several culturally relevant theories, critical race theory (CRT) was deemed appropriate based on the first of its three propositions, that race continues to be a significant factor determining inequity in the United States. CRT is a product of the movement in critical legal studies that seeks to analyze the effects of race and racism in U. S. justice system (Delgado, 1995). Early contributors to the development of CRT were legal scholars Derrick Bell, Richard Delgado, and Alan Freeman, as a result of their dissatisfaction with the sluggish progress of racial reform in this country (Delgado & Stefancic, 2001; Ladson-Billings, 1998). CRT focused on the intersection of race and property to understand social inequality (Ladson-Billings & Tate, 1995), and has expanded beyond its initial application in legal studies to become an effective lens with which to examine the field of education (Ladson-Billings, 1998).

CRT operated under the assumptions that race continues to be pervasive in defining inequality in the United States and that rights this country are based on property rights rather

than human rights. According to Bell (1992), racism is an integral, permanent, and indestructible component of this society. As a product of White supremacy, the apparent presence of racism in the United States manifests throughout its major institutions. Therefore, applying CRT to student evaluations in the COE such as the structured and standardized evaluation form (items on the end-of-course faculty evaluation), will allow one to glean the impact of race in student ratings. The ultimate goal of CRT is to unmask and expose racism in its various permutations (Ladson-Billings, 1998). The assumption that race continues to be a significant factor in determining inequity in the United States (Bell, 1988, 1992; Delgado, 1995; Delgado & Stefancic, 2001; Ladson-Billings, 1998) is the assumption that the examination of student ratings of teaching effectiveness is viewed. Therefore, one might pose the question: What do the data indicate about the effects of faculty race, especially given that 87% of the undergraduate population is White?

REVIEW OF LITERATURE

Much research has taken place on student evaluations of teaching effectiveness spanning three or more decades. A variety of characteristics related to student evaluations of teaching effectiveness have been examined. However, this review will only focus on research and the characteristics related to evaluation of faculty race and undergraduate course level.

Faculty Race

There is a relatively small amount of quantitative data from researchers who have explored the effects of race/ethnicity and student evaluations in higher education. Five studies were found that included faculty race/ethnicity (Anderson & Smith, 2005; DiPietro & Faye, 2005; Hameresh & Parker, 2005; Smith, 2007; Smith & Anderson, 2005). However, only one of those included Black faculty. In their studies on course evaluations, Anderson and Smith (2005) and Smith and Anderson (2005) compared Hispanic and Anglo faculty and found that female Hispanic faculty received the lowest course evaluation ratings. Hameresh and Parker (2005) also found that Hispanic faculty received lower course evaluation ratings than White faculty. DiPietro and Faye (2005) studied the evaluation ratings of African American, Hispanic, Asian American, and White faculty and found that Hispanic faculty received the lowest ratings. Asian American faculty received slightly better course evaluations than their Hispanic colleagues, but their scores were still lower than those of White faculty. The number of African American faculty in DiPietro and Faye's study was too small to draw any conclusions. Smith (2007) described and compared student ratings of teaching effectiveness on end-of-course evaluations for three groups of faculty (White, Black, and "Other" racial groups, which included Asians, Latinos, and Native Americans). Mean ratings indicated that White faculty and faculty identified as "Other" received significantly higher mean scores on all items measured than Black faculty. In fact, Black faculty rated lower than White faculty and faculty identified as "Other."

Level of Course

Several studies have examined the rating of instructors according to level of course, undergraduate, or graduate. Course level was not a variable of investigation for Wollert and West (2000), when no significant differences were found in ratings between academic departments. However, some of the ratings were related to whether the course was graduate or undergraduate level. In Wachtel's (1998) review of the existing research on evaluations of the teaching performance of college and university faculty and level of course, he found that most studies reported higher level courses tend to receive higher ratings. He further stated that no explanation for this relationship has been put forth. Sailor, Worthen, and Shin (1997) examined the correlations between mean instructor rating and level of course. Findings showed the mean instructor rating was higher for graduate level classes than for lower and upper division

undergraduate classes. It is speculated that graduate students are both better students and more critical evaluators of instruction, but replication and extension with different samples are needed before this tentative explanation can be accepted with confidence. Huang, Guo, Druva-Roush, and Moore (1995) used the generalizability theory to examine the sources of variability present in a teacher and course evaluation instrument. Three levels of courses used were (a) undergraduate, (b) intermediate, and (c) graduate, whereby findings showed that course evaluations from graduate level students were more reliable than those of lower and intermediate levels.

From the above findings, it can be concluded that minority or non-White faculty received lower evaluation scores than White faculty, and Black faculty received even lower scores. Furthermore, Black faculty scored the lowest regardless of the variables under investigation. Also, graduate students evaluated faculty higher than undergraduate students. The findings from this study will add to the body of literature addressing the social implications of student evaluations as they are employed within the structure of higher education departments, generally and departments within the colleges of education, specifically. It will also contribute to the overall body of literature addressing critical race theory and its implementation in the field of education as the primary theoretical lens for understanding racialized discrepancies. The use of CRT in this current study is a means to place it in a broader context of education and more specifically to use it in examining Black faculty teacher ratings.

METHOD

Purpose

The major purpose of this study was to describe the undergraduate student ratings of teaching effectiveness based on the traditional 36-item end-of-course evaluation form used in the COE. The secondary purpose was to compare the teaching effectiveness for the tenure-track faculty in this study based on race (White, Black, and other racial groups including Asians, Latinos, and Native Americans) and undergraduate level courses.

Procedure

Data were obtained from the university's Office of Test Scoring and Reporting Services (TSARS). The TSARS Office is responsible for scanning student evaluation forms and storing the data. Approval to analyze this data was secured from several sources, including the university's Institutional Review Board. The TSARS Office provided files for courses taught by tenure-track faculty who used the traditional 36 item end-of-course evaluation for three academic years. A database was created for a three-year period, which totaled 13,702 undergraduate evaluation forms. Therefore, undergraduate student ratings data for tenure-track faculty who used the 36-item student evaluation form adapted by the college were included in the database for this study.

Current and previous COE directories were used to identify faculty characteristics of interest (race). Additionally, each department's home page was used to identify characteristics of faculty that were in question or not easily identifiable. Each faculty was given a designation according to race for data analysis purposes. Race was coded as White, Black, and all other faculty (Asian, Latino, or Native American) were identified as "Other." Historically, the presence of under-represented faculty in the COE has not been reflective of their representation in the general population. At the time of this study, the COE had approximately 216 tenure-track and non-tenure-track faculty members. Of the 190 tenure-track faculty in this study, 156 (82%) were White, 24 (13%) Black, and 10 (5%) were identified as "Other." At this southeastern Research Extensive university, the COE has the largest number of Black faculty; it is the home of 20% (24) of the university's Black faculty, which creates a critical mass.

The COE at this Research Extensive institution is one of the largest in the nation and offers comprehensive degree programs with 20 majors for Bachelor of Science. Undergraduate students in the COE comprise approximately 3,000 or 60% of the total student population. The university is located in close proximity to a metropolitan area that is highly diverse, and as a result, has influenced the demographic composition of the College of Education. The largest student ethnic groups in the COE are Caucasian and African American; with the Caucasian student population at 87% while the African American population is 6.5% (Office of Institutional Research, 2004). Included among the student population are a small percentage of Asian and Hispanic students, 2% and 4%, respectively.

Instrumentation

Effective instruction in the COE is based on a 36-item course evaluation form that is completed by students. Many student evaluation forms contain both multidimensional and global items (d'Apollonia & Abrami, 1997). Multidimensional items measure a single aspect of teaching such as organization, preparation, or interest in subject matter while global items measure general impressions such as overall value of course and overall teaching ability (d'Apollonia & Abrami, 1997). The COE course evaluation form contains both multidimensional and global items on a 5-point Likert scale. The course evaluation form is a Scantron sheet (a machine-readable page on which students mark answers) with 36 pre-printed multiple-choice items. The first 8 items request demographic information related to the student. Items 9 thru 34 are multidimensional items of student opinions about the instructor and the course; the possible responses and anchors are: 1 = Almost Never, 2 = Infrequently, 3 = Occasionally, 4 = Often, and 5 = Almost Always. Items 35 and 36 are global items inquiring student opinions about overall value of course and overall teaching ability, respectively. Possible responses and anchors for Items 35 and 36 are: 1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, and 5 = Excellent. Four negatively worded items (12, 13, 26, and 32) were reverse coded and "did not" was added to the statements so that all items were scored and read in the same direction.

Data were analyzed using descriptive and inferential statistics. The overall teaching effectiveness in undergraduate courses in the COE was presented along with a description of faculty race (White, Black, and other racial groups that included Asian, Latino, and Native American). In this study, consideration was not given to characteristics of faculty such as academic rank or gender. Similarly, characteristics of courses such as subject matter or course option (required or elective) were not under investigation. Course level was considered as only undergraduate student ratings are included in this analysis.

RESULTS

Undergraduate Student Ratings

Undergraduate ratings of teaching effectiveness in the COE are shown in Table 1 for all faculty. These findings reflect faculty who taught undergraduate courses for the three-year period. The evaluation forms allow students to select undergraduate or graduate as their course level. Data from those identifying themselves as undergraduate students revealed that 4 of the 26 multidimensional items were rated slightly below 4.0 for the faculty in this study. The undergraduate ratings on the multidimensional items ranged from a low of 3.80 to a high of 4.63, which indicated "occasionally" and "often" on the evaluation form. The lowest mean score for the multidimensional item was 3.80 on question 31, "the instructor explained text materials that were confusing to students" and the highest mean score of 4.63 was received on item 33, "the instructor seemed well informed about the material presented."

For both global items, 35 and 36 ("overall value of course" and "overall value of teaching"), mean scores were slightly below 4.00 for faculty who taught undergraduate courses (see Table

1). Mean scores of 3.82 and 3.97, respectively for “overall value of course” and “overall teaching ability” indicate “good.” These ratings for “overall value of course” and “overall teaching ability,” “good,” provide a global view of undergraduate teaching effectiveness in the college.

Table 1

Undergraduate Student Ratings of Teaching Effectiveness for Faculty Race and Gender for Items 9-36

| Item | Statement | Undergrad ^a | White ^b | Black ^c | Other ^d |
|------|--|------------------------|--------------------|--------------------|--------------------|
| 9 | The instructor was willing to give individual assistance outside of class. | | | | |
| | <i>M</i> | 4.28 | 4.39 | 3.73 | 4.59 |
| | <i>SD</i> | .97 | .84 | 1.25 | .71 |
| 10 | The instructor encouraged students to think for themselves. | | | | |
| | <i>M</i> | 4.42 | 4.50 | 4.10 | 4.51 |
| | <i>SD</i> | .86 | .77 | 1.09 | .70 |
| 11 | The instructor gave tests that were reasonable in length. | | | | |
| | <i>M</i> | 4.22 | 4.28 | 3.95 | 4.34 |
| | <i>SD</i> | 1.08 | 1.03 | 1.23 | .97 |
| 12 | The instructor spent time on unimportant and irrelevant materials. | | | | |
| | <i>M</i> | 3.95 | 4.03 | 3.53 | 4.33 |
| | <i>SD</i> | 1.16 | 1.08 | 1.36 | .99 |
| 13 | The instructor pitched the presentation above the heads of the students. | | | | |
| | <i>M</i> | 4.28 | 4.32 | 4.16 | 4.24 |
| | <i>SD</i> | .99 | .98 | 1.03 | .95 |
| 14 | The instructor encouraged students to ask questions. | | | | |
| | <i>M</i> | 4.40 | 4.45 | 4.20 | 4.44 |
| | <i>SD</i> | .87 | .82 | .98 | .87 |
| 15 | The instructor tried to get you to see beyond the limits of the course. | | | | |
| | <i>M</i> | 4.36 | 4.45 | 4.07 | 4.38 |
| | <i>SD</i> | .88 | .81 | 1.06 | .82 |
| 16 | The instructor was well prepared each day. | | | | |
| | <i>M</i> | 4.51 | 4.61 | 4.10 | 4.68 |
| | <i>SD</i> | .85 | .74 | 1.13 | .66 |

Table 1 continues

| Item | Statement | Undergrad ^a | White ^b | Black ^c | Other ^d |
|------|--|------------------------|--------------------|--------------------|--------------------|
| 17 | The instructor clearly described the grading procedures. | | | | |
| | <i>M</i> | 4.23 | 4.29 | 3.93 | 4.42 |
| | <i>SD</i> | 1.08 | 1.03 | 1.18 | 1.00 |
| 18 | Test content was representative of assigned material. | | | | |
| | <i>M</i> | 4.33 | 4.39 | 4.03 | 4.51 |
| | <i>SD</i> | .97 | .91 | 1.18 | .77 |
| 19 | The instructor stimulated the intellectual curiosity of the students. | | | | |
| | <i>M</i> | 4.09 | 4.17 | 3.72 | 4.21 |
| | <i>SD</i> | 1.04 | .97 | 1.22 | .95 |
| 20 | The instructor was enthusiastic about the subject. | | | | |
| | <i>M</i> | 4.54 | 4.62 | 4.16 | 4.70 |
| | <i>SD</i> | .76 | .67 | .98 | .62 |
| 21 | The instructor was clear about basic principles. | | | | |
| | <i>M</i> | 4.29 | 4.38 | 3.90 | 4.41 |
| | <i>SD</i> | .94 | .85 | 1.15 | .87 |
| 22 | The instructor clearly indicated what materials the tests would cover. | | | | |
| | <i>M</i> | 4.21 | 4.29 | 3.84 | 4.38 |
| | <i>SD</i> | 1.07 | 1.00 | 1.28 | .88 |
| 23 | The instructor kept the course moving at a steady pace. | | | | |
| | <i>M</i> | 4.29 | 4.36 | 3.95 | 4.52 |
| | <i>SD</i> | .94 | .88 | 1.13 | .80 |
| 24 | The instructor tried to stimulate creative abilities. | | | | |
| | <i>M</i> | 4.08 | 4.18 | 3.72 | 4.07 |
| | <i>SD</i> | 1.09 | 1.03 | 1.24 | 1.06 |
| 25 | The instructor gave advice on how to study for the course. | | | | |
| | <i>M</i> | 3.82 | 3.90 | 3.47 | 3.98 |
| | <i>SD</i> | 1.25 | 1.21 | 1.35 | 1.17 |
| 26 | The instructor assigned a lot of burdensome busy work. | | | | |
| | <i>M</i> | 4.03 | 4.06 | 3.80 | 4.28 |
| | <i>SD</i> | 1.17 | 1.16 | 1.26 | 1.03 |
| 27 | The instructor gave presentations that were logically arranged. | | | | |

Table 1 continues

| Table 1 continued | | | | | |
|--------------------------|--|------|------|------|------|
| | <i>M</i> | 4.26 | 4.33 | 3.87 | 4.50 |
| | <i>SD</i> | 1.02 | .96 | 1.21 | .85 |
| 28 | The instructor tried to increase the interests of class members in the subject. | | | | |
| | <i>M</i> | 4.29 | 4.38 | 3.94 | 4.40 |
| | <i>SD</i> | .94 | .86 | 1.15 | .83 |
| 29 | The instructor's information seemed up-to-date. | | | | |
| | <i>M</i> | 4.60 | 4.66 | 4.32 | 4.71 |
| | <i>SD</i> | .68 | .60 | .88 | .61 |
| 30 | In this class, I felt free to express my opinions. | | | | |
| | <i>M</i> | 4.27 | 4.35 | 3.96 | 4.27 |
| | <i>SD</i> | 1.03 | .94 | 1.26 | 1.01 |
| 31 | The instructor explained text materials that were confusing to students. | | | | |
| | <i>M</i> | 3.80 | 3.86 | 3.48 | 4.07 |
| | <i>SD</i> | 1.29 | 1.27 | 1.34 | 1.19 |
| 32 | The instructor demands an unreasonably large amount of work. | | | | |
| | <i>M</i> | 4.04 | 4.07 | 3.82 | 4.25 |
| | <i>SD</i> | 1.15 | 1.13 | 1.21 | 1.07 |
| 33 | The instructor seemed well informed about the material presented. | | | | |
| | <i>M</i> | 4.63 | 4.72 | 4.23 | 4.71 |
| | <i>SD</i> | .71 | .57 | 1.02 | .59 |
| 34 | The instructor recognized student's difficulties in understanding new materials. | | | | |
| | <i>M</i> | 3.99 | 4.08 | 3.59 | 4.16 |
| | <i>SD</i> | 1.05 | .99 | 1.19 | .99 |
| 35 | Overall value of course | | | | |
| | <i>M</i> | 3.82 | 3.94 | 3.31 | 4.05 |
| | <i>SD</i> | 1.12 | 1.05 | 1.28 | 1.01 |
| 36 | Overall teaching ability | | | | |
| | <i>M</i> | 3.97 | 4.08 | 3.44 | 4.22 |
| | <i>SD</i> | 1.14 | 1.05 | 1.35 | 1.05 |

Note. Data (*M* = Mean, *SD* = Standard Deviation) are generated for items 9-36 from the evaluation forms completed by students. Each evaluation form completed is a case where: ^a *n* = 13702; ^b *n* = 10206; ^c *n* = 2792; ^d *n* = 704. Mean scores are based on a 5-point Likert scale: 1 = Almost never, 2 = Infrequently, 3 = Occasionally, 4 = Often, and 5 = Almost always for items 9-34 and 1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, and 5 = Excellent for items 35-36.

Undergraduate Student Ratings on Multidimensional Items for Faculty Race

Mean scores of undergraduate student ratings on the multidimensional items (9–34) for faculty race revealed that faculty identified as “Other” received the highest mean score on 20 of the 26 items. White faculty received a higher mean score on the remaining 6 of 26 items; 13, 14, 15, 24, 30, and 33.

Faculty identified as “Other” received mean scores on the multidimensional items that ranged from 3.98 to 4.71; the lowest mean was received on item 25 (“the instructor gave advice on how to study for the course”) and the highest mean on items 29 and 33 (“the instructor’s information seemed up-to-date” and “the instructor seemed well informed about the material presented”). On the multidimensional items, mean scores ranged for White faculty from a low of 3.86 to a high of 4.72 on items 31 (“the instructor explained text materials that were confusing to students”) and 33 (“the instructor seemed well informed about the material presented”), respectively. Black faculty received mean scores that ranged from a low of 3.47 to a high of 4.32 on items 25 (“the instructor gave advice on how to study for the course”) and 29 (“the instructor’s information seemed up-to-date”). As designated by descriptors on the end of course evaluation form, mean scores indicate “occasionally” and “often” for White faculty, Black faculty, and faculty identified as “Other.”

When students acknowledged the multidimensional items, faculty in each racial group received ratings below 4.0. Faculty identified as “Other” received a rating below 4.0 on 1 item, White faculty 2 items, and Black faculty 17 items. The highest rated faculty identified as “Other” was at least .5 of a percent higher than the lowest rated Black faculty on 11 of the 26 multidimensional items (9, 12, 16, 20, 21, 22, 23, 25, 27, 31, and 34). However, less than .1 of a percent existed between the highest rated faculty identified as “Other” and White faculty.

Undergraduate Student Ratings on Global Items for Faculty Race

On the two overall items (35 and 36) for undergraduate ratings, faculty identified as “Other” received higher mean scores than White faculty and Black faculty. Student ratings on overall items indicated “good” for Black faculty (3.31 and 3.44), “good” and “very good” for White faculty (3.94 and 4.08), and “very good” for faculty identified as “Other” (4.05 and 4.22).

Differences in Undergraduate Student Ratings Based on Faculty Race

To determine if differences exist for faculty based on race and the multidimensional items and the two global items, *t* tests were performed. The number of faculty who were identified as “Other” was too small for comparisons and thereby not included in the analysis. A composite of the 26 multidimensional items (9 to 34) was formed to provide a holistic view of teaching items. Once formed, analyses were performed using the composite of multidimensional items and the two global items (“overall value of course” and “overall teaching ability”) based on faculty race (White, Black). The alpha level was set at .05. White faculty ($M = 3.96$, $SD = .46$) reported significantly higher mean scores than Black faculty ($M = 3.64$, $SD = .67$), $t(3564.767) = 23.099$, $p = .000$, $d = .055$ on the composite of multidimensional items. Also, significant differences were also noted on both “overall value of course” $t(3758.675) = 23.567$, $p = .000$, $d = .053$ and “overall teaching ability” $t(3670.644) = 22.864$, $p = .000$, $d = .052$. White faculty had significantly higher mean scores than Black faculty on the two global items, “overall value of course” and “overall teaching ability.”

When traditional statistics are used with a large sample, almost any difference will be significant (Fan, 2001; Thompson, 1996); the sample in this study was large (14,321 undergraduate evaluation forms). However, statistical significance does not necessarily mean that the difference is large or important. Therefore, an indication of the practical significance of a finding should be considered, which entails the calculation of an effect size such as Cohen’s *d* or

eta-square (Thompson, 1996). For the comparison of two groups, Cohen's d is appropriate and performed in this study to determine the practical significance of the findings.

Cohen (1988) determined that an effect size of .05 has a medium effect. All of the calculations in this study indicate a medium effect. Fan (2001) provides practical guidelines for combining statistical significance testing and effect-size measures for decision making. Based on Fan's guidelines, it can be concluded that the statistically significant differences found in this study are also differences of practical concern. This suggests that it is meaningful and should be considered for personnel decisions.

DISCUSSION

The purpose of this study was twofold. First, to describe the undergraduate student ratings of teaching effectiveness based on the traditional 36-item end-of-course evaluation form used in the College of Education at a southeastern Research Extensive predominantly White institution. Second, using critical race theory to compare the teaching effectiveness for the tenure-track faculty in this study based on race (White, Black, and other racial groups including Asians, Latinos, and Native Americans). In this study, data were analyzed for each of the 28 items that related to instruction on the end-of-course evaluation form. Three major findings emerged from this study. First, faculty in this study, regardless of race, had above average mean scores on teaching for most items. Above average or high ratings are viewed as 4.0 or above on a 5-point scale; only 6 of the 28 items were below 4.0 on a 5-point scale. Second, of the three groups of faculty (White, Black, and "Other"), Faculty identified as "Other" received the highest while Black faculty received the lowest mean scores on the 26 multidimensional items and the two global items; Black faculty received higher mean scores on multidimensional items, but lower mean scores on the two global items. Third, faculty was different regarding race on the composite of multidimensional items and the two global items.

An examination of the items on the end-of-course evaluation form reveals that the multidimensional items can be categorized as behavioral and objective while the global items are affective and subjective; the multidimensional items are concrete whereas the global items are abstract. Therefore, students may be able to observe and evaluate the multidimensional items more readily than they can the global items.

It is worth noting that faculty identified as "Other," White faculty, and Black faculty received relatively small (less than .5 and .1 of a point) variation in mean ratings with respect to the 26 multidimensional items. The similarity of faculty ratings on multidimensional items suggests that faculty in the COE regardless of race have similar skill sets to carry out the responsibilities of teaching. Researchers (Gore; 1993; Rich, 1997; Traina, 1999) maintained that good teachers know the subject, involve students in learning, care about students and their success, and demonstrate enthusiasm for teaching. Many of these traits are among the multidimensional items on the end-of-course evaluation form where faculty had above average mean scores.

Although faculty members were similar on multidimensional items, they were very different on the two global items. In fact, for Black faculty, the lowest mean score on the multidimensional item (3.47) was slightly higher than both mean scores on "overall value of course" (3.31) and "overall teaching ability" (3.44). This finding supports the contention of Black faculty that their student ratings are lower on the global items ("overall value of course" and "overall teaching ability"), but higher on the multidimensional items. The discrepancy in higher mean scores on multidimensional items and lower mean scores on global items for Black faculty becomes questionable especially when researchers declare that traits of good teaching are among the multidimensional items.

According to Nast (1999), student evaluations assume that students possess a breadth of knowledge about the subject matter which the faculty member teaches. It is also believed that student evaluations can express resistance, hostility, and disapproval of a faculty member rather

than providing an assessment of teaching effectiveness (Nast, 1999). Furthermore, Steward and Phelps (2000) maintain that discrepancies in teaching ratings may be the result of external issues or factors outside of the faculty member's control. Faculty of color are encouraged to attend to factors that are within their control such as class preparation, organization and planning, selection of current and relevant course content, and degree of interest in and passion for course content (Steward & Phelps, 2000). However, when factors within faculty of color's control have been attended to and negative teaching evaluations persist, then a part of the students' evaluation of teaching may rest in the students' biases, prejudices, misperceptions, and lack of interaction and experience with such faculty (Steward & Phelps, 2000). Mean scores indicating "occasionally" and "often" on multidimensional items are evidence that Black faculty are attending to factors within their control.

Based on statistical and practical tests of significance, one summary statement can be made concerning a composite of the multidimensional items and the global items based on faculty race. Faculty race was a factor in student ratings of teaching effectiveness and had practical use in the real world (in this case, the real world is personnel or career matters in higher education) and perhaps overall ratings may be attributed to the student's factors considering that an overwhelming majority of undergraduate students in the college of education is White.

As a means to unmask racism, CRT is a useful tool in the explanation of student ratings of teaching effectiveness among Black faculty. Therefore, race should not be ignored as a factor when student evaluations are used as a criterion for administrative decisions regarding merit, promotion, and tenure. This practice should not alarm but reaffirm the first tenet of CRT, that race matters, and according to Delgado and Stefancic (2001):

... that racism is ordinary, not aberrational—"normal science," the usual way society does business, the common, everyday experience of most people of color in this country. (p. 7)

CONCLUSION

Researchers are just beginning to examine the issue of race bias in student evaluations (Huston, 2005). The results of the current study concur with previous research that examined discrepancies in student evaluations of faculty based on the race of the instructor (DiPietro & Faye, 2005; Hamermesh & Parker, 2005). The findings illustrate that, of the three faculty racial groups, Black faculty mean scores were the lowest on items 9-34, which addressed specific topics about instruction. On items 35 and 36, which addressed overall teaching quality and overall value of the course, Black faculty mean scores were again the lowest of the faculty groups analyzed.

The few studies that have been conducted have found that Hispanic and Asian American faculty received lower ratings than White faculty (Anderson & Smith, 2005; DiPietro & Faye, 2005; Hamermesh & Parker, 2005; Smith & Anderson, 2005); little research has focused specifically on Black faculty. The data in this study indicate that Black faculty received lower ratings than White faculty and faculty identified as "Other" on both multidimensional and global items.

The findings from this study are significant because they provide empirical data about student evaluations of Black faculty and contribute to the dialogue about the use of student end-of-course evaluations in making decisions about promotion, tenure, merit increases, and teaching awards. The lower student ratings for Black faculty on the global items—used in the COE to make personnel decisions—are especially troublesome because these ratings have the power to affect faculty merit increases and careers.

There is public debate regarding whether the United States has shifted into a post-racial era. Evolving to a post-racial and post-racist society is paramount for a multi-racial and multi-cultural nation. This is especially a necessity when certain racial groups are marginalized while others are privileged because of their race as indicated by CRT. Moreover, at this point in the evolutionary

process, race clearly matters in this nation's social institutions. CRT has exposed the significance of race in legal studies, initially, and has extended to other areas: such as education (Ladson-Billings & Tate, 1995; Solórzano, 1997; Solórzano & Yosso, 2002) and women studies (Wing, 2003). Institutions of higher education are not immune to the significance of race and how it impacts the experiences of Black faculty.

The data indicate that there are racial differences in how students evaluate Black faculty at this PWI. Using critical race theory as a lens to explain this occurrence, one can conclude that race does matter in how students evaluate both faculty and the value of the courses faculty teach at this institution and therefore, matters when examining teaching effectiveness.

REFERENCES

- Advertising Age's: American Demographics. (2006). *New census data details major changes in language landscape*. Retrieved from <http://adage.com/americandemographics/>
- Allison, D. C. (2008). Free to be me? Black professors, White institutions. *Journal of Black Studies*, 38, 641-662.
- Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences*, 2, 184-201.
- Bell, D. A. (1988). White supremacy in America: Its legal legacy, its economic costs. *Villanova Law Review*, 33, 767-779.
- Bell, D. A. (1992). Faces at the bottom of the well: The permanence of racism. New York: Basic.
- Bromberg, M. S. (1993). Harvard Law School's war over faculty diversity. *The Journal of Blacks in Higher Education*, 1, 75-82.
- Cashin, W. E. (1988). *Student ratings of teaching: A summary of the research*. (IDEA Paper No. 20). Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Cashin, W. E. (1989). *Defining and evaluating college teaching*. (IDEA Paper No. 21). Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Chizhik, E. W., & Chizhik, A. W. (2005). Are you privileged or oppressed? Students' conceptions of themselves and others. *Urban Education*, 40, 116-143.
- Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, 31, 16-30.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cooper, J. E., Massey, D., & Graham, A. (2006). Being "Dixie" at a historically Black university: A White faculty member's exploration of Whiteness through the narratives of two Black faculty members. *Negro Educational Review*, 57, 117-135.
- d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52, 1198-1208.
- DeFina, A. (1996). *An effective alternative to faculty evaluation: The use of the teaching portfolio*. (ERIC Document Reproduction Service No. ED 394 561). Columbus, OH: Clearinghouse on Adult, Career, and Vocational Education.
- Delgado, R. (Ed.). (1995). *Critical race theory: The cutting edge*. Philadelphia: Temple University Press.
- Delgado, R., & Stefancic, J. (2001). *Critical race theory: An introduction*. New York: New York University Press.
- DiPietro, M., & Faye, A. (2005, October). *Online student-ratings-of-instruction (SRI) mechanisms for maximal feedback to instructors*. Paper presented at the 30th Annual Meeting of the Professional and Organizational Development Network, Milwaukee, WI.
- Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *The Journal of Educational Research*, 94, 275-282.

- Gore, G. R. (1993). What is good teaching? *Physics Teacher*, 31, 482.
- Grant, C. (1994). Challenging the myths about multicultural education. *Multicultural Education*, 2, 4-9.
- Hamermesh, D. S., & Parker, A. M. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24, 369-376.
- Hendrix, K. G. (Ed.). (2007). *Neither White nor male: Female faculty of color: New directions for teaching and learning*. San Francisco: Jossey-Bass.
- Hobson, S. M., & Talbot, D. M. (2001). Understanding student evaluations. *College Teaching*, 49, 26-31.
- Huang, C., Guo, S., Druva-Roush, C., & Moore, J. (1995). *A generalizability theory approach to examining teaching evaluation instruments completed by students*. (ERIC Document Reproduction Service No. ED 394984). Columbus, OH: Clearinghouse on Adult, Career, and Vocational Education.
- Huston, T. (2005). *Research report: Race and gender bias in student evaluations of teaching*. Retrieved from http://www.seattleu.edu/cetl/cetl_006.htm
- Ladson-Billings, G. (1998). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education*, 11, 7-24.
- Ladson-Billings, G., & Tate, W. F., IV. (1995). Toward a critical race theory of education. *Teachers College Record*, 97, 47-68.
- Li, G., & Beckett, G. H. (Eds.) (2006). "Strangers" of the academy: Asian women scholars in higher education. Herndon, VA: Stylus.
- Nast, H. J. (1999). 'Sex', 'race' and multiculturalism: Critical consumptions and the politics of course evaluations. *Journal of Geography and Higher Education*, 23, 102-115.
- Office of Institutional Research. (2004). *Hope scholars*. Retrieved from <http://www.oir.uga.edu>
- Orlans, H. (1992). Affirmative action in higher education. *Annals of the American Academy of Political and Social Science*, 523, 144-158.
- Rich, D. (1997). 7 habits of good teachers today. *Education Week*, 40, 53-57.
- Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30, 387-415.
- Sailor, P., Worthen, B. R., & Shin, E. (1997). Class level as a possible mediator of the relationship between grades and student ratings of teaching. *Assessment & Evaluation in Higher Education*, 22, 261-269.
- Smith, B. P. (2007). Student ratings of teaching effectiveness: An analysis of end-of-course faculty evaluations. *College Student Journal*, 41, 788-800.
- Smith, B. P., & Womble, M. N. (2000). African-American women in the academy: Exemplary teaching experiences. *Journal of the Research Association of Minority Professors*, 4, 11-28.
- Smith, G., & Anderson, K. J. (2005). Students' ratings of professors: The teaching style contingency for Latino/a professors. *Journal of Latinos and Education*, 4, 115-136.
- Solórzano, D. G. (1997). Images and words that wound: Critical race theory, racial stereotyping, and teacher education. *Teacher Education Quarterly*, 24, 5-19.
- Solórzano, D. G., & Yosso, T. J. (2002). Critical race methodology: Counter-storytelling as an analytical framework for education research. *Qualitative Inquiry*, 8, 23-44.
- Stanley, C. A. (Ed.). (2006). *Faculty of color: Teaching in predominantly White colleges and universities*. Boston, MA: Anchor.
- Steward, R. J., & Phelps, R. E. (2000). Faculty of color and university students: Rethinking the evaluation of faculty teaching. *Journal of the Research Association of Minority Professors*, 4, 49-56.
- Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, 25, 26-30.
- Traina, R. P. (1999). What makes a good teacher? *Education Week*, 18, 34.
- U. S. Census Bureau. (2006). *Georgia quickfacts from the U.S. Census Bureau*. <http://quickfacts.census.gov/qfd/states/13000.html>

- Vargas, L. (Ed). (2002). *Women faculty of color in the White classroom: Narratives on the pedagogical implications of teacher diversity*. New York: Peter Lang.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 34, 191-211.
- Wagenaar, T. C. (1995). Student evaluation of teaching: Some cautions and suggestions. *Teaching Sociology*, 64, 64-68.
- Wing, A. K. (Ed). (2003). *Critical race feminism: A reader* (2nd ed.). New York: New York University Press.
- Wollert, M. H., & West, R. F. (2000). *Differences in student ratings of instructional effectiveness based on the demographic and academic characteristics of instructors*. (ERIC Document Reproduction Service No. ED 448182). Columbus, OH: Clearinghouse on Adult, Career, and Vocational Education.

AUTHORS

BETTYE P. SMITH is Professor in Workforce Education, Leadership, and Social Foundation at the University of Georgia in Athens. BILLY HAWKINS is Associate Professor of Kinesiology in the College of Education at the University of Georgia.

All comments and queries regarding this article should be addressed to smithb@uga.edu

Copyright of Journal of Negro Education is the property of Howard University and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.