

Article

Evolving Impressions: Undergraduate Perceptions of Graduate Teaching Assistants and Faculty Members over a Semester

K. Denise Kendall and Elisabeth E. Schussler

Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996

Submitted July 27, 2012; Revised October 8, 2012; Accepted October 14, 2012

Monitoring Editor: Diane K. O'Dowd

Undergraduate experiences in lower-division science courses are important factors in student retention in science majors. These courses often include a lecture taught by faculty, supplemented by smaller sections, such as discussions and laboratories, taught by graduate teaching assistants (GTAs). Given that portions of these courses are taught by different instructor types, this study explored student ratings of instruction by GTAs and faculty members to see whether perceptions differed by instructor type, whether they changed over a semester, and whether certain instructor traits were associated with student perceptions of their instructors' teaching effectiveness or how much students learned from their instructors. Students rated their faculty instructors and GTAs for 13 instructor descriptors at the beginning and near the end of the semester in eight biology classes. Analyses of these data identified differences between instructor types; moreover, student perception changed over the semester. Specifically, GTA ratings increased in perception of positive instructional descriptors, while faculty ratings declined for positive instructional descriptors. The relationship of these perception changes with student experience and retention should be further explored, but the findings also suggest the need to differentiate professional development by the different instructor types teaching lower-division science courses to optimize teaching effectiveness and student learning in these important gateway courses.

INTRODUCTION

Increased employment of, and reliance on, contingent instructors is a new reality at many institutions of higher education. The term "contingent instructor" typically refers to part-time, non-tenure track faculty or graduate teaching assistants (GTAs; Johnson, 2011). Contingent instructors may constitute nearly half of the instructors used by undergraduate institutions (Jaeger, 2008; Baldwin and Wawrzynski, 2011), revealing the great dependence higher education can have on contingent instructor employment.

Despite this increased reliance on contingent instructors, research regarding their impact on teaching and student learning is relatively sparse and inconclusive (Benjamin, 2002; Umbach, 2007). One study documented that contingent instructors typically give higher grades (Johnson, 2011), yet another found that contingent instructors do not differentially impact final student grades (Bolge, 1995). It has been documented that out-of-class interactions are beneficial for student learning. Despite these benefits contingent instructors are generally less available to students than their noncontingent colleagues (Benjamin, 2002; Jaeger, 2008) and, correspondingly, students are less likely to complete their degree when taught by contingent instructors (Jaeger, 2008).

The issue of degree completion is important because science departments in higher education typically suffer high attrition rates from the majors (Seymour and Hewitt, 1997; President's Council of Advisors on Science and Technology, 2012). Factors such as student loss of interest in the major, course demands, instructors, and inadequate educational background and preparation of the student have been cited as impacting student attrition (Seymour and Hewitt, 1997). Seymour and Hewitt (1997) also documented attrition effects that were

DOI: 10.1187/cbe.12-07-0110

Address correspondence to: K. Denise Kendall (kkendall@utk.edu).

© 2013 K. D. Kendall and E. E. Schussler. CBE—Life Sciences Education
© 2013 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

specific to instructors, including poor teaching, lack of approachability, lack of availability for help, language barriers, and the high burden of instruction placed on TAs. However, studies directly exploring instructor impact on student retention are rare, and ones investigating differential effects of faculty members and GTAs on retention even rarer. Johnson (2011) found that instructor type does not differentially impact student retention, yet O'Neal *et al.* (2007) documented that GTAs can positively impact student retention by creating a positive laboratory atmosphere. The lack of research on this topic indicates that academia still does not fully understand the impact of instructor type on student learning (Jaeger, 2008), which lends credence to calls for further research.

Also in need of research are strategies for targeted professional development for different instructor types (Baldwin and Wawrzynski, 2011), and this study specifically addresses this need for one subgroup of contingent instructors, GTAs. A survey of 4114 graduate students by Golde and Dore (2001) found that many graduate programs (53.6%) require teaching by their graduate students. A survey in the biological sciences discipline found that 71% of laboratories at comprehensive universities ($n = 15$), and 91% of laboratories at research universities ($n = 34$) were taught by GTAs (Sundberg *et al.*, 2005). Similarly, Rushin *et al.* (1997) surveyed 153 graduate school programs and found that 97% of these programs reported using GTAs to teach laboratories and/or lecture class sections. While the dependence may vary from institution to institution, there is substantial reliance in academia on GTAs instructing undergraduates (Rushin *et al.*, 1997; Sundberg *et al.*, 2005).

GTA Use: Benefits and Concerns

Several studies have explored perceptions of GTAs, particularly with regard to GTA preparation for, and quality of, instruction (Park, 2002; Dudley, 2009; Muzaka, 2009; Kendall and Schussler, 2012). These studies have captured the perceptions of GTA instruction from the viewpoints of undergraduate students, GTAs, and professors from both the United States and the United Kingdom, providing multiple perspectives within higher education systems about the impact of GTA instruction on the educational environment and student learning. Throughout this article, we use the term "descriptor" (or its derivatives) to refer to words that are often used to describe perceptions of instructors (e.g., "boring," "engaging," or "organized"). These words have also been referred to as instructor characteristics, traits, or qualities in other studies.

In studies of student perception of instructors, GTAs have been identified with positive descriptors of being engaging, approachable, informal, relaxed, interactive, relatable, understanding, and able to personalize teaching (Park, 2002; Dudley, 2009; Muzaka, 2009; Kendall and Schussler, 2012). Yet, they have also been identified with negative descriptors, including lacking confidence, knowledge, experience, and authority (Park, 2002; Dudley, 2009; Muzaka, 2009; Kendall and Schussler, 2012). Meanwhile, professors have been described with positive descriptors such as respected and organized, but negative descriptors such as strict, boring, and distant (Kendall and Schussler, 2012). Interestingly, these studies have reported contradictory results for enthusiasm; Kendall and Schussler (2012) found that professors are considered

more enthusiastic, while Park (2002) and Muzaka (2009) found that GTAs are thought of as more enthusiastic.

The existence of negative descriptors for GTAs has raised concerns about using GTAs as instructors at universities. Specifically, undergraduate students report being concerned about a lack of overall subject knowledge demonstrated by their GTAs, limited communication between GTAs and professors teaching the same course, variability of standards used by GTAs to assess undergraduates, and GTAs appearing nervous and not confident (Park, 2002; Muzaka, 2009). GTAs themselves have expressed concern about contradicting professors they teach for, having minimal teaching training, and lacking authority in the classroom (Park, 2002; Dudley, 2009; Muzaka, 2009). Similarly, professors have expressed their own concerns about GTAs' minimal instructional training and lack of confidence negatively affecting GTAs' ability to foster student learning (Park, 2002; Muzaka, 2009). Professors also noted the inability to guarantee consistency in GTA teaching and the possibility that GTAs limit undergraduate student access to "real academics" as other concerns about GTAs as instructors (Park, 2002; Muzaka, 2009).

Students' Evaluations of Teaching

Instructor teaching effectiveness is often evaluated by institutions for the purposes of instructor retention, but these evaluations also provide feedback to instructors who wish to modify their teaching to foster student learning. A key strength of these evaluations is that the feedback comes from students directly, allowing students to express their perceptions of their instructors and their instructors' ability to foster learning (Baird, 1987; Emery *et al.*, 2003; Clayson *et al.*, 2006; Zabaleta, 2007; Helterbran, 2008; Kogan *et al.*, 2010). However, some question whether factors such as the grades students receive from their instructors, instructor age, or instructor gender have a stronger influence on the ratings instructors receive than actual instructional effectiveness (d'Apollonia and Abrami, 1997; Greenwald and Gillmore, 1997; Marsh and Roche, 1997; Emery *et al.*, 2003; Clayson and Haley, 2011).

Studies exploring these concerns have provided some insight; for instance, Clayson *et al.* (2006) found that students systematically changed evaluations based on the grades they expected. Likewise, Baird (1987) and Greenwald and Gillmore (1997) noted a positive relationship between student evaluation ratings and course grades received. Research investigating instructor age conducted by Bos *et al.* (1980) noted that students rated teaching assistants in their later twenties superior to teaching assistants in their early twenties. Zabaleta (2007) found similar age influences on student evaluations; however, Zabaleta offered the caution that age is inferred by the rater and therefore may be an erroneous factor. Inconclusive results have been found regarding the influence of instructor gender; for example, Zabaleta (2007) and Bos *et al.* (1980) found no significant role of instructor gender on evaluation ratings. In contrast, Feldman's (1993) meta-analysis revealed that students favor female instructors, and Basow (1995) and Centra and Gaubatz (2000) reported differences between male and female instructor ratings.

Project Rationale

Previous work has found that undergraduates have different perceptions of GTAs and faculty members with regard to

several instructional descriptors, some potentially more beneficial to learning than others. Kendall and Schussler (2012) documented this in a study on perceived differences between hypothetical professors and GTAs. The current study compares student ratings of their biology instructors (faculty members and GTAs) at the same institution to test whether the differences in instructor descriptors hold true when students are asked to rate their current, nonhypothetical instructors. The faculty members recruited for this study were all PhD-level, full-time lecturers or tenure track-instructors of biology lecture courses, while the GTAs were all biology graduate students pursuing either a master's or PhD and were employed part-time to teach laboratory sections associated with the lecture sections.

The goal of this study was to investigate undergraduate perceptions of faculty and GTAs at the beginning and near the end of one semester of instruction with the following questions in mind:

- Q1: Will the differences between hypothetical instructor types identified by Kendall and Schussler (2012), and supported by Dudley (2009), Muzaka (2009), and Park (2002), hold true for their current, nonhypothetical, instructors?
- Q2: Will student ratings of their instructors change over the course of the semester?
- Q3: Will student ratings of their instructors be related to the students' perceptions of how effective their instructors are at teaching or how much they learned from their instructors?
- Q4: Will demographic variables, such as instructor teaching experience, instructor gender, and instructor age, affect student ratings of their instructors?

Hypotheses for this study were:

- H1: Ratings of GTAs and faculty members will differ, given the previous research by Park (2002), Dudley (2009), Muzaka (2009), and Kendall and Schussler (2012).
- H2: Undergraduate ratings of their instructors, as well as the relative ratings between instructor types, will change over the semester, as students become more aware of the behaviors and attitudes of their instructors (Nussbaum, 1992; Helterbran, 2008; Pattison *et al.*, 2011).
- H3: Student descriptor ratings for their instructors will be correlated to students' perceptions of instructors' teaching effectiveness and students' perceptions of how much they learned from their instructors (Nussbaum, 1992; Helterbran, 2008; Pattison *et al.*, 2011).
- H4: Teaching experience, gender, and age of the instructors will be correlated with student ratings of these instructors for some of the descriptors in this study (Bos *et al.*, 1980; Baird, 1987; Feldman, 1993; Basow, 1995; Greenwald and Gillmore, 1997; Centra and Gaubatz, 2000; Dunkin, 2002; Clayson *et al.*, 2006; Zabaleta, 2007).

The documentation of undergraduate student perceptions of teaching and learning by these two instructor types will reveal aspects students consider important for effective instruction and which should be the focal point of professional development. This in turn will lead to a more thorough understanding of how to maximize student experience and re-

tention in gateway undergraduate science courses taught by GTAs and faculty members.

MATERIALS AND METHODS

Data Collection

Data were collected by means of "initial" and "final" course surveys administered to undergraduate students in Spring 2011 at a large, southern research university in the United States. For the purposes of this study, "initial" refers to surveys students completed at the beginning of the second class meeting of the semester, while "final" refers to surveys students completed 2–3 wk prior to the end of the semester.

Undergraduate participants were recruited from semester-long majors and nonmajors general biology courses (nonmajors: Introductory Biology; majors: Biodiversity, Cellular Biology, Genetics, and Ecology) in which the instructors had agreed to participate in the study. Each of the selected courses had a lecture (ranging in size from 32 students to 225 students) and a laboratory (maximum of 25 students per section) component; the laboratories were all taught by GTAs, who typically taught two or three laboratory sections per semester, and the lecture classes were all taught by PhD-level faculty members (tenure-line [$n = 3$] or lecturers [$n = 3$]).

Although in some contexts lecturers may be considered contingent instructors, at this university they are highly qualified (PhD-level), full-time, and permanent (some have worked for more than 10 yr at the university) faculty members most likely indistinguishable from a tenure-line faculty member to the majority of undergraduates. The only truly contingent instructors at this institution are part-time adjunct lecturers and postdoctoral associates (who were not a part of this study) or GTAs. Throughout the study, the verbiage used when students were asked to rate their faculty member or GTA was "instructor." This prevented us from using terms that would potentially bias student responses. Throughout the methods and results, the faculty members and GTAs will be referred to as different "instructor types."

Survey Design

The initial and final surveys consisted of 15 items that were identical, except in terms of tense (Table 1). Thirteen items were created based on descriptors found to differ between instructor types in a study by Kendall and Schussler (2012), during which students rated hypothetical instructors. Of those 13 items, 12 were created by selecting two descriptors for GTAs and two descriptors for faculty from each of the three instructional themes (delivery technique, classroom atmosphere, and relationship) identified in Kendall and Schussler (2012). The descriptor "enthusiasm," which was attributed to different instructor types in different studies (Park, 2002; Muzaka, 2009; Kendall and Schussler, 2012), was also chosen as an item for this study. The wording of 11 of the items was as follows: "This instructor is . . ." with the appropriate descriptor inserted (e.g., "boring," "enthusiastic"). The wording for the other two descriptor items ("respect" and "relate") was "I . . . (to) this instructor." The final two items on the surveys asked students to rate the instructor's effectiveness in teaching the material (a statement taken directly from the university student evaluations) and how much the student

Table 1. Survey items asking students to rate their instructor for 13 descriptors as well as teaching effectiveness and student learning using a rating scale of “excellent,” “very good,” “good,” “fair,” “poor,” and “very poor.”

This instructor is confident.
This instructor is organized.
This instructor is relaxed.
This instructor is engaging.
This instructor is uncertain.
This instructor is nervous.
This instructor is strict.
This instructor is distant.
I relate to this instructor.
This instructor is understanding.
I respect this instructor.
This instructor is boring.
This instructor is enthusiastic.
The instructor’s effectiveness in teaching material.
The amount you will learn in the course.

Initial and final survey differed only in tense (i.e., The amount you will learn [initial] and The amount you learned in the course [final]).

learned (a modification of a statement from the university student evaluations).

At the top of each survey, undergraduate participants created a unique code (their birth date and the first four letters of their mother’s maiden name), allowing the researcher to match initial and final responses as well as student responses for each instructor type. Undergraduates also responded to demographic questions, including gender, enrollment status, major, native language, current Introductory Biology course enrollment, and previous Introductory Biology courses completed. After the survey was created, the researchers asked undergraduates, graduate students, administrative staff, and faculty members to review the survey and provide feedback about its face validity; these individuals expressed that the survey was suitable for the study and that the instructions were clear.

The survey instructed undergraduate participants to rate their instructors based on how well each item described him/her by circling one choice of a six-point scale including: very poor, poor, fair, good, very good, or excellent. This six-point scale was chosen to match the rating scale on the university’s student evaluation of teaching form, which undergraduates complete each semester. Some undergraduates expressed confusion regarding the rating scale, so students were verbally informed during the survey administration that they could also think of “excellent” as being “strongly agree” and “very poor” as being “strongly disagree” with a continuum in between.

Initial surveys were administered in a paper–pencil format by the researchers (authors of this paper) or a trained assistant in the lecture classes at the start of the second class meeting of the semester. Thus, students had anywhere from 50 to 75 min of exposure to the faculty member prior to completing the survey. Students were informed about the purpose of the study and were told that it was voluntary, not part of their grade, and the results would not be seen by their instructors (instructors were asked to leave the room for the administration of the surveys). Survey completion took no more than 5 min. All surveys for the lecture instructors were completed prior to the surveys in the laboratory sections because the

laboratory surveys were administered at the beginning of the second week of lab. Thus, students may have had up to 2 h and 50 min of exposure to a GTA prior to completing the survey. Because of the number of concurrent laboratory sections, the survey administration was done slightly differently for laboratories. GTAs were provided with a script that they read introducing the study, and then they chose a student to read the rest of the instructions while the GTA was out of the room. The surveys were collected in an envelope, sealed, and delivered to the main administrative office by one of the students in the lab section. The administration of the final surveys followed the same procedures and occurred between weeks 12 and 13 of the 15-wk semester.

Instructor demographics were collected from faculty members and GTAs by means of a questionnaire sent via email during week 11 of the semester. This questionnaire asked the instructor for his/her course number, section number(s), semesters of university teaching experience (not including the current semester), age, gender, native language, highest degree completed, and degree sought (if GTA).

An exemption from written consent was obtained for the study and no incentives were offered for participation. All procedures were reviewed and approved by the Institutional Review Board for Human Subjects.

Survey Validity and Reliability

The survey used in this study was designed utilizing descriptors from three instructional themes (delivery technique, classroom atmosphere, and relationship) identified from qualitative analysis of student responses from a previous study at the same institution (Kendall and Schussler, 2012). The validity of these descriptors rests with the strength of this prior research, as well as research by others who have identified similar themes and descriptors in their own research (Bos *et al.*, 1980; Park, 2002; Arnon and Reichel, 2007; Walker, 2008; Muzaka, 2009). Face validity, as previously described, was also used.

We verified the reliability of the items within each theme by calculating Cronbach’s alpha for both initial and final surveys, although it is important to note that analyses of the results was for individual items and not for themes. The delivery technique theme consisted of five items, including nervous, uncertain, organized, enthusiastic, and confident. The classroom atmosphere and relationship themes consisted of four items each (relaxed, engaging, distant, and strict for classroom atmosphere; relate, understanding, respect, and boring for relationship). Cronbach’s alpha results for the initial survey were: 0.731 for delivery technique, 0.631 for classroom atmosphere, and 0.761 for relationship. Cronbach’s alpha results for the final survey were 0.816, 0.711, and 0.831, respectively, for delivery technique, classroom atmosphere, and relationship. These values represent a valid internal consistency for the items chosen from each theme.

Data Analysis

Only undergraduate respondents ($n = 255$) who completed both the initial and final surveys for both their faculty member and GTA were included in the data analysis. At the beginning of the analyses, the Likert-scale ratings for each survey item for each instructor were converted to numeric responses in the following manner: excellent = 6, very good = 5, good = 4, fair = 3, poor = 2, and very poor = 1.

Data from the instructor demographics were also converted to numeric responses for analysis. University teaching experience was grouped as follows: 0–1 semesters (coded as “0”), 2–4 semesters (coded as “1”), 5–8 semesters (coded as “2”), and more than 9 semesters (coded as “3”). Because the study took place in the Spring semester, few instructors had 0 semesters of experience; thus 0–1 semesters of teaching experience indicated instructors with little to no teaching experience. The ranges of 2–4 semesters, 5–8 semesters, and more than 9 were chosen based on the time it typically takes to obtain a master’s degree (2–4 semesters) and PhD (5–8 semesters). However, not all GTAs teach while obtaining their degrees, and students complete graduate degrees at various paces. Gender was converted to a 1 for females and a 2 for males. Instructor age was grouped as follows: 25 and under (coded as “1”), 26–30 (coded as “2”), 31–40 (coded as “3”), 41–50 (coded as “4”), and 51–60 (coded as “5”). These age ranges were chosen based on results from Bos *et al.* (1980), in which differences in perception of teaching were found for instructors in their early twenties as compared with those in their later twenties.

The analyses were performed using nonparametric methodology because the Likert-type choices on the surveys were ordinal, and there is no guarantee that students perceive the difference between intervals on the point scale equally (e.g., Excellent and Very Good may not be the same distance to a student as the distance from Very Good and Good). This theoretical lack of equal distances between responses violates assumptions for parametric methodology, requiring the data to be analyzed via nonparametric tests (Huck, 2008). The analyses followed the main questions of the study as outlined below.

“Will student ratings differ between instructor types?” and “Do these ratings change over a semester?”

Wilcoxon signed-rank tests were conducted to evaluate whether student ratings on the initial surveys and final surveys differed significantly ($\alpha = 0.05$) by instructor type (faculty vs. GTAs). Wilcoxon signed rank tests were also conducted for each instructor type to determine whether student ratings changed significantly ($\alpha = 0.05$) over the course of the semester (SPSS 19.0; IBM).

“Will student ratings of their instructors be related to the students’ perceptions of how effective their instructors’ teaching is or how much they learned from their instructors?” and “Will student ratings be related to demographic characteristics of their instructors?”

Spearman rank correlations were performed on pooled data (GTAs and faculty together) to determine whether university teaching experience, age, or gender correlated with student initial or final ratings for each of the 13 descriptors ($\alpha = 0.05$; SPSS 19.0). Spearman rank correlations were also performed on the pooled data to determine whether student perceptions of their instructors’ teaching effectiveness or student perceptions of how much they learned from their instructors, as rated on the final surveys, correlated with any of the 13 student ratings of descriptors on the final surveys ($\alpha = 0.05$; SPSS 19.0).

RESULTS

Participants

Undergraduate Students. The surveys were administered to a potential pool of 1170 students in eight lecture sections and 43 associated laboratory sections. After removing students who did not complete both initial and final surveys for both their professor and GTA, 255 participants remained. This response rate was a result of the survey being voluntary, offering no incentives, dropping of courses by students, or absences during any of the survey administrations.

Twenty-two percent of the individuals who completed the surveys were enrolled in second-semester Introductory Biology for nonmajors, 36% in majors Biodiversity, 13% in majors Cellular Biology, and 29% in majors Genetics. These respondents were mostly female (62%), native English speakers (98%), first-year students (49%), and non-biology majors (67%). Second- and third-year students comprised 25 and 19%, respectively, of the respondents, while 7% were fourth-year students or beyond. Thirty-two percent of students were biology majors, with 3% declaring a concentration in Ecology and Evolutionary Biology, 5% in Microbiology, and 15% in Biochemistry, Cellular and Molecular Biology. Thirty-one percent of respondents had not completed another biology course at the university; however, 39% had completed one other semester-long lecture/laboratory biology course. Those who completed two or three courses comprised 20 and 5%, respectively; an additional 3% had completed four or more other biology courses. Of the participants who completed other courses, 31% had completed first-semester Introductory Biology, 9% second-semester Introductory Biology, 32% Biodiversity, 30% Cellular Biology, 3% Ecology, 2% first-semester Botany, and 1% second-semester Botany.

Instructors. The instructors of the students who participated in the study were six faculty members teaching eight lecture sections and 25 GTAs teaching 43 laboratory sections. The eight lecture sections consisted of three (38%) second semester nonmajors’ Introductory Biology classes, two (25%) Biodiversity, one (13%) Cellular Biology, and two (25%) Genetics classes.

Faculty Members. The faculty members in this study had from 3 to 45 semesters of university-level teaching experience, with an average of 16 semesters. Sixty-seven percent of the faculty members were male, while 33% were female. Seventeen percent of faculty members were between the ages of 31 and 40, 50% were between 41 and 50, and 33% were between 51 and 60. The majority of the faculty members were native English speakers (83%). All faculty members held a PhD in a life sciences field.

GTAs. Teaching experience for the GTAs in this study ranged from 0 to 12 semesters of university-level instruction, with an average of four semesters. Of the 25 GTAs, 28% were male, and 72% were female. The majority of the GTAs were under the age of 30, with 32% being under 25 years old and 48% between 26 and 30 years old. The remaining 20% of GTAs were between the ages of 31 and 40 (12%) and 41 and 50 (8%). The majority of GTAs were native English speakers (64%). Forty percent of GTAs were teaching labs associated with second-semester Introductory Biology, 28% Biodiversity, 12% Cellular Biology, and 20% Genetics. Thirty-six percent of GTAs had

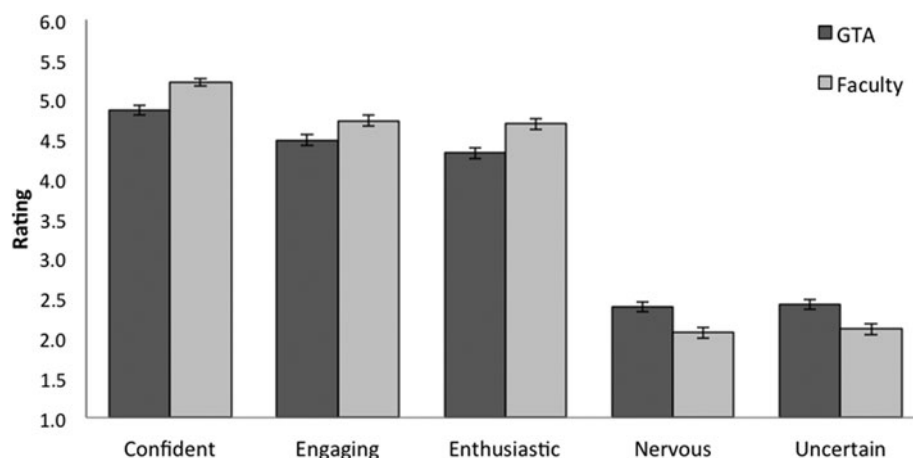


Figure 1. Mean \pm standard error for descriptors (in alphabetical order) in which GTAs differed significantly from faculty members for initial ratings (Wilcoxon signed-rank tests [$\alpha = 0.005$]).

earned a master's and bachelor's degree already, while the remaining 64% of GTAs had only completed a bachelor's degree. Of the participating GTAs, 84% were in graduate school to pursue a PhD, while 16% were seeking a master's degree.

Will Student Ratings Differ between Instructor Types? Do These Ratings Change over a Semester?

Initial Ratings. The responses provided by students indicated that they initially perceived differences between GTAs and faculty members for the following descriptors: confident ($Z = -4.623, p = 0.000$), engaging ($Z = -2.998, p = 0.003$), enthusiastic ($Z = -4.343, p = 0.000$), nervous ($Z = -4.053, p =$

0.000), and uncertain ($Z = -3.828, p = 0.000$; Figure 1). These ratings indicated that students perceived GTAs to be more nervous and uncertain, and faculty members to be more confident, engaging, and enthusiastic. Descriptive statistics by instructor type for initial and final ratings are shown in Table 2.

Final Ratings. By the end of the semester, participants indicated that GTAs differed from faculty members for the following descriptors: boring ($Z = -2.918, p = 0.004$), engaging ($Z = -2.446, p = 0.014$), organized ($Z = -4.546, p = 0.000$), relate ($Z = -2.602, p = 0.009$), respect ($Z = -2.075, p = 0.038$), and understanding ($Z = -3.311, p = 0.001$; Figure 2). These final ratings indicated a student perception that their

Table 2. Descriptive statistics including mean, standard deviation (SD), median (med), minimum (min), and maximum (max) by instructor type for initial and final ratings of descriptors (in alphabetical order).

Rating	Descriptor	Graduate teaching assistants					Faculty members				
		Mean	SD	Med	Min	Max	Mean	SD	Med	Min	Max
Initial	Boring	2.72	1.18	2	1	6	2.62	1.10	2	1	6
	Confident	4.86	0.90	5	1	6	5.22	0.79	5	2	6
	Distant	2.48	1.11	2	1	6	2.35	1.12	2	1	6
	Engaging	4.49	1.11	5	1	6	4.73	1.09	5	1	6
	Enthusiastic	4.32	1.08	4	1	6	4.69	1.07	5	1	6
	Nervous	2.39	1.07	2	1	5	2.07	1.11	2	1	6
	Organized	4.84	0.96	5	1	6	4.82	0.90	5	2	6
	Relate	3.69	1.14	4	1	6	3.69	1.12	4	1	6
	Relaxed	4.69	1.08	5	2	6	4.81	1.05	5	1	6
	Respect	4.77	0.98	5	1	6	4.90	1.04	5	1	6
	Strict	3.25	1.21	3	1	6	3.31	1.16	3	1	6
Final	Uncertain	2.42	1.05	2	1	6	2.11	1.10	2	1	6
	Understanding	4.27	1.00	4	1	6	4.29	1.03	4	1	6
	Boring	2.68	1.29	2	1	6	3.00	1.35	3	1	6
	Confident	5.04	0.95	5	2	6	5.04	0.93	5	1	6
	Distant	2.32	1.19	2	1	6	2.48	1.16	2	1	6
	Engaging	4.62	1.21	5	1	6	4.34	1.27	4	1	6
	Enthusiastic	4.41	1.15	4	1	6	4.54	1.08	5	1	6
	Nervous	2.10	1.09	2	1	6	2.10	1.05	2	1	6
	Organized	4.89	1.13	5	1	6	4.40	1.22	5	1	6
	Relate	3.74	1.37	4	1	6	3.42	1.33	3	1	6
	Relaxed	4.80	1.22	5	1	6	4.65	1.14	5	1	6
	Respect	4.87	1.15	5	1	6	4.67	1.26	5	1	6
	Strict	3.24	1.34	3	1	6	3.10	1.26	3	1	6
	Uncertain	2.16	1.06	2	1	6	2.29	1.16	2	1	6
	Understanding	4.56	1.21	5	1	6	4.24	1.22	4	1	6

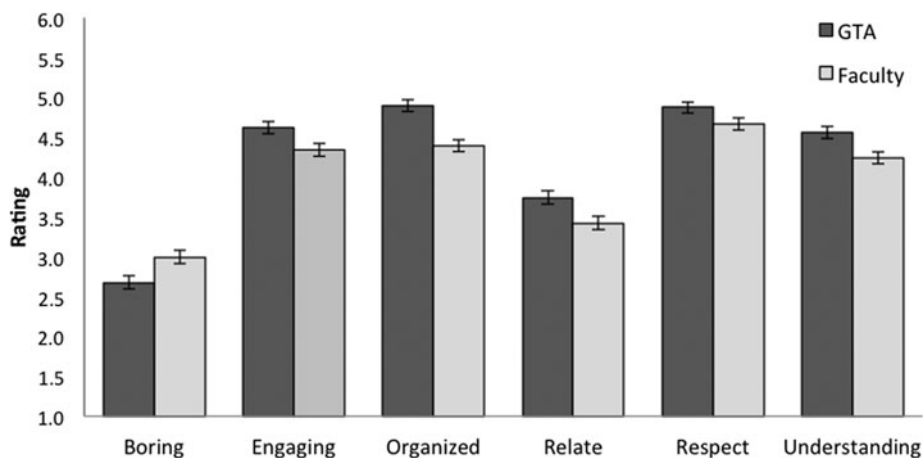


Figure 2. Mean \pm standard error for descriptors (in alphabetical order) in which GTAs differed significantly from faculty members for final ratings (Wilcoxon signed-rank tests [$\alpha = 0.005$]).

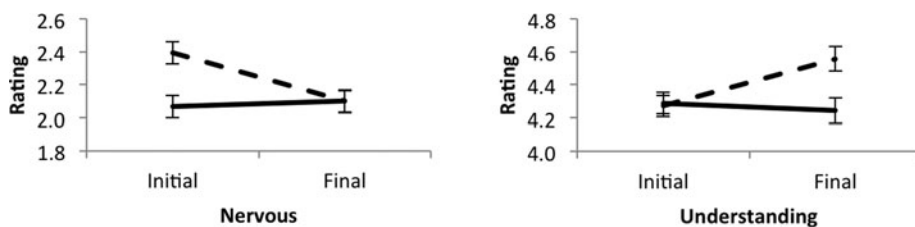


Figure 3. Descriptors (in alphabetical order) in which significant changes in mean student ratings (from 1 to 6 \pm standard error) were found over the semester for GTAs only. GTAs are depicted with dashed lines and faculty with solid lines.

faculty members were more boring than GTAs and that their GTAs were more engaging, organized, relatable, respected, and understanding than faculty members.

Change in Ratings. The initial to final changes in student ratings are organized by those that occurred only for GTAs (Figure 3), those that occurred only for faculty members (Figure 4), and those that occurred for both faculty and GTAs

(Figure 5). GTAs had changes in ratings for the descriptors of nervous ($Z = -3.808, p = 0.000$) and understanding ($Z = -4.278, p = 0.000$). GTAs decreased in perceptions of nervousness and increased in perceptions of understanding from the initial to final surveys (Figure 3). Meanwhile, faculty members had changes in ratings for the descriptors of boring ($Z = -3.952, p = 0.000$), organized ($Z = -5.138, p = 0.000$), relate ($Z = -2.707, p = 0.007$), relaxed ($Z = -2.598, p = 0.009$), and

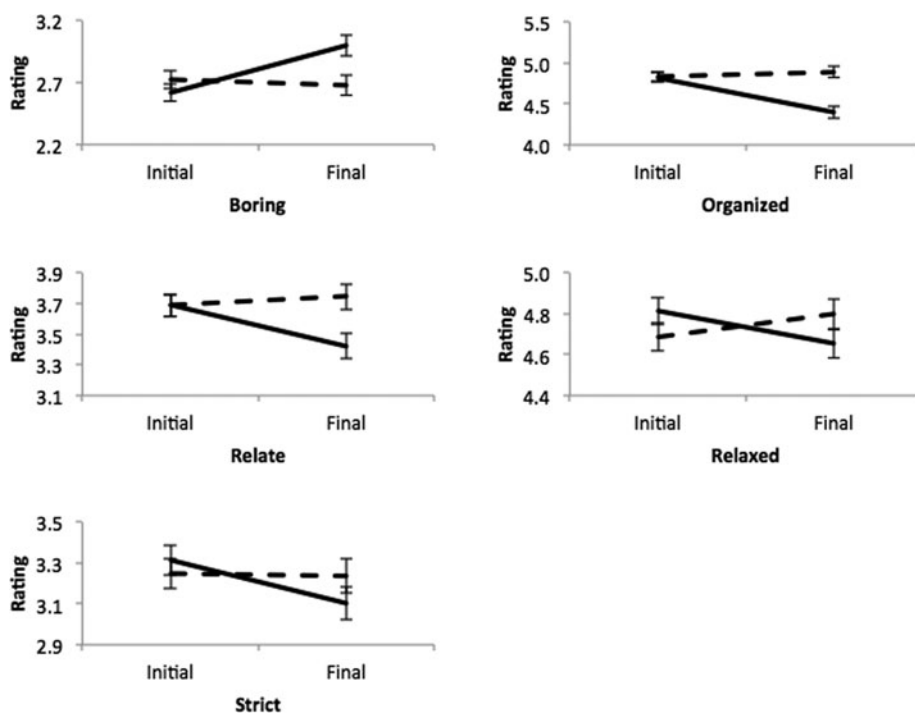


Figure 4. Descriptors (in alphabetical order) in which significant changes in mean student ratings (from 1 to 6 \pm standard error) were found over the semester for faculty only. GTAs are depicted with dashed lines and faculty with solid lines.

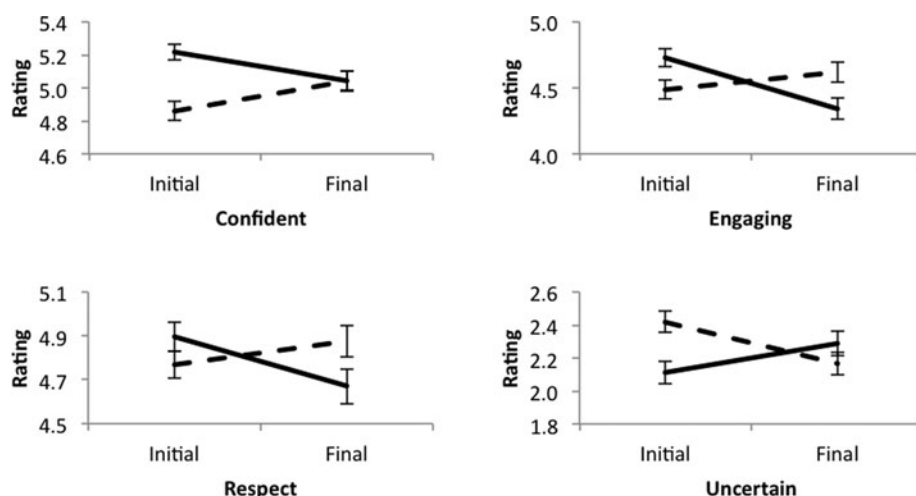


Figure 5. Descriptors (in alphabetical order) in which significant changes in mean student ratings (from 1 to 6 \pm standard error) were found over the semester for GTAs and faculty. GTAs are depicted with dashed lines and faculty with solid lines.

strict ($Z = -2.408$, $p = 0.016$; Figure 4). The student participants indicated that faculty members were more boring by the end of the semester, and less organized, relatable, relaxed, and strict than GTAs (Figure 4).

Ratings that changed for both faculty and GTAs over the semester included confident, engaging, respect, and uncertain. Significance values for GTAs were: confident ($Z = -3.547$, $p = 0.000$), engaging ($Z = -1.972$, $p = 0.049$), respect ($Z = -2.035$, $p = 0.042$), and uncertain ($Z = -3.034$, $p = 0.002$); for faculty members: confident ($Z = -2.657$, $p = 0.009$), engaging ($Z = -4.245$, $p = 0.000$), respect ($Z = -2.598$, $p = 0.009$), and uncertain ($Z = -2.130$, $p = 0.033$; Figure 5). For each descriptor, GTA and faculty changes were in the opposite direction. For the descriptors of confident, engaging, and respect, GTAs increased in their ratings from the initial to final surveys, whereas faculty members decreased in their ratings. For the descriptor of uncertain, GTAs decreased in their ratings, while faculty members increased (Figure 5).

Will Student Ratings of Their Instructors Be Related to the Students' Perceptions of How Effective Their Instructors' Teaching Is or How Much They Learned from Their Instructors?

Teaching Effectiveness. For the final survey, pooled undergraduate ratings of their instructor's effectiveness in teaching the material found that these ratings correlated significantly with all final survey ratings of instructor descriptors ($\alpha = 0.05$). All p values were 0.000, except strict, which was 0.005. There were strong positive associations between final survey ratings of instructor effectiveness and ratings of confident ($r = 0.568$), engaging ($r = 0.696$), enthusiasm ($r = 0.588$), organized ($r = 0.667$), relate ($r = 0.653$), relaxed ($r = 0.612$), respect ($r = 0.686$), and understanding ($r = 0.619$). Contrarily, there was a strong negative association between perceptions of teaching effectiveness and ratings of boring ($r = -0.522$); moderately negative associations for distant ($r = -0.476$), nervous ($r = -0.342$), and uncertain ($r = -0.472$); and a weak negative association with strict ($r = -0.120$).

Student Learning. Pooled results of undergraduate perceptions of how much they learned in the course on the final survey correlated significantly with final student ratings of

all descriptors, except strict ($\alpha = 0.05$, for all significant variables $p = 0.000$). These analyses indicated strong positive associations between ratings of student learning and ratings of confident ($r = 0.515$), engaging ($r = 0.595$), enthusiasm ($r = 0.525$), organization ($r = 0.570$), relate ($r = 0.574$), relaxed ($r = 0.528$), respect ($r = 0.611$), and understanding ($r = 0.517$). Contrarily, moderate negative associations were found between ratings of student learning and perceptions of boring ($r = -0.446$), distant ($r = -0.410$), nervous ($r = -0.306$), and uncertain ($r = -0.441$).

Will Student Ratings Be Related to Demographic Characteristics of Their Instructors?

Initial Ratings. Weak positive associations were identified between university teaching experience and initial student ratings for confident ($r = 0.269$, $p = 0.000$), engaging ($r = 0.154$, $p = 0.001$), relaxed ($r = 0.145$, $p = 0.001$), and strict ($r = 0.095$, $p = 0.031$), while a moderately positive association was identified for the rating of enthusiastic ($r = 0.414$, $p = 0.001$). There were also weak negative associations between university teaching experience and ratings of nervous ($r = -0.224$, $p = 0.000$) and uncertain ($r = -0.193$, $p = 0.000$). The highest mean student ratings for confident, relaxed, and engaging were found for the two more experienced instructor groups (5–8 and 9+ semesters). Instructors with the least teaching experience (0–1 and 2–4 semesters) had the highest means for uncertain and nervous. Interestingly, mean student ratings for strict and enthusiastic were highest for instructors with the least and most teaching experience (0–1 and 9+ semesters). For the descriptors in which significant associations were found with teaching experience, descriptive statistics are included in Table 3. In terms of gender, there were weak negative associations for the ratings of organized ($r = -0.153$, $p = 0.001$) and uncertain ($r = -0.100$, $p = 0.024$), with females being rated higher for both (descriptive statistics in Table 3). Finally, instructor age had weak positive associations with initial student ratings for confident ($r = 0.222$, $p = 0.000$), engaging ($r = 0.128$, $p = 0.004$), enthusiasm ($r = 0.202$, $p = 0.000$), and relaxed ($r = 0.103$, $p = 0.020$). There were weak negative associations between instructor age and ratings of distant ($r = -0.112$, $p = 0.012$), nervous ($r = -0.230$, $p = 0.000$), and uncertain ($r = -0.214$, $p = 0.000$). Instructors

Table 3. Mean \pm standard error for all initial ratings of descriptors (in alphabetical order) that had significant associations with teaching experience, instructor gender, or instructor age (Spearman rank correlations, $\alpha = 0.05$).

	0–1 semesters	2–4 semesters	5–8 semesters	9+ semesters	
A. Teaching experience					
Confident	4.83 ± 0.13	4.68 ± 0.09	5.00 ± 0.07	5.34 ± 0.05	
Engaging	4.49 ± 0.14	4.40 ± 0.10	4.54 ± 0.09	4.82 ± 0.08	
Enthusiastic	4.46 ± 0.14	4.31 ± 0.10	4.39 ± 0.09	4.72 ± 0.08	
Nervous	2.55 ± 0.15	2.50 ± 0.10	2.19 ± 0.09	1.99 ± 0.08	
Relaxed	4.69 ± 0.14	4.49 ± 0.10	4.72 ± 0.09	4.94 ± 0.08	
Strict	3.26 ± 0.15	3.17 ± 0.12	3.17 ± 0.09	3.44 ± 0.09	
Uncertain	2.45 ± 0.14	2.47 ± 0.10	2.31 ± 0.09	2.05 ± 0.08	
	Female	Male			
B. Instructor gender					
Organized	4.94 ± 0.06	4.70 ± 0.06			
Uncertain	2.36 ± 0.06	2.16 ± 0.07			
	Under 25	26–30	31–40	41–50	51–60
C. Instructor age					
Confident	4.73 ± 0.11	4.91 ± 0.08	4.86 ± 0.10	5.32 ± 0.06	5.22 ± 0.08
Distant	2.85 ± 0.15	2.43 ± 0.09	2.23 ± 0.11	2.27 ± 0.08	2.45 ± 0.14
Engaging	4.37 ± 0.13	4.47 ± 0.10	4.72 ± 0.11	4.76 ± 0.09	4.68 ± 0.13
Enthusiastic	4.00 ± 0.13	4.36 ± 0.09	4.66 ± 0.11	4.77 ± 0.09	4.57 ± 0.13
Nervous	2.79 ± 0.14	2.23 ± 0.08	2.31 ± 0.11	2.09 ± 0.09	1.90 ± 0.12
Relaxed	4.63 ± 0.13	4.66 ± 0.09	4.79 ± 0.11	4.66 ± 0.09	5.10 ± 0.10
Uncertain	2.74 ± 0.14	2.32 ± 0.09	2.26 ± 0.10	2.08 ± 0.09	2.07 ± 0.13

aged 26–30, 41–50, and 51–60 had the highest mean ratings for the descriptor confident. Instructors classified in age groups of 26–30, 31–40, and 51–60 had the highest mean ratings for the descriptor relaxed. The highest mean student ratings for engaging and enthusiastic were for instructors in the oldest three age groups (31–40, 41–50, and 51–60). However, instructors belonging to the three youngest age groups (under 25, 26–30, and 31–40) had the highest mean student ratings for uncertain and nervous. Finally, instructors classified in the youngest two groups (under 25 and 26–30) and the oldest age group (51–60) had the highest mean ratings for distant. Descriptive statistics for these significant associations with instructor age can be found in Table 3.

Final Ratings. At the end of the semester, there were weak positive associations between university teaching experience and student ratings of confident ($r = 0.148$, $p = 0.001$) and strict ($r = 0.097$, $p = 0.029$). There were weak negative associations between teaching experience and ratings of nervous ($r = -0.148$, $p = 0.001$), organized ($r = -0.089$, $p = 0.044$), uncertain ($r = -0.122$, $p = 0.006$), and understanding ($r = -0.150$, $p = 0.001$). Mean student ratings for the descriptor confident were highest for the two more experienced instructor groups (5–8 and 9+ semesters of teaching experience). Meanwhile, instructors with the least teaching experience (0–1 and 2–4 semesters) had the highest means for uncertain and nervous. Mean student ratings for strict were highest for instructors with the least and most teaching experience (0–1 and 9+ semesters). Meanwhile, the highest mean ratings for organized and understanding were found for instructors with 2–4 and 5–8 semesters of teaching experience. Descriptive statistics can be found in Table 4 for all descriptors with significant associations with

teaching experience. For gender, there were weak negative associations between gender and ratings for organized ($r = -0.129$, $p = 0.004$), and understanding ($r = -0.107$, $p = 0.015$), whereby females were perceived as more organized and understanding (descriptive statistics in Table 4). Finally, instructor age had a weak positive association with final ratings for boring ($r = 0.109$, $p = 0.014$) and weak negative associations with final ratings for engaging ($r = -0.107$, $p = 0.015$), organized ($r = -0.227$, $p = 0.000$), relate ($r = -0.134$, $p = 0.003$), and understanding ($r = -0.149$, $p = 0.001$). Instructors belonging to the three youngest age groups (under 25, 26–30, and 31–40) had the highest mean ratings for organized, engaging, and relate, while instructors classified in the groups of 26–30, 31–40, and 51–60 had the highest mean ratings for understanding. Instructors in the youngest age group (under 25) and the oldest two age groups (41–50 and 51–60) had the highest mean student ratings for boring. Descriptive statistics can be found in Table 4 for all descriptors that had significant associations with instructor age.

DISCUSSION

Although some descriptors associated with hypothetical instructors (Kendall and Schussler, 2012) held true in nonhypothetical educational settings, it is apparent that reality did not always match with the theoretical. Interestingly, however, student perceptions of their instructors evolved over the course of the semester, with GTAs being more likely than faculty members to gain in student perception of positive instructional characteristics by the end of the semester. Weak associations were identified between instructor demographics and student ratings, suggesting that the age, gender, or

Table 4. Mean \pm standard error for all final ratings of descriptors (in alphabetical order) that had significant associations with teaching experience, instructor gender, or instructor age in final ratings (Spearman rank correlations, $\alpha = 0.05$).

	0–1 semesters	2–4 semesters	5–8 semesters	9+ semesters	
A. Teaching experience					
Confident	4.68 ± 0.12	4.85 ± 0.10	5.21 ± 0.66	5.14 ± 0.06	
Nervous	2.38 ± 0.13	2.36 ± 0.11	1.91 ± 0.08	1.99 ± 0.07	
Organized	4.63 ± 0.15	4.73 ± 0.10	4.88 ± 0.09	4.41 ± 0.09	
Strict	3.38 ± 0.17	2.89 ± 0.13	2.98 ± 0.10	3.41 ± 0.10	
Uncertain	2.66 ± 0.14	2.39 ± 0.12	1.94 ± 0.08	2.20 ± 0.08	
Understanding	4.35 ± 0.16	4.67 ± 0.11	4.59 ± 0.09	4.11 ± 0.09	
	Female	Male			
B. Instructor gender					
Organized	4.79 ± 0.07	4.47 ± 0.08			
Understanding	4.51 ± 0.08	4.28 ± 0.08			
	Under 25	26–30	31–40	41–50	51–60
C. Instructor age					
Boring	2.91 ± 0.14	2.58 ± 0.11	2.67 ± 0.16	2.98 ± 0.11	3.12 ± 0.15
Engaging	4.40 ± 0.14	4.77 ± 0.10	4.60 ± 0.14	4.25 ± 0.11	4.32 ± 0.13
Organized	4.99 ± 0.10	4.99 ± 0.09	4.68 ± 0.12	4.18 ± 0.12	4.51 ± 0.12
Relate	3.61 ± 0.15	3.79 ± 0.12	4.00 ± 0.16	3.19 ± 0.11	3.43 ± 0.14
Understanding	4.39 ± 0.14	4.61 ± 0.10	5.01 ± 0.11	3.82 ± 0.10	4.40 ± 0.12

teaching experience of an instructor may not impact student ratings as much as individual instructional abilities. This is evidenced by the moderate to strong associations between student ratings of instructor teaching effectiveness and perception of learning and all student ratings of the instructor descriptors, except *strict*.

Student Expectations

On the basis of previous research, we expected faculty members to be rated as more boring, confident, distant, enthusiastic, organized, respected, and strict than GTAs, and we expected GTAs to be rated as more engaging, nervous, relatable, relaxed, uncertain, and understanding than faculty members at the beginning of the semester (indicating that these may be stereotypes of each instructor type; Park, 2002; Dudley, 2009; Muzaka, 2009; Kendall and Schussler, 2012). In reality, only a few of these expectations held true for actual instructors. As expected, undergraduates rated faculty members higher initially for being *confident* and *enthusiastic*, while GTAs were rated higher for being *nervous* and *uncertain*. However, undergraduate students also rated faculty members higher for *engaging*, which was contrary to previous research. No differences between instructor types were seen for initial ratings of *boring*, *distant*, *organized*, *relate*, *relaxed*, *respect*, *strict*, and *understanding*, suggesting that reality trumps stereotypes for these characteristics.

These results indicate that either initial instructor impressions were enough to overcome any pre-existing expectations students might have about their instructors or that many of the previously identified descriptor differences are not strongly engrained stereotypes. However, for the descriptors that did hold true, faculty members benefited from more positive descriptions than GTAs. Therefore, if stereotypes do exist, they seem to differentially benefit faculty members.

Change in Ratings

Ambady and Rosenthal (1992) noted that brief exposure to behavior provides enough information for an individual to predict with significant accuracy their perception of others. This suggests that student ratings of their instructors should be stable over a semester, yet student perceptions of both GTAs and faculty evolved over the semester in this study. A possible explanation is that it may depend on the descriptor; those related to more overt qualities, such as *nervous* and *uncertain*, may be easier to quickly judge than qualities such as *relate* and *understanding* (Ambady and Rosenthal, 1992). These types of descriptors may only be judged when students see how the instructor deals with periodic, situational issues over the semester (such as an unfair exam questions or missing class because of a car breakdown). Birch *et al.* (2012) found that cues such as voice clarity, preparation, and classroom control impact student impressions and expectations of lecturers; however, these cues may represent only a subset of the behaviors that students use to judge an instructor over a semester. Thus, future work should explore what factors are influencing student perceptions of their instructors, as well as changes in these perceptions, and whether instructors can direct these changes through the display of specific behaviors or attitudes, as proposed by Helterbran (2008), Nussbaum (1992), and Pattison *et al.* (2011).

When undergraduate student perception changed, the trend of these changes was that GTAs gained in ratings of more positive descriptors over the course of the semester, whereas faculty decreased in student perceptions of positive descriptors. This may reflect students entering the classroom with certain expectations for their instructors (Helterbran, 2008; Pattison *et al.*, 2011) and having higher expectations for a faculty member than they do for a GTA. If this is the case, it appears that faculty members were not able to meet all of the expectations of students who participated in this study,

and GTAs were more easily able to meet and exceed student expectations for them. However, these differences may also be contextual, such as differences in classroom setting (lecture vs. laboratory). For example, students may expect instructors to be more engaging in smaller laboratory settings than in lecture settings; therefore, engaging lecture instructors may more easily meet or exceed student expectations. Student ratings on teaching evaluations have been found to be influenced by student grades (Baird, 1987; Greenwald and Gillmore, 1997; Clayson *et al.*, 2006); therefore, faculty ratings may decline over the semester because lecture course work is typically worth more than laboratory course work, so lower grades in lecture would disproportionately affect the student's overall course grade.

The faculty and GTA descriptors that changed in opposite directions (*confident*, *engaging*, *respect*, and *uncertain*) may be the most susceptible to these instructor or contextual expectations and also the easiest to change through classroom interactions. This is important for both GTAs and faculty to understand, so they can consider why students rate them lower or higher on these aspects over the semester, and actively work either to maintain or gain in the positive aspects. These aspects should also be a point of discussion in professional development and mentoring of university instructors.

Teaching Effectiveness and Student Learning

All of the student descriptor ratings were significantly associated (moderately or strongly, except *strict*, which was weak) with undergraduate perceptions of instructor teaching effectiveness, and all (except *strict*) with perceived student learning. This intimate relationship between student perception of each descriptor and ratings of teaching effectiveness and student learning is an important reminder to instructors and administrators of the many facets—some of which may not be thought of as instructional skills—that influence student perception of teaching. For instance, an instructor who is not rated well in teaching effectiveness by students may simply not relate well to students. Zabaleta (2007) cautioned that factors such as lifestyle likeness or age group similarity (e.g., descriptors such as *ability to relate* or *understand*) may be influencing student ratings on evaluations. Translating instructor ratings into specific behaviors students use to judge them would make it easier for instructors to make changes to their instructional delivery that could ultimately also improve student perception of instructor teaching effectiveness.

We note that *strict* was unusual in this study because it was not viewed by students the same for both teaching and learning (student ratings of *strict* were negatively correlated with teaching effectiveness, whereas there was a positive, but not significant, correlation between student ratings of *strict* and their perceptions of how much they learned). In addition, correlations between student ratings of *strict* and student perceptions of teaching effectiveness and learning were not as strong as with the other descriptors. In another study (Kendall and Schussler, 2013), students explained that *strict* had both negative and positive connotations to students. For instance, a strict instructor may be unbending in their rules to the point of being unfair. However, *strict* can also be characteristic of a respected instructor because the instructor enforces polite behavior in a large classroom or applies the same rules to all students. This potentially explains why *strict* had lower correlations with teaching effectiveness and was not correlated

to student learning; however, future work is necessary to elucidate facets underlying, and perhaps confounding, student perceptions of effective instruction and learning.

Instructor Demographic Variables

Although there were associations among student ratings of their instructors and instructor demographics, these associations were weak, with the exception of university teaching experience and enthusiasm. The results for some of the descriptors support the findings by Zabaleta (2007) and Bos *et al.* (1980) that instructor age may influence student ratings and gender can influence student ratings (Feldman, 1993; Basow, 1995; Centra and Gaubatz, 2000). However, there were also results indicating that experience (novice vs. experienced) and instructor gender (Bos *et al.*, 1980; Dunkin, 2002; Zabaleta, 2007) had no impact on student ratings of certain descriptors.

It may be more instructive to identify which descriptor ratings and instructor demographics were correlated with each other on both the initial and final surveys (Table 5). For example, *confident*, *nervous*, *strict*, and *uncertain* were significantly associated with teaching experience for both initial and final ratings, with *confident* and *strict* being weak positive associations and *nervous* and *uncertain* being weak negative associations. *Organized* was associated with gender on both the initial and final surveys, with females being perceived as being more organized than males. Finally, *engaging* was consistently perceived as being associated with age, but the association changed (according to shifts in student ratings from the initial to final survey) from a weak positive to a weak negative association. Linking these results with the consistent differences between instructor types (faculty are *confident* and *enthusiastic* and GTAs are *uncertain* and *nervous*) may indicate that teaching experience is one of the most important distinguishing features between faculty and GTAs for students.

The notion that student ratings are influenced by these instructor demographics, however, must be treated with caution, since student perception of teaching experience and age may not be accurate (also noted in Zabaleta, 2007). For instance, a young-looking faculty member may have more teaching experience and be older than students perceive. So, some of the differences identified between instructor types may result from GTAs being perceived as having less teaching experience than faculty, which may or may not always be true. A follow-up study could document actual teaching experience and student perceptions of teaching experience to relate to student descriptor ratings. Regardless, these results once again highlight the influence confounding factors, such as age, teaching experience and gender of instructors, can have on student perceptions, and that these factors should be considered when administrators and instructors review evaluations of teaching.

Also unclear is whether student ratings of the descriptors are accurate reflections of classroom practice. While we quantitatively identified positive and negative aspects of instructors (similar to Pattison *et al.* [2011] and Helterbran [2008] identifying effective and ineffective instructional aspects), we recommend that there be further exploration regarding student perceptions and actual instructor practice. For instance, students in this study reported that GTA *confidence* increased over the semester, but it is unknown whether GTAs were actually becoming more confident or were merely

Table 5. Significant Spearman ($\alpha = 0.05$) associations of descriptors with instructor demographics for initial and final ratings.

	Initial ratings	Final ratings		Initial ratings	Final ratings
Teaching experience	Confident $r = 0.269$ $p = 0.000$	Confident $r = 0.148$ $p = 0.001$	Age	Confident $r = 0.222$ $p = 0.000$	Boring $r = 0.109$ $p = 0.014$
	Engaging $r = 0.154$ $p = 0.001$	Nervous $r = -0.148$ $p = 0.001$		Distant $r = -0.112$ $p = 0.012$	Engaging $r = -0.107$ $p = 0.015$
	<i>Enthusiastic</i> $r = 0.414$ $p = 0.001$	Organized $r = -0.089$ $p = 0.044$		Engaging $r = 0.128$ $p = 0.004$	Organized $r = -0.227$ $p = 0.000$
	Nervous $r = -0.224$ $p = 0.000$	Strict $r = 0.097$ $p = 0.029$		Enthusiastic $r = 0.202$ $p = 0.000$	Relate $r = -0.134$ $p = 0.003$
	Relaxed $r = 0.145$ $p = 0.001$	Uncertain $r = -0.122$ $p = 0.006$		Nervous $r = -0.230$ $p = 0.000$	Understanding $r = -0.149$ $p = 0.001$
	Strict $r = 0.095$ $p = 0.031$	Understanding $r = -0.150$ $p = 0.001$		Relaxed $r = 0.103$ $p = 0.020$	
	Uncertain $r = -0.193$ $p = 0.000$			Uncertain $r = -0.214$ $p = 0.000$	
Gender	Organized $r = -0.153$ $p = 0.001$	Organized $r = -0.129$ $p = 0.004$			
	Uncertain $r = -0.100$ $p = 0.024$	Understanding $r = -0.107$ $p = 0.015$			

Nonbold descriptors have weak associations, italic descriptors have moderate associations, and bold descriptors remain significant throughout the semester.

exceeding undergraduate expectations, or undergraduate students were interpreting them to be more confident (without GTAs feeling more confident). It is also possible that discipline, or content within a discipline, impacts student perception of instructors. For instance, students may indicate that being *engaging* is important for biology but not chemistry. Additional studies could further clarify the types of professional development to offer to instructors of different courses and even disciplines.

CONCLUSION

This study has revealed the complex perceptions students have about GTAs and faculty members as instructors. While some demographic characteristics impact undergraduate student perception of GTAs and faculty members, it appears that student expectations and the nature of individual instructors themselves play a more vital role in student perception. Even when differences between instructor types were identified, GTAs increased in ratings for many descriptors viewed as positive by undergraduates, while they decreased in some negative descriptors over the course of a semester. This indicates that student learning may not be as hindered by the use of contingent instructors in academia as previously suspected. While future studies are still necessary to identify

whether specific instructor behaviors affect rating changes over the semester, the results of this study provide support for Baldwin and Wawrzynski's (2011) call for targeted development strategies for different instructor types.

To enhance instructional abilities of GTAs, for instance, institutions or departments could help GTAs identify their strengths and weaknesses based on the descriptors of this study and use these results to form groups of GTAs who already have, or need to develop, particular instructional capacities. GTAs can also be given the opportunity to microteach, or observe, laboratories prior to teaching their own sections in order to obtain feedback and help them plan for their own teaching. This may also decrease GTA uncertainty and enhance GTA confidence in their instruction. While this study found that GTAs increased in perceptions of descriptors positively associated with teaching effectiveness over the semester, professional development may also need to emphasize that students do see value in some less positive characteristics. For instance, it may be instructive to review the positive and negative aspects of strict with GTAs to highlight the subtle dynamics that make a difference to students in the classroom. GTAs can be encouraged to self-reflect on the teaching they are providing to undergraduates each class and identify areas of instructional weakness. Institutions can support this self-reflection by providing GTAs with the opportunity

to video-record class sessions, enabling GTAs to observe their instructional behaviors more objectively. Given that student ratings can change over the course of the semester, we also recommend that GTAs encourage students to provide them with regular feedback so GTAs can make strategic modifications to their behaviors throughout the semester (Keutzer, 1993).

Undergraduates who find their introductory sciences courses to be enjoyable and educational may be more likely to pursue a degree in the sciences (Seymour and Hewitt, 1997). Accordingly, introductory science instructors are an important component in fostering student retention and learning in the sciences and the instructional abilities of faculty and GTAs should be fostered and developed by their institutions. Through researching student perceptions, instructors and professional development coordinators can better understand what undergraduates expect from instructors, and how this relates to student perceptions of teaching effectiveness and learning in their courses. Through this research, targeted professional development opportunities for each instructor type involved in gateway science courses can be developed to maximize student retention and learning in all aspects of these courses, regardless of instructor type.

ACKNOWLEDGMENTS

We thank the Department of Ecology and Evolutionary Biology and Division of Biology at the University of Tennessee for support and funding for this project. Ms. Raina Akin, Ms. Kate Weaver, and graduate students of the Department of Ecology and Evolutionary Biology assisted with the survey administration. We also thank the instructors and students who participated in the project, and M. L. Niemiller for reviewing drafts of the manuscript.

REFERENCES

- Ambady N, Rosenthal R (1992). Thin slices of behavior as predictors of interpersonal consequences: a meta-analysis. *Psychol Bull* 2, 256–274.
- Arnon S, Reichel N (2007). Who is the ideal teacher? Am I? Similarity and difference in perception of education regarding the qualities of a good teacher and of their own qualities as teachers. *Teachers and Teaching* 13, 441–464.
- Baird JS (1987). Perceived learning in relation to student evaluation of university instruction. *J Educ Psychol* 79, 90–91.
- Baldwin RG, Wawrzynski MR (2011). Contingent faculty as teachers: what we know; what we need to know. *Am Behav Sci* 55, 1485–1509.
- Basow SA (1995). Student evaluations of college professors: when gender matters. *J Educ Psychol* 87, 656–665.
- Benjamin E (2002). How over-reliance upon contingent appointments diminishes faculty involvement in student learning. *Peer Rev* 5, 4–10.
- Birch PDJ, Batten J, Manley AJ, Smith MJ (2012). An exploratory investigation examining the cues that students use to form initial impressions and expectancies of lecturers. *Teach Higher Educ* iFirst 17, 1–13.
- Bolge RD (1995). Examination of Student Learning as a Function Of Instructor Status (Full-Time vs. Part-Time) at Mercer Community College, West Windsor, NJ: Mercer Community College.
- Bos RR, Zakrajsek DB, Wolf V, Stoll S (1980). Teaching assistant traits: their influence on student ratings. *Improv Coll Univ Teach* 28, 179–185.
- Centra JA, Gaubatz NB (2000). Is there gender bias in student evaluations of teaching? *J Higher Educ* 71, 17–33.
- Clayson DE, Frost TR, Sheffet MJ (2006). Grades and the student evaluation of instruction: a test of the reciprocity effect. *Acad Mgmt Learn Educ* 5, 52–65.
- Clayson DE, Haley DA (2011). Are students telling us the truth? A critical look at student evaluation of teaching. *Marketing Educ Rev* 21, 101–112.
- d'Apollonia S, Abrami RC (1997). Navigating student ratings of instruction. *Am Psychol* 52, 1198–1208.
- Dudley M (2009). Jumping out of an airplane: a TA's perspective on teaching effectiveness. *East Educ J* 38, 1–10.
- Dunkin MJ (2002). Novice and award-winning teachers' concepts and beliefs about teaching in higher education. In: *Teacher Thinking, Beliefs, and Knowledge in Higher Education*, ed. N Hativa and P Goodyear, Dordrecht, The Netherlands: Kluwer Academic.
- Emery CR, Kramer TR, Tian RG (2003). Return to academic standards: a critique of students' evaluations of teaching effectiveness. *Qual Assur Educ* 11, 37–46.
- Feldman KA (1993). College students' views of male and female college teachers: Part IIL. Evidence from students' evaluations of their classroom teachers. *Res Higher Educ* 34, 151–211.
- Golde CM, Dore TM (2001). *At Cross Purposes: What the Experiences of Doctoral Students Reveal about Doctoral Education*, Philadelphia, PA: Pew Charitable Trusts.
- Greenwald AG, Gillmore GM (1997). Grading leniency is a removable contaminant of student ratings. *Am Psychol* 52, 1209–1217.
- Helterbran VR (2008). The ideal professor: student perceptions of effective instructor practices, attitudes, and skills. *Education* 129, 125–138.
- Huck S (2008). *Reading Statistics and Research*, Boston: Pearson Education.
- Jaeger AJ (2008). Contingent faculty and student outcomes. *Academe* 94, 42–43.
- Johnson IY (2011). Contingent instructors and student outcomes: an artifact or fact? *Res Higher Educ* 52, 761–785.
- Kendall KD, Schussler EE (2012). Does instructor type matter? Undergraduate student perception of graduate teaching assistants and professors. *CBE Life Sci Educ* 11, 187–199.
- Kendall KD, Schussler EE (2013). More than words: probing the terms undergraduate students use to describe their instructors. *Int J Teach Learn Higher Educ* 25(2).
- Keutzer CS (1993). Midterm evaluation of teaching provides helpful feedback to instructors. *Teach Psychol* 20, 238–240.
- Kogan LR, Schoenfeld-Tacher R, Hellyer PW (2010). Student evaluations of teaching: perceptions of faculty based on gender, position, and rank. *Teach Higher Educ* 15, 623–636.
- Marsh HW, Roche LA (1997). Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. *Am Psychol* 52, 1187–1197.
- Muzaka V (2009). The niche of graduate teaching assistants (GTAs): perceptions and reflections. *Teach Higher Educ* 14, 1–12.
- Nussbaum JF (1992). Effective teacher behaviors. *Commun Educ* 41, 167–180.
- O'Neal C, Wright M, Cook C, Perorazio T, Purkiss J (2007). The impact of teaching assistants on student retention in the sciences: lessons for TA training. *J Coll Sci Teach* 36, 24–29.
- Park C (2002). Neither fish nor fowl? The perceived benefits and problems of using graduate teaching assistants (GTAs) to teach undergraduate students. *High Educ Rev* 35, 50–62.
- Pattison P, Hale JR, Gowens P (2011). Mind and soul: connecting with students. *J Legal Stud Educ* 28, 39–66.

President's Council of Advisors on Science and Technology (PCAST) (2012). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President. www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf (accessed 3 April 2012).

Rushin JW, Saix JD, Lumsden A, Streubel DP, Summers G, Bernson C (1997). Graduate teaching assistant training: a basis for improvement of college biology teaching and faculty development? *Am Biol Teach* 59, 86–90.

Seymour E, Hewitt N (1997). *Talking about Leaving: Why Undergraduates Leave the Sciences*, Boulder, CO: Westview.

Sundberg MD, Armstrong JE, Wischusen EW (2005). A reappraisal of the status of introductory biology laboratory education in U.S. colleges & universities. *Am Biol Teach* 67, 525–529.

Umbach PD (2007). How effective are they? Exploring the impact of contingent faculty on undergraduate education. *Rev Higher Educ* 30, 91–123.

Walker R (2008). Twelve characteristics of an effective teacher: a longitudinal qualitative, quasi-research study of in-service and pre-service teachers' opinions. *Educ Horizons* 87, 61–68.

Zabaleta F (2007). The use and misuse of student evaluations of teaching. *Teach Higher Educ* 12, 55–76.