ResearchGate

See discussions, stats, and author profiles for this publication at:
https://www.researchgate.net/publication/284478514

Person Perception in the College Classroom: Accounting for Taste in Students' Evaluations of Teaching Effectiveness

Article · January 2009

CITATION
1

READS
12

4 authors, including:

Edward Orehek
University of Pittsburgh

44 PUBLICATIONS **402** CITATIONS

SEE PROFILE

Jennifer Gross
Grand Valley State Univer…

11 PUBLICATIONS **195** CITATIONS

SEE PROFILE

All content following this page was uploaded by Edward Orehek on 24 November 2015.

The user has requested enhancement of the downloaded file.

Person Perception in the College Classroom: Accounting for Taste in Students' Evaluations of Teaching Effectiveness

Jennifer Gross,[1] Brian Lakey, and Kristin Edinger
Grand Valley State University

Edward Orehek
University of Maryland

Denise Heffron
Grand Valley State University

An important applied aspect of person perception occurs when college students evaluate their professors' teaching. Student evaluations of teaching typically are conceptualized as reflecting the characteristics of professors. Yet, this view overlooks the possibility that teaching evaluations also reflect the personal tastes of students, manifested as systematic disagreement among students. Large effects of personal tastes are routinely observed in person perception research and, therefore, should be expected in students' evaluations of teaching. This article describes 3 studies in which students evaluated the same professors' teaching effectiveness. In each study, students' evaluations were strongly influenced by their personal tastes regarding teaching. Moreover, personal tastes in teaching were related in meaningful ways to students' positive affect and memory for lectures.

Nearly all colleges and universities in the United States use students' evaluations of teaching effectiveness as part of the evidence for tenure and promotion decisions (Wilson, 1998). The psychometric properties of such scales are impressive. The scales are reliable (Cashin, 1995; Marsh, 1984) and when rating the same professors, there is reasonable agreement among current students, other faculty, administrators, and alumni (Centra, 1975; Feldman, 1989b).

Further, professors' scores on students' evaluations correlate substantially with students' learning, when learning is assessed by standardized exams in multi-section courses using a common syllabus (Cohen, 1981; Feldman, 1989a). Scores derived from these measures typically are averages taken across different observers and thereby capture the aspect of teaching effectiveness that reflects interrater agreement. This is appropriate when one is interested only in the aspect of teaching effectiveness that reflects agreement. Yet, there is likely other meaningful variance in teaching evaluations

[1]Correspondence concerning this article should be addressed to Jennifer Gross, Department of Psychology, One Campus Drive, Grand Valley State University, Allendale, MI 49401. E-mail: grossj@gvsu.edu

Journal of Applied Social Psychology, 2009, **39**, 7, pp. 1609–1638.
© 2009 Copyright the Authors
Journal compilation © 2009 Wiley Periodicals, Inc.

that would be useful and interesting to understand. For example, research using the social relations model (Kenny, 1994) has shown that in addition to interrater agreement, person perception also reflects the personal tastes of perceivers to a very large degree. Thus, one might expect that students' evaluations of college teaching might also involve personal tastes. If so, it would be important to understand the size of the influence of such tastes, as well as their correlates.

Understanding the role of personal tastes in teaching evaluations should help users better understand the psychological constructs that these measures reflect, as well as potentially improve the effectiveness of college teaching. In the studies described here, we estimate the magnitude of students' personal tastes when evaluating teaching effectiveness and partially account for tastes by linking them to students' affect and memory for lectures. To do this, we applied the little known, yet powerful techniques from Cronbach, Gleser, Nanda, and Rajaratnam's (1972) multivariate generalizability theory.

The apocryphal expression "There's no accounting for taste" is commonly taken to mean that people's personal tastes (especially tastes with which we disagree) cannot be understood in any rational way and, therefore, are beyond psychological analysis. This may explain why there appears to be no psychological literature on personal tastes. Yet, the present studies show how generalizability (G) theory (Cronbach et al., 1972) and Kenny's (1994) social relations model (SRM) provide a framework for analyzing personal tastes psychologically.

First, G/SRM approaches provide a quantitative definition of personal tastes: the extent to which people have different profiles of preferences for the same set of stimuli. Second, G/SRM approaches can isolate the component of preferences that reflects personal tastes, from the component that reflects perceivers' tendency to see all stimuli as good or bad, and from the component that reflects interrater agreement. Third, multivariate G theory provides a method for investigating the correlates of personal tastes, thereby permitting the establishment of the construct validity of personal tastes (Cronbach & Meehl, 1955). In the present studies, we focus on personal tastes in teaching because such tastes can play an important role in tenure and promotion decisions, and might also play a role in developing more effective instruction.

The important role of personal tastes in person perception has been documented most thoroughly by research on the SRM (Kenny, 1994), inspired in part from Cronbach's (1955) work on accuracy in person perception, and Cronbach et al.'s (1972) G theory (Kenny, Mohr, & Levesque, 2001). When perceivers judge the same targets, SRM and G theory typically partition variance into three distinct effects: perceiver, target, and relationship. Relationship and perceiver effects reflect systematic disagreement among observers. Relationship effects capture well the concept of personal tastes because tastes imply that different perceivers display different profiles in evaluating the same stimuli. For example, one perceiver may like classical music, hate country music, and be indifferent to jazz; whereas another perceiver may display the opposite pattern. As applied to perceptions of teaching, personal tastes would be reflected when Student A perceives Professor A as more effective than Professor B, whereas Student B prefers Professor B.

Personal tastes are different from the second source of systematic disagreement: perceiver effects. Perceiver effects reflect the extent to which perceivers differ in their ratings averaged across the same targets. As applied to teaching evaluations, Student A may be consistently more generous in rating professors than is Student B. Target effects reflect interrater agreement. As applied to teaching evaluations, target effects reflect the extent to which observers agree that some professors are more effective than others.

There are strong effects of personal tastes in the perception of target personality, including each of the Big Five dimensions (Kenny, 1994; Paulhus & Reynolds, 1995), as well as interpersonal characteristics (e.g., target supportiveness; Lakey, McCabe, Fisicaro, & Drew, 1996), and the extent to which targets elicit secure attachment (Cook, 2000). However, to our knowledge, the role of personal tastes in students' teaching evaluations has not yet been investigated. Beyond estimating the magnitude of any such effects, it is important to begin to develop an understanding of the psychology of personal tastes in teaching by estimating some of the correlates of taste.

Some observers wonder if the very strong effects for personal tastes in SRM studies truly reflect meaningful psychological variance, or merely the momentary whims of study participants, or perhaps only error variance. Whim and error do not reflect personal tastes in the sense that we mean it. Instead, the concept of *personal tastes* implies that tastes are related to other aspects of psychological experience. To elaborate on an earlier example, one might question the psychological meaning of expressed tastes in music, if a stated preference for classical versus country music was not related to concert attendance and CD purchases. Thus, a major goal of the studies described here is to link personal tastes regarding teaching effectiveness to students' affective reactions to and memory for professors' lectures.

Several recent studies have used Cronbach et al.'s (1972) multivariate G analyses to investigate the nomological network of personal tastes as applied to social support. For example, Neely et al. (2006) found that personal tastes regarding which providers were supportive were related to perceivers' positive affect when conversing with those providers. Similarly, Branje, van Lieshout, and van Aken (2005), as well as Lakey, Lutz, and Scoboria (2004) found that personal tastes regarding which providers were supportive were related to perceivers' personal tastes regarding which providers were more agreeable. Such findings are inconsistent with the idea that personal tastes, as reflected in relationship effects, reflects mere whim or error, and instead indicates that personal tastes reflect psychologically meaningful phenomena in how perceivers differ in evaluating the same stimuli.

In the present research, we used multivariate G analyses to investigate the extent to which personal tastes in teaching effectiveness were linked to students' memory for lectures and affect during lectures. Research in social cognition has suggested that links between teaching evaluations and affect should occur. Affect is related to a wide range of evaluations, including life satisfaction, consumer goods, other people (Clore, Schwarz, & Conway, 1994; Forgas, 2001; Forgas & Bower, 1987), as well as college teaching effectiveness (Fortunato & Mincy, 2003). Thus, if personal tastes play an important role in teaching evaluations, tastes should be reflected in students' affective reactions to professors. That is, Professor A might elicit favorable affect in Student A, but not Student B; whereas Professor B might elicit favorable affect in Student B, but not Student A.

In addition, we expected that affect would play a role in any link between personal tastes in teaching effectiveness and memory for lectures. Here, we assumed that students' affective reactions to professors occur both when students listen to lectures, as well as when students think about the material later: either when studying or taking an exam. If so, affect might promote memory through one of two general mechanisms (Bower & Forgas, 2001). First, in mood-dependent retrieval, material is remembered well when the context at retrieval corresponds to the context at encoding (Barnett & Ceci, 2002; Bower, 1981; Eich, 1995), according to the encoding specificity principle (Tulving & Thomson, 1973). For example, Bower (1981) reported that students who learned words in happy or sad moods showed better memory for the words when students' moods at testing matched the moods during learning. As applied to memory for lectures, taking an exam in the presence of a given professor might stimulate favorable affect in some students. Memory for lectures should be enhanced insofar as the affect experienced during the exam was also experienced when students listened to or studied lectures.

Second, in mood-congruent processing (Bower & Forgas, 2001), mood-congruent information is processed more deeply, and depth of processing promotes memory (Craik & Lockhart, 1972). For example, when a professor elicits positive affect in some students, this affect might be associated with lecture content; therefore, those students might think about the material more deeply. We expect that one or both of these mechanisms operate in college teaching; thus, personal tastes regarding teaching should be related to more favorable affect and to better memory for favored professors, compared to disfavored professors. We expect that these processes will be most applicable to positive affect (Forgas & Bower, 1987; Fortunato & Mincy, 2003). Although negative affect might also promote memory through mood-congruent processes, negative affect in testing situations also increases the accessibility of off-task thoughts that impair performance on tests of memory (Sarason, Pierce, & Sarason, 1996).

Overview of Current Studies

The goal of Study 1 is to provide an initial estimate of the magnitude of personal tastes in students' evaluations of college teaching. In Study 1, two samples of students rated professors (from whom all students had taken courses) on a widely used measure of teaching effectiveness. The goal of Studies 2 and 3 is to examine the extent to which personal tastes in teaching are related to memory for and affect during lectures.

In Study 2, guest lecturers visited an introductory psychology class and students completed measures of affect, teaching evaluations, and memory for lectures. Study 3 examined the same constructs as Study 2, but provided a higher degree of experimental control. Although not our primary interest, the current studies also provided estimates for students' tendency to perceive all professors as good or bad (i.e., student effects), as well as for interrater agreement among students on the effectiveness of specific professors' teaching. We emphasize the role of personal tastes in the present studies because of the large magnitude of personal tastes in person perception (Kenny, 1994; Kenny et al., 2001) and because, to our knowledge, personal tastes have not yet been studied in research on teaching evaluations.

Study 1

Method

Participants

Study 1 consists of two samples. Although both samples were small in an absolute sense, G studies typically have high power because they use repeated-measures experimental designs. Smaller samples have yielded significant effects of personal tastes for other measures of person perception (Lakey, Drew & Sirl, 1999). None of the current investigators were among the faculty rated in the study. All participants received compensation of $5.

Sample A was composed of 17 undergraduate psychology majors (14 women, 3 men) from a regional Midwestern state university. The 17 students represented approximately 50% of the students who had taken at least one course from the same three professors who routinely taught courses required of all majors. It was necessary to have all students rate the same professors because isolating the effects of personal tastes requires a fully crossed design. The potential pool of students was determined by checking the registrar's records for seniors and alumni (within 1 year of graduation) to identify those who had taken courses from all three professors. We focused on senior students and recent graduates because one of the professors taught a course taken predominantly by seniors.

On average, students had taken between 1.1 and 1.4 courses from each professor, and between 1.1 and 1.6 years had elapsed since taking their last course from a given professor. Although the lag between completing a course and rating faculty may seem substantial, Feldman (1989b) found that ratings just after completing a course corresponded well to ratings by alumni years after ($r = .69$).

Sample B was composed of 13 graduate students (9 women, 4 men), who were enrolled in a doctoral program in clinical psychology at a large, urban Midwestern state university. Students rated the same three professors from whom all had taken courses. The 13 students represented approximately 25% of the entire body of students who had taken courses from the three faculty members. On average, students had taken between 1.4 and 2.4 courses from each faculty member, and rated professors between 1.1 and 1.5 years after taking their last course with a given professor.

Procedure

Instructions and consent forms emphasized that students' ratings would remain confidential, and that faculty would not learn of the ratings. Before rating each professor, students answered a series of questions about the physical and social context of the classroom in which the class met. Because knowledge has many contextual determinants (e.g., physical, social, and semantic contexts; Barnett & Ceci, 2002), we hoped that answering these questions would render more accessible participants' memory for their evaluations of each professor. Sample questions are "Where did you sit in class?"; "Whom did you sit next to in class?"; "Did the room have traditional desks or nontraditional seating?"; "Approximately how many students attended class?"; and "Was the classroom equipped with a chalkboard or a whiteboard?"

Measures

Students rated professors' teaching effectiveness using the Students' Evaluations of Educational Quality (SEEQ; Marsh, 1982a). The SEEQ is among the most widely used measures of teaching effectiveness. Extensive evidence attests to the measure's reliability and validity (Marsh, 1984). Internal consistency (i.e., generalizability across items) of the scale in Sample A was .97, .95, and .00 for personal tastes, interrater agreement, and students, respectively.[2] Generalizability across items was .00 for students because, as will be described in the Results, there was no student variance in Sample A.

In Sample B, items on the SEEQ that belonged to the exams/grading subscale were not administered because these items were less applicable to the classes taught in the doctoral program from which students were sampled. Generalizability across the SEEQ items used in Sample B was .98, .96, and .94 for personal tastes, interrater agreement, and students, respectively.

Analytical Strategy

Scores on the SEEQ (Marsh, 1982a) were analyzed as a fully crossed G study measuring Students × Professors × Items (Cronbach et al., 1972) using ordinary least squares ANOVA. The small samples argued against the use of restricted maximum likelihood estimation (DeShon, Ployhart, & Sacco, 1998), as used in Studies 2 and 3. Each student served as a level of the between-subjects Students factor, and each professor served as a level of the within-subjects Professors factor. The Professors factor had three levels in both samples, whereas the Students factor had 17 levels in Sample A and 13 levels in Sample B. In both samples, we combined odd and even numbered items into two aggregates to simplify the design and to reduce measurement error, resulting in two levels for the within-subjects Items factor.

The Students × Professors interaction, indicating the extent to which students systematically disagreed on the relative effectiveness of the same professors, captured the effects of personal tastes. In addition, the design also

[2]In most psychological research, internal consistency (i.e., generalizability across items) is only estimated for differences among persons (i.e., students in the present studies). However, because the present research estimated effects for students, interrater agreement (i.e., professors), and personal tastes (i.e., Students × Professors), we calculated generalizability across items for each component. The formulas were

$$\alpha_S = (\sigma^2{}_S/((\sigma^2{}_S + (\sigma^2{}_{S\times I}/n_i) + (\sigma^2{}_{S\times P\times I}/n_i))$$

$$\alpha_P = (\sigma^2{}_P)/((\sigma^2{}_P + (\sigma^2{}_{P\times I}/n_i) + (\sigma^2{}_{S\times P\times I}/n_i))$$

$$\alpha_{S\times P} = (\sigma^2{}_{S\times P})/((\sigma^2{}_{S\times P} + (\sigma^2{}_{S\times P\times I}/n_i))$$

in which S indicates students, P indicates professors, and I indicates items. These generalizability estimates presage the study results because the terms in the equations refer to the study effects.

yielded two other substantive effects. First, interrater agreement was reflected in the Professors main effect (also known as Target and Partner effects in SRM). The Students main effect (also known as Actor and Perceiver effects in SRM) indicate the extent to which students differed in their ratings of professors, averaged across professors and items. The design also yielded a variety of effects for items (e.g., Students × Items and Professors × Items interactions), but such terms typically are construed as measurement error.

The typical G study has one observation per cell and, therefore, has no within-subjects error term. Instead, the highest order interaction is used as the error term (Kenny, 1994), which was the Students × Professors × Items interaction in the present design. Quasi *F*s were used to test the significance of main effects because in three-factor random designs, the mean squares for main effects are confounded with 2 two-way interactions. However, the mean squares for the Students × Professors interaction are confounded only with the error term. Therefore, true *F* tests are appropriate for such effects. Formulas for degrees of freedom for quasi-*F*s typically result in fractions, so we rounded degrees of freedom down to the nearest whole number.

Results and Discussion

Both Samples A and B showed very strong and significant effects of personal tastes, as displayed in Table 1. In addition, both samples displayed substantial, statistically significant interrater agreement. There was no significant Students effect for Sample A, whereas the Students effect was significant in Sample B. Analyses for each SEEQ subscale yielded results nearly identical to the results of analyses of all items.

Thus, Study 1 provided initial evidence for strong effects of students' personal tastes in teaching evaluations. The magnitude of this effect was quite large: larger than typically reported for ratings of targets' personality characteristics (Kenny, 1994; Kenny et al., 2001), and more similar to ratings of liking (Kenny, 1994) and target supportiveness (Lakey et al., 1996). Study 1 also replicated previous findings of substantial interrater agreement in teaching evaluations (Centra, 1975; Feldman, 1989b; Gillmore, Kane, & Naccarato, 1978; Marsh, 1982b), but the magnitudes observed in Study 1 were somewhat larger than those in previous studies, which typically accounted for approximately 25% of the variance (correlations in the $r = .50$ range). Thus, although students' evaluations of professors' teaching effectiveness partly reflected interrater agreement, such perceptions were also strongly a matter of personal tastes. The implications of these findings for the use of student ratings of professors in personnel evaluation will be considered in the General Discussion.

Table 1

Source, Variance Components, and Proportion of Variance Accounted for in Teaching Evaluations: Study 1

Source	Variance component	F or F′	df	Proportion of variance
Sample A				
Students	−.027	—[a]	—[a]	.000
Professors (interrater agreement)	.191	325.00*	1, 32	.365
Items	.003	7.50	1, 32	.005
Personal tastes (Students × Professors)	.311	33.68*	32, 32	.594
Students × Items	−.001	—[a]	—[a]	.000
Professors × Items	−.001	—[a]	—[a]	.000
Students × Professors × Items (error)	.019			.036
Sample B				
Students	.106	65.00*	3, 24	.170
Professors (interrater agreement)	.249	649.00*	1, 24	.401
Items	−.002	—[a]	—[a]	.000
Personal Tastes (Students × Professors)	.251	46.55*	24, 24	.403
Students × Items	.000	1.00	12, 24	.001
Professors × Items	.005	7.00	2, 24	.008
Students × Professors × Items (error)	.011			.018

Note. In calculating proportion of variance explained, negative variances were rounded to zero.
[a]Not calculated because the variance estimate was negative.
*$p < .05$.

Study 2

Given Study 1's strong evidence for students' personal tastes in teaching evaluations, Study 2 was designed to test for links among personal tastes in teaching evaluations, and students' memory for and affect during lectures. As applied to student affect, personal tastes mean that students differ in their profiles of affective reactions to the same professors. For example, Student A might respond with positive affect to Professor A's lectures, but not to Professor B's lectures. In contrast, Student B might display the opposite pattern.

Some studies have found effects of personal tastes for both positive (Neely et al., 2006) and negative affect (Ingraham & Wright, 1987) in non-teaching contexts, so we expected to find similar effects when students listened to lectures.[3] An effect of personal tastes on memory occurs when perceivers systematically differ in how well they remember the same details of each stimulus within a sequence of stimuli. As applied to teaching, Student A might have better memory for Professor A's lectures than for Professor B's lectures, but Student B might have better memory for Professor B's lectures than for Professor A's lectures.

To our knowledge, there is only one published SRM paper on person memory (Bond, Dorsky, & Kenny, 1992), but the study was not conducted in a teaching context, and it did not find evidence for an effect of personal tastes on memory. However, given that a common goal among students is to remember lectures—and given that Study 1 identified strong effects of personal tastes in teaching evaluations—we expect a significant effect for tastes in memory for lectures as well.

Method

Participants

Study participants were 74 students who enrolled in an introductory psychology course at a regional, Midwestern state university. They participated in the study to complete a course requirement for research participation.

Procedure

The students rated each of four professors' (2 female, 2 male) teaching effectiveness and affect immediately after hearing each professor's lecture during a regularly scheduled class meeting. Memory for lectures was tested by a quiz administered at the beginning of the following class period, as well

[3]Although Kenny (1994) discussed relationship effects for *affect*, his use of the term is roughly synonymous with *liking*. When we use the term *affect*, we refer to constructs similar to mood and emotion.

as by items on the next regularly scheduled exam. Exam items that referred to lectures that were part of the study counted toward students' grades, just as other exam items.

Guest professors presented 50-min lectures on topics commonly covered in introductory psychology classes. The lectures were on infant cognitive development, the development of reasoning, eating disorders, and substance abuse. None of the current investigators were among the faculty rated. Instructions and consent forms emphasized that students' ratings would remain confidential, and that faculty would not learn of the ratings.

Measures

Students' evaluations of teaching. As in Study 1, students completed the Students' Evaluations of Educational Quality (Marsh, 1982a). Items on the questionnaire were modified to reflect the fact that students rated professors on the basis of a single lecture, rather than an entire course. For example, the item "You found the course intellectually challenging and stimulating" was rewritten to read "You found the lecture intellectually challenging and stimulating." Only the 23 items that were appropriate for rating a single lecture were administered. Generalizability across items was .95, .75, and .90 for personal tastes, interrater agreement, and students, respectively.

Affect. Participants completed the Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988), a widely used 20-item scale that reflects two mostly independent, factor-analytically derived dimensions of positive and negative affect. Sample adjectives include "interested" and "excited" for positive affect, and "distressed" and "nervous" for negative affect. Instructions asked participants to indicate the extent to which they experienced the affect described by each adjective "during the lecture." For positive affect, generalizability across items was .83, .56, and .82 for personal tastes, interrater agreement, and students, respectively. For negative affect, generalizability across items was .82, .46, and .71 for personal tastes, interrater agreement, and students, respectively.

Memory. Students responded to 11 multiple-choice quiz items and 3 multiple-choice test items for each lecture, written by each guest professor (and modified by the first author). The 3 test items were a subset of the quiz items. Scores for each item were dichotomous (i.e., correct or incorrect). Therefore, we created aggregates of items by averaging the number of correct answers across items.

To isolate the effect of personal tastes from error, we needed a separate Items factor with at least two levels. Therefore, we constructed one level of the Items factor by aggregating across all quiz items, and a second level by aggregating across all test items. For memory, items were nested within professors because each professor covered different material; therefore, test items were unique to each professor. This required a Students × Items:Professors design, which yielded an internal consistency estimate only for generalizability across items for the students component.[4] Generalizability across items was .14 for students. This low generalizability estimate was not a concern, however, as the study was designed to detect Students × Professors interactions (i.e., personal tastes), rather than Students effects. An internal consistency estimate for personal tastes is not available in this design because there is no Students × Professors × Items interaction term.

Analytical Strategy

Tests of hypotheses about the effect of personal tastes on teaching evaluations and affect were conducted as in Study 1, except that the larger sample size permitted the use of restricted maximum likelihood estimation, using the variance components procedure within SPSS Version 13.0 (2005). As just described, memory items were nested within professors, yielding five variance components: Students; Items:Professors; Professors; Students × Professors; and Students × Items:Professors. The last variance component (i.e., Students × Items:Professors) served as the error term because it was the highest-order interaction.

Tests of hypotheses about correlations among the personal tastes components of constructs were conducted using multivariate G analyses (Brennan, 2001a; Cronbach et al., 1972; Strube, 2000). Following previous research (Lakey et al., 2004; Lakey & Scoboria, 2005; Neely et al., 2006), we calculated multivariate G correlations using the computer program mGENOVA (Brennan, 2001b). We analyzed the data as a $p^\bullet \times i^\bullet$ multivariate G design, as described by Brennan (2001a), for which students were treated as p, and professors were treated as i.

To reduce the complexity of the design, items were not treated as factors in these analyses. The variables were teaching evaluations, memory for lectures, and students' affect. Significance tests for multivariate G correlations (ρ) were conducted using the normal approximation bootstrap method (Mooney & Duval, 1993), because traditional parametric significance tests

[4]The formula for generalizability across items for the student component of the memory measure was

$$(\sigma^2_S/((\sigma^2_S + (\sigma^2_{S\times I:P}/n_i)$$

in which S indicates students, and I:P indicates items nested within professors.

are not available for these correlations (Lakey et al., 2004; Lakey & Scoboria, 2005; Neely et al., 2006).

Results

We first examined the extent to which there were effects of personal tastes on teaching evaluations, affect, and memory (Table 2). As in Study 1, there was a strong and significant effect of personal tastes on teaching evaluations. In addition, there were significant effects of tastes for both positive and negative affect experienced by students, as well as for memory for lectures.[5] That is, students differed in their profiles of affective responses and memory for lectures across the same professors. Although not the focus of the present study, there were also significant students effects for teaching evaluations, positive affect, and negative affect, indicating that students differed in their characteristic evaluations and affect across the four lectures. Effects for interrater agreement were typically small and not significant for all constructs.

Next, we examined the extent to which the effects of personal tastes on teaching evaluations, affect, and memory were interrelated. As displayed in Table 3, there was a strong and significant correlation between the personal tastes components of positive affect and teaching evaluations. That is, the student–professor dyads that elicited the most positive affect in students also elicited the most favorable teaching evaluations. In addition, there was a marginally significant correlation ($p < .10$) between the effects of personal tastes on memory and teaching evaluations, such that the student–professor dyads that elicited better memory for lectures also elicited more favorable teaching evaluations. No other correlations were significant for personal tastes.

Although not the focus of the study, we should note that teaching evaluations and positive affect were also related when both reflected individual differences among students. That is, the students who characteristically

[5]To investigate the extent to which the effects for personal tastes might be explained by a pattern by which students of a given gender responded more favorably to professors of a given gender (e.g., female students might prefer female professors), we conducted Student Gender × Professor Gender ANOVAs with fixed factors for each of our dependent variables. There were small but significant Student Gender × Professor Gender interactions for teaching evaluations and positive affect, accounting for 1.8% and 2.4% of the variance, respectively. Interactions for negative affect and memory were not significant. For positive affect, t tests reveal that female professors elicited more positive affect in female students than did male professors, whereas male and female professors did not differentially influence male students. For teaching evaluations, the significant Student Gender × Professor Gender interaction was driven by a single professor. For female students only, one professor had much lower scores than did the other three, among whom there were no significant differences.

Table 2

Source, Variance Components, and Proportion of Variance Accounted for in Teaching Evaluations: Study 2

Source	Variance component	*SE*	Proportion of variance
Teaching evaluations			
Students	.084*	.021	.301*
Professors (interrater agreement)	.027	.024	.096
Items	.000	.000	.000
Personal tastes (Students × Professors)	.148*	.015	.528*
Students × Items	.003*	.001	.012*
Professors × Items	.002	.001	.006
Students × Professors × Items (error)	.016	—	.058
Memory for lectures			
Students	.002	.001	.041
Professors (interrater agreement)	.016	.013	.296
Items:Professors	.000	.000	.000
Personal tastes (Students × Professors)	.009*	.002	.168*
Students × Items:Professors (error)	.027	—	.494
Positive affect			
Students	.266*	.058	.370*
Professors (interrater agreement)	.062	.057	.086
Items	.038	.057	.053
Students × Professors (personal tastes)	.224*	.026	.312*
Students × Items	.031*	.009	.043
Professors × Items	.007	.007	.010
Students × Professors × Items (error)	.090	—	.126
Negative affect			
Students	.083*	.022	.276*
Professors (interrater agreement)	.024	.021	.079
Items	.000	.000	.000
Students × Professors (personal tastes)	.125*	.015	.417*
Students × Items	.013*	.005	.043*
Professors × Items	.000	.000	.000
Students × Professors × Items (error)	.056	—	.185

*$p < .05$.

Table 3

Multivariate Generalizability Correlations: Study 2

	1	2	3
1. Teaching evaluations			
Students	—		
Interrater agreement	—		
Personal tastes	—		
2. Memory for lectures			
Students	−.18 (.30)	—	
Interrater agreement	NC	—	
Personal tastes	.10† (.06)	—	
3. Positive affect			
Students	.31* (.13)	NC	—
Interrater agreement	NC	NC	—
Personal tastes	.53* (.05)	.10 (.07)	—
4. Negative affect			
Students	−.23 (.15)	NC	.07 (.15)
Interrater agreement	NC	NC	NC
Personal tastes	−.05 (.06)	−.03 (.07)	−.01 (.05)

Note. Standard errors appear in parentheses. NC = not calculated because there were no univariate effects for one or both of the variables.
†$p < .10$. *$p < .05$.

experienced the most positive affect also characteristically evaluated teaching most favorably. There were no other significant correlations involving students components. We did not calculate correlations among the components of constructs that reflected interrater agreement because there was no significant interrater agreement for any of the constructs (Kenny, 1994). It is not meaningful to estimate correlations among components that likely reflect only randomness. In Table 3, "NC" indicates correlations that were not calculated for this reason.

Discussion

The primary goal of Study 2 was to investigate links among teaching evaluations, memory, and affect when these constructs reflected the effects of personal tastes specifically. Consistent with predictions, each professor elicited positive affect in some students, but not others; and when a professor elicited positive affect in a given student, that student also evaluated the professor's teaching more favorably. There was also a trend whereby when Student A evaluated Professor A's lecture more favorably than Professor B's lecture, Student A also displayed better memory for Professor A's lecture than for Professor B's lecture. However, contrary to expectations, the extent to which a given student experienced positive affect in response to a given professor was not related to the student's memory for that particular professor's lecture.

The link between positive affect and teaching evaluations is consistent with a large body of work indicating that affect is related to a range of evaluative judgments (Clore et al., 1994; Forgas, 2001; Forgas & Bower, 1987), including teaching evaluations (Fortunato & Mincy, 2003). However, Study 2 went beyond these findings by documenting links between affect and judgment when the constructs specifically reflected personal tastes. This finding is consistent with the results of Neely et al. (2006), who found that recipients' personal tastes regarding provider supportiveness were linked to recipients' affect when interacting with specific providers.

Study 2 was silent regarding the mechanism for the correlations between student affect and teaching evaluations. Previous research has indicated several possibilities, however. Substantial research has indicated that participants use their current affective state as information in making judgments (Clore et al., 1994). In this case, students might have used a "How do I feel about this professor?" heuristic when judging professors. For example, "If I feel good, this professor must have been an effective teacher." Forgas' (2001) affect infusion model describes the conditions under which such a process is likely to occur. The model predicts that affect is most likely to influence judgments when there is no previously made judgment available to retrieve from memory, and the response requires "more constructive and open-ended information search strategies" (Forgas, 2001, p. 298). Study 2 was designed to encourage such processes by presenting professors with whom students had no previous contact, thereby prohibiting the retrieval of previous judgments from memory.

As predicted, there was also a significant effect of personal tastes on memory for lectures. We should note, however, that this effect was observed only for analyses that involved both quiz and test scores. There was no evidence for personal tastes on memory when only quiz scores were analyzed. Test scores alone could not be analyzed, because there were too few test items to create aggregates that had appropriate distributional properties. Moreover, there was only a weak and marginally significant link between teaching evaluations and memory, and no evidence for our hypothesized link between affect and memory for lectures.

These weak effects involving memory for lectures might have resulted from Study 2's realism, which also introduced substantial imprecision in the design, especially with regard to the assessment of personal tastes. For example, professors were confounded with order, topic, and item difficulty. That is, professors were presented to students in a specific order, taught only one topic, and wrote their own quiz and test items. Students might have differed systematically in how their memories were influenced by these factors. Thus, our estimates of the effects of tastes on memory (i.e., Students × Professors interactions) were confounded with Students × Orders, Students × Topics, and Students × Item Difficulty interactions. These other effects might have obscured the effects of tastes on memory. Nonetheless, the marginal link between teaching evaluations and memory for lectures suggests that further study is warranted. Study 3 provided a more highly controlled assessment of these relations.

Consistent with previous studies (Ingraham & Wright, 1987; Neely et al., 2006), Study 2 found no significant interrater agreement on the extent to which different stimulus people elicited the same affective response among perceivers. Instead, the extent to which a given stimulus person elicited positive or negative affect appeared to be largely a matter of the personal tastes of perceivers. Study 2 also found significant individual differences among students in the extent to which they experienced positive and negative affect across the four professors. These findings are consistent with both Neely et al. and Lakey and Scoboria (2005), who also found similar effects when support recipients rated support providers.

An important question is the extent to which professors' elicitation of affect in a given student is related in an obvious way to behaviors that one normally thinks of as effective teaching. For example, a given professor might elicit favorable affect in a given student because the student likes the way the professor approaches a given topic. One student might prefer a more quantitative approach while another student might prefer a more qualitative approach, and a given professor might elicit positive affect and favorable perceptions because the professor matches the preferences of a given student.

However, it is also possible that the students' affect is elicited by factors irrelevant to what we think of as teaching. For example, a given professor's physical appearance, manner of speech, or expressed interests in music or politics might elicit favorable affect in some students, but not others; and this affect might drive perceptions of teaching. Future research should attempt to distinguish among these mechanisms, and the multivariate G approach used in the current study would be a useful tool.

Study 3

Study 2 found effects of personal tastes on teaching evaluations, memory for lectures, and students' affect, as well as a strong correlation between the tastes components of perceivers' positive affect and teaching evaluations. There was also a marginally significant link between the tastes components of teaching evaluations and memory for lectures. Study 3 is designed to replicate these links in a study with greater experimental control. We expect that greater experimental control will capture stronger links involving the effects of tastes on memory.

Method

Participants

Study participants were 41 undergraduate students (26 women, 15 men) who were enrolled in introductory psychology classes at a regional, Midwestern state university. They participated in the study to complete a course requirement for research participation.

Procedure

Participants viewed professional-quality videotapes of three college-level lectures projected onto a large movie screen. The three lectures were selected from a commercially available series, and included lectures on Isaac Newton's cosmology, John Locke's theory of politics, and Thomas More's *Utopia*. Professors at well-known colleges and universities delivered the lectures. The videos were edited to be similar in length and amount of content. Presentation order was counterbalanced.

Immediately after viewing each of the video lectures, participants completed measures in the following order: (a) evaluations of teaching effectiveness; (b) affect; and (c) memory for lecture. Participants completed the teaching evaluation measure first, in an effort to minimize purely artifactual influences in which participants inferred their evaluations directly from observing their own responses to the memory or affect measures.

Measures

Teaching evaluations. Participants completed the SEEQ (Marsh, 1982a), as in Study 2. Items that could not be modified to reflect a video-recorded lecture were not administered (e.g., "Students were encouraged to participate in class discussions"). Generalizability across items was .93, .86, and .79 for personal tastes, interrater agreement, and students, respectively.

Memory for lectures. We created 10-item multiple-choice tests for each of the three lectures. Following Williams and Ware (1978), we rewarded students with $0.10 for every correct answer to provide increased incentive for students to attend to the lectures. Students were informed of this contingency prior to viewing the videos. A colorful poster-board reminder of the reward contingency was in full view at all times.

Unlike Study 2, the investigators wrote the test items to be of approximately equal difficulty and to sample the entire lecture. As in Study 2, test items were nested within professors because each professor covered different material; therefore, the test items were unique to each professor. This required a Students × Items:Professors design that yields an internal consistency estimate only for generalizability across items for the students component. Generalizability across items was .45 for students. This low generalizability estimate was not a concern, however, as the study was designed to detect Students × Professors interactions (i.e., personal tastes).

Affect. Participants completed the PANAS (Watson et al., 1988), as in Study 2. For positive affect, generalizability across items was .88, .58, and .78 for personal tastes, interrater agreement, and students, respectively. For negative affect, generalizability across items was .74, .17, and .66 for personal tastes, interrater agreement, and students, respectively.

Results

Data were analyzed using the same procedures as in Study 2. For all measures, odd and even items were averaged to create an Items factor with two levels.

Consistent with Studies 1 and 2, there was a strong, significant effect of personal tastes on teaching evaluations (Table 4). Also consistent with Study 2, there were significant effects of tastes on both positive affect and negative affect. As expected, there was also a large, significant effect of tastes on memory. Consistent with Study 2, there were also significant students effects for positive and negative affect. In addition, there was also a students effect for memory. There was no significant interrater agreement for any construct.

Multivariate G analyses indicate that the student–professor dyads that elicited the most positive affect also elicited the most favorable teaching evaluations and the best memory for lectures (Table 5). In contrast, the student–professor dyads that elicited the most negative affect also elicited the least positive affect and the least favorable evaluations, but not poorer

Table 4

Source, Variance Components, and Proportion of Variance Accounted for in Teaching Evaluations: Study 3

Source	Variance component	*SE*	Proportion of variance
Teaching evaluations			
Students	.069	.035	.163
Professors (interrater agreement)	.100	.106	.238
Items	.000	.000	.000
Personal tastes (Students × Professors)	.216*	.037	.512*
Students × Items	.004	.004	.009
Professors × Items	.000	.001	.001
Students × Professors × Items (error)	.033	—	.079
Memory for lectures			
Students	.306*	.124	.224*
Professors (interrater agreement)	.000	.000	.000
Items:Professors	.021	.027	.015
Personal tastes (Students × Professors)	.300*	.116	.219*
Students × Items:Professors (error)	.742	—	.542
Positive affect			
Students	.101*	.040	.257*
Professors (interrater agreement)	.041	.049	.104
Items	.008	.015	.020
Students × Professors (personal tastes)	.183*	.033	.462*
Students × Items	.005	.006	.012
Professors × Items	.006	.007	.014
Students × Professors × Items (error)	.052	—	.130
Negative affect			
Students	.049*	.018	.342*
Professors (interrater agreement)	.003	.004	.021
Items	.000	.001	.000
Students × Professors (personal tastes)	.042*	.009	.293*
Students × Items	.020*	.007	.138*
Professors × Items	.000	.000	.000
Students × Professors × Items (error)	.030	—	.207

*$p < .05$.

Table 5

Multivariate Generalizability Correlations: Study 3

	1	2	3
1. Teaching evaluations			
Students	—		
Interrater agreement	—		
Personal tastes	—		
2. Memory for lectures			
Students	NC	—	
Interrater agreement	NC	—	
Personal tastes	.27* (.12)	—	
3. Positive affect			
Students	NC	.21 (.29)	—
Interrater agreement	NC	NC	—
Personal tastes	.76* (.04)	.32* (.09)	—
4. Negative affect			
Students	NC	−.16 (.37)	.06 (.39)
Interrater agreement	NC	NC	NC
Personal tastes	−.50* (.10)	−.12 (.13)	−.44* (.13)

Note. Standard errors appear in parentheses. NC = not calculated because there were no univariate effects for one or both of the variables.
*$p < .05$.

memory for lectures. That is, as in Study 2, the personal tastes components of teaching evaluations and positive affect were significantly and strongly correlated. The personal tastes components of teaching evaluations and memory were also significantly correlated, replicating the marginal effect in Study 2. However, in addition, unlike Study 2, the personal tastes components of positive affect and memory were significantly correlated.

Our final hypothesis was that the effects of personal tastes on positive affect could account for the correlation between the tastes components of teaching evaluations and memory for lectures. If so, then controlling for positive affect should have diminished the correlation between teaching evaluations and memory for lectures. Treating each student–professor dyad as the unit of analysis, we first computed standardized residuals for teaching effectiveness and memory, removing each variable's shared variance with positive affect. We then calculated multivariate G correlations on the residualized scores. As hypothesized, controlling for positive affect eliminated the significant correlation between the personal tastes components of teaching effectiveness and memory for lectures ($\rho = .04$, $SE = .18$), and there was a significant difference between the original correlation and the residualized correlation ($\rho\Delta = .23$, $SE = .11$). We did not conduct parallel analyses for negative affect because the effects of personal tastes on negative affect and memory were not significantly related.

Discussion

Study 3 provides additional evidence that the student–professor dyads that elicit the most positive affect also elicit the most favorable teaching evaluations and the best memory for lectures. Positive affect appeared to play a key role in the link between teaching evaluations and memory for lectures. When positive affect was controlled, the link between the effects of personal tastes on memory for lectures and teaching evaluations was significantly reduced and eliminated.

Compared to Study 2, the stronger experimental control of Study 3 captured stronger effects of personal tastes on memory, as well as stronger intercorrelations among the personal tastes components of teaching evaluations, memory, and affect. To our knowledge, Studies 2 and 3 present the first evidence for the effects of personal tastes on memory. Thus, not only are there individual differences in memory ability, and not only are some people more memorable than others (Bond et al., 1992), but memory also can depend on specific student–professor dyads. However, our studies were not designed to identify the mechanisms by which the effects of tastes were linked to memory. Future research could integrate G/SRM methods with more traditional experimental methods to investigate such factors as encoding (Craik & Lockhart, 1972), arousal (Clark, Milberg, & Ross, 1983), and aspects of context other than affect.

Both Studies 2 and 3 provide evidence for a link between the personal tastes components of teaching evaluations and affect. As discussed previously, this finding is consistent with a large body of research that links affect to a wide range of evaluative judgments (Clore et al., 1994; Forgas, 2001; Forgas & Bower, 1987). Yet, some readers might wonder whether the strong correlation between teaching evaluations and positive affect might reflect item overlap, given that measures of both positive affect and teaching evaluations used some similar terms (e.g., "enthusiastic"). We do not believe that the correlation reflects only overlapping item content because the shared variance between the personal tastes components of teaching evaluations and positive affect was linked to a behavioral measure (i.e., memory) that had no item overlap with positive affect.

Study 3 also provides a fourth replication of effects of personal tastes on teaching evaluations. These effects were strong and consistent, with a median effect size of 41% of the variance and a range of 25% to 53%. Moreover, across the four samples, these effects were obtained with 13 professors and a range of contexts, including graduate and undergraduate courses; single lectures and entire semesters; as well as live and video-recorded lectures. Across the studies, the magnitude of influence for personal tastes compared favorably to the magnitude of interrater agreement.

In the current studies, the median effect size for interrater agreement was 31% of the variance, with a range from 10% to 40%. These magnitudes were similar to interrater agreement for teaching evaluations reported by other investigators, which have typically been approximately .50, reflecting approximately 25% of the variance (Centra, 1975; Feldman, 1989b; Gillmore et al., 1978; Marsh, 1982b). Across our four estimates, the largest values for interrater agreement were observed when students rated faculty after completing entire courses. The smallest values occurred when students rated professors immediately after hearing single lectures. This may mean that interrater agreement for professors' teaching effectiveness increases as students have exposure beyond a single lecture, although the studies differed in other respects, besides amount of exposure. This may be noteworthy because increased consensus among observers with increased exposure to targets has not been commonly observed in studies of person perception (Kenny, 1994; Kenny, Albright, Malloy, & Kashy, 1994).

The nonsignificant interrater agreement in the present Studies 2 and 3 may appear to conflict with the results of Ambady and Rosenthal (1993), who found significant interrater agreement in rating teachers among observers with very brief exposures. However, differences between the current Studies 2 and 3 and those of Ambady and Rosenthal most likely resulted from the low power of the current studies to detect interrater agreement, rather than a difference in effect sizes. For example, the magnitude of interrater agreement for teaching effectiveness in the current Study 3 was comparable in size to the magnitude observed by Ambady and Rosenthal. We would like to emphasize, however, that the current studies were not designed to detect interrater agreement.

General Discussion

It is commonly believed that personal tastes are beyond rational and psychological analysis. However, the present studies demonstrate how

G/SRM approaches make it possible to account for tastes in a psychological sense. This approach permitted the quantification of taste, which permitted estimating the magnitude of its effects. Most importantly, multivariate G analyses (Cronbach et al., 1972) permitted initial mapping of the nomological network of personal tastes in college teaching.

In the present studies, personal tastes in teaching behaved in a meaningful way, as tastes were linked to affect and memory for lectures. Thus, rather than being a mere whim or a reflection of error, personal tastes—as operationalized with the G/SRM tradition—reflects meaningful psychological variance. Of course, these correlates of tastes in teaching may not correspond to any other domain of taste (e.g., music, art). Investigators interested in the role of personal tastes in these judgments will need to map the nomological networks separately for those judgments. The multivariate G approach appears to be especially useful in this regard. We focused on beginning to understand personal tastes in teaching because such tastes can have an important role in tenure and promotion decisions, and might also play a role in developing more effective instruction.

In personnel decisions, students' evaluations of teaching effectiveness are intended to reflect the stable properties of professors; that is, differences among professors that are broadly generalized across students, classes, and time. Consistent with this interpretation is research indicating substantial interobserver agreement in teaching evaluations (Centra, 1975; Feldman, 1989b; Gillmore et al., 1978; Marsh, 1982b). However, the current studies identified strong effects of personal tastes that are likely ignored in most personnel evaluations. This is a problem because a professor who receives below average scores on students' evaluations might, in fact, be extremely effective with a given subgroup of students, yet this information is not preserved when teaching evaluations are tabulated as mean differences among professors, averaged across students.

Furthermore, unless professors are randomly assigned to courses or teach all courses, a given professor's average score might reflect the effects of personal tastes more so than the effectiveness of the professor. That is, professors' scores on teaching evaluations might be confounded with the effects of students' personal tastes and Students × Topics interactions. If professors regularly teach courses (e.g., quantitative vs. clinical) that attract different types of students with different tastes, Students × Topics as well as Students × Professors interactions will likely influence students' evaluations of these professors. For example, a given professor might be especially effective at teaching students drawn to quantitative courses, but less effective at teaching students drawn to clinical courses. Yet, if this faculty member is assigned to teach many sections of clinical courses, the professor's mean evaluation scores would be substantially lower than if he or she were assigned to teach mostly quantitative courses; not because the professor is less effective, but because the professor has been badly matched with courses and students.

The large effects of personal tastes on teaching evaluations and memory for lectures also have implications for improving teaching. To harness the effects of tastes, universities would need a procedure for forecasting which students would do well with which professors. Many departments already attempt to do this in an informal way by selecting one professor to teach honors sections and another professor to teach large introductory sections. Similarly, some students try to take as many courses as possible from a few professors. To expand and formalize such practices, one would need a fast, inexpensive way to forecast effective matches for large numbers of students and professors. For example, in Study 3, the extent to which specific student–professor dyads elicited positive affect was related to perceptions of teaching effectiveness and students' memory for lectures. Perhaps one could forecast the effectiveness of particular student–professor dyads over the course of an entire semester from the extent to which small samples of professors' behaviors elicited positive affect in some students, but not others.

It might be possible to use Internet technology to provide students with brief samples of professors' teaching and collect data on students' reactions. Then, based on students' reactions to these brief samples, students could be matched with specific professors. Ambady and Rosenthal's (1993) research supports the plausibility of this approach by finding that independent judges could accurately forecast teaching evaluations from personality judgments based on 6-s video samples of teachers' behaviors.

Before closing, it is important to acknowledge the limitations of the present studies. First, the design of our studies confounded personal tastes (i.e., Students × Professors interactions) with Students × Topics interactions (Gillmore et al., 1978). Therefore, the effects of personal tastes observed in the present study might reflect student–topic matching as much as student–professor matching. Disentangling the effects of student–professor matching from student–topic matching would require a design in which professors and topics were fully crossed and each professor taught each topic. However, exposing students to different faculty delivering the same lecture would present other methodological problems. Second, each of our studies assessed constructs at a single point in time, and we have assumed that the effects of tastes are stable across occasions. However, recent research has suggested that such influences are only partly stable over time (Neely et al., 2006). Third, the samples of professors in each study were small, and this seriously limited the power to detect interrater agreement. However, detecting interrater agreement was not an important goal of the current studies. Fourth, in Studies 2 and 3, students were assessed after a single lecture, which is quite different from how teaching evaluations are typically used in applied settings. Thus, the phenomena involving affect and memory in Studies 2 and 3 might not generalize to how teaching evaluations are used in personnel decisions. The comparatively small estimates for interrater agreement in Studies 2 and 3 might indicate poor generalizability for those studies.

It is also important to consider the potential role of students' mood state immediately before participating in the study. Students in good moods just before the study began might have rated professors more favorably and might have reported more positive and less negative affect than students in bad moods. Yet, how students' moods might have influenced the study findings is complex. Students' mood states would have had different influences on study findings depending on whether students completed measures in a single session on the same day (Studies 1 and 3), or completed measures for different professors on different days (Study 2). When ratings were made on the same day, students in bad moods might have rated all professors more negatively and might have reported more negative and less positive affect than students in good moods. Such an effect would have inflated the students effects, which capture differences among students, averaged across professors and items.

In contrast, mood would not have inflated interrater agreement or the effects of personal tastes (both of which are based on differences among professors), as long as the effects of mood were applied uniformly across professors. However, mood would have operated differently if students made their ratings on different days (Study 2). In this case, mood state would not have influenced the magnitude of students effects because mood state would not have been constant across all 4 days. Instead, mood would have created differences in ratings of professors, depending on which professor was rated on a given day. For example, Student 1 might have been in a bad mood on Day 1 when he rated Professor 1, and his mood might have led to a more negative than expected evaluation of Professor 1 and more negative than expected reports of Student 1's own affect. Yet, Student 2 might have been in a bad mood on Day 2 when she rated Professor 2, and her mood might have led to a more negative than expected evaluation of Professor 2 and more negative than expected reports of Student 2's own affect. Such a pattern would have magnified the extent to which students disagreed in their ratings of the two professors, inflating the effects of personal tastes, compared to differences among students or interrater agreement. Thus, although previous mood state is a viable alternative hypothesis for the effects of personal tastes in Study 2, mood does not appear to be able to explain the effects of personal tastes in Studies 1 or 3.

In summary, this article describes how G/SRM approaches provide the analytical framework for developing a psychological accounting of taste. We focused on personal tastes in teaching because of the important role students' evaluations of teaching play in tenure and promotion. In fact, there were very strong effects for personal tastes in evaluating teaching. Yet, far from being mere whim or error, personal tastes were related meaningfully to students' memory for and affect during lectures. These findings might have implications for using students' evaluations of teaching for tenure and promotion decisions, as well as for improving the effectiveness of college teaching.

References

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*, 431–441.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for transfer. *Psychological Bulletin*, *128*, 612–637.

Bond, C. F., Dorsky, S. E., & Kenny, D. A. (1992). Person memory and memorability: A round-robin analysis. *Basic and Applied Social Psychology*, *13*, 285–302.

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*, 129–148.

Bower, G. H., & Forgas, J. P. (2001). Mood and social memory. In J. P. Forgas (Ed.), *Affect and social cognition* (pp. 95–120). Mahwah, NJ: Lawrence Erlbaum.

Branje, S. J. T., van Lieshout, C. F. M., & Van Aken, M. A. G. (2005). Relations between agreeableness and perceived support in family relationships: Why nice people are not always supportive. *International Journal of Behavioral Development*, *29*, 120–128.

Brennan, R. L. (2001a). *Generalizability theory*. New York: Springer.

Brennan, R. L. (2001b). *Manual for mGENOVA* (Version 2.1). Iowa City, IA: Iowa Testing Programs.

Cashin, W. E. (1995). *Student ratings of teaching: The research revisited* (IDEA Paper No. 32). Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Centra, J. A. (1975). Colleagues as raters of classroom instruction. *Journal of Higher Education*, *46*, 327–337.

Clark, M., Milberg, S., & Ross, J. (1983). Arousal cues arousal-related material in memory: Implications for understanding effects of mood on memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 633–649.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 2, pp. 323–417). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, *51*, 281–309.

Cook, W. L. (2000). Understanding attachment security in family context. *Journal of Personality and Social Psychology*, *78*, 285–294.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, *52*, 177–193.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

DeShon, R. P., Ployhart, R. E., & Sacco, J. M. (1998). The estimation of reliability in longitudinal models. *International Journal of Behavioral Development*, *22*, 493–515.

Eich, E. (1995). Searching for mood dependent memory. *Psychological Science*, *6*, 67–75.

Feldman, K. A. (1989a). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, *30*, 583–645.

Feldman, K. A. (1989b). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, *30*, 137–194.

Forgas, J. P. (2001). Affect, cognition, and interpersonal behavior: The mediating role of processing strategies. In J. P. Forgas (Ed.), *Affect and social cognition* (pp. 293–318). Mahwah, NJ: Lawrence Erlbaum.

Forgas, J. P., & Bower, G. H. (1987). Mood effects on person–perception judgments. *Journal of Personality and Social Psychology*, *53*, 53–60.

Fortunato, V. J., & Mincy, M. D. (2003). The interactive effects of dispositional affectivity, sex, and a positive mood induction on student evaluations of teachers. *Journal of Applied Social Psychology*, *33*, 1945–1972.

Gillmore, G. M., Kane, M. T., & Naccarato, R. W. (1978). The generalizability of student ratings of instruction: Estimation of the teacher and course components. *Journal of Educational Measurement*, *15*, 1–13.

Ingraham, L. J., & Wright, T. L. (1987). A social relations model test of Sullivan's anxiety hypothesis. *Journal of Personality and Social Psychology*, *52*, 1212–1218.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the Big Five. *Psychological Bulletin*, *116*, 245–258.

Kenny, D. A., Mohr, C. D., & Levesque, M. J. (2001). A social relations variance partitioning of dyadic behavior. *Psychological Bulletin*, *127*, 128–141.

Lakey, B., Drew, J. B., & Sirl, K. (1999). Clinical depression and perceptions of supportive others: A generalizability analysis. *Cognitive Therapy and Research*, *23*, 511–533.

Lakey, B., Lutz, C. J., & Scoboria, A. (2004). The information used to judge supportiveness depends on whether the judgment reflects the personality of perceivers, the objective characteristics of targets, or their unique relationships. *Journal of Social and Clinical Psychology*, *23*, 796–814.

Lakey, B., McCabe, K. M., Fisicaro, S., & Drew, J. B. (1996). Environmental and personal determinants of social support: Three generalizability studies. *Journal of Personality and Social Psychology*, *70*, 1270–1280.

Lakey B., & Scoboria A. (2005). The relative contribution of trait and social influences to the links among perceived social support, affect, and self-esteem. *Journal of Personality*, *73*, 361–388.

Marsh, H. M. (1982a). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, *52*, 77–95.

Marsh, H. M. (1982b). The use of path analysis to estimate teacher and course effects in student ratings of instructional effectiveness. *Applied Psychological Measurement*, *6*, 47–59.

Marsh, H. M. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, *76*, 707–754.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Thousand Oaks, CA: Sage.

Neely, L., Lakey, B., Cohen, J., Barry, R., Orehek, E., Abeare, C., & Meyer, W. (2006). Trait and social processes in the link between social support and affect: An experimental laboratory investigation. *Journal of Personality*, *74*, 1015–1046.

Paulhus, D. L., & Reynolds, S. (1995). Enhancing target variance in personality impressions: Highlighting the person in person perception. *Journal of Personality and Social Psychology*, *69*, 1233–1242.

Sarason, I. G., Pierce, G. R., & Sarason, B. R. (1996). *Cognitive interference: Theories, methods, and findings*. Hillsdale, NJ: Lawrence Erlbaum.

SPSS, Inc. (2005). *SPSS for Windows* (Release 13.0). Chicago: Author.

Strube, M. J. (2000). Reliability and generalizability theory. In G. G. Laurence & P. R. Yarnold (Eds.), *Reading and understanding MORE multivariate statistics* (pp. 23–66). Washington, DC: American Psychological Association.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

Watson, D., Clark, L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Williams, R. G., & Ware, J. E., Jr. (1978). Validity of student ratings of instruction under different incentive conditions: A further study of the Dr. Fox effect. *Journal of Educational Psychology*, *68*, 48–56.

Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *Chronicle of Higher Education*, *44*, A12–A14.

View publication stats