Exploring Qualitative Training Reactions: Individual And Contextual Influences On Trainee Commenting

By: Reanna Poncheri Harman, **J. Kemp Ellington**, Eric A. Surface, and Lori Foster Thompson

Abstract

Training reactions are the most common criteria used for training evaluation, and reaction measures often include opportunities for trainees to provide qualitative responses. Despite being widely used, qualitative training reactions are poorly understood. Recent trends suggest commenting is ubiquitous (e.g., tweets, texting, Facebook posts) and points to a currently untapped resource for understanding training reactions. In order to enhance the interpretation and use of this rich data source, this study explored commenting behavior and investigated 3 broad questions: who comments, under what conditions, and how do trainees comment? We explore both individual difference and contextual influences on commenting and char- acteristics of comments in 3 studies. Using multilevel modeling, we identified significant class-level variance in commenting in each of the 3 samples of trainees. Because commenting has only been considered at the individual level, our findings provide an important contribution to the literature. The shared experience of being in the same class appears to influence commenting in addition to individual differences, such as interest in the topic (Studies 1 and 2), satisfaction (Studies 2 and 3), and entity beliefs (Study 3). Furthermore, we demonstrated that item wording may have an impact on commenting (Study 3) and should be considered as a potential lever for training professionals to influence commenting behavior from trainees. Training professionals, particularly those who regularly administer training evaluation surveys, should be aware of nonresponse to open-ended items and how that may impact the information they collect, use, and present within their organizations.

Harman, R. P., **Ellington, J. K.**, Surface, E. A., & Thompson, L. F. (2015). Exploring qualitative training reactions: Individual and contextual influences on trainee commenting. Journal of Applied Psychology, 100(3), 894-916. http://dx.doi.org/10.1037/a0038380. Publisher version of record available at: http://psycnet.apa.org/record/2014-55443-001?doi=1

Exploring Qualitative Training Reactions: Individual and Contextual Influences on Trainee Commenting

Reanna Poncheri Harman
SWA Consulting Inc., Raleigh, North Carolina

J. Kemp Ellington
Illinois Institute of Technology

Eric A. Surface
SWA Consulting Inc., Raleigh, North Carolina

Lori Foster Thompson
North Carolina State University

Training reactions are the most common criteria used for training evaluation, and reaction measures often include opportunities for trainees to provide qualitative responses. Despite being widely used, qualitative training reactions are poorly understood. Recent trends suggest commenting is ubiquitous (e.g., tweets, texting, Facebook posts) and points to a currently untapped resource for understanding training reactions. In order to enhance the interpretation and use of this rich data source, this study explored commenting behavior and investigated 3 broad questions: who comments, under what conditions, and how do trainees comment? We explore both individual difference and contextual influences on commenting and characteristics of comments in 3 studies. Using multilevel modeling, we identified significant class-level variance in commenting in each of the 3 samples of trainees. Because commenting has only been considered at the individual level, our findings provide an important contribution to the literature. The shared experience of being in the same class appears to influence commenting in addition to individual differences, such as interest in the topic (Studies 1 and 2), satisfaction (Studies 2 and 3), and entity beliefs (Study 3). Furthermore, we demonstrated that item wording may have an impact on commenting (Study 3) and should be considered as a potential lever for training professionals to influence commenting behavior from trainees. Training professionals, particularly those who regularly administer training evaluation surveys, should be aware of nonresponse to open-ended items and how that may impact the information they collect, use, and present within their organizations.

Keywords: commenting behavior, survey comment nonresponse, training evaluation, training reactions, open-ended comments

Learning and development are big investments for U.S. organizations, with expenditures for 2013 estimated at $164.2 billion (Miller, 2013), and training reactions are by far the most commonly used criterion for training evaluation (Goldstein & Ford, 2002), with an American Society of Training and Development (2009; ASTD) study finding that 91.6% of participating organizations used reactions to evaluate learning. Despite widespread use, reliance on reactions has been criticized, primarily on the grounds that reactions have not shown consistently strong relationships with other training criteria (i.e., learning, behavior, and results; Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). However, Kraiger (2002) argued that trainee reactions have value beyond predicting additional levels of training criteria, such as decision making, feedback, and marketing. ASTD (2009) found

This article was published Online First December 15, 2014.

Reanna Poncheri Harman, SWA Consulting Inc., Raleigh, North Carolina; J. Kemp Ellington, Department of Psychology, Illinois Institute of Technology; Eric A. Surface, SWA Consulting Inc., Raleigh, North Carolina; Lori Foster Thompson, Department of Psychology, North Carolina State University.

This research used data from an original longitudinal training effectiveness study sponsored by the Special Operations Forces Language Office (SOFLO), U.S. Special Operations Command (USSOCOM), and was conducted by SWA Consulting Inc. (SWA) as a prime contractor or subcontractor under the following contract numbers: H92239-04-P-0210, USZA22-02-D-0015, and N65236-08-D6805. Throughout the article, the terms study and research are used interchangeably. The usage of these words is not meant to imply that the investigation drew upon any particular source of funding within or outside of the organization. Furthermore, the views expressed in this article are those of the authors and do not necessarily reflect the positions or policies of the SOFLO, the USSOCOM, the U.S. government, or SWA. Eric A. Surface was principal investigator for the aforementioned contracts and has a financial interest in SWA. An earlier version of Study 2 was presented at the annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, Louisiana, and is based on Reanna Poncheri Harman's dissertation. We thank Jack Donnelly and Mark Roemer for supporting the use of SOFLO project data for research purposes. We thank SWA team members who supported the original project work that serves as the foundation for this research as well as North Carolina State University students who assisted with the coding. Finally, we thank Eduardo Salas and the reviewers for their guidance.

Correspondence concerning this article should be addressed to Reanna Poncheri Harman, 311 S. Harrington Street, Suite 200, Raleigh, NC 27603. E-mail: reannaharman@gmail.com

that 35.9% of "high-level business, HR and learning professionals" in its study viewed reactions as having *high* or *very high value*, with only 1.2% indicating reactions have *no value*. As Brown (2005) points out, it is necessary to understand the factor structure and underlying nomological network of training reactions in order to make full use of this information. It is important for training researchers to look beyond the initial criticisms of measuring training reactions and gain a deeper understanding of the value of this information in practice as well as research.

One important research avenue is the exploration of qualitative training reactions, which are not widely researched and are poorly understood. Although our focus is on training evaluation, commenting is becoming ubiquitous in society. Tweets, Facebook posts, texting, and blogs are all forms of commenting in the digital age. The increased emphasis on analyzing "big data," text mining, and social media mining also signals interest in understanding qualitative inputs (Aggarwal & Zhai, 2012). Likewise, commenting opportunities are frequently included on training evaluation surveys and increasingly on review websites for instruction or training products. For example, the website www.ratemyprofessor.com allows students to provide ratings as well as comments about experiences with their college professors. In addition, the increasing prevalence of massive open online courses (MOOCs; Baggaley, 2013) and the emergence of MOOC rating sites (e.g., CourseTalk, Mooctivity, and Moodadvisor), which allow for extensive commenting, will focus more attention on qualitative training reactions and their usefulness. Recent trends, such as the addition of word or tag clouds as an analysis option in many survey software solutions (e.g., Survey Monkey or Qualtrics), also suggest an interest in qualitative reactions. The pervasiveness of commenting opportunities, coupled with limited commenting research, reveals an unexplored avenue for gaining a deeper understanding of training reactions.

Research has shown that most survey participants do not comment when given the opportunity. Commenting rates reported in the literature range from 34% to 40% (Borg, 2005; Poncheri, Lindberg, Thompson, & Surface, 2008; Siem, 2005). These percentages are low but are not necessarily indicative of bias. The issue of bias in survey response has been extensively explored in the context of unit or survey nonresponse, but not in commenting behavior (e.g., Gannon, Nothern, & Carroll, 1971; Rogelberg, Spitzmüller, Little, & Reeve, 2006). Bias is indicated when there are relevant differences between those who participate in surveys (or in our case, those who comment) compared with those who do not (Rogelberg & Stanton, 2007).

If training comments are going to be used effectively by decision makers for feedback, marketing, and decision making (Kraiger, 2002), it is imperative that we understand more about commenting. If those who comment on training surveys have a different perspective than those who do not, there is a potential for bias in the data, which can ultimately have damaging consequences for trainees, instructors, and the training program, if decision makers attend to and act upon these biased comments.

We view commenting on a survey as a type of workplace behavior. In an organizational training context, survey or item response, including commenting, can be considered performance, as it is an action that contributes to organizational goals (i.e., Borman, 1991; Campbell, 1999). Because there is no existing theory or framework specifically related to commenting behavior, we considered commenting more broadly as a workplace behavior and used theories of workplace behavior, such as job performance, to develop our hypotheses and research questions. According to Campbell's (1990) conceptualization, performance is an observable behavior that has determinants and happens in a context. Like any other behavior, we argue that commenting on a survey is determined by both individual and situational factors (Johns, 2006; Lewin, 1936). Campbell and colleagues identified declarative knowledge, procedural knowledge and skill, and motivation as the direct determinants of job performance, but other variables (i.e., ability, personality, and education) as indirect determinants (Campbell, McCloy, Oppler, & Sager, 1993). In our studies, we explore several of these indirect or distal predictors of performance (e.g., verbal aptitude, education, personality) to determine what drives trainees to make comments and what may predict the quality of comments provided.

In addition to considering theories of behavior and job performance, we incorporate more specific motivational theories, such as affective events theory (AET; Weiss & Cropanzano, 1996). Providing a comment on a training survey is a behavior that is driven by affective reactions to the training event or some aspect(s) of the training (i.e., instructors, materials, physical environment, and assessment). Those affective reactions and the commenting behavior itself are influenced by individual dispositions and by features of the training context. In addition to the more distal individual differences, we use specific motivational theories as the basis for examining some proximal affective predictors (e.g., interest and satisfaction). We also draw upon research related to similar behaviors, such as unit nonresponse (e.g., Rogelberg & Stanton, 2007) and closed-ended item nonresponse (e.g., Craig & McCann, 1978; Marcus & Schütz, 2005) to develop hypotheses related to the influence of individual differences on commenting.

In our studies, we argue that whether or not a person comments, and the type of comment provided, is determined by relevant individual differences (both proximal and distal) and contextual factors. Both conceptual and existing empirical research suggest that commenting should largely be influenced by individual factors, but the fact that our research is conducted in the context of multiple training classes warrants investigation of potential situational influences (i.e., class-level variance) on the behavior, as trainees in a class share many of the same experiences in the learning environment. In fact, training researchers have specifically identified a need for multilevel analysis in training research (Mathieu & Tesluk, 2010). In order to fully understand a behavior, influences from at least one level "outside" and one level "within" the behavior must be investigated (Hackman, 2003). We draw upon existing training evaluation and effectiveness research to investigate potentially relevant contextual influences on commenting.

We conducted three field studies to investigate commenting behavior in a work-related training context with multiple simultaneous classes. First, across all three studies, we investigate whether the shared experience of being in the same training class influences individual commenting for the aforementioned reasons. Study 1 explores individual and contextual factors to determine whether there are relevant differences between commenters and noncommenters, thereby exploring the possibility of bias in this specific case of item nonresponse. Study 2 seeks to replicate Study 1 to establish generalizability of our findings, to further expand our

exploration of individual difference and contextual variables, and to explore the characteristics of comments and commenters. Study 3 investigates the influences of an additional individual difference variable, a different conceptualization of satisfaction, a behaviorally focused measure of instructor behavior to explore trainer effects, and the influence of item wording on commenting.

Study 1

Hypotheses and Rationale

In order to explore the individual and situational influences on qualitative reactions, this study examines the impact of education, verbal aptitude, interest in the topic, conscientiousness, satisfaction, and the training context on commenting.

Education and verbal aptitude. Individuals with relatively low education levels are less likely to respond to surveys (e.g., Gannon et al., 1971; Rogelberg & Luong, 1998). The findings are less clear for item nonresponse. Craig and McCann (1978) found that individuals with relatively less education skipped more closed-ended items on mail surveys. More recently, two studies (Clayton, Rogers, & Stuifbergen, 1999; McNeely, 1990) examined, but failed, to find a significant relationship between education and commenting. However, in both studies, the education level for commenters was higher than that of noncommenters.

Green (1996) offers several explanations for the positive relationship between education and response: "access to more information, greater cognitive skill, greater fluidity in translating opinions and ideas verbally, and more varied experience with answering questions in different item formats" (p. 174). The latter two are particularly relevant when studying commenting.

Hypothesis 1a (H1a): Education will be positively related to commenting.

Although there is no existing research, Green's (1996) explanation provides theoretical support for a positive link between verbal aptitude and commenting.

Hypothesis 1b (H1b): Verbal aptitude will be positively related to commenting.

Conscientiousness. It has been effectively argued that voluntary participation in organizational surveys can be considered in the broader domain of employee performance as a component of organizational citizenship behavior (OCB) or contextual performance (Tomaskovic-Devey, Leiter, & Thompson, 1994). Much research has examined the link between personality factors and OCB. When examining personality factors according to the Big Five framework, conscientiousness has emerged as the strongest predictor of OCB. Conscientiousness is associated with dependability, self-discipline, carefulness, perseverance, planfulness, and achievement orientation (Barrick, Mount, & Judge, 2001; George & Jones, 2003). Rogelberg and colleagues (e.g., Rogelberg et al., 2003; Rogelberg, Luong, Sederburg, & Cristol, 2000) found that nonrespondents are less conscientious than survey respondents (Rogelberg et al., 2003), and argue that these findings are supported by theory related to OCB. Conscientious individuals are theoretically more likely to respond because they are more willing to engage in helping behaviors and are less likely to forget to respond.

Hypothesis 1c (H1c): Conscientiousness will be positively related to commenting.

Interest in the topic. Rogelberg and Luong (1998) indicated that interest in the topic is positively related to survey participation. Martin (1994) conducted a study with members of an amateur bowling tournament in which participants were randomly assigned to a higher interest condition (bowling survey) or a lower interest condition (restaurant survey), and found that higher interest participants were twice as likely to respond to the survey and less likely to skip individual items when responding than lower interest participants. More recently, Groves, Presser, and Dipko (2004) used a similar approach to examine topic interest and participation in a telephone survey, and concluded that those who are interested in the topic are more likely to participate than those who are not interested. We extend this rationale to commenting.

Hypothesis 1d (H1d): Interest in the topic will be positively related to commenting.

Satisfaction. Studies have also linked satisfaction to commenting. Clayton et al. (1999) found that individuals who were less satisfied with their economic resources were more likely to provide unsolicited narrative comments. Meanwhile, McNeely (1990) found that commenters were less satisfied with their jobs than were noncommenters.

In response to an organizational climate survey in which open-ended comments were specifically solicited, Poncheri et al. (2008) found dissatisfied employees were more likely to comment than satisfied employees. This finding is supported by the general psychological phenomenon that "bad is stronger than good" (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001, p. 323) and the positive–negative asymmetry (PNA) effect in particular (Peeters, 1971). In describing the PNA effect in the context of forming impressions or opinions, Baumeister et al. (2001) assert "in general, and apart from a few carefully crafted exceptions, negative information receives more processing and contributes more strongly to the final impression than does positive information" (pp. 323–324). Dissatisfied individuals are, therefore, more likely to deeply process information about their dissatisfaction, whereas satisfied individuals are less likely to do so. As such, satisfied individuals should have less to communicate when asked to comment.

Hypothesis 1e (H1e): Satisfaction will be negatively related to commenting.

Training context. The importance of exploring contextual influences on behavior has been widely promoted in the psychological literature (e.g., Johns, 2006). In addition, the concept of an interaction between person and environment to influence behavior has been explored in the training literature specifically (i.e., aptitude–treatment interactions; Gully & Chen, 2010). Johns (2006) describes two levels of analysis for classifying context—omnibus and discrete (which focuses on task, social, and physical contexts). Although the omnibus context for this training program is the same across classes, there is potentially important variability across classes in terms of discrete context. Given salient differ-

ences in the learning task between classes, we will focus on two aspects of the discrete task context (Johns, 2006) to account for variability across classes, if it exists. Both training length and task difficulty have been shown to impact training results (e.g., Cole, 2008; Driskell, Willis, & Copper, 1992; Orvis, Horn, & Belanich, 2008). This study will, therefore, investigate the following research question.

> *Research Question 1 (RQ1):* Does the training context account for variability in commenting, and if so, do elements of the task context (i.e., training length and task difficulty) explain the between-context variance?

Sample and Method

Data were collected as part of a large-scale training effectiveness project. Trainees ($N = 160$) were members of a large military organization who were participating in sustainment or enhancement job-related foreign language training. They were experienced job incumbents. There were 25 classes covering 10 languages, with the number of trainees per class[1] ranging from one to 17. Trainees completed pre- and posttraining surveys.

Education. Education was measured with a single item on both the pre- and posttraining surveys: "Please indicate the highest level of education that you have attained." Trainees chose from the following response options: high school, some college, BA or BS degree, MA or MS degree, and PhD or EdD degree. For the purposes of our analyses, we created two categories: high school and some college or higher.

Verbal aptitude. Verbal aptitude was measured with the Armed Services Vocational Aptitude Battery (U.S. Department of Defense, 1984). We used a composite measure, verbal expression, which is composed of paragraph comprehension and word knowledge.

Personality. Goldberg's (1999) International Personality Item Pool (IPIP; 50 items) was used in this study. Trainees responded to these items on the pretraining survey using a 5-point Likert-type scale ranging from 1 (*very inaccurate*) to 5 (*very accurate*).

Interest in the topic. In this study, interest in the topic was operationalized as motivation to participate in foreign language training. A four-item measure of motivation developed for use in this study was included on the posttraining survey. Items were rated on a 7-point Likert-type scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). A confirmatory factor analysis (CFA) was conducted on this measure to confirm a single factor. The model indicated good fit according to procedures outlined in Drewes (2009; see Table 1), and the estimate of maximal scale reliability (R_{Max}) was acceptable (see Table 2).

Satisfaction. Satisfaction was measured with 12 items developed for this study. These items were developed based on research related to training reactions (e.g., Morgan & Casper, 2000) and written to reflect four dimensions of satisfaction: (a) satisfaction with the instructor, (b) satisfaction with the utility of training, (c) satisfaction with language training in the unit, and (d) satisfaction with learning materials and environment. Trainees responded to these items on the posttraining survey using a 7-point Likert-type scale ranging from 1 (*extremely dissatisfied*) to 7 (*extremely satisfied*). Table 1 provides the CFA results and Table 2 provides the reliability estimates.

Training context. Two aspects of task context (Johns, 2006) were explored: task difficulty and training length. Task difficulty was operationalized using a classification system established by the U.S. government to indicate how difficult languages are to learn for native English speakers (Silva & White, 1993), consisting of four difficulty categories in which French is an example of a Category (CAT) I language, German is CAT II, Russian is CAT III, and Arabic is CAT IV. Training length was operationalized as the number of training hours.

Open-ended item. One open-ended item was provided at the end of the posttraining survey: "Please use the space below to provide any additional information or to make comments on the language training you received at [this training location]."

Results and Discussion

Means, standard deviations, reliability estimates, and correlations are presented in Table 2. Fifty-five trainees (34%) provided comments. In order to address Hypotheses 1a through 1e, two-level hierarchical generalized linear models (HGLMs) were estimated with a Bernoulli sampling model and logit link function, as the criterion in these models was dichotomous (i.e., 0 = no comment, 1 = comment). Multilevel modeling followed a staged approach (Raudenbush & Bryk, 2002), with the first stage including an unconditional or "null" model in order to test for between-class variance in commenting behavior (RQ1), and a subsequent conditional random-intercept (fixed slope) model that included the trainee (H1a through H1e) and class-level (RQ1) predictors. All predictors were centered on their grand mean (Hofmann & Gavin, 1998).

The unconditional HGLM results indicated significant class-level variance associated with commenting behavior ($T_{00} = 1.25$, $df = 24$, $x^2 = 57.20$, $p < .001$), with 28% of the variability in the log-odds of commenting residing between-classes, addressing RQ1 and indicating that the training context impacts commenting. Conditional HGLM results (presented in Table 3) suggested that only interest in the topic was related to commenting, supporting H1d. Education, verbal aptitude, and conscientiousness were not related to commenting behavior either in the HGLM or bivariate correlations, so H1a, H1b, and H1c were not supported. Although the HGLM results did not show a statistically significant relationship between satisfaction and commenting, the bivariate correlations for three of the four dimensions of satisfaction were significant, providing partial support for H1e. Neither of the class-level predictors (i.e., training length and task difficulty) explored were significant. Variance explained (R^2) was calculated using the approach described by Snijders and Bosker (2012), and indicated that the collective set of predictors explained 27% of the variance in the odds of commenting.

Because commenting has only been considered at the individual level, our findings provide an important contribution to the literature. In addition to individual-level factors, the shared experience of being in the same class appears to influence commenting (i.e., significant class-level variance). Our findings have implications for data representativeness across classes and individuals. More research is needed to understand the influences on commenting as only one of our

[1] This reflects the number of trainees per class who had complete data for inclusion in this study. There were no classes with only one student on the roster.

Table 1
Confirmatory Factor Analyses

Model	x^2	*df*	x^2/df	*p*	CFI	TLI	RMSEA [90% CI]	SRMR
Study 1								
Interest in the topic	20.13	2	10.07	0.00	0.825	0.474	0.213 [.135, .303]	0.065
Satisfaction – Four-factor model	144.10	48	3.00	0.00	0.949	0.930	0.101 [.082, .120]	0.056
Study 2								
Interest in the topic	2.65	2	1.33	0.27	0.998	0.994	0.028 [.000, .105]	0.014
Satisfaction – Four-factor model	368.28	48	7.67	0.00	0.910	0.876	0.126 [.114, .138]	0.106
Study 3								
Interest in the topic	0.674	2	0.337	0.72	1.000	1.002	0.000 [.000, .038]	0.003
Satisfaction – Five-factor model	1222.5	242	5.05	0.00	0.841	0.819	0.100 [.095, .106]	0.076
Classroom management	182.2	9	20.21	0.00	0.977	0.962	0.124 [.109, .140]	0.018

Note. We only present confirmatory factor analyses for scales that were developed for these studies. Several fit indices were examined, including the chi-square (x^2) statistic, the CFI, the TLI, the RMSEA, and the SRMR. Because the chi-square statistic is sensitive to sample size and often found to be significant (Yu, 2002), the other indices were examined. For the CFI and TLI, values above 0.90 and closer to 1.00 are considered acceptable (Hu & Bentler, 1999). For the RMSEA and the SRMR, values below 0.08 are considered good fit (Millsap, 2002; Vandenberg & Lance, 2000). *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = square root mean residual; CI = confidence interval.

individual-level, and none of our class-level, predictors were significant in the HGLM analysis. Study limitations include sample size (for both trainees and classes) and limited contextual variables available to explore the nature of class-level variance. In order to address these limitations, as well as to explore some additional antecedents and further explore the nature of comments provided, we conducted a second study. The goals of Study 2 are to replicate and expand Study 1 findings and to explore comment characteristics with a larger sample in a different military foreign language training context, which tests an element of generalizability.

Study 2

Hypotheses and Rationale

Building on Study 1 findings, Study 2 explores the influence of individual and contextual characteristics on commenting and adds an additional focus on the characteristics of comments, which has not been previously explored in the literature.

Individual differences. Like Study 1, we hypothesize that education (Hypothesis 2a; H2a), verbal aptitude (Hypothesis 2b; H2b), conscientiousness (Hypothesis 2c; H2c), and interest in the topic (Hypothesis 2d; H2d) will be positively related to commenting. Satisfaction will be negatively related (Hypothesis 2e; H2e).

We also explore the influence of trainee learning on commenting. It is expected that trainees who perform poorly on their end-of-course assessment will be more likely to provide comments on a posttraining evaluation survey, as they are likely to be dissatisfied with the training. Drawing upon the rationale supporting the hypotheses related to satisfaction, it is expected that trainees who learn less than other trainees will make more comments.

Hypothesis 2f (H2f): Trainee learning will be negatively related to commenting.

Training context. In Study 1, we found significant class-level variance in commenting, but neither of the two class-level predictors we explored was related to commenting. In this study, we

Table 2
Study 1 – Descriptive Statistics and Zero-Order Correlations

	M	*SD*	1	2	3	4	5	6	7	8	9
Individual-level (L1) variables											
1. Education	1.80	.40	—								
2. Verbal aptitude	55.93	4.82	.09	—							
3. Interest in the topic	5.22	1.00	.17*	.09	(.87)						
4. Conscientiousness	3.87	.61	.15	-.09	.27**	(.92)					
5. Satisfaction – Instructor	5.90	1.10	-.11	.19*	.28**	.03	(.97)				
6. Satisfaction – Utility of training	5.34	1.28	-.06	.06	.25**	-.03	.75**	(.96)			
7. Satisfaction – Language training in unit	5.00	1.32	-.12	.03	.06	-.06	.47**	.55**	(.97)		
8. Satisfaction – Learning materials and environment	4.86	1.11	-.10	-.06	.24**	.01	.44**	.55**	.51**	(.85)	
9. Commenting behavior	—	—	-.03	.00	.15	.01	-.07	-.17*	-.17*	-.19*	—
Class-level (L2) variables											
1. Task difficulty	3.12	.91	—								
2. Training hours	178.92	197.27	.37	—							

Note. R_{Max} is provided on the diagonal for items where relevant. L1, *n* = 160; L2, *n* = 25. Education: 1 = *high school*, 2 = *some college or higher*; Interest in the topic: 1 = *strongly disagree*, 7 = *strongly agree*; Conscientiousness: 1 = *very inaccurate*, 5 = *very accurate*; Satisfaction scales: 1 = *extremely dissatisfied*, 7 = *extremely satisfied*; Comment behavior: 0 = *no comment made*, 1 = *comment made*. L1 = Level 1; L2 = Level 2.
* *p* < .05. ** *p* < .01.

Table 3
Study 1 – HGLM Results for Commenting Behavior

Fixed effects	Coefficient	*SE*	*t*	Odds ratio	95% confidence interval
Individual-level (L1) predictor					
Intercept	-.770*	.324	-2.380	.463	.237, .906
Education	-.821	.626	-1.312	.440	.128, 1.518
Verbal aptitude	-.028	.043	-.658	.972	.894, 1.058
Interest in the topic	.576*	.246	2.337	1.778	1.092, 2.895
Conscientiousness	-.102	.330	-.309	.903	.470, 1.734
Satisfaction – Instructor	.074	.273	.271	1.077	.628, 1.847
Satisfaction – Utility of training	-.487	.302	-1.612	.614	.338, 1.118
Satisfaction – Language training in unit	-.101	.185	-.549	.904	.627, 1.303
Satisfaction – Learning materials and environment	-.198	.238	-.833	.820	.513, 1.313
Class-level (L2) predictor					
Task difficulty	.016	.320	.051	1.017	.524, 1.972
Training hours	.001	.002	.438	1.001	.997, 1.005

Note. L1, $n = 160$; L2, $n = 25$. HGLM = hierarchical generalized linear models; *SE* = robust standard error; L1 = Level 1; L2 = Level 2.
* $p < .05$. ** $p < .01$.

hypothesize that there will be class-level variance associated with commenting, and we will explore three predictors of class-level variance to include task difficulty, class size, and class learning. Both class size and class learning could be considered as components of the discrete task and social contexts (Johns, 2006), as class size impacts the selection and execution of the learning activities as well as the social dynamics (e.g., Finn, Pannozzo, & Achilles, 2003), and class learning can impact the structure of instruction (e.g., pace, coverage) and the social interactions in the class. Both class size and learning influence the learning process (Pickett & Fraser, 2010). If the class is large, the instructor has less time to dedicate to each individual trainee, compared with when the class is small. If the class is composed of high-ability and high-performing trainees who are learning quickly, then the instructor can move at a faster pace than would be possible if trainees are struggling. Because these influences on the learning process will be evident to trainees who are part of a social group (i.e., class), and who likely voice their opinions about the training to one another, these contextual influences create a shared experience that can lead to reactions that affect commenting.

Research Question 2 (RQ2): Does the training context account for variability in commenting, and if so, do elements of the task context (i.e., task difficulty, class size, and learning) explain the between-context variance?

Comment characteristics. Beyond understanding what drives a trainee to comment, it is also important to consider the types of comments provided and how trainee individual differences impact the types of comments provided. Comments can vary with regard to tone (negative vs. positive), scope (general vs. specific), and purpose (prescriptive vs. descriptive). Although comment tone has received limited empirical attention, this study is the first to examine comment scope and purpose—noteworthy attributes that are based on Kulesa and Bishop's (2006) categorization of survey questions.

Comment tone. Comment tone is the most frequently studied comment characteristic and has been explored in a variety of survey contexts, including climate surveys (Poncheri et al., 2008), 360-degree surveys (Smither & Walker, 2004), public opinion polls (Garcia, Evans, & Reshaw, 2004), and general employee opinion surveys (Borg, 2005). With the exception of the Smither and Walker (2004) study, comments tended to be negative in tone.

Although no past research has explored individual characteristics that affect comment tone, three characteristics are conceptually relevant: satisfaction, agreeableness, and emotional stability. Research suggests a negative relationship between satisfaction and the tendency to provide comments (Clayton et al., 1999; McNeely, 1990; Poncheri et al., 2008), and most open-ended comments tend to be negative (Borg, 2005; Garcia et al., 2004; Poncheri et al., 2008). Presumably, most comments are negative because most commenters are dissatisfied.

Hypothesis 3a (H3a): Satisfaction will be positively related to comment tone.

Agreeableness is associated with cooperation, trust, compliance, and sympathy (Barrick et al., 2001), and "agreeable individuals generally are easy to get along with and are 'team players'" (George & Jones, 2003, p. 9). Agreeable individuals should provide positive comments because of their tendency to support the status quo (LePine & Van Dyne, 2001).

Hypothesis 3b (H3b): Agreeableness will be positively related to comment tone.

Individuals with low emotional stability are prone to negative emotions, and are critical and pessimistic when compared with individuals with high emotional stability (Barrick et al., 2001; George & Jones, 2003). These qualities should be reflected in comment tone.

Hypothesis 3c (H3c): Emotional stability will be positively related to comment tone.

Comment scope. Comment scope (general vs. specific) is a second characteristic worth considering. A general comment is broad in nature and does not offer details about precise issues, whereas a specific comment provides information targeted at particular issues. Specific comments are often desirable because they offer in-depth information, which decision makers can use to

determine if and what type of action is needed. In all likelihood, those who are dissatisfied will provide detailed information about the cause of their dissatisfaction, whereas those who are relatively satisfied will not. Poncheri et al. (2008) found that comment length increases as comments become more negative in tone, which may be caused by individuals providing detailed information about their dissatisfaction. As Taylor (1991) notes, "other things being equal, negative events appear to elicit more physiological, affective, cognitive, and behavioral activity and prompt more cognitive analysis than neutral or positive events" (p. 67). This should lead to greater specificity in comments from dissatisfied individuals.

Hypothesis 4a (H4a): Satisfaction will be negatively related to comment specificity.

Conscientiousness also has a conceptually relevant relationship with comment specificity. Conscientious individuals are organized and detail-oriented (George & Jones, 2003). This attention to detail may lead highly conscientious individuals to provide specific information when making a point either verbally or in writing.

Hypothesis 4b (H4b): Conscientiousness will be positively related to comment specificity.

Comment purpose. Purpose (prescriptive vs. descriptive) is a third dimension upon which survey comments vary. A prescriptive comment provides a suggestion for change or improvement, whereas a descriptive comment provides a narrative of the current state. Although no research has explored the individual differences that impact the tendency to provide prescriptive comments, there is an extensive literature on voice behavior that is relevant. Voice behavior is a type of OCB and is defined as "constructive change-oriented communication intended to improve the situation" (LePine & Van Dyne, 2001, p. 326). LePine and Van Dyne (2001) describe voice behavior as encapsulating behaviors that have been included in descriptions of OCB, such as "suggesting organizational improvements; making constructive suggestions; suggesting ideas for how others in the unit should proceed; and persuading others to accept ideas, opinions, and directions" (p. 327).

Although they did not study surveys, LePine and Van Dyne (2001) conducted a study to explore, in part, individual differences that are related to voice behavior. They hypothesized that conscientious individuals would be likely to invest effort in providing suggestions for change, and found a positive relationship between conscientiousness and voice behavior.

Hypothesis 5a (H5a): Conscientiousness will be positively related to comment prescriptiveness.

Extraversion is associated with positive emotions, sociability, gregariousness, warmth, dominance, ambition, and excitement-seeking. LePine and Van Dyne (2001) argued that extraverts would be better at communicating their thoughts and more willing to express opinions that would challenge the status quo than introverts. They found that extraversion was positively related to voice behavior. Furthermore, respondents who exhibit dominance and ambition, two facets of extraversion, will be likely to assert their opinions in a prescriptive manner compared with those who do not exhibit these traits (DeYoung, Quilty, & Peterson, 2007).

Hypothesis 5b (H5b): Extraversion will be positively related to comment prescriptiveness.

Sample and Method

Data were collected as part of the same overarching training effectiveness project as Study 1. The data reported in this article were collected as part of a larger study that covered training events in an organization during a specific period of time and data from this larger study have been presented at several conferences (Bauer, Orvis, Ely, & Surface, 2013; Bhavsar, Hess, & Surface, 2008; Bhavsar, Poncheri, & Surface, 2006; Watson, Thompson, & Surface, 2011). Participants (N = 419) were members of a large military organization participating in mandatory, long-term, job-related foreign language training. This was an initial acquisition language program conducted as part of the institutionalized training prior to unit assignment, so they were not incumbents. There were 77 classes covering nine languages, with one to eight trainees per class.[2] No instructor taught more than two classes, and only eight instructors taught more than one class. Trainees completed pre- and posttraining surveys.

This study used the same measures as Study 1 for education, verbal aptitude, personality, interest in the topic, satisfaction, and the open-ended item (see Table 1 for CFAs related to these measures for the Study 2 sample).

Trainee learning. Learning was measured using the Defense Language Proficiency Test (DLPT), a high-stakes, end-of-course assessment that measures language proficiency (Silva & White, 1993). The DLPT tests listening and reading proficiency. Raw scores were averaged for the two tests to create a composite proficiency variable (Dierdorff & Surface, 2008).

Training context. Three aspects of training context were explored in this study: task difficulty, class size, and class learning. Task difficulty was operationalized in the same way as Study 1, using the CAT I-IV coding system (Silva & White, 1993). Training length was not explored in this study because it was confounded with task difficulty (all CAT I/II courses were the same length, and all CAT III/IV courses were the same length at this institution). Class size was calculated based on course rosters obtained at the beginning of the course. Class learning was operationalized by creating an aggregate (mean) DLPT composite score for each class. The intraclass correlation (ICC1,k) for the aggregated DLPT scores was .88, indicating the reliability of the class means (Shrout & Fleiss, 1979).

Discrete comment coding. According to Smither and Walker (2004), "the first step in coding written comments is determining the unit of analysis (e.g., a word, a phrase, a sentence, several sentences)" (p. 577). In this study, the unit of analysis is topic areas or themes. Some sentences may include multiple topics, whereas other topics may span several sentences. Parts of sentences were not separated when coding for discrete comments.

Discrete comment coding was conducted by the lead researcher and two independent raters. Rater training included a review of the definition of a discrete comment and the unit of analysis, a review of examples of discrete coding, and practice. After the two raters determined the number of discrete comments per open-ended response and separated the comments into discrete parts, interrater reliability was assessed. In the event of disagreement, the lead

[2] This reflects the number of trainees per class who had complete data for inclusion in this study. There were no classes with only one student on the roster.

researcher made a final decision about the number of discrete comments.

Coding for comment tone, scope, and purpose. The lead researcher and two independent raters read each discrete comment and rated comment tone (1 = *very negative*, 2 = *somewhat negative*, 3 = *neither negative nor positive*, 4 = *somewhat positive*, 5 = *very positive*), scope (1 = *not at all specific*, 2 = *somewhat specific*, 3 = *very specific*), and purpose (1 = *not at all prescriptive*, 2 = *somewhat prescriptive*, 3 = *very prescriptive*). All raters were trained using a sample of 302 comments that could not be included in this study because those trainees had not completed a presurvey. Raters were provided with definitions for each criterion and for each scale point, and were provided with six example comments for each scale point to use as a reference. The training included a review of definitions and examples, practice with rating 20 comments, and then a follow-up assignment to rate 100 comments as preliminary coding practice. Preliminary interrater reliability statistics were calculated before raters were able to proceed with actual coding.

Interrater reliability. Interrater reliability for all coding was calculated using two-way mixed (ICCs; Shrout & Fleiss, 1979; von Eye & Mun, 2005). As shown in Tables 4 and 5, all ICC estimates exceeded 0.75, which is considered to be excellent (Fleiss, Levin, & Paik, 2003). After reliability was confirmed, discrete tone, scope, and purpose ratings were averaged across raters, yielding a single average rating for each discrete comment. Then, discrete comment ratings were averaged to create an overall rating of tone, scope, and purpose.

Results and Discussion

Means, standard deviations, reliability estimates, and correlations for all study variables are presented in Tables 4 and 5. Two hundred twenty-two trainees (53% of the total sample) provided comments in response to the open-ended question. The average comment tone rating was "somewhat negative" (M = 2.13, SD = .60), the average comment scope rating was "somewhat specific" (M = 1.95, SD = .47), and the average comment purpose rating was between "not at all prescriptive" and "somewhat prescriptive" (M = 1.60, SD = .62).

HGLM was used to explore the hypotheses related to commenting behavior (see Table 6). The unconditional model showed significant class-level variance (τ_{00} = .46, df = 70, χ^2 = 103.71, p < .01), with 12% of the variability attributable to classes, addressing RQ2. The conditional HGLM results indicated that verbal aptitude and conscientiousness were not related to commenting, so H2b and H2c were not supported. Interest in the topic (supporting H2d) and satisfaction with the instructor were the only significant individual-level predictors of commenting, although satisfaction with the instructor was positively related to commenting behavior, which was in the opposite direction as expected (H2e not supported). Although HGLM results did not show a relationship between education and commenting, the bivariate correlation showed a significant and positive relationship providing partial support for H2a.

Although individual learning was not predictive of commenting (H2f not supported), class-level learning was negatively associated with commenting. As class-level learning decreases, commenting by members of the class increases, possibly suggesting that a shared sense of dissatisfaction with the learning process can impact qualitative reactions. The other class-level predictors, task difficulty and class size, were not significant. The collective set of predictors explained 15% of the variance in the odds of commenting.

In terms of comment characteristics, hierarchical linear modeling (HLM) indicated that no class-level variance was found for comment tone or scope; therefore, multiple regression was used to explore the hypothesized relationships. Although all four dimensions of satisfaction were positively correlated with comment tone, satisfaction with the utility of training was the only satisfaction dimension with a significant relationship in the regression results, partially supporting H3a (see Table 7). Agreeableness (H3b) and

Table 4
Study 2 – Descriptive Statistics and Zero-Order Correlations for Commenting Behavior Hypotheses

	M	*SD*	1	2	3	4	5	6	7	8	9	10
Individual-level (L1) variables												
1. Education	1.76	.43	—									
2. Verbal aptitude	57.63	3.91	.24**	—								
3. Interest in the topic	4.84	1.17	.11	.13*	(.92)							
4. Conscientiousness	3.99	.60	.02	.03	.12*	(.92)						
5. Satisfaction – Instructor	5.35	1.81	-.01	.09	.18**	.03	(.97)					
6. Satisfaction – Utility of training	4.05	1.61	-.02	-.03	.27**	.10	.55**	(.95)				
7. Satisfaction – Language training in unit	4.34	1.18	-.04	.03	.19**	.11	.39**	.55**	(.97)			
8. Satisfaction – Learning materials and environment	3.83	1.25	.03	.03	.16**	.06	.34**	.48**	.62**	(.81)		
9. Learning	3.91	5.99	.08	.30**	.25**	-.03	.17**	.21**	.09	.11*	—	
10. Commenting behavior	—	—	.13*	.10	.16**	.02	.04	-.11	-.10	-.10	-.05	(.96)[a]
Class-level (L2) variables												
1. Task difficulty	2.63	1.38	—									
2. Class size	7.66	.83	-.32**	—								
3. Class learning	36.58	5.76	-.57**	.07	—							

Note. R_{Max} is provided on the diagonal for items where relevant. L1, n = 288; L2, n = 71. Education: 1 = *high school*, 2 = *some college or higher*; Interest in the topic: 1 = *strongly disagree*, 7 = *strongly agree*; Conscientiousness: 1 = *very inaccurate*, 5 = *very accurate*; Satisfaction scales: 1 = *extremely dissatisfied*, 7 = *extremely satisfied*; Comment behavior: 0 = *no comment made*, 1 = *comment made*. L1 = Level 1; L2 = Level 2.
[a] Number of comments = 222. Interrater reliability was calculated using the two-way mixed intraclass correlation (Shrout & Fleiss, 1979).
* p < .05. ** p < .01.

Table 5
Study 2 – Descriptive Statistics and Zero-Order Correlations for Commenting Characteristics Hypotheses

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11
Individual-level (L1) variables													
1. Conscientiousness	3.98	.56	(.91)										
2. Extraversion	3.44	.75	.06	(.93)									
3. Emotional stability	3.87	.62	.53**	.10	(.93)								
4. Agreeableness	3.75	.54	.25**	.30**	.27**	(.87)							
5. Satisfaction – Instructor	5.43	1.83	.05	-.06	.08	.06	(.97)						
6. Satisfaction – Utility of training	3.93	1.72	.10	.04	.04	.19**	.57**	(.95)					
7. Satisfaction – Language training in unit	4.23	1.22	.12	.07	.05	.17*	.39**	.54**	(.97)				
8. Satisfaction – Learning materials and environment	3.78	1.28	.13	-.04	.02	.07	.36**	.56**	.59**	(.81)			
9. Comment tone	2.13	.60	.01	-.04	.00	.04	.32**	.40**	.31**	.32**	(.94)[a]		
10. Comment scope	1.95	.47	.23**	.02	.26**	.08	.10	.10	.09	.15*	-.04	(.84)[a]	
11. Comment purpose	1.60	.62	.08	.03	.09	-.01	.13	.08	.08	.14*	.16*	.20**	(.95)[a]
Class-level (L2) variables													
1. Task difficulty	2.65	1.38	—										
2. Class size	7.66	.82	-.30**	—									
3. Class learning	36.92	5.68	-.49**	-.01	—								

Note. R_{Max} is provided on the diagonal for items where relevant. L1, $n = 215$; L2, $n = 70$. Personality scales: 1 = *very inaccurate*, 5 = *very accurate*; Satisfaction scales: 1 = *extremely dissatisfied*, 7 = *extremely satisfied*; Comment tone: 1 = *very negative*, 5 = *very positive*; Comment scope: 1 = *not at all specific*, 3 = *very specific*; Comment purpose: 1 = *not at all prescriptive*, 3 = *very prescriptive*. L1 = Level 1; L2 = Level 2.
[a] Number of comments = 614; interrater reliability was calculated using the two-way mixed intraclass correlation (Shrout & Fleiss, 1979).
* p :S .05. ** p :S .01.

emotional stability (H3c) were not related to comment tone in the bivariate or regression results. Conscientiousness was positively related to comment scope in both the bivariate and regression results, supporting H4b (see Table 8), but satisfaction (H4a) was not related to comment scope. HLM results indicated class-level variance (19%) for comment purpose (τ_{00} = .08, df = 69, x^2 = 120.24, p < .001). None of the trainee or class-level predictors were significantly related to comment purpose, so H5a and H5b were not supported (see Table 9).

The type of qualitative coding used in this study has not been widely used in the literature, which may be perceived as a limitation, but was based on a process developed by Smither and Walker (2004), and on best practices in qualitative coding. One unique element of our approach involved averaging comment tone, scope, and purpose ratings across items and raters. This had many methodological benefits (creating one score for each trainee for each variable) and, in the case of averaging across raters, was justified based on the interrater reliability calculations. The main limitation is that this could mask variability in an individual's response if a trainee provided a "mixed" comment (i.e., one very positive and one very negative commenting averaged to equal the scale midpoint, suggesting a neutral opinion). However, in our study, this only happened in 10 cases. Furthermore, this limitation would apply to a traditional coding approach as well, in which a coder mentally averages the ratings across the different types of comments. We argue that rating each comment independently and then averaging across comments is preferable to the more cognitively challenging task of mentally averaging ratings to assign a holistic rating to a comment block.

Table 6
Study 2 – HGLM Results for Commenting Behavior

Fixed effects	Coefficient	*SE*	*t*	Odds ratio	95% confidence interval
Individual-level (L1) predictor					
Intercept	.254	.154	1.651	1.289	.948, 1.753
Education	.439	.303	1.448	1.551	.853, 2.822
Verbal aptitude	.016	.040	.406	1.016	.939, 1.100
Interest in the topic	.342**	.123	2.778	1.408	1.104, 1.794
Conscientiousness	-.078	.217	-.358	.925	.603, 1.420
Satisfaction – Instructor	.175*	.087	2.025	1.191	1.005, 1.413
Satisfaction – Utility of training	-.184	.118	-1.556	.832	.658, 1.051
Satisfaction – Language training in unit	-.043	.167	-.257	.958	.690, 1.331
Satisfaction – Learning materials and environment	-.157	.151	-1.041	.855	.635, 1.151
Learning	.037	.041	.892	1.037	.956, 1.125
Class-level (L2) predictor					
Task difficulty	-.053	.148	-.355	.949	.705, 1.276
Class size	.102	.198	.516	1.108	.746, 1.644
Class learning (mean)	-.122*	.055	-2.228	.885	.793, .987

Note. L1, $n = 288$; L2, $n = 71$. HGLM = hierarchical generalized linear models; *SE* = robust standard error; L1 = Level 1; L2 = Level 2.
* p :S .05. ** p :S .01

Table 7
Study 2 – Regression Results for Comment Tone

Predictor	*B*	*SE B*		*t*
Satisfaction – Instructor	.039	.025	.120	1.553
Satisfaction – Utility of training	.084**	.031	.242	2.704
Satisfaction – Language training in unit	.035	.041	.072	.867
Satisfaction – Learning materials and environment	.048	.039	.102	1.226
Agreeableness	-.033	.074	-.029	-.441
Emotional stability	-.013	.063	-.014	-.211
R^2	.188			

Note. $N = 215$.
* $p < .05$. ** $p < .01$.

We successfully replicated two of the main findings from Study 1—the existence of class-level variance in commenting and the positive relationship between interest in the topic and commenting. Several new findings emerged, including a negative relationship between class-level learning and commenting. This is particularly interesting considering that individual learning did not impact commenting, and suggests a potential influence of either the task or social context on the learning process that impacts commenting. For example, low-performing classes may commiserate about their performance compared with others and then comment. Furthermore, satisfaction with the instructor was positively related to commenting, providing a counterpoint to the negativity bias frequently observed in commenting research. Previous research (e.g., Cerrito, 2000; Clement, 1982; Greenwald & Gillmore, 1997a and 1997b) on the relationship between student quantitative ratings of their instructors and academic achievement (e.g., course grades) might inform our understanding of the qualitative results. For example, easier professors get higher ratings (e.g., Cohen, 1981; Greenwald & Gillmore, 1997a, 1997b).

Our findings from Studies 1 and 2 suggest that commenting may be largely influenced by more proximal or "state-like" predictors (i.e., interest and satisfaction), or by the social context. For example, in Study 2, we did not find any significant relationships between the more distal, stable individual differences (i.e., education, verbal aptitude, and conscientiousness) and commenting behavior, but found that interest and satisfaction were influential. This suggests that the specific act of commenting in response to a survey is most likely influenced by proximal or state-like individual differences, rather than more distal or stable individual differences. Study 3 seeks to replicate some of the previous findings from Studies 1 and 2, but also adds additional distal and proximal individual differences variables to further explore the nature of commenting and the speculation that commenting is influenced by more proximal individual differences.

Study 3

Hypotheses and Rationale

Study 3 extends our investigation of distal and proximal individual differences and contextual factors, and adds implicit person theory (IPT; Dweck, 1986) as a new individual difference variable, as well as a different conceptualization of training satisfaction (Lee & Pershing, 1999), in order to shed light on previous mixed satisfaction results. We also explore a behaviorally focused measure of instructor behavior—classroom management—to investigate potential trainer effects. Finally, we investigate item wording—specifically, item specificity—which is something training professionals and researchers can manipulate to influence commenting.

Individual differences. As in Studies 1 and 2, we hypothesize that education (Hypothesis 6a; H6a), verbal aptitude (Hypothesis 6b; H6b), conscientiousness (Hypothesis 6c; H6c), and interest in the topic (Hypothesis 6d; H6d) will be positively related to commenting.

To expand on the previous findings, we also explore IPT as a predictor of commenting. Dweck (1986) originally conceptualized IPT as beliefs that individuals have regarding the malleability of intelligence. There are two main beliefs that a person may hold. Entity theorists believe intelligence is fixed, whereas incremental theorists believe intelligence is malleable. Although this individual difference has not been explored in relation to commenting, it is expected that entity theorists would be less likely to provide comments because of their belief in the fixed nature of intelli-

Table 8
Study 2 – Regression Results for Comment Scope

Predictor	*B*	*SE B*		*t*
Satisfaction – Instructor	.015	.021	.060	.734
Satisfaction – Utility of training	-.002	.026	-.009	-.091
Satisfaction – Language training in unit	-.009	.034	-.024	-.272
Satisfaction – Learning materials and environment	.044	.033	.119	1.347
Conscientiousness	.180**	.057	.215	3.188
R^2	.070			

Note. $N = 215$.
* $p < .05$. ** $p < .01$.

Table 9
Study 2 – HLM Results for Comment Purpose

Fixed effects	Coefficient	*SE*	*t*
Individual-level (L1) predictor			
Intercept	1.595**	.051	31.400
Conscientiousness	.096	.060	1.590
Extraversion	.020	.052	.392
Class-level (L2) predictor			
Task difficulty	.049	.042	1.171
Class size	.034	.062	.560
Class learning	-.008	.008	-1.105

Note. L1, $n = 212$; L2, $n = 68$. HLM = hierarchical linear modeling; *SE* = robust standard error; L1 = Level 1; L2 = Level 2.
* $p < .05$. ** $p < .01$.

gence. It is likely that someone who holds an entity theory of intelligence would be reluctant to comment because of their belief that people (i.e., in this case the instructor or program managers) cannot change. Therefore, the following is proposed:

Hypothesis 6e (H6e): IPT will be negatively related to commenting such that entity theorists will be less likely to comment.

In Studies 1 and 2, we hypothesized negative relationships between satisfaction and commenting, but findings were mixed. For example, the bivariate correlations for satisfaction with the instructor and commenting were nonsignificant in both studies. In Study 2, the HGLM showed a positive relationship between satisfaction with the instructor and commenting, whereas the bivariate correlations for the other three satisfaction dimensions were negative, with mixed significance. Our original satisfaction-commenting hypothesis was based on the PNA effect (Peeters, 1971), "bad is stronger than good" (Baumeister et al., 2001; p. 323), and empirical research on organizational surveys (e.g., McNeely, 1990; Poncheri et al., 2008). To explore this hypothesis further, we use a different conceptualization of training satisfaction in Study 3 to determine whether or not our previous results were caused by an instrumentation effect (Shadish, Cook, & Campbell, 2002). First, single-item indicators of overall instructor effectiveness and overall utility of the training were added. Although there are known criticisms of single-item measures, there is some evidence to suggest their value (Wanous, Reichers, & Hudy, 1997), and these items were added as standardized measures across many language training programs at the request of the client organization. This change is consistent with ASTD's (2009) recommendation to standardize training evaluation across an organization. Second, we investigate additional dimensions of satisfaction based on a framework developed by Lee and Pershing (1999). These researchers conducted a comprehensive review of existing literature related to dimensions of training reactions (e.g., Forsyth, Jolliffe, & Stevens, 1995; Phillips, 1997; Sanderson, 1995; Warr & Bunce, 1995), and identified 11 specific facets of satisfaction. We included five of these facets—satisfaction with (a) program materials, (b) instructional activities, (c) program time and length, (d) training environment, and (e) delivery methods and technology. Regardless of previous mixed findings, we will maintain our original hypothesis to investigate the new instrumentation.

Hypothesis 6f (H6f): Satisfaction will be negatively related to commenting.

In order to explore potential trainer effects on commenting behavior, we investigate trainee ratings of the instructor's classroom management behaviors (e.g., states the goals and objectives for each class session; e.g., Stronge, Ward, Tucker, & Hindman, 2007). It is expected that a trainee who perceives his instructor to be effectively managing the classroom will be less likely to comment than a trainee who perceives his instructor as less effectively managing the classroom—similar to what we hypothesize for general satisfaction with the instructor.

Hypothesis 6g (H6g): Classroom management behavior will be negatively related to commenting.

Training context. In both Studies 1 and 2, we found significant class-level variance in commenting and class learning as a significant class-level predictor in Study 2, but class size, training length, and task difficulty were not significant class-level predictors. We continue to investigate the presence of class-level variance as well as explore available[3] class-level predictors.

Research Question 3 (RQ3): Does the training context account for variability in commenting, and if so, do elements of the task context (i.e., task difficulty and class size) explain the between-context variance?

Comment characteristics. Building upon the findings from Study 2 related to comment characteristics, we explore comment tone, scope, and purpose. We hypothesize a positive relationship between satisfaction and tone. We also hypothesize a positive relationship between ratings of instructor's classroom management and tone.

Hypothesis 7a (H7a): Satisfaction will be positively related to comment tone.

Hypothesis 7a (H7b): Classroom management will be positively related to comment tone.

Although we did not find a positive relationship between comment scope and satisfaction in Study 2, we will explore this hypothesis again using a new instrumentation of satisfaction, and also explore the relationship between classroom management and scope. Finally, we seek to replicate the Study 2 finding that conscientiousness is positively related to scope.

Hypothesis H8a (H8a): Satisfaction will be negatively related to comment specificity.

Hypothesis H8b (H8b): Classroom management will be negatively related to comment specificity.

Hypothesis H8c (H8c): Conscientiousness will be positively related to comment specificity.

In Study 2, we found evidence of class-level variance for comment purpose, but no individual-level or class-level predictors

[3] Trainee and class learning data were not available for Study 3.

were significant. In Study 3, we will once again investigate the relationship between conscientiousness and purpose.

Hypothesis H9a (H9a): Conscientiousness will be positively related to comment prescriptiveness.

Item wording. One of the most important contextual influences on commenting may be item wording, particularly because it is a proximal cue for the behavior. In Studies 1 and 2, the open-ended item was general and descriptive (Kulesa & Bishop, 2006), which provided no cues about the type of comments to provide. As noted by Kulesa and Bishop (2006), "the type of question selected certainly influences the answers received and can drive the choice of methods applied to interpret those answers" (p. 239). This study includes two items (one general and one specific) to determine whether the pattern of results differs with regard to item specificity. The specific item provides a stronger situation (salient cues about performance expectations) compared with the general item (i.e., strong v. weak situations; Mischel, 1968). Both items are prescriptive in nature (i.e., soliciting recommendations) to standardize commenting purpose in a way that was plausible to respondents.

Research Question 4 (RQ4): How will item wording impact commenting behavior?

Research Question 5 (RQ5): Does the training context account for variability in commenting for both types of items and, if so, do elements of the task context (i.e., task difficulty and class size) explain the between-context variance?

Research Question 6 (RQ6): Will the pattern of relationships between individual and contextual factors with commenting be the same across item types?

Sample and Method

Data were collected as part of the same training effectiveness project as in Studies 1 and 2. Trainees (N = 314) were members of a large military organization participating in mandatory, long-term, job-related foreign language training. This was an initial acquisition language program conducted as part of the institutionalized training prior to unit assignment. These trainees were not job incumbents. There were 59 classes covering 10 languages, with two to eight trainees per class.[4] The gender and age profile is the same as Studies 1 and 2.

This study used the same measures as Study 1 for education and verbal aptitude, as well as the same training context variables of class size and language difficulty.

Personality. The four-item measure of conscientiousness on the mini-IPIP (Donnellan, Oswald, Baird, & Lucas, 2006) was used. Trainees responded on a 5-point Likert-type scale ranging from 1 (*very inaccurate*) to 5 (*very accurate*).

Interest in the topic. Interest in the topic was operationalized in the same way as in Studies 1 and 2, but the scale was revised for Study 3 and measured on the presurvey (as opposed to the postsurvey[5]). The four-item measure was rated on a 7-point Likert-type agreement scale. CFA results are presented in Table 1.

IPT. Three items were used to measure IPT and were taken from a scale developed by Levy, Stroessner, & Dweck (1998). The three items were written to reflect entity beliefs (i.e., belief that people do not change). Trainees responded on a 6-point Likert-type scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*).

Satisfaction. Satisfaction was measured with two global items developed for use in this study: (a) overall effectiveness of the instructor, and (b) overall usefulness of the training. These indicators of overall satisfaction were measured on the postsurvey. In addition, the posttraining survey included a measure of satisfaction, with five dimensions developed based on Lee and Pershing's (1999) framework. These dimensions are satisfaction with (a) program materials, (b) instructional activities, (c) program time and length, (d) training environment, and (e) delivery methods and technology. Trainees responded on a 5-point Likert-type scale ranging from 1 (*dissatisfied*) to 5 (*satisfied*). Table 1 provides the CFA results for this study.

Classroom management. Trainees rated their instructor's classroom management behavior on a six-item scale developed and validated for use in language training settings (SWA Consulting Inc., 2009). Trainees responded on a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Table 1 provides study-specific CFA results.

Open-ended items. Two open-ended prescriptive items were included to investigate the impact of the specificity of item wording. Client stakeholders wanted an exclusive focus on prescriptive items in order to request more actionable qualitative results from learners. The first is a specific prescriptive item related to the instructor: "Use the space below to provide any information or comments related to the instructor. Be sure to indicate things your instructor does that are particularly effective, as well as indicate anything the instructor can do to be more effective." The second is a general prescriptive item: "Please provide any additional comments about how your language training course can be improved or made more effective in the future."

Comment coding and interrater reliability. The coding process was the same as in Study 2, except that two raters were used for each phase. Interrater reliability for all coding was considered excellent (see Tables 10, 11, and 12) and was calculated using ICC(K) with two raters (LeBreton & Senter, 2008).

Results and Discussion

Means, standard deviations, reliability estimates, and correlations for all study variables are presented in Tables 10 through 12. Eighty-seven trainees (28% of the sample) provided comments in response to the instructor-specific item, 94 (30%) provided comments in response to the general item, and of those, 49 (16%) trainees provided comments in response to both items. Although commenting rates did not differ between the two item types, the fact that some trainees only responded to the instructor-specific item or the general item suggests that item wording does affect who comments (RQ4). The average comment tone rating was approaching "somewhat positive" (M = 3.88, SD = 1.25) for the instructor-specific item and "somewhat negative" (M = 2.23, SD = .73) for the general item. The

[4] This reflects the number of trainees per class who had complete data for inclusion in this study.

[5] This was an operational decision made to achieve a client objective.

Table 10

Study 3 – Descriptive Statistics and Zero-Order Correlations for Commenting Behavior Hypotheses

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Individual-level (L1) variables																	
1. Education	1.76	.42	—														
2. Verbal aptitude	57.64	5.04	.25**	—													
3. Interest in the topic	6.19	.76	.01	-.07	(.92)												
4. Conscientiousness	3.87	.68	.02	-.14*	.19**	(.87)											
5. Implicit person theory (entity)	3.51	1.20	-.07	-.01	-.13*	-.10	(.96)										
6. Satisfaction – Instructor	4.41	.87	-.03	.00	.12*	.07	.01	—									
7. Satisfaction – Utility	4.20	.88	-.07	-.07	.19**	.10	-.04	.50**	—								
8. Satisfaction – Program materials	3.76	.88	-.10	-.04	.10	.07	.01	.42**	.54**	(.95)							
9. Satisfaction – Instructional activities	3.98	.81	-.10	-.09	.11*	.08	.05	.65**	.52**	.61**	(.95)						
10. Satisfaction – Program time and length	4.02	.85	-.12*	-.04	.04	.09	.00	.41**	.47**	.50**	.52**	(.94)					
11. Satisfaction – Training environment	3.91	.87	-.01	-.04	.16**	.09	-.04	.46**	.52**	.53**	.53**	.46**	(.91)				
12. Satisfaction – Delivery methods and technology	3.99	.83	-.05	-.06	.18**	.04	-.03	.51**	.60**	.71**	.65**	.52**	.65**	(.95)			
13. Classroom management	4.36	.72	-.05	-.08	.13*	.08	-.02	.74**	.43**	.33**	.52**	.29**	.35**	.40**	(.98)		
14. Post specific comment	—	—	.06	.02	.05	-.02	-.12*	-.09	.03	-.09	-.11	-.17**	-.01	-.03	-.07	(.99)[a]	
15. Post general comment	—	—	.04	.19**	-.04	.03	-.12*	.04	-.05	-.12*	-.17**	-.08	-.08	-.14*	.03	.36**	(.94)[b]
Class-level (L2) variables																	
1. Task difficulty	2.71	1.23	—														
2. Class size	7.76	1.44	-.23	—													

Note. R_{Max} is provided on the diagonal for items where relevant. L1, *n* = 314; L2, *n* = 59. Education: 1 = *high school*, 2 = *some college or higher*; Interest in the topic: 1 = *strongly disagree*, 7 = *strongly agree*; Conscientiousness: 1 = *very inaccurate*, 5 = *very accurate*; Implicit *Person Theory*: 1 = *strongly disagree*, 6 = *strongly agree*; Satisfaction scales: 1 = *extremely dissatisfied*, 7 = *extremely satisfied*; Comment behavior: 0 = *no comment made*, 1 = *comment made*. L1 = Level 1; L2 = Level 2.

[a] Number of comments = 87; interrater reliability was calculated using ICC(K) with two raters (LeBreton & Senter, 2008). [b] Number of comments = 94; interrater reliability was calculated using ICC(K) with two raters (LeBreton & Senter, 2008).

* *p* :S .05. ** *p* :S .01.

Table 11

Study 3 – Descriptive Statistics and Zero-Order Correlations for Instructor-Specific Comment Characteristic Hypotheses

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
Individual-level (L1) variables														
1. Satisfaction – Instructor	4.30	1.06	—											
2. Satisfaction – Utility	4.20	.90	.45**	—										
3. Satisfaction – Program materials	3.59	.81	.53**	.47**	(.95)									
4. Satisfaction – Instructional activities	3.80	.84	.73**	.54**	.68**	(.95)								
5. Satisfaction – Program time and length	3.79	.96	.32**	.36**	.32**	.39**	(.94)							
6. Satisfaction – Training environment	3.87	.85	.53**	.51**	.50**	.58**	.42**	(.91)						
7. Satisfaction – Delivery methods and technology	3.92	.82	.59**	.50**	.55**	.65**	.29**	.69**	(.95)					
8. Classroom management	4.27	.88	.79**	.46**	.46**	.61**	.24**	.44**	.51**	(.98)				
9. Conscientiousness	3.88	.66	.01	.15	.09	.10	-.02	.16	.01	-.02	(.87)			
10. Comment tone	3.83	1.23	.65**	.33**	.45**	.58**	.32**	.43**	.41**	.65**	.01	(.97)[a]		
11. Comment scope	2.04	.62	-.24**	-.12	-.26**	-.34**	-.12	-.23*	-.30**	-.33**	.26**	-.48**	(.88)[a]	
12. Comment purpose	1.34	.66	-.27**	-.22*	-.28**	-.33**	-.10	-.23*	-.27**	-.29**	-.04	-.59**	.37**	(.97)[a]
Class-level (L2) variables														
1. Task difficulty	2.77	1.23	—											
2. Class size	7.77	1.49	-.24	—										

Note. R_{Max} is provided on the diagonal for items where relevant. L1, n = 107; L2, n = 52. Satisfaction scales: 1 = *extremely dissatisfied*, 7 = *extremely satisfied*; Personality scales: 1 = *very inaccurate*, 5 = *very accurate*; Comment tone: 1 = *very negative*, 5 = *very positive*; Comment scope: 1 = *not at all specific*, 3 = *very specific*; Comment purpose: 1 = *not at all prescriptive*, 3 = *very prescriptive*. L1 = Level 1; L2 = Level 2.

[a] Number of comments = 87; interrater reliability was calculated using ICC(K) with two raters (LeBreton & Senter, 2008).

* p < .05. ** p < .01.

Table 12

Study 3 – Descriptive Statistics and Zero-Order Correlations for General Recommendation Comment Characteristic Hypotheses

Individual-level variables	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Satisfaction – Instructor	4.45	.88	—											
2. Satisfaction – Utility	4.09	1.01	.46**	—										
3. Satisfaction – Program materials	3.54	.85	.52**	.52**	(.95)									
4. Satisfaction – Instructional activities	3.72	.84	.62**	.54**	.63**	(.95)								
5. Satisfaction – Program time and length	3.90	.86	.27**	.40**	.42**	.40**	(.94)							
6. Satisfaction – Training environment	3.80	.88	.42**	.49**	.37**	.44**	.26**	(.91)						
7. Satisfaction – Delivery methods and technology	3.78	.85	.53**	.50**	.57**	.62**	.28**	.57**	(.95)					
8. Classroom management	4.39	.73	.80**	.44**	.41**	.50**	.22*	.34**	.45**	(.98)				
9. Conscientiousness	3.92	.64	.11	.16	.15	.19*	.08	.17	.16	.12	(.87)			
10. Comment tone	2.23	.72	.01	.26**	.14	.12	.02	.12	.01	.05	.08	(.92)[a]		
11. Comment scope	2.48	.47	.01	-.05	-.11	-.06	-.06	.06	.01	.01	.13	-.22*	(.91)[a]	
12. Comment purpose	2.20	.79	.06	-.01	-.12	-.10	-.09	.05	-.14	.05	.04	-.12	.02	(.88)[a]

Note. N = 116. R_{Max} is provided on the diagonal for items where relevant. Satisfaction scales: 1 = *extremely dissatisfied*, 7 = *extremely satisfied*; Personality scales: 1 = *very inaccurate*, 5 = *very accurate*; Comment tone: 1 = *very negative*, 5 = *very positive*; Comment scope: 1 = *not at all specific*, 3 = *very specific*; Comment purpose: 1 = *not at all prescriptive*, 3 = *very prescriptive*.

[a] Number of comments = 94; interrater reliability was calculated using ICC(K) with two raters (LeBreton & Senter, 2008).

* p < .05. ** p < .01.

average comment scope rating was "somewhat specific" for both the instructor-specific item ($M = 2.01$, $SD = .63$) and the general item ($M = 2.46$, $SD = .45$), although slightly higher for the general item. The average comment purpose rating was "not at all prescriptive" ($M = 1.30$, $SD = .62$) for the instructor-specific item and "somewhat prescriptive" ($M = 2.19$, $SD = .80$) for the general item.

HGLM was used to explore the hypotheses related to commenting behavior (see Tables 13 and 14). Unconditional models showed significant class-level variance for instructor-specific ($T_{00} = .24$, $df = 58$, $\chi^2 = 76.34$, $p = .05$) and general ($T_{00} = .39$, $df = 58$, $\chi^2 = 81.24$, $p < .05$) commenting behavior, with 7% and 11% of the variance attributable to classes, respectively, addressing RQ5. For the instructor-specific item, the significant individual-level predictors in the HGLM were entity theory (supporting H6e), satisfaction with the utility of training (in the opposite direction as expected), and satisfaction with program time and length (supporting H6f). The predictors explained 12% of the variability in the odds of instructor-specific commenting. Bivariate correlations for entity theory and satisfaction with program time and length were also significant in the direction expected. Neither of the class-level predictors examined was significantly related to instructor-specific commenting behavior.

For the general item, the significant individual-level predictors in the HGLM were verbal aptitude (supporting H6b), entity theory (supporting H6e), and satisfaction with instructional activities (supporting H6f). The predictors explained 20% of the variance in general commenting. Bivariate correlations for verbal aptitude, entity theory, and satisfaction with instructional activities were also significant, consistent with the HGLM results. Two additional correlations were also significant: satisfaction with program materials and satisfaction with delivery methods and technology. The class-level predictors were not associated with commenting behavior on the general recommendation item.

In terms of comment characteristics for the instructor-specific item, there was significant between-class variance for tone ($T_{00} = .57$, $df = 51$, $\chi^2 = 117.75$, $p < .001$), scope ($T_{00} = .07$, $df = 51$, $\chi^2 = 73.31$, $p < .05$), and purpose ($T_{00} = .23$, $df = 51$, $\chi^2 = 154.53$, $p < .001$), with 38%, 18%, and 51% of the variance attributable to class, respectively. Table 15 includes the HLM results examining predictors of instructor-specific comment tone, scope, and purpose. Significant individual-level predictors for comment tone were satisfaction with the instructor (supporting H7a) and classroom management (supporting H7b). All dimensions of satisfaction and classroom management were positively correlated with comment tone in the bivariate results. With regard to the class-level variables, task difficulty was negatively related to comment tone (addressing RQ4). The predictors explained 23% of the individual-level variance, and 93% of the class-level variance in instructor-specific comment tone.

Significant individual-level predictors for comment scope were conscientiousness (supporting H8c) and classroom management (H8b). In the bivariate results, scope was negatively correlated with five dimensions of satisfaction, and was positively correlated with conscientiousness. Task difficulty was positively related to comment scope (addressing RQ4). The predictors explained 7% of the individual-level variance, and 98% of the class-level variance in instructor-specific comment scope. For comment purpose, although bivariate results showed negative relationships between seven dimensions of satisfaction and comment purpose, in the HLM analyses, only the class-level predictors were significant (addressing RQ4). Both class size and task difficulty were positively associated with comment purpose, explaining 13% of the class-level variance.

With regard to the comment characteristics for the general recommendations item, there was no significant class-level variance for tone, scope, or purpose; therefore, multiple regression was used to explore the hypothesized relationships (see Table 16). There were no significant predictors in either the regression or bivariate results for comment scope or purpose for the general item. However, satisfaction with the utility of training was signif-

Table 13
Study 3 – HGLM Results for Instructor-Specific Commenting Behavior

Fixed effects	Coefficient	*SE*	*t*	Odds ratio	95% confidence interval
Individual-level (L1) predictor					
Intercept	-1.053**	.160	-6.574	.349	.253, .481
Education	.186	.309	.601	1.204	.655, 2.213
Verbal aptitude	.016	.028	.563	1.016	.961, 1.073
Interest in the topic	.033	.185	.179	1.034	.718, 1.487
Conscientiousness	-.094	.242	-.389	.910	.565, 1.466
Implicit person theory (entity)	-.208*	.099	-2.108	.812	.669, .986
Satisfaction – Instructor	-.048	.288	-.166	.953	.540, 1.682
Satisfaction – Utility of training	.446*	.192	2.323	1.562	1.070, 2.281
Satisfaction – Program materials	-.246	.225	-1.094	.782	.502, 1.218
Satisfaction – Instructional activities	-.146	.216	-.676	.864	.565, 1.323
Satisfaction – Program time and length	-.579**	.177	-3.271	.561	.396, .794
Satisfaction – Training environment	.173	.200	.864	1.188	.802, 1.761
Satisfaction – Delivery methods and technology	.127	.279	.455	1.135	.655, 1.967
Classroom management	-.181	.321	-.565	.834	.443, 1.570
Class-level (L2) predictor					
Task difficulty	-.139	.134	-1.042	.870	.666, 1.137
Class size	.028	.109	.258	1.029	.826, 1.281

Note. L1, $n = 314$; L2, $n = 59$. HGLM = hierarchical generalized linear model; *SE* = robust standard error; L1 = Level 1; L2 = Level 2
* $p \leq .05$. ** $p \leq .01$.

Table 14
Study 3 – HGLM Results for General Recommendation Commenting Behavior

Fixed effects	Coefficient	*SE*	*t*	Odds ratio	95% confidence interval
Individual-level (L1) predictor					
Intercept	-.997**	.165	-6.041	.369	.265, .514
Education	-.164	.336	-.487	.849	.438, 1.647
Verbal aptitude	.084**	.031	2.671	1.088	1.022, 1.157
Interest in the topic	-.163	.147	-1.109	.849	.636, 1.135
Conscientiousness	.206	.201	1.026	1.229	.827, 1.827
Implicit person theory (entity)	-.250*	.110	-2.278	.779	.627, .967
Satisfaction – Instructor	.542	.285	1.899	1.719	.980, 3.017
Satisfaction – Utility of training	.050	.193	.261	1.052	.719, 1.538
Satisfaction – Program materials	-.051	.188	-.271	.950	.656, 1.377
Satisfaction – Instructional activities	-.776**	.214	-3.630	.460	.302, .701
Satisfaction – Program time and length	.007	.185	.040	1.007	.699, 1.451
Satisfaction – Training environment	.039	.253	.155	1.040	.632, 1.712
Satisfaction – Delivery methods and technology	-.297	.299	-.994	.743	.413, 1.338
Classroom management	.283	.262	1.080	1.327	.792, 2.222
Class-level (L2) predictor					
Task difficulty	.087	.150	.578	1.091	.808, 1.473
Class size	.035	.139	.251	1.036	.784, 1.368

Note. L1, $n = 314$; L2, $n = 59$. HGLM = hierarchical generalized linear model; *SE* = robust standard error; L1 = Level 1; L2 = Level 2.
* $p < .05$. ** $p < .01$.

icant and positively related to comment tone, which was the opposite direction as hypothesized (H6f).

Of note, the consistent finding from Studies 1 and 2 that interest in the topic influences commenting was not replicated here. This may be related to slight revisions to the scale or the timing of measurement (measured on the presurvey, which is more distal than the previous measure). This suggests that interest in the topic at the time of commenting may be more important than pretraining interest. The contextual factors—task difficulty and class size—were not significantly related to commenting for either type of item, informing RQ3.

Considering that some aspects of satisfaction were positively related to commenting, whereas others were negatively related, our findings highlight the importance of a multidimensional view of training reactions. In addition, not all dimensions of satisfaction were consistently related to commenting on the two items, suggesting that item wording (i.e., specific vs. general) may have an impact as well.

It is interesting that the positive relationship between satisfaction with the instructor and commenting holds in this study when the item is general and prescriptive, but is not significant for the specific, instructor item in which the prompt asked for both positive (things that are effective) and negative (areas for improvement) feedback. One possible interpretation is that the specific wording of the item eliminated the effect of satisfaction on commenting, so that both those who were dissatisfied and those who were satisfied with the instructor commented.

Table 15
Study 3 – HLM Results for Instructor-Specific Comment Tone, Scope, and Purpose

	Tone			Scope			Purpose[a]		
Fixed effects	Coefficient	*SE*	*t*	Coefficient	*SE*	*t*	Coefficient	*SE*	*t*
Individual-level (L1) predictor									
Intercept	3.812**	.080	47.368	2.059**	.048	42.743	1.386**	.077	18.035
Satisfaction – Instructor	.267*	.122	2.194	.160	.092	1.734			
Satisfaction – Utility of training	-.199	.114	-1.752	.090	.065	1.379			
Satisfaction – Program materials	.108	.170	.636	-.031	.087	-.351			
Satisfaction – Instructional activities	.236	.145	1.633	-.203	.124	-1.634			
Satisfaction – Program time and length	.114	.086	1.328	.019	.056	.345			
Satisfaction – Training environment	.142	.110	1.293	-.070	.071	-.989			
Satisfaction – Delivery methods and technology	-.204	.140	-1.459	-.060	.081	-.746			
Classroom management	.544**	.123	4.412	-.252**	.091	-2.761			
Conscientiousness				.275**	.070	3.953	-.034	.090	-.384
Class-level (L2) predictor									
Task difficulty	-.177**	.049	-3.586	.115**	.038	3.025	.139*	.066	2.097
Class size	-.101	.069	-1.466	-.007	.032	-.221	.137*	.068	2.016

Note. L1, $n = 107$; L2, $n = 52$. HLM = hierarchical linear modeling; *SE* = robust standard error; L1 = Level 1; L2 = Level 2.
[a] Satisfaction and classroom management were not hypothesized to relate to comment purpose, so they were not included in the regression analyses.
* $p < .05$. ** $p < .01$.

Table 16
Study 3 – Regression Results for General Recommendation Comment Tone, Scope, and Purpose

Predictor	Tone				Scope				Purpose[a]			
	B	*SE B*		*t*	*B*	*SE B*		*t*	*B*	*SE B*		*t*
Satisfaction – Instructor	-.142	.135	-.174	-1.049	.046	.095	.087	.487				
Satisfaction – Utility of training	.237**	.084	.350	2.841	-.022	.060	-.048	-.377				
Satisfaction – Program materials	.118	.103	.146	1.143	-.081	.073	-.146	-1.098				
Satisfaction – Instructional activities	.054	.110	.067	.495	-.051	.081	-.093	-.634				
Satisfaction – Program time and length	-.092	.082	-.114	-1.128	-.009	.059	-.017	-.155				
Satisfaction – Training environment	.061	.091	.078	.666	.049	.065	.092	.757				
Satisfaction – Delivery methods and technology	-.179	.107	-.226	-1.676	.035	.077	.063	.455				
Classroom management	.032	.148	.033	.216	-.012	.104	-.019	-.118				
Conscientiousness					.094	.070	.129	1.345	.044	.114	.036	.389
R^2	.125				.047				.001			

Note. $N = 116$.
[a] Satisfaction and classroom management were not hypothesized to relate to comment purpose, so they were not included in the regression analyses.
* $p < .05$. ** $p < .01$.

In addition, there was a new finding for the influence of entity beliefs on commenting across both item types (the only significant individual or contextual factor for both the general and specific items; RQ6), suggesting that commenting might be, in part, a function of a dispositional view that change is not possible. This may impact what the trainees expect can and will be done with the feedback. If there is an expectation that the training will not or cannot be modified or improved (e.g., instructor's performance), there may be no motivation to provide input. Given evidence suggesting that receiving survey feedback and observing actions resulting from a survey influence future survey responses (Thompson & Surface, 2009), our findings regarding entity beliefs should be further explored to investigate the interplay of participating in evaluation resulting in programmatic change, implicit beliefs, and commenting.

In terms of comment characteristics, we did not find between-class variance for the general recommendations item (consistent with Study 2 findings). However, we did find significant between-class variance for the instructor-specific item for comment tone, scope, and purpose (addressing RQ5). It makes sense that there would be significant between-class variance related to the types of comments provided about the instructor, because instructor behavior and quality varies across classes. For example, trainees who were satisfied with their instructors and who thought their instructors effectively managed the class provided more positive comments.

General Discussion

As noted by Kraiger (2002), feedback, marketing, and decision making are the three main goals of training evaluation. Capturing trainee reactions related to the quality of training and instruction can provide valuable information for feedback (program or instructional improvement) as well as for program marketing and decision making. Although the value of collecting trainee reactions is debated (e.g., Long, DuBois, & Faley, 2008), reactions are widely used, and many organizational stakeholders find them of at least some value (ASTD, 2009). Many training evaluation surveys ask trainees to respond to open-ended questions as well as provide quantitative ratings of reaction items. The response rate is typically very different for quantitative and qualitative items with fewer trainees providing comments. This raises potential representativeness issues with comments (e.g., Rogelberg & Stanton, 2007), and suggests caution should be taken until more is known about commenting, such as who makes comments, what individual and contextual factors influence commenting, and the nature of comments made.

It has been our experience in working with training directors, managers, and supervisors for several years that they are always keenly interested in reading comments from trainees, but they do not necessarily interpret comments with appropriate caution. Training professionals should be aware of nonresponse to open-ended items and how that may impact the information they collect, use, and present within their organizations. Our studies show that in training contexts, aggregating commenting data across classes could hide potentially important class-level differences. It is important not to overgeneralize comments beyond the appropriate level.

Our research should be considered an important first step in gaining a deeper understanding of trainee commenting on training evaluation surveys by providing empirical evidence that can inform the use of qualitative reactions in organizations for the three goals noted by Kraiger (2002). Previous work looking at commenting has been done with organizational climate surveys or 360-degree feedback, but the training context needs to be investigated because it differs in important ways that may impact commenting. Related to AET, training comments are often made within the context of a discrete training event (i.e., event focused), not the job or the overall organizational experience. Training can be totally separated from other work responsibilities—one's manager and organization's climate cannot be so easily separated. Because trainees are being asked to provide feedback posttraining (typically) to inform future training that will not directly benefit them (they typically will not participate in the same training course again), they may be less likely to provide in-depth feedback, as if there were a direct future benefit for them. Because the results likely would have no impact on their day-to-day jobs, the commenting and the characteristics of comments provided may differ substantially.

Additionally, organizational training is increasingly being delivered online through vendors (Burgess & Russell, 2003), and

learners are increasingly going online for self-directed learning (Brown, 2001). As commenting has become ubiquitous online (e.g., Facebook), instructional or training review sites have emerged with user comments in addition to quantitative ratings (e.g., Ratemyprofessor, MoocAdvisor). Just as comments on training evaluation surveys can impact decisions related to training, comments on social media or review sites have the same potential to impact the adoption or selection of learning opportunities. Comments made via social media and review sites should be investigated empirically, and the findings from our three studies can inform this needed research.

Findings and Implications

We conducted three studies in a work-related training context to further understanding of commenting by studying both individual and situational factors related to the behavior (Johns, 2006; Lewin, 1936). Because we consider commenting to be performance (Borman, 1991; Campbell, 1999), we used Campbell's (1990) conceptualization of performance and more specific motivational theories (i.e., AET; Weiss & Cropanzano, 1996) to explore distal and proximal individual difference predictors. In addition, we leveraged the training context and multilevel analysis to explore situational influences. Although commenting should largely be influenced by individual factors, the training context offers an opportunity to investigate situational influences as well as individual differences using a multilevel approach, as there are numerous classes and each one creates a micro context. Taken together, our three studies address a need for more multilevel analysis in training research (Mathieu & Tesluk, 2010).

A major contribution of this research is our consistent finding across three studies that a trainee's class accounts for variability in commenting on end-of-course training surveys, which has important implications for how qualitative training reactions should be analyzed and presented. Table 17 provides a summary of the multivariate findings across all three studies.

In terms of overall commenting, we found 28% to 53% commenting rates across our three studies, which is comparable with commenting research in other areas, such as organizational climate surveys (although 53% is slightly higher than previous research). This means 50% to 70% of trainees who respond to closed-ended items do not comment on training evaluation surveys. These respondents' opinions about the training may or may not be captured by the closed-ended items. There is also the possibility that the comments made may not be representative of the nonresponders opinions. From a practical standpoint, training professionals may consider collecting data using alternative techniques (i.e., focus

Table 17
Comparison of Findings Across Studies

Variables	Measured in study	Comment behavior	Comment tone	Comment scope	Comment purpose
Class-level variance					
Study 1 – General item		Yes			
Study 2 – General item		Yes	No	No	Yes
Study 3a – General item		Yes	No	No	No
Study 3b – Specific item		Yes	Yes	Yes	Yes
Individual-level predictors					
Education	1, 2, 3				
Verbal aptitude	1, 2, 3	3a			
Interest in the topic	1, 2, 3	1, 2			
Conscientiousness	1, 2, 3		2	2, 3b	
Extraversion	2				
Emotional stability	2				
Agreeableness	2				
Implicit person theory	3	3a, 3b			
Satisfaction – Instructor	1, 2	2[a]			
Satisfaction – Utility of training	1, 2		2		
Satisfaction – Language training in unit	1, 2				
Satisfaction – Learning materials and environment	1, 2				
Satisfaction – Instructor	3		3b		
Satisfaction – Utility of training	3	3b	3a		
Satisfaction – Program materials	3				
Satisfaction – Instructional activities	3	3a			
Satisfaction – Program time and length	3	3b			
Satisfaction – Training environment	3				
Satisfaction – Delivery methods and technology	3				
Classroom management	3		3b	3b	
Class-level predictors					
Training hours	1				
Task difficulty	1, 2		3b	3b	3b
Class size	2				3b
Class learning	2	2			

Note. The numbers in this table refer to the Study (1, 2, or 3). Because there were two open-ended items explored in Study 3, 3a refers to findings related to the general item and 3b refers to findings related to the specific item.
[a] Relationship was in the opposite direction as hypothesized.

groups) to supplement surveys and/or changing open-ended item wording to more strongly encourage responding. More research is needed to determine what types of open-ended questions engender the most response. Finally, analyzing responses to closed-ended items to assess whether commenters differ from noncommenters on important variables of interest may provide a rough indication of whether there might be bias in the comments, and, therefore, how to aggregate, analyze, and present the qualitative results. In some cases, acknowledging that nonresponse bias potentially exists and that the comments should be viewed with caution may be sufficient. In other cases, it may be important to consider the level of nonresponse (i.e., individual, class, program) in determining the appropriate level of aggregation of information and generalizability. Regardless, nonresponse should be considered when interpreting comments. Research on how comments are used for decisions and how nonresponse bias impacts the quality of those decisions needs to be conducted.

In terms of item wording, Study 3 provides some limited evidence that item wording may reduce biased responding. Although satisfaction with the instructor was associated with commenting in response to general open-ended items for Studies 2 and 3, the effect was not observed for the more specific open-ended item wording, which was focused on the instructor. The more specific item directed trainees to provide both positive and constructive feedback, which may have encouraged dissatisfied trainees to comment equally as often as the satisfied trainees. Although not conclusive, this is worth additional exploration. Future research should investigate item wording (specificity) and item purpose experimentally, while taking into account relevant individual and contextual factors. As survey items create a context for response, it is also important to consider research related to situational cues, such as situation strength (Mischel, 1968) or trait relevance and activation (Tett & Burnett, 2003), when investigating item wording. As noted by Hattrup and Jackson (1996), to understand behavior, individual differences, situational factors, and criteria (measures of behavior) must be considered jointly.

Overall, our findings were mixed in terms of identifying individual differences that influenced commenting. Previous research has focused on commenting as an individual-level phenomena exclusively. Although we found significant class-level variation across the studies, most of the variation in commenting was associated with the individual level (81% to 93%). We found little consistent support across studies for the relationship between distal or trait-like individual differences (i.e., education, verbal aptitude, conscientiousness) in commenting. We found more consistent relationships with the more proximal or state-like predictors, such as interest in the topic and satisfaction. In Studies 1 and 2, interest in the topic was positively related to commenting, which means those who were not interested in the training did not comment. If this happens, training supervisors are likely missing valuable feedback from the audience they need to reach the most (i.e., those who are not interested). This is particularly relevant for mandatory training, like that studied here, but may not generalize to voluntary training. Various dimensions of satisfaction were related to commenting in Studies 2 and 3, but not always in the directions hypothesized. Finally, entity theory was related to commenting across two different item types in Study 3. Taken collectively, with the exception of entity theory, these individual differences that had consistent prediction can be classified as more "state-like" than "trait-like," and, therefore, may be influenced to change behavior (e.g., aptitude-treatment interaction; Cronbach, 1957). Future research should examine whether the state–trait distinction matters in this line of research and explore additional "state" variables, such as mood, in an attempt to gain a great understanding of commenting. Furthermore, the mixed findings related to satisfaction suggest the need to conceptualize and study training reactions from a multidimensional perspective in training research. Alvarez, Salas, and Garofano (2004) identified reaction measures as one of three "most salient areas in need of further development" in the training literature (p. 407). More research is needed to explore other dimensions of satisfaction and additional individual difference variables. In addition, training satisfaction might be better conceptualized using a formative model as opposed to reflective model (Law & Wong, 1999; MacKenzie, Podsakoff, & Jarvis, 2005), which is also an important area for future research.

This study explored several contextual factors (i.e., class size, length, task difficulty, and class learning) and found few significant predictors of the class-level variance in commenting. Class learning (only investigated in Study 2) was the only significant class-level predictor of commenting among those we explored in our three studies. It is likely that commenting is affected by proximal contextual factors (elements of the discrete context), just as proximal individual differences have more influence than distal predictors. Future commenting research should explore other variables in the task, social, and physical contexts (Johns, 2006).

Finally, we explored the nature of comments in Studies 2 and 3. For the more general item in both studies, there were no differences across classes in tone and scope, suggesting that the type of comment provided is largely driven by the individual and not the context. However, there was class-level variance in tone, scope, and purpose for the instructor-specific item in Study 3, suggesting that the type of comment provided depends on the type of question asked. These findings also make sense from a situational strength perspective, as weak situations (i.e., responding to a general comment) allow for the expression of individual differences, whereas strong situations (i.e., responding to a specific item) constrain individual differences (Mischel, 1968). Future research should apply the coding model used in our study to evaluate its merit and explore these questions in other contexts. Additionally, future research should use open-ended item wording and instructions to determine whether organizations can cue the type of response needed for its feedback, marketing, and decision-making purposes.

Overall, this study can be considered a first step in beginning to understand what drives commenting behavior and the quality of comments provided. Our findings that show that more proximal individual difference and contextual factors influence commenting suggest the importance of exploring more complex models (i.e., mediation and moderation) of commenting behavior. It is likely that some of the more distal individual differences are mediated by more proximal predictors and moderated by factors in the omnibus and discrete contexts.

Limitations

Study-specific limitations included small sample size (Study 1), limited contextual variables to explore (Study 1), use of single-item measures (Study 3), and change in measurement (between Studies 2 and 3). All three studies may have been impacted by mono-source/mono-method bias, given that most data were collected via surveys. However, in our studies, most relationships investigated were measured on multiple surveys with a substantial amount of time between administrations.

As is true in any applied context, another limitation is related to the generalizability of these findings to other populations, administration contexts, survey types, and question types. Although we were able to investigate our hypotheses using three separate samples and provide some evidence for generalizability, the norms for commenting in this sample may differ from those found in non-military and more gender-balanced populations. The strong organizational culture of the military and the mandatory nature of training may lead to higher commenting rates compared with voluntary training that occurs in many other organizations. The commenting rate was higher for Study 2 compared commenting rates on other types of surveys, but the commenting rates for Studies 1 and 3 were consistent with previous research. Therefore, it is unclear whether the omnibus context (Johns, 2006) of the military might create a generalization issue. At present, it is unknown as to whether or not this study's findings extend to alternative training configurations and formats (e.g., shorter duration of the training program, shorter length of the survey, different discrete context, and different content). More research is needed to understand the impact of training context (e.g., Mathieu & Tesluk, 2010), especially on learning and learning outcomes, such as qualitative reactions.

Commenting, like other behaviors, is a function of the individual in his or her environment (Lewin, 1936), and is also a function of how commenting is measured (Hattrup & Jackson, 1996). Comment representativeness has implications for the three goals of evaluation (Kraiger, 2002)—feedback, marketing, and decision making. We know stakeholders who have used comments or exemplar comments for all three. Essentially, this study provides a warning to decision makers who may become intrigued by recent trends in qualitative analysis (i.e., word clouds, text analysis) and focus only on the rich and vivid feedback provided by respondents, without considering where those comments came from or whether they are representative of the population. In particular, in the context of training, it is important to determine the appropriate level of generalizability or aggregation of comment data to avoid masking important class-level differences. Although our studies provide some initial information about commenting behavior in the context of training evaluation, more research is needed to help guide best practices in collecting and using this data.

References

Aggarwal, C. C., & Zhai, C. (Eds.). (2012). *Mining text data*. New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-3223-4

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50,* 341–358. http://dx.doi.org/10.1111/j.1744-6570.1997.tb00911.x

Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review, 3,* 385–416. http://dx.doi.org/10.1177/1534484304270820

American Society of Training and Development. (2009). *The value of evaluation: Making training evaluations more effective*. Retrieved from http://www.astd.org/Publications/Research-Reports/2009/2009-Value-of-Evaluation-Making

Baggaley, J. (2013). MOOC rampant. *Distance Education, 34,* 368–378. http://dx.doi.org/10.1080/01587919.2013.835768

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9,* 9–30. http://dx.doi.org/10.1111/1468-2389.00160

Bauer, K. N., Orvis, K. A., Ely, K., & Surface, E. A. (2013, May). *Reexamining the construct of motivation in learning contexts: A meta-analytic path analysis.* Poster presented at the 25th Annual Convention of the Association for Psychological Science, Washington, DC.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, D. K. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323–370. http://dx.doi.org/10.1037/1089-2680.5.4.323

Bhavsar, K. S., Hess, C. E., & Surface, E. A. (2008, April). *Validation of a learning styles instrument*. Paper presented at the 23rd annual meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Bhavsar, K. S., Poncheri, R. M., & Surface, E. A. (2006, February). *Investigating the construct validity of a learning styles inventory.* Paper presented at the annual Industrial-Organizational/Organizational Behavior (IOOB) Graduate Student Conference, Fairfax, VA.

Borg, I. (2005, April). Who writes what kinds of comments? Some new findings. In A. I. Kraut (Chair), *Grappling with write-in comments in a web-enabled survey world*. Symposium conducted at the 20th annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Borman, W. C. (1991). Job behavior, performance, and effectiveness. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, pp. 271–326). Palo Alto, CA: Consulting Psychologists Press.

Brown, K. G. (2001). Using computers to deliver training: Which employees learn and why? *Personnel Psychology, 54,* 271–296. http://dx.doi.org/10.1111/j.1744-6570.2001.tb00093.x

Brown, K. G. (2005). An examination of the structure and nomological network of trainee reactions: A closer look at "smile sheets." *Journal of Applied Psychology, 90,* 991–1001. http://dx.doi.org/10.1037/0021-9010.90.5.991

Burgess, J. R. D., & Russell, J. E. A. (2003). The effectiveness of distance learning initiatives in organizations. *Journal of Vocational Behavior, 63,* 289–303. http://dx.doi.org/10.1016/S0001-8791(03)00045-9

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 687–732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P. (1999). The definition and measurement of performance in the new age. In D. R. Ilgen & E. D. Pulakos (Eds.), *The changing nature of performance: Implications for staffing, motivation, and development* (pp. 399–429). San Francisco, CA: Jossey-Bass.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco, CA: Jossey-Bass.

Cerrito, P. (2000). An examination of college-wide student teaching evaluations. *College Student Journal, 34,* 165–171.

Clayton, D. K., Rogers, S., & Stuifbergen, A. (1999). Answers to unasked questions: Writing in the margins. *Research in Nursing & Health, 22,* 512–522. http://dx.doi.org/10.1002/(SICI)1098-240X(199912)22:6<512::AID-NUR8>3.0.CO;2-G

Clement, R. W. (1982). Testing the hierarchy theory of training evaluation: An expanded role for trainee reactions. *Public Personnel Management, 12,* 176–184.

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. http://dx.doi.org/10.3102/00346543051003281

Cole, N. (2008). How long should a training program be? A field study of "rules-of-thumb." *Journal of Workplace Learning, 20,* 54–70. http://dx.doi.org/10.1108/13665620810843647

Craig, C. S., & McCann, J. M. (1978). Item nonresponse in mail surveys: Extent and correlates. *Journal of Marketing Research, 15,* 285–289. http://dx.doi.org/10.2307/3151264

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671– 684. http://dx.doi.org/10.1037/h0043943

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93,* 880–896. http://dx.doi.org/10.1037/0022-3514.93.5.880

Dierdorff, E. C., & Surface, E. A. (2008). If you pay for skills, will they learn? Skill change and maintenance under a skill-based pay system. *Journal of Management, 34,* 721–743. http://dx.doi.org/10.1177/0149206307312507

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18,* 192–203. http://dx.doi.org/10.1037/1040-3590.18.2.192

Drewes, D. W. (2009). Subject-centered scalability: The sine qua non of summated ratings. *Psychological Methods, 14,* 258–274. http://dx.doi.org/10.1037/a0016621

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77,* 615–622. http://dx.doi.org/10.1037/0021-9010.77.5.615

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41,* 1040–1048.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. *Review of Educational Research, 73,* 321–368. http://dx.doi.org/10.3102/00346543073003321

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/0471445428

Forsyth, I., Jolliffe, A., & Stevens, D. (1995). *Evaluating courses: Practical strategies for teachers, lecturers and trainers*. London, UK: Kogan Page.

Gannon, M. J., Nothern, J. C., & Carroll, S. J. (1971). Characteristics of nonrespondents among workers. *Journal of Applied Psychology, 55,* 586–588. http://dx.doi.org/10.1037/h0031907

Garcia, J., Evans, J., & Reshaw, M. (2004). "Is there anything else you would like to tell us"—Methodological issues in the use of free-text comments from postal surveys. *Quality & Quantity: International Journal of Methodology, 38,* 113–125. http://dx.doi.org/10.1023/B:QUQU.0000019394.78970.df

George, J. M., & Jones, G. R. (2003). Psychological dimensions of organizational behavior. In B. M. Staw (Ed.), *Psychological dimensions of organizational behavior* (Vol. 3, pp. 3–23). Upper Saddle River, NJ: Prentice Hall.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, Netherlands: Tilburg University Press.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations* (4th ed.). Belmont, CA: Wadsworth.

Green, K. E. (1996). Sociodemographic factors and mail survey response. *Psychology & Marketing, 13,* 171–184. http://dx.doi.org/10.1002/(SICI)1520-6793(199602)13:2<171::AID-MAR4>3.0.CO;2-C

Greenwald, A. G., & Gillmore, G. M. (1997a). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52,* 1209–1217. http://dx.doi.org/10.1037/0003-066X.52.11.1209

Greenwald, A. G., & Gillmore, G. M. (1997b). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology, 89,* 743–751. http://dx.doi.org/10.1037/0022-0663.89.4.743

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68,* 2–31. http://dx.doi.org/10.1093/poq/nfh002

Gully, S., & Chen, G. (2010). Individual differences, attribute-treatment interactions, and training outcomes. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 3–64). New York, NY: Routledge/Taylor & Francis Group.

Hackman, J. (2003). Learning more by crossing levels: Evidence from airplanes, hospitals, and orchestras. *Journal of Organizational Behavior, 24,* 905–922. http://dx.doi.org/10.1002/job.226

Hattrup, K., & Jackson, S. E. (1996). Learning about individual differences by taking situations seriously. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 507–547). San Francisco, CA: Jossey-Bass.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24,* 623–641. http://dx.doi.org/10.1177/014920639802400504

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Johns, G. (2006). The essential impact of context on organizational behavior. *The Academy of Management Review, 31,* 386–408. http://dx.doi.org/10.5465/AMR.2006.20208687

Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice* (pp. 331–376). San Francisco, CA: Jossey-Bass.

Kulesa, P., & Bishop, R. J. (2006). What did they really mean? New and emerging methods for analyzing themes in open-ended comments. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies, and applications* (pp. 238–263). San Francisco, CA: Jossey-Bass.

Law, K. S., & Wong, C. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management, 25,* 143–160. http://dx.doi.org/10.1016/S0149-2063(99)80007-5

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11,* 815–852. http://dx.doi.org/10.1177/1094428106296642

Lee, S. H., & Pershing, J. A. (1999). Effective reaction evaluation in evaluating training programs. *Performance Improvement, 38,* 32–39.

LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with big five personality characteristics and cognitive ability. *Journal of Applied Psychology, 86,* 326–336. http://dx.doi.org/10.1037/0021-9010.86.2.326

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personal and Social Psychology, 74,* 1421–1436.

Lewin, K. (1936). *Principles of topological psychology*. New York, NY: McGraw-Hill. http://dx.doi.org/10.1037/10019-000

Long, L. K., DuBois, C. Z., & Faley, R. H. (2008). Online training: The value of capturing trainee reactions. *Journal of Workplace Learning, 20*, 21–37. http://dx.doi.org/10.1108/13665620810843629

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology, 90*, 710–730. http://dx.doi.org/10.1037/0021-9010.90.4.710

Marcus, B., & Schütz, A. (2005). Who are the people reluctant to participate in research? Personality correlates of four different types of nonresponse as inferred from self- and observer ratings. *Journal of Personality, 73*, 959–984. http://dx.doi.org/10.1111/j.1467-6494.2005.00335.x

Martin, C. L. (1994). The impact of topic interest on mail survey response behaviour. *Journal of the Market Research Society, 36*, 327–338.

Mathieu, J. E., & Tesluk, P. E. (2010). A multilevel perspective on training and development effectiveness. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 405–440). New York, NY: Routledge/Taylor & Francis Group.

McNeely, R. L. (1990). Do respondents who pen comments onto mail surveys differ from other respondents? A research note on the human services job satisfaction literature. *Journal of Sociology and Social Welfare, 17*, 127–137.

Miller, L. (2013). *ASTD's 2013 state of the industry report*. Alexandria, VA: American Society of Training & Development.

Millsap, R. E. (2002). Structural equation modeling: A user's guide. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 257–301). San Francisco, CA: Jossey-Bass.

Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.

Morgan, R. B., & Casper, W. J. (2000). Examining the factor structure of participant reactions to training: A multidimensional approach. *Human Resource Development Quarterly, 11*, 301–317. http://dx.doi.org/10.1002/1532-1096(200023)11:3<301::AID-HRDQ7>3.0.CO;2-P

Orvis, K. A., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior, 24*, 2415–2433. http://dx.doi.org/10.1016/j.chb.2008.02.016

Peeters, G. (1971). The positive-negative asymmetry: On cognitive consistency and positivity bias. *European Journal of Social Psychology, 1*, 455–474. http://dx.doi.org/10.1002/ejsp.2420010405

Phillips, J. J. (1997). *Handbook of training evaluation and measurement methods*. Houston, TX: Gulf.

Pickett, L., & Fraser, B. (2010). Creating and assessing positive classroom learning environments. *Childhood Education, 86*, 321–326. http://dx.doi.org/10.1080/00094056.2010.10521418

Poncheri, R. M., Lindberg, J. T., Thompson, L. F., & Surface, E. A. (2008). A comment on employee surveys: Negativity bias in open-ended responses. *Organizational Research Methods, 11*, 614–630. http://dx.doi.org/10.1177/1094428106295504

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology, 88*, 1104–1114.

Rogelberg, S. G., & Luong, A. (1998). Nonresponse to mailed surveys: A review and guide. *Current Directions in Psychological Science, 7*, 60–65. http://dx.doi.org/10.1111/1467-8721.ep13175675

Rogelberg, S. G., Luong, A., Sederburg, M. E., & Cristol, D. S. (2000). Employee attitude surveys: Examining the attitudes of noncompliant employees. *Journal of Applied Psychology, 85*, 284–293. http://dx.doi.org/10.1037/0021-9010.85.2.284

Rogelberg, S. G., Spitzmüller, C., Little, I., & Reeve, C. L. (2006). Understanding response behavior to an online special topics organizational satisfaction survey. *Personnel Psychology, 59*, 903–923. http://dx.doi.org/10.1111/j.1744-6570.2006.00058.x

Rogelberg, S. G., & Stanton, J. M. (2007). Understanding and dealing with organizational survey nonresponse. *Organizational Research Methods, 10*, 195–209. http://dx.doi.org/10.1177/1094428106294693

Sanderson, G. (1995). Objectives and evaluation. In S. Truelove (Ed.), *Handbook of training and development* (2nd ed., pp. 113–114). Oxford, UK: Blackwell.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton, Mifflin and Company.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420

Siem, F. M. (2005, April). History of survey comments at the Boeing Company. In K. J. Fenlason (Chair), *Comments: Where have we been? Where are we going?* Symposium conducted at the 20th annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Silva, J. M., & White, L. A. (1993). Relation of cognitive aptitudes to success in foreign language training. *Military Psychology, 5*, 79–93. http://dx.doi.org/10.1207/s15327876mp0502_1

Smither, J. W., & Walker, A. G. (2004). Are the characteristics of narrative comments related to improvement in multirater feedback ratings over time? *Journal of Applied Psychology, 89*, 575–581. http://dx.doi.org/10.1037/0021-9010.89.3.575

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education, 20*, 165–184. http://dx.doi.org/10.1007/s11092-008-9053-z

SWA Consulting Inc. (2009, March). *Development of an instructor behavioral evaluation for foreign language instructors* (Tech. Rep. No. #2010010603). Raleigh, NC: Author.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*, 67–85. http://dx.doi.org/10.1037/0033-2909.110.1.67

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500–517. http://dx.doi.org/10.1037/0021-9010.88.3.500

Thompson, L. F., & Surface, E. A. (2009). Promoting favorable attitudes toward personnel surveys: The role of follow-up. *Military Psychology, 21*, 139–161. http://dx.doi.org/10.1080/08995600902768693

Tomaskovic-Devey, D., Leiter, J., & Thompson, S. (1994). Organizational survey nonresponse. *Administrative Science Quarterly, 39*, 439–457. http://dx.doi.org/10.2307/2393298

U.S. Department of Defense. (1984). *Test manual for the Armed Services Vocational Aptitude Battery (DoD 1340.12AA)*. North Chicago, IL: U.S. Military Entrance Processing Command.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70. http://dx.doi.org/10.1177/109442810031002

von Eye, A., & Mun, E. Y. (2005). *Analyzing rater agreement: Manifest variable methods*. Mahwah, NJ: Erlbaum.

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82*, 247–252. http://dx.doi.org/10.1037/0021-9010.82.2.247

Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology, 48*, 347–375. http://dx.doi.org/10.1111/j.1744-6570.1995.tb01761.x

Watson, A. M., Thompson, L. F., & Surface, E. A. (2011, April). *Cognitive and motivational influences on training performance: A longitudinal study*. Poster presented at the 26th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Weiss, H. M., & Cropanzano, R. (1996). Affective events theory: A theoretical discussion of the structure, causes, and consequences of affective experiences at work. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 18, pp. 1–74). Greenwich, CT: JAI Press.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Unpublished doctoral dissertation). University of California, Los Angeles.