ABSTRACT

HARMAN, REANNA PONCHERI. Understanding Survey Comment Nonresponse and the Characteristics of Nonresponders. (Under the direction of Lori Foster Thompson.)

This study explored survey commenting behavior and investigated the two broad questions of “who” comments and “how” respondents comment. Participants were military personnel (N=419) who were given the opportunity to respond to an open-ended question on a training evaluation survey after completing a foreign language course. The first set of hypotheses examined the question of “who” comments. As expected, results showed that dissatisfied individuals and those interested in the topic were especially inclined to comment. Additionally, partial support was found for a positive relationship between education and commenting behavior, such that those with some college education but no four-year degree or higher were more likely to comment than those with no college experience. Unexpectedly, however, no significant difference was found between respondents with a four-year degree and higher than those with either some or no college experience. In contrast with the study hypotheses, conscientious respondents were not especially inclined to comment.

The second set of hypotheses explored the question of “how” respondents comment. As hypothesized, dissatisfied respondents tended to provide negative comments, and conscientious respondents provided specific comments. In contrast to the study hypotheses, respondents scoring high on agreeableness and emotional stability did not provide particularly positive comments, those who were satisfied with course materials and the learning environment were more (not less) likely to provide specific comments, and those
high in extraversion and conscientiousness were not particularly likely to provide suggestions.

The vividness of verbatim comments may lead managers and training evaluators to place more weight on comments than quantitative data (i.e., ratings) when following up on survey results. This study suggests that those who do so may be basing their decisions on input from an unrepresentative segment of respondents. Steps should be taken to inform managers and training evaluators of the potential for nonresponse bias in open-ended comments. In addition, survey research and practice would benefit from interventions aimed at increasing the likelihood of hearing from the voices that often go unheard.
Understanding Survey Comment Nonresponse and the Characteristics of Nonresponders

by
Reanna Poncheri Harman

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Psychology

Raleigh, North Carolina
2008

APPROVED BY:

Dr. Lori Foster Thompson  Dr. S. Bartholomew Craig
Committee Chair

Dr. Samuel B. Pond, III  Dr. Mark A. Wilson
BIOGRAPHY

Reanna Poncheri Harman was born November 28, 1981 in Washington, DC. She graduated from Saint Mary's Ryken High School in Leonardtown, MD in 1999 and began her undergraduate studies at Mount Saint Mary's University in the fall of that year. She graduated summa cum laude with a Bachelor of Science degree in Business and a minor in Psychology in 2003.

Reanna began graduate school at North Carolina State University in 2003. She worked as a tutor during her first year with the Academic Support Program for Student Athletes. Reanna has taught Introduction to Psychology and Social Psychology at NC State and she participated in Preparing the Professoriate, a year-long mentoring program. Reanna is a Senior Consultant at SWA Consulting Inc., where she has been employed for four years. Reanna works on projects primarily related to training evaluation for various clients. Her research interests include survey research, training evaluation, and theoretical and empirical exploration of contextual predictors in these areas of research and in the domain of job performance and job analysis.
ACKNOWLEDGMENTS

When I was a child, one of my favorite books was a Sesame Street classic titled, *I Can Do It Myself*. The main characters showcased a spirit of autonomy when it came to some of the more challenging tasks encountered in childhood. What I have come to realize is that while there was a lot to learn from that book when I was young, achieving my Ph.D. was not something that I could do myself. There are so many people who made this accomplishment possible and I would like to take this opportunity to thank some very special people who helped me along my journey.

First and foremost, I would like to thank Dr. Lori Foster Thompson who was my guiding light and constant source of encouragement during this process. There are always some moments when you doubt your own ability to succeed, but she never once doubted me and gave me the support that I needed to make progress each day. I always left our meetings feeling confident and determined that I could achieve my goal.

I would like to thank my committee members, Dr. Mark Wilson, Dr. Bart Craig, and Dr. Bob Pond who provided me with guidance and support during this process, both at formal dissertation meetings and informally through email and in-person meetings. Their individual expertise helped to strengthen my manuscript and I am very grateful. I have learned so much from each of them both in and out of the classroom.

A special thank you to Dr. Eric Surface who has been my mentor for many years and who made it possible for me to work on this project. Without Eric’s support, guidance, and
feedback this research would not have been possible. I am very grateful for this and all other opportunities which he has made possible.

A very special thank you to two gentlemen, Mr. Jack Donnelly and Mr. Mark Roemer. Your support of this research is most appreciated and I am very glad to have had this opportunity to work with you.

I would like to acknowledge the excellent work of my data coding team: Jennifer London, Daniel Stanhope, Will Stoughton, and Tom Whelan. Thank you for your diligence and dedication to this project.

There are so many colleagues and friends who I would like to thank. I would like to thank Dr. Stephen Ward whose constant encouragement and advice to “do a little every day” truly had an impact on the way I approached my dissertation. I would like to thank Kartik Bhavsar, Clara Hess, and Aaron Watson for listening and providing support during the process. I would like to thank Tara Behrend who attended both my practice dissertation proposal and defense meetings and provided invaluable feedback. Finally, a very special thank you to Jenn Lindberg McGinnis, who read drafts, attended practice meetings, and helped me think through ideas. Jenn was always on the other end of the phone to listen when I needed to talk and I am so grateful for her friendship.

Without my family, I would never have had the opportunity to pursue my Ph.D. and I could not have asked for a more supportive and loving family. I would like to thank my grandparents, Betty and Jerry Finamore and Connie and Art Poncheri, who have always encouraged me to follow my dreams. I would like to thank Nicki Smith who has always been like family to me. Her will to succeed is contagious and I’m just glad that I’ve been around to
catch some of it when I needed it most. I would like to thank my brother, Jason Poncheri, for being a good friend and helping me to put things into perspective every now and then. Most of all, I would like to thank my parents, Debbie and Rick Poncheri, who have made many sacrifices so that I could achieve my goals. Their love, encouragement, and support have been unconditional and I couldn’t possibly thank them enough for this opportunity.

Last, but in no way least, I would like to thank my husband, Aaron, for his love and support. He is truly my best friend and I could not have done this without him by my side.
# TABLE OF CONTENTS

LIST OF TABLES........................................................................................................................................ vii

Literature Review....................................................................................................................................... 1

- Open-Ended Comments: The Context ........................................................................................................ 2
- Unit and Item Nonresponse ....................................................................................................................... 5
- Commenting Behavior: Who Comments? ...................................................................................................... 8

- Open-Ended Comments: A Classification Framework ............................................................................ 15
- How Participants Comment: Who Makes What Type of Comments? ............................................................ 18

Method..................................................................................................................................................... 23

- Study Background and Training Context ................................................................................................. 23
- Participants ........................................................................................................................................... 24
- Measures ............................................................................................................................................ 25
- Coding Comments ................................................................................................................................. 26

Results.................................................................................................................................................... 31

- Preliminary Analyses .............................................................................................................................. 31

Discussion.................................................................................................................................................. 41

- Coding Implications ................................................................................................................................. 44
- Limitations ........................................................................................................................................... 46
- Future Research .................................................................................................................................... 48
- Practical Implications ............................................................................................................................. 51

References.................................................................................................................................................. 55

Appendices.................................................................................................................................................. 73

- Appendix A. Interest in the Topic Items .................................................................................................... 74
- Appendix B. Personality Items: IPIP (Goldberg, 1999) ............................................................................. 75
- Appendix C. Satisfaction Items .................................................................................................................. 78
- Appendix D. Rater Training Protocol for Identifying Discrete Comments .................................................. 79
- Appendix E. Definitions of Comment Tone, Scope, and Purpose ................................................................. 81
- Appendix F. Rater Training Protocol for Coding Comment Tone, Scope, and Purpose ................................ 84
LIST OF TABLES

Table 1. Interrater Reliability and Agreement............................................... 64
Table 2. Exploratory Factor Analysis: Interest in the Topic.......................... 65
Table 3. Exploratory Factor Analysis: Satisfaction...................................... 66
Table 4. Descriptive Statistics and Correlation Matrix................................. 67
Table 5. Percentage of Respondents who Commented by Education Level....... 68
Table 6. Logistic Regression for Analysis Predicting Commenting Behavior..... 69
Table 7. Multiple Regression Predicting Comment Tone............................... 70
Table 8. Multiple Regression Predicting Comment Scope............................ 71
Table 9. Multiple Regression Predicting Comment Purpose.......................... 72
Literature Review

Survey response behavior is an important concern for practitioners that has received recent research attention (e.g., Marcus & Shutz, 2005; Rogelberg, Spitzmüller, Little, & Reeve, 2006). Organizational surveys are key communication tools between employees and decision-makers, but like all other forms of communication only provide the perspective of those who choose to participate. As such, it is important to understand the characteristics of those who participate in this form of communication as well as the characteristics of those who do not. Researchers have been working toward this goal by investigating the differences between individuals who choose to participate in survey research and those who do not, the aptly labeled nonresponders (e.g., Gannon, Nothern, & Carroll, 1971; Rogelberg et al., 2006).

While there is a well-developed research literature that focuses on understanding differences between responders and nonresponders, there has been less attention devoted to nonresponse which occurs when individuals skip items on organizational surveys. Both closed-ended and open-ended items are susceptible to skipping, but research suggests that nonresponse to open-ended items is a widespread, yet poorly understood phenomenon (Poncheri, Lindberg, Thompson, & Surface, 2008). Open-ended questions play an important role in organizational surveys because they provide a rich source of data by communicating employee opinions in the employees’ own words. Kulesa and Bishop (2006) emphasize the value of this type of information: “The rich and diverse content that can result from the intelligent application of comment questions creates a virtual focus group that management
can use to have a virtual conversation with employees on nearly any topic of interest” (p. 238).

It is clear that open-ended comments are valuable to organizational decision-makers, but it is probable that many employee opinions will be unheard as a result of nonresponse to these open-ended items. Even when comments are provided, it is important to acknowledge that these comments will have unique characteristics. Furthermore, it is probable that individual differences influence the types of comments provided.

The purpose of this study is to explore the issue of nonresponse as it applies to open-ended items on a post-training questionnaire and to explore characteristics of those who provide comments. This study considers two broad questions concerning “who” comments and “how” respondents comment. Therefore, the first goal of this study will be to examine the individual differences that distinguish those who provide comments (commenters) from those who do not provide comments (noncommenters) to gain a deeper understanding of who comments. A second goal will be to explore the individual differences that influence the types of comments that are provided to begin to understand characteristics related to how respondents comment.

Open-Ended Comments: The Context

There has been very little research attention aimed at understanding open-ended comments in the organizational survey literature. As such, it is important to begin by defining the context in which open-ended comments are explored. Several characteristics should be considered when studying open-ended comments, including the type of survey being administered and the types of open-ended questions on the survey.
There are many different types of surveys, some that are work-related and others that are not. Church and Waclawski (1998) provide a classification for surveys which are not work-related: political (public opinion polling, voting processes, election polling, party affiliation, census), economic (census, market/product research, advertising testing, economic behavior), and social (attitudes and opinions, leadership, religious values, social issues). There is no similar classification in the domain of organizational surveys, although Kraut (1996) identifies two main reasons for conducting organizational surveys: assessment and change.

An organizational survey is defined as “a systematic process of data collection designed to quantitatively measure specific aspects of organizational members’ experience as they relate to work” (Church & Waclawski, 1998, p. 4). A consideration of this definition suggests that there are three main types of organizational surveys regularly administered in the workplace: employee attitude surveys (including climate surveys, general employee opinion surveys, and special topic surveys), performance reviews (including 360-degree feedback), and training evaluation surveys. This distinction is important because the purpose and function of these surveys are unique and may lead to different types of response behavior.

Research has been conducted to explore open-ended comments in the context of employee attitude surveys (Poncheri et al., 2008; Rogelberg, Luong, Sederburg, & Cristol, 2000; Rogelberg et al., 2006) and 360-degree feedback (Smither & Walker, 2004), but there is no research which has explicitly examined the role of open-ended comments in the context of training evaluation. The current study will address this gap in the literature.
Types of questions. Survey items can be categorized into two main groups: closed-ended items and open-ended items. Closed-ended items are those that have pre-determined response options, while open-ended items allow respondents to use their own words when providing answers (Church & Waclawski, 1998; Edwards, Thomas, Rosenfeld, & Booth-Kewley, 1997; Fink, 1995; Krosnick, 1999).

Surveys often contain a mixture of closed-ended and open-ended questions (Church & Waclawski, 1998) and the wording of the question stem is an important concern when writing and analyzing both types of items (Dillman, 2000). Kulesa and Bishop (2006) offer a categorization of open-ended survey questions. Their categorization includes four different types of questions that vary along two dimensions: scope (general/specific) and purpose (prescriptive/descriptive). General questions are broad in focus and do not inquire about a particular issue, while specific questions are focused on targeted issues. In addition, questions can be descriptive (encouraging statements about a particular issue) or prescriptive (encouraging suggestions or recommendations for change). These types of questions may elicit very different responses and therefore, must be acknowledged when analyzing responses to open-ended questions. In this study, the open-ended question that will be investigated can be categorized as general and descriptive.

In summary, the hypotheses proposed in this study will be explored within the context of a training evaluation survey using a general, descriptive open-ended question. The next section provides a review of theory and research related to survey nonresponse and serves as the foundation for hypotheses related to the first broad question (i.e., “who” comments) explored in this study.
Unit and Item Nonresponse

There are two main types of nonresponse that have been identified in the literature: unit (or complete) nonresponse and item nonresponse. Unit nonresponse occurs when a potential participant does not respond to (i.e., fails to return; Dollinger & Leong, 1993; Rogelberg et al., 2006) a survey, whereas item nonresponse occurs when respondents omit individual items on a survey (Bosnjak & Tuten, 2001; Bosnjak, Tuten, & Wittman, 2005; Marcus & Shutz, 2005). Nonresponse is an important concern for both researchers and practitioners at both the unit- and item-levels.

Unit nonresponse. Rogelberg (2006) identifies several reasons why unit nonresponse is problematic including smaller data samples, reduced statistical power, larger confidence intervals, the inability to use certain statistical techniques, reduced face validity, and bias. Rogelberg and Luong (1998) define unit nonresponse bias as occurring when “the individuals responding to a survey differ from nonrespondents on variables relevant to the survey topic” (p. 60-61). Nonresponse bias is quantified according to the following equation: Nonresponse bias = P_{NR} (X_{Res} – X_{Pop}), where the proportion of nonrespondents is multiplied by the difference between the respondent mean and the population mean (if it were known) on a survey relevant variable. As this equation demonstrates, low response rates do not necessarily lead to bias and high response rates do not guarantee the absence of bias (Rogelberg & Stanton, 2007). The important piece of information to consider is whether or not respondents differ from nonrespondents in relevant ways.

Most of the research on nonresponse bias has examined unit nonresponse and there is evidence that the occurrence of unit nonresponse is on the rise. Baruch (1999) examined
survey response rates in the management and behavioral sciences literatures over a 20 year time-span (1975-1995) and determined that there has been a steady decline in response rates (i.e., from 64% to approximately 50%).

Rogelberg and colleagues have conducted several studies to examine the issue of unit nonresponse (Rogelberg et al., 2003; Rogelberg et al., 2000; Rogelberg et al., 2006; Spitzmüller, Glenn, Barr, Rogelberg, & Daniel, 2006). These researchers have explored different reasons for unit nonresponse and have identified two main forms: active and passive. Active nonrespondents are those who consciously choose not to respond to a survey request, while passive nonrespondents do not make a conscious choice to withhold participation (e.g., they may fail to respond because they forget or misplace the survey; Rogelberg et al., 2003).

Research in this area has shown that active nonrespondents differ from both passive nonrespondents and respondents (Rogelberg et al., 2003; Spitzmüller et al., 2006). For example, Rogelberg et al. (2003) found that active nonrespondents differed from passive nonrespondents in terms of satisfaction with the survey sponsor and conscientiousness. In Rogelberg et al.’s study, passive nonrespondents were more satisfied and conscientious than active nonrespondents. Therefore, bias may occur as a result of a failure to capture the opinions of active nonrespondents. However, it is important to note that active nonresponse is not necessarily indicative of negative attitudes, as implied by Rogelberg et al.’s (2003) findings. In a study which explored unit nonresponse in the context of an employee web survey, Thompson and Surface (2007) found that some active nonresponse may be the result of satisfaction with the current state of the organization.
Item nonresponse. It is difficult, and often impossible, to evaluate differences between responders and unit nonresponders because survey researchers often do not have access to information about participants who choose not to respond. However, the issue of item nonresponse can be more easily addressed. Item nonresponse is an important concern to the extent that participants who respond to individual survey items differ from those who skip items.

Research examining the issue of item nonresponse has focused almost exclusively on closed-ended items (i.e., Craig & McCann, 1978; Marcus & Shutz, 2005). There is an extensive literature which explores appropriate methods for handling missing data, but this research focuses on items with quantitative response options, not on qualitative, open-ended responses. It is clear that many of the strategies, such as imputation, used to handle missing quantitative data cannot be applied to open-ended responses. Therefore, other methods for evaluating and handling open-ended item nonresponse are needed.

Although most item nonresponse research has focused on closed-ended items, there have been some recent studies aimed at understanding nonresponse to open-ended items. Research has shown that most survey participants do not provide comments when presented with open-ended items on organizational surveys. Commenting rates reported in the literature range from 34% to 40% (Borg, 2005; Poncheri et al., 2008; Siem, 2005). Although these percentages are low, they do not necessarily indicate the presence of bias in the comments. As Rogelberg and Stanton (2007) demonstrated at the level of unit nonresponse, bias would be present to the extent that noncommenters differ from commenters in ways that are relevant to the study. Therefore, the first goal of this study is to examine characteristics of
commenters and noncommenters to see if any important differences emerge that could be indicative of item nonresponse bias.

*Commenting Behavior: Who Comments?*

In order to understand what distinguishes commenters from noncommenters, it is important to explore the criterion of commenting behavior. Commenting behavior has not been examined extensively in the literature, but as suggested previously the broader issue of survey response behavior has received considerable research attention (e.g., Rogelberg et al., 2006; Rogelberg, et al., 2000). 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). OCB was originally defined by Organ (1988) as “behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective functioning of the organization” (p. 4). Rogelberg et al. (2000) argue that voluntary survey response behavior can be conceptualized as OCB to the extent that participation in the survey contributes to effective organizational functioning.

By extension, it can be argued that commenting behavior in response to organizational surveys is also appropriately conceptualized as a form of OCB or contextual performance. This is particularly relevant in cases in which participation in the organizational survey is formally mandated. For example, surveys administered before or after an organizational training event are sometimes required of all participants. In this case, the survey response behavior is not voluntary, but the provision of open-ended comments may be
perceived and treated as optional. In fact, recent research has shown that open-ended questions are viewed as more optional when compared to closed-ended questions on a web-based survey (Poncheri & Thompson, 2007). Since unit nonresponse is unlikely in the case of a required survey, nonresponse that would have typically been observed as failure to return a survey may occur at the item-level instead. Therefore, research that has been conducted on unit nonresponse provides important insights when predicting commenting behavior in the context of a required survey.

The unit nonresponse and OCB literatures allow for hypotheses to be developed in relation to commenting behavior despite the limited amount of research in this area. In addition, the few studies that have been conducted to explore nonresponse to open-ended items provide an important foundation for some of the hypotheses proposed in this study. In this section, hypotheses are only proposed for relationships with strong theoretical and previous empirical support. Hypothesized predictors of commenting behavior discussed next include education, interest in the topic, conscientiousness, and satisfaction. The four hypotheses presented on the following pages examine the same criterion (i.e., commenting behavior) and address the broad question of “who” comments.

**Education.** The impact of education on response behavior has been widely supported in previous research related to unit nonresponse. In a review of the literature on unit nonresponse, Rogelberg and Luong (1998) cite several studies which found that individuals with relatively low education levels were more likely to be nonrespondents. Rosenthal and Rosnow (1975) reviewed 26 studies which explored the relationship between education and survey response or intentions to participate. According to their review, significant positive
relationships were found in all but two studies. This led the researchers to conclude that “a summary of the relationship between volunteering and education can be unusually unequivocal. Better-educated people are more likely to participate in behavioral (usually survey) research…” (Rosenthal & Rosnow, 1975, p. 70). Although the relationship between education and unit nonresponse is well-supported, not all studies have found significant results. For example, Green (1991) failed to find an effect for education. Green (1996) attributed this to range restriction on the lower end of the scale because all participants in the study had a bachelor’s degree or higher.

The findings are less clear in the area of item nonresponse. With regard to closed-ended item nonresponse, Craig and McCann (1978) found that individuals with relatively less education skipped more items when responding to mail surveys on a variety of topics. More recently, two studies (Clayton, Rogers, & Stuifbergen, 1999; McNeely 1990) have examined item nonresponse to open-ended questions. Both failed to find a significant relationship between education and commenting behavior. However, in both studies the education level for commenters was higher than the education level of noncommenters, suggesting a trend and the need for further exploration.

In addition to empirical evidence, the relationship between education and response is also supported by theoretical rationale that extends to commenting behavior. Green (1996) identifies several factors that may explain why higher levels of education are associated with higher levels of response behavior: “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 in the context of commenting behavior. Therefore, the following hypothesis is proposed:

H1a: Higher levels of education will be positively related to commenting behavior, such that more educated individuals will be more likely to provide comments.

Interest in the topic. Rogelberg and Luong (1998) identify interest in the topic as another predictor of unit nonresponse. They indicate that interest in the topic is positively related to survey participation. Martin (1994) provides a review of previous published research which explored the impact of interest in the topic on response behavior. He identifies three broad categories which describe previous conceptualizations of topic interest: 1) surrogate measures or single-item self-reports of interest, 2) researchers’ judgment about whether respondents were interested, and 3) speculation or inference that interest in the topic accounts for observed differences. Martin (1994) criticizes previous research for lacking “methodological rigour” (p. 329) and a closer review of these studies reveals a plethora of ways in which topic interest has been labeled and operationalized [e.g., signature indicating willingness to participate in future research related to the topic (Pace, 1959); single-item measuring strength of opinion (Pearl & Fairely, 2001); familiarity with the topic, personal experience with the topic, and importance of the topic (Suchman & McCandless, 1940)].

In order to address limitations associated with previous research in this area, Martin (1994) employed an experimental design to investigate the impact of topic interest on survey response. In this study, members of an amateur bowling tournament were surveyed, such that participants were randomly selected to receive a bowling survey (higher-interest condition) or a restaurant survey (lower-interest condition). Martin (1994) found that those who
received the bowling survey were twice as likely to respond than those who received the restaurant version of the survey. Martin (1994) also found evidence that those who were interested in the topic were less likely to omit individual items. More recently, Groves, Presser, and Dipko (2004) used a similar approach to examine the impact of topic interest on 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. In terms of commenting behavior, it is reasonable to suggest that those who are interested in the topic will be more likely to provide comments. Arguably, those who are interested in the topic have thought more about the topic, and therefore, have more to say when asked to provide their opinions when compared with those who are indifferent. The following hypothesis will be tested:

H1b: Interest in the topic will be positively related to commenting behavior, such that individuals who are more interested in the topic will be more likely to provide comments.

Conscientiousness. There has been a great deal of research that has focused on the link between personality factors and OCB. It is widely accepted that personality factors are more strongly related to OCB than to task performance (e.g., Motowidlo, Borman, & Schmit, 1997). When examining personality factors according to the Big Five framework (i.e., conscientiousness, extraversion, openness to experience, emotional stability, and agreeableness), 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., 2000; Rogelberg et al., 2003) have found that both active and passive nonrespondents are less conscientious than
survey respondents (Rogelberg et al., 2003). Rogelberg et al. 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, unlike active nonrespondents, and they are less likely to forget or misplace a survey, unlike passive nonrespondents. The following hypothesis will be examined to see if these findings extend to commenting behavior:

H1c: Conscientiousness will be positively related to commenting behavior, such that more conscientious individuals will be more likely to provide comments.

Satisfaction. Studies that have examined individual differences with respect to commenting behavior have found that satisfaction plays an important role. Clayton et al. (1999) conducted a study to explore the characteristics of individuals who provided unsolicited narrative comments on a survey related to living with multiple sclerosis (MS). The researchers found that 25% of respondents provided unsolicited narrative comments in the margins of the questionnaire that was administered. The researchers investigated several demographic, illness-related, and attitudinal constructs, but found few significant differences between commenters and noncommenters. They found that individuals who perceived their economic resources as less adequate (i.e., less satisfied with economic resources) and those who lived in non-metropolitan counties were more likely to provide comments.

McNeely (1990) examined demographic and satisfaction differences between commenters and noncommenters on a job satisfaction survey. Unlike the Clayton et al. (1999) study, these comments were specifically solicited by the provision of one open-ended question. The main finding from McNeely’s study was that commenters were less satisfied
with their jobs than noncommenters. The researchers also examined several demographic characteristics (i.e., age, race, gender, education, tenure in job, and marital status). Only two of the demographic variables, age and race, were significantly related such that younger workers were more likely to provide comments than older workers and Caucasians were more likely to provide comments than minorities. The researchers argued that these significant demographic findings could be explained by the findings related to satisfaction because younger workers and Caucasian workers tend to be less satisfied than older and minority employees, respectively.

Finally, Poncheri et al. (2008) found that in response to an organizational climate survey in which open-ended comments were specifically solicited, dissatisfied employees were more likely to provide comments 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. 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” (p. 323-324). Dissatisfied individuals are, therefore, more likely to deeply process information related to the source of their dissatisfaction, while satisfied individuals are less likely to do so. As such, satisfied individuals should have less to communicate when asked to provide comments than their dissatisfied counterparts. Based on the findings from these studies and theoretical support for the relationship, it is hypothesized that:
H1d: Satisfaction will be negatively related to commenting behavior, such that more satisfied individuals will be less likely to provide comments.

The hypotheses proposed in this section will provide answers to the question of who provides comments. The first three hypotheses have not been previously explored and if supported will provide evidence for the connection between the unit nonresponse literature and nonresponse to open-ended items. The fourth hypothesis has been explored in the open-ended comments literature and if supported will provide evidence that the finding generalizes to a different type of survey – training evaluation.

The preceding predictions focus somewhat generally on the inclination to provide or withhold any type of comment on a training evaluation survey. Simply knowing whether or not someone provided a comment is an important first step. At the same time, it is necessary to recognize that all comments are not created equal. Improving our understanding of commenting behavior will therefore benefit from a finer-grained analysis, one in which survey comments are classified and examined by type. The next section presents an overview of methods for categorizing/coding comments and describes a framework to consider when investigating the second broad question driving this research (i.e., “How” do respondents comment?). Drawing from this framework, hypothesized antecedents of different types of comments are then considered and proposed.

*Open-Ended Comments: A Classification Framework*

There are many ways that comments can be coded or classified. The most popular approach for analyzing open-ended comments is content analysis. Church and Waclawski (1998) define content analysis of comments as “taking the individual idiosyncratic comments
and opinions and converting them into a number of main categories or themes for interpretation purposes” (p. 133). The information obtained from content analysis provides a summary of the key points made in response to open-ended comments and as such, can be very useful to recipients of survey results. However, it is important to note that content analysis is dependent on the type of survey and the type of questions that are asked. Content codes developed to analyze written responses in one survey context may not be appropriate for completely different survey contexts. In addition, content codes developed to analyze responses to a particular question on a survey may not be applicable to other questions on the same survey.

While the content of a comment may be dependent on survey or question type, other characteristics of comments are not dependent on survey or question type. This means that regardless of the survey type or the question wording, these characteristics can be coded and studied. There are two main categories of comment characteristics that have been explored across many types of surveys and questions: tone and length.

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), general employee opinion surveys (Borg, 2005), 360-degree surveys (Smither & Walker, 2004), and public opinion polls (Garcia, Evans, & Reshaw, 2004). Comment tone is most often rated on a continuum of negative-neutral-positive. In most of the studies just cited, comments tended to be negative in tone. Borg found that most comments provided in response to a general employee opinion survey were negative. Similarly, Poncheri et al. found that across seven survey dimensions on a climate survey, 56% of the comments were
negative in tone. Finally, Garcia et al. conducted two public opinion survey studies and found that in both cases, the majority of comments provided were negative. The only study to find a different pattern of results was conducted by Smither and Walker in the context of multirater (or 360-degree) feedback. In their study, the majority of comments were positive in tone.

Comment length is another characteristic that has been explored in the survey literature. Unlike comment tone, comment length can be evaluated in a much more objective manner by calculating the number of words in a response (i.e., word count). There are two main findings that have emerged in studies exploring comment length. The first is that comments provided in response to web surveys tend to be longer than comments provided in response to paper-and-pencil surveys (Fenlason & Suckow-Zimberg, 2006). The other main finding is that comment length tends to increase as comments become more negative in tone (Borg, 2005; Poncheri et al., 2008). This research helps us to gain a better understanding of the circumstances under which we can expect to find lengthier comments, but does not provide an indication of whether or under what conditions lengthier comments are desired.

Besides tone and length, there are two additional categories of comment characteristics that have not been explored empirically, but are theoretically relevant. These are borrowed from Kulesa and Bishop’s (2006) categorization of survey questions: scope (general/specific) and purpose (prescriptive/descriptive). With regard to scope, a general comment is broad in nature and does not provide details about precise issues, whereas a specific comment provides information targeted at particular issues. It should be noted that Garcia et al. (2004) also indicate the importance of scope, although they do not use this term per se. Instead, they describe comments as shallow or deep, which arguably correspond to
comments that are general or specific, respectively. With regard to purpose, a prescriptive comment provides suggestions for change or improvement, whereas a descriptive comment provides a narrative of the current state.

In this study, the comment characteristics to be investigated include comment tone, comment scope, and comment purpose. Although it has been examined in past research, comment length is excluded from this study intentionally. It is unclear how to interpret the value of a lengthy comment. Perhaps comment length is a proxy for comment qualities, such as scope. That is, relatively lengthy comments may provide more details than short comments, indicating a specific scope. On the other hand, lengthy comments could simply be repetitive. There is little value in assessing features of comments that are proxies for substantive characteristics that can be measured. If length is indeed a proxy for scope, then it is more valuable to assess comment scope than to assess comment length. Therefore, this study will not contain hypotheses related to comment length, and instead will focus on comment tone, scope, and purpose.

*How Participants Comment: Who Makes What Type of Comments?*

Whereas the hypotheses proposed in the first section focused on the difference between commenters and noncommenters, the remaining hypotheses focus exclusively on commenters and investigate the characteristics which predict “how” respondents comment. This section explores the individual differences that impact the tendency to provide comments that vary in terms of tone, scope, and purpose. Since there is very little previous empirical research to support the hypotheses proposed in this section, hypotheses are proposed where the strongest theory-based arguments can be made.
Comment tone. No research has been conducted to explore individual characteristics that impact the tone of responses to open-ended questions. However, there are three characteristics that are theoretically related to tone. The first characteristic is satisfaction. Research supports the negative relationship between satisfaction and the tendency to provide comments (Clayton et al., 1999; McNeely, 1990; Poncheri et al., 2008). Furthermore, there is evidence that most open-ended comments tend to be negative in tone (Borg, 2005; Garcia et al., 2004; Poncheri et al., 2008). Presumably, most comments are negative because most commenters are dissatisfied, suggesting that people use the survey comment block to express their satisfaction level in words. Based on this rationale, the following hypothesis is proposed:

H2a: Satisfaction will be positively related to the tendency to provide comments that are positive in tone.

The second characteristic that is theoretically related to comment tone is the personality characteristic of agreeableness. Agreeableness is associated with cooperation, trust, compliance, and sympathy (Barrick et al., 2001; George & Jones, 2003). George and Jones note that “agreeable individuals generally are easy to get along with and are ‘team players’” (p. 9). Agreeable individuals will most likely provide comments that are positive in tone because of their tendency to support the status quo (LePine & Van Dyne, 2001). Therefore, the following hypothesis is proposed:

H2b: Agreeableness will be positively related to the tendency to provide comments that are positive in tone.
The final characteristic that is theoretically related to comment tone is emotional stability. Individuals with low levels of emotional stability have a tendency to experience negative emotions and are critical and pessimistic when compared with individuals who have high levels of emotional stability (Barrick et al., 2001; George & Jones, 2003). The tendency to be pessimistic and critical should be reflected in the content of the comments individuals provide. The following hypothesis is proposed:

H2c: Emotional stability will be positively related to the tendency to provide comments that are positive in tone.

Comment scope. There are several characteristics that may impact the tendency to provide comments that differ in scope or specificity. Specific information is arguably more desirable in the context of open-ended comments because specific comments provide detailed information to recipients of survey results. This information can then be used to determine if action is needed, and if so, what type of action may be needed. Despite a lack of research in this area, there are two potential predictors with relevance: satisfaction and conscientiousness.

In all likelihood, individuals who are relatively satisfied will not be motivated to provide a lot of details related to why they are satisfied. On the other hand, individuals who are dissatisfied will most likely provide detailed information about the cause of their dissatisfaction. Poncheri et al. (2008) found that comment length increases as comments become more negative in tone and indicate that this may be the result of individuals attempting to provide detailed information about their dissatisfaction. Taylor (1991) notes that “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 may lead to higher levels of specificity in comments provided by dissatisfied individuals. As such, it can be argued that relatively dissatisfied individuals are especially inclined to provide specific comments. The following hypothesis is proposed:

H3a: Satisfaction will be negatively related to the tendency to provide specific comments.

Conscientiousness also has a theoretically relevant relationship with comment specificity or scope. Conscientious individuals are organized and detail-oriented. This attention to detail may lead highly conscientious individuals to provide very specific information when making a point either verbally or in writing. Therefore, conscientious individuals should provide comments that are specific in nature. The following hypothesis is proposed:

H3b: Conscientiousness will be positively related to the tendency to provide specific comments.

*Comment purpose.* Comment purpose refers to whether the comment is prescriptive (providing a suggestions or recommendations for change) or descriptive (statements related to the way things currently are). Although no research has explicitly 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 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. The researchers conducted a laboratory study in which they asked teams of four individuals to participate in a computer-mediated decision-making task. Team members communicated exclusively via text messages and all communication was recorded. The transcripts were later coded for various characteristics, including voice behavior which was operationalized as communication that was both change-oriented and constructive. The authors then correlated their measure of voice behavior with several individual differences, including the Big Five personality factors.

LePine and Van Dyne (2001) hypothesized that conscientious individuals would be likely to invest effort in providing suggestions for change and “are more willing to engage in conversations about ideas intended to improve the situation” (p. 328). In their study, they found a positive relationship between conscientiousness and voice behavior ($r=0.26$, $p<.05$). Therefore, in the survey context it is expected that:

H4a: Conscientiousness will be positively related to the tendency to provide prescriptive comments.

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 \((r=0.30, p<.05)\). Therefore, it is expected that:

**H4b:** Extraversion will be positively related to the tendency to provide prescriptive comments.

None of the hypotheses proposed in this section have been explored previously. If supported, these findings will provide valuable information related to how different types of individuals comment. As a whole, this study offers insight into who provides what types of comments in the context of a post-training evaluation survey.

**Method**

*Study Background and Training Context*

The data used in this study were collected as part of a large-scale training effectiveness project related to foreign language training in the military. Participants in this study were members of a large military organization who were participating in foreign language training. Foreign language courses ranged in length depending on the difficulty of the language taught. Languages that are considered to be easier for a native English speaker to learn are classified as Category (CAT) I/II languages (e.g., French and Spanish; U.S. Department of the Army, 1996). Languages that are considered to be more difficult for a native English speaker to learn are classified as CAT III/IV languages (e.g., Arabic and Japanese; U.S. Department of the Army, 1996). CAT I/II classes lasted for 18 weeks, while CAT III/IV classes lasted for 24 weeks. Participants spent approximately six hours per day in training during this time. As part of participation in training, participants were asked to complete a pre-training questionnaire and a post-training questionnaire for the purposes of
training evaluation. Participants were asked to provide responses to both closed-ended and open-ended items using paper-and-pencil scannable questionnaires. Responses to open-ended items were typed prior to analysis. This questionnaire was not anonymous (i.e., identifying information was collected).

Participants

Three cohorts (i.e., individuals who started and completed training at the same time) of military personnel were included in this study (N=621). Although the participants’ age and gender were not recorded, the population of trainees from which these cohorts were drawn is known to be primarily male, with most trainees in their 20’s and 30’s. These cohorts were selected for inclusion in this study because all three had the opportunity to complete the same version of both the pre- and post-training questionnaires. Of these, 182 personnel completed only the pre-training questionnaire, 20 completed only the post-training questionnaire, and 419 completed both the pre- and post-training questionnaires. In this population and for this training context it is standard to have fewer personnel complete the post-training questionnaire than the pre-training questionnaire for a variety of logistical reasons. In all of the analyses performed in this study, the matched sample of 419 personnel was used.

In order to ensure that these 419 individuals were not unusual in any study-relevant ways, mean comparisons (i.e., a series of t-tests for continuous variables) and chi-square tests of independence for the education variable were conducted between the study participants and the excluded cases. Results revealed no significant differences between included and excluded cases for the variables examined in this study.
For the analyses conducted to test study hypotheses, the final sample of 419 ensured there was adequate (i.e., .8) statistical power to detect a .14 (i.e., small to medium; Cohen, 1992) or larger effect.

Measures

Education. Education was measured with a single item on the post-training questionnaire: “Please indicate the highest level of education that you have attained.” Respondents chose from the following response options: high school; some college; B.A. or B.S. degree; M.A. or M.S. degree; and Ph.D. or Ed.D. degree. Some participants did not provide education data on the post-training questionnaire. For these participants, response to the same item collected on the pre-training questionnaire was used. Only a small number of respondents indicated that they had a Master’s degree (N=7) or a Doctoral degree (N=2). Therefore, these two categories were collapsed with the third category (B.A. or B.S. Degree) and relabeled “Four-year degree or higher.”

Interest in the topic. Because this study focused on responses to a questionnaire administered to evaluate foreign language training, interest in the topic was operationalized as motivation to participate in foreign language training. This operationalization of interest in the topic is novel, but shares features of operationalizations used in previous research. For example, this measure can be categorized as a self-report, surrogate measure, which is one of the categories identified by Martin (1994). Motivation to participate in foreign language training was measured with five items on the post-training questionnaire. These items were

---

1 An “other” option was also included. Respondents who indicated “other” provided details about their highest level of education which were used to recode the response into one of the existing categories. Those who indicated “Associate’s degree” were recoded into the “some college” category.
developed for use in this study. Participants responded on a 7-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). An example item is: “I would like to learn as many languages as possible.” See Appendix A for these items.

**Personality.** Goldberg’s (1999) 50-item measure of personality (10 items for each of the Big Five personality factors), the International Personality Item Pool (IPIP), was used in this study. Participants responded to these items on the pre-training questionnaire using a 5-point Likert-type scale ranging from 1 (very inaccurate) to 5 (very accurate). See Appendix B for these items.

**Satisfaction.** Satisfaction was measured with 18-items developed for use in this study. These items were developed based on research related to training reactions (e.g., Morgan & Casper, 2000). Participants responded to these items on the post-training questionnaire using a 7-point Likert-type scale ranging from 1 (extremely dissatisfied) to 7 (extremely satisfied). See Appendix C for these items.

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

**Coding Comments**

**Commenting behavior.** Commenting behavior was operationalized by coding whether a comment was made (1) or not made (0) in response to the open-ended item. Comments which were not substantive (i.e., “none,” “no comment,” “not applicable”) were coded as zero. In this study, there were two cases where comments were coded as zero because they
were not substantive. Due to the objective nature of this coding decision, one rater coded for commenting behavior.

*Discrete comments.* Before coding for comment tone, scope, and purpose, it was necessary to define and identify discrete comments to aid subsequent coding. When coding a particularly lengthy comment, it may become difficult for raters to keep track of variations in tone, scope, and purpose. For example, it is possible that a respondent may use a single open-ended item to provide multiple comments, related to different topics that vary substantially in regards to tone. Although theoretically, there should not be a difference between a rater’s judgment of overall comment tone, scope, and purpose and the average tone, scope, and purpose of multiple discrete comments within a given comment block, there is a methodological difference in how this would be determined. An overall comment rating entails a single judgment which may be subject to rater error due to the complexity of rating a multi-faceted comment. By breaking multi-faceted comments into discrete components, raters should be able to more accurately evaluate the distinct elements of the comment. These ratings can then be averaged together mathematically rather than forcing raters to take a “mental average” of the overall comment characteristic.

Smither and Walker (2004) note that “the first step in coding written comments is determining the unit of analysis to be coded (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, while other topics may span several sentences (Smither & Walker, 2004). A decision was made to keep sentences complete when coding
for discrete comments. In other words, parts of sentences were not separated from one another.

Prior to discrete comment coding, all raters were trained. Three raters (the lead researcher and two additional raters – both Industrial/Organizational (I/O) psychology doctoral students) discussed the definition of a discrete comment and identified the unit of analysis. This step was followed by a review of examples for classifying comments as discrete and coding practice. Two raters coded all comments. When coding for discrete comments, both raters read all of the responses to the open-ended questions (organized in Microsoft Excel so that one column represented response to the open-ended question and each row represented a participant) in order to determine the number of comments provided by each participant. As part of this process, both raters independently recorded the number of discrete comments in one cell adjacent to the comment and marked the discrete comments in the original cell (using brackets).

After the two raters completed coding for discrete comments, interrater reliability and agreement were assessed. In the event of a disagreement, a third rater (the lead researcher) reviewed the coding decisions and made a final determination of the total number of discrete comments provided by each participant. The third rater then separated the discrete comments by copying and pasting them into separate cells in the Microsoft Excel file. Each discrete comment was listed on its own row along with an associated participant identifier number so that comment codes could be aggregated after coding was complete. See Appendix D for more details related to rater training and the process of identifying discrete comments; this document was provided to raters during the training session.
Comment tone, scope, and purpose. Three independent raters (the lead researcher and two I/O psychology doctoral students who were different from the raters used for discrete comment coding) read each discrete comment and evaluated comment tone, scope, and purpose. Raters were asked to evaluate comment tone with the following prompt: What is the tone of this comment? Raters were asked to provide codes for comment tone based on the following scale: 1 = very negative, 2 = somewhat negative, 3 = neither negative nor positive, 4 = somewhat positive, 5 = very positive. In order to evaluate comment specificity/scope, raters were prompted by the following question: How specific is this comment? They were asked to respond according to the following scale: 1 = not at all specific, 2 = somewhat specific, 3 = very specific. Finally, raters evaluated comment purpose and were provided with the following prompt: How prescriptive is this comment (i.e., to what extent does this comment provide recommendations or suggestions)? Raters responded according to the following scale: 1 = not at all prescriptive, 2 = somewhat prescriptive, 3 = very prescriptive.

The use of Likert-type scales to rate open-ended comments has been used in previous research (Smither & Walker, 2004). Comment tone was rated on a 5-point Likert-type scale because two opposite ends of a continuum (i.e., positive and negative) were being evaluated. Comment scope and purpose were rated on a 3-point Likert-type scale because for these characteristics, the evaluation focuses on the magnitude of a unidimensional comment characteristic. While comment tone measures a comment’s characteristic (+) and its opposite (-), scope and prescription arguably require fewer scale points because they do not require the assessment of both a comment’s characteristic and its opposite.
The lead researcher developed materials for training coders by coding a sample of open-ended responses, which included 302 discrete comments provided by individuals not included in this study. After coding this sample, the lead researcher developed definitions for each of the scale points on the tone, scope, and purpose scales (See Appendix E). The lead researcher then selected six examples for each scale point from this training file to provide to coders. Prior to coding for tone, scope, and purpose, raters were trained on coding protocol; raters were provided with this document during the training session and referred to it during coding (see Appendices E and F).

As part of training, the lead researcher reviewed the purpose of the project as well as the definitions, rating scales, and examples of comment tone, scope, and purpose. Using the sample of 302 discrete comments, all three coders rated a sample of 20 comments for tone, scope, and purpose. The three coders then discussed their assigned ratings to establish a shared mental model. All three coders then rated a sample of 100 comments for practice before beginning the actual coding. Preliminary interrater reliability and agreement statistics were assessed after this stage of practice coding and were deemed acceptable.

Ratings were provided in Microsoft Excel and were entered in three cells adjacent to the discrete comments for comment tone, scope, and purpose, respectively. Raters made all three ratings for each comment before moving to the next comment. After coding was completed, interrater reliability and agreement were assessed (see the Results section for additional details) and were considered acceptable. The discrete comment tone, scope, and

---

2 These comments were not used for analysis in this study because these participants did not have an opportunity to take the pre-training questionnaire. Data collection for this study began after this cohort of students started training.
purpose ratings were first averaged across raters so that each discrete comment had a single averaged rating for tone, scope, and purpose. Then, the discrete comment ratings were averaged together to represent an overall rating of tone, scope, and purpose for each participant/trainee.

Results

Preliminary Analyses

Interrater agreement and reliability. There is an important distinction made in the literature between interrater reliability and interrater agreement (James, Demaree, & Wolf, 1984, 1993; Kozlowski & Hattrup, 1992). Interrater reliability assesses the consistency or correlation between raters, while interrater agreement assesses the degree to which raters assign the same ratings. In this study, the primary concern is achieving interrater reliability. However, it is possible for interrater reliability to be high even when interrater agreement is low. It is valuable to assess both interrater reliability and agreement to determine both the consistency and absolute agreement between raters.

Interrater reliability for discrete comment coding (i.e., the number of discrete items in the comment block) and coding related to comment tone, scope, and purpose was calculated using intraclass correlations (ICC; Shrout & Fleiss, 1979). The intraclass correlation is a method for determining the amount of between-rater variability when there are two or more raters of interest (von Eye & Mun, 2005). There are six forms of the ICC described by Shrout and Fleiss (1979) and the choice of which form to use depends on assumptions about the raters and the unit of analysis. In this study the model used was the two-way mixed ICC. The equation for this form of the ICC is: \( \text{ICC}(3, k) = \frac{(M_{\text{between}} - M_{\text{within}})}{M_{\text{between}}} \).
ICCs were interpreted according to the following commonly accepted levels: excellent (>0.75), good (0.60–0.75), moderate (0.40–0.59), and poor (<0.40; Fleiss, Levin, & Paik, 2003). ICCs for both discrete comment coding and coding for comment tone, scope, and purpose are presented in Table 1. All ICC estimates exceeded 0.75, and therefore are considered to be excellent.

Interrater agreement for discrete comment coding was assessed by calculating the percentage of absolute agreement between the two raters as well as the percentage of agreement within one level. Calculating these metrics entailed looking at the number of discrete components each rater identified per comment. Absolute agreement was determined by computing the percentage of comments for which the two raters identified exactly the same number of discrete components. Meanwhile, “agreement within one level” was determined by computing the percentage of comments where the absolute difference between the number of discrete comments identified by the two raters was less than or equal to one. Results showed that the raters agreed perfectly 68% of the time, and agreed within one level 94% of the time. Interrater agreement for comment tone, scope, and purpose was assessed using $r_{wg}$, an index developed by James et al. (1984). According to James et al., $r_{wg}$ values greater than or equal to 0.70 are considered acceptable. All median $r_{wg}$ estimates presented in Table 1 exceeded 0.70. Since acceptable levels of interrater reliability (ICC ≥ 0.60) and interrater agreement ($r_{wg} ≥ 0.70$) were obtained, mean ratings were calculated across raters for each discrete comment.

After the completion of coding and assessment of reliability and agreement, discrete comment ratings were aggregated across raters and then back to the comment level. It is
important to note that due to the use of Likert-type scales and the method used for averaging comment ratings, the middle point on the scale (e.g., values approaching “3” on the comment tone scale) has two different interpretations after averaging. In the context of comment tone, for example, a value around the midpoint may represent either a neutral attitude or multiple opposing attitudes balancing one another (e.g., two very negative and two very positive discrete comments). In both cases, the overall comment is considered to be neither “positive” nor “negative.”

**Exploratory factor analyses.** Whereas pre-established scales were used to measure some of the constructs examined in this study (e.g., conscientiousness, agreeableness), new scales were created to measure two constructs of interest – satisfaction and interest in the topic. Prior to hypothesis testing, the dimensionality of these new measures was examined. Specifically, two exploratory factor analyses (EFAs) were conducted to assess the factor structure of items used to measure interest in the topic and satisfaction. Confirmatory factor analysis (CFA) was also considered, but was determined to be inappropriate since the item development was not modeled on a specific theoretical framework or based on previous empirical work (Fabrigar, Wegener, MacCallum, & Strahan, 1999). As Fabrigar et al. (1999) indicate, EFA is more appropriate than CFA “in situations in which a researcher has relatively little theoretical or empirical basis to make strong assumptions about how many common factors exist or what specific measured variables these common variables are likely to influence” (p. 277). For both EFA analyses, maximum likelihood estimation was used with oblique (promax) rotation, which allowed the factors to correlate (Fabrigar et al., 1999).
The first EFA was used to assess the factor structure of the five items developed to measure interest in the topic, which in this context is represented by motivation to participate in foreign language training. In the training literature, measures of motivation to learn are often developed for use in specific studies (e.g., Baldwin, Magjuka, & Loher, 1991; Birdi, Allan, & Warr, 1997; Noe & Schmitt, 1986; Quinones, 1995) in an attempt to customize the items to the training context. Although it is common practice to develop items to measure this construct, it is important to assess the extent to which the items measure a single underlying latent construct.

Results of this analysis showed support for a one-factor model (see Table 2). In order to determine the number of factors to retain, the following criteria were used 1) Kaiser criterion (number of factors with eigenvalues that exceed 1.00; Kaiser, 1960), 2) percentage of variance explained, and 3) the scree test (Cattell, 1966). The first eigenvalue was the only one to exceed 1.00 and explained 50.5% of the variance. The scree plot also clearly showed a break between the first and second factor. In addition to these criteria, interpretability was also considered and a one-factor solution was conceptually supported.

A second EFA was used to assess the factor structure of the 18 items developed to measure satisfaction. Trainee reactions, such as satisfaction, are frequently measured as part of training evaluation (Bassi, Benson, & Cheney, 1996; Goldstein & Ford, 2002; Saari, Johnson, McLaughlin, & Zimmerle, 1988). Despite the frequency of their use, 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). Due to the limited
understanding of the dimensionality of satisfaction and because these items were developed for use in this study, EFA was used to determine the factor structure of this measure.

Results of this analysis showed support for a four-factor model of satisfaction (see Table 3). The eigenvalues of the first four factors exceeded 1.00 and explained 47.40%, 13.91%, 6.45%, and 6.83% of the variance, respectively. The scree plot showed a break between the fourth and fifth factors. The four factor solution is also supported from a conceptual perspective based on item content. Five items loaded most highly on the first factor which was labeled satisfaction with the instructor. Four items loaded most highly on the second factor which was labeled satisfaction with the utility of training. Six items loaded most highly on the third factor which was labeled satisfaction with course materials and learning environment. Finally, three items loaded most highly on the fourth factor which was labeled satisfaction with language training in the unit. One of the six items that loaded most highly on satisfaction with course materials and learning environment also had a comparable cross-loading on the factor labeled satisfaction with the instructor. This cross-loader can be explained by the content of the item which contains reference to both the course materials and the instructor’s use of the materials (“Indicate your level of satisfaction with the quality of audio and visual aids used by the instructor”). Therefore, this item was dropped from subsequent analyses (i.e., omitted when computing composites for satisfaction), resulting in a 5-item measure of satisfaction with course materials and learning environment.

Since a one-factor solution resulted from the first factor analysis examining interest in the topic (i.e., motivation to participate in foreign language training), the mean of those items was computed and used in subsequent analyses. Because support was found for four factors
in the case of satisfaction (i.e., satisfaction with the instructor, satisfaction with the utility of training, satisfaction with language training in the unit, and satisfaction with course materials and learning environment), the means of these factors were computed and each were treated as separate predictors in subsequent analyses.\(^3\)

*Checking for non-normal distributions.* Prior to testing the study hypotheses, data were checked for violations of the normality assumptions by examining the skewness and kurtosis of all relevant continuous variables. One variable (satisfaction with the instructor) showed an unacceptable (absolute value greater than 1.00) skewness value and therefore required a transformation to address the problem. Since this variable was negatively skewed, the variable was reflected prior to transformation. A square root transformation was performed on this variable and the skewness value was reexamined and deemed appropriate. The transformed variable was then unreflected to ease interpretation (Osborne, 2002). The unreflected transformed variable was used in all analyses that follow.

*Descriptive statistics and correlations.* Means, standard deviations, reliability estimates, and correlations for all study variables are presented in Table 4. Although not collected for the purposes of this study, openness to experience is included in the table as a point of interest for the reader. A total of 222 respondents (53% of the sample) provided comments in response to the open-ended question. The average comment tone rating was “somewhat negative” \((M=2.11, SD=.61)\), the average comment scope rating was “somewhat

\(^3\) It should be noted that both mean composites and factor scores were considered for aggregation. Mean composites were selected for aggregation to enhance generalizability. Incidentally, the mean composites and factor scores for all of the factors were almost perfectly correlated (i.e., the lowest correlation was .98).
specific” \((M=1.96, \text{SD}= .47)\), and the average comment purpose rating was between \(\text{“not at all prescriptive”} \) and \(\text{“somewhat prescriptive”} \) \((M=1.60, \text{SD}= .63)\).

**Hypothesis Testing**

Both bivariate and multiple regression analyses were considered as options for hypothesis testing in this research. Since this study’s predictions focus on the independent influence of the predictors on the criteria of interest, bivariate analyses (e.g., Pearson product-moment correlation coefficients) were deemed most appropriate for the purpose of hypothesis testing. Multiple regressions were subsequently conducted for exploratory purposes to examine the unique influence of each variable when considered alongside the other predictors of interest.

**Commenting behavior.** Table 5 shows the percentage of commenters within each of the three education levels examined. In order to test H1a, a chi-square test of independence was conducted. The omnibus test comparing all three levels of education with commenting behavior approached statistical significance \(\chi^2(2)=5.51, p= .06\). Follow-up 2x2 chi-square tests of independence were conducted. As expected, participants with some college education but no four-degree or higher were significantly more likely to comment than were those with no college experience \(\chi^2(1)=5.39, p= .02\). However, individuals with a four-year degree or higher were not significantly more likely to comment than those with some college experience \(\chi^2(1)=0.28, p= .60\). Additionally, individuals with a four-year degree or higher were not significantly more likely to comment than those with no college experience \(\chi^2(1)=2.64, p= .10\). Therefore, H1a was partially upheld but did not receive full support.
Point-biserial correlation coefficients were used to assess H1b-H1d (See Table 4). There were three significant correlations between the predictor variables and commenting behavior. Interest in the topic was positively related to commenting behavior (providing support for H1b), while satisfaction with the utility of training and satisfaction with language training in the unit were negatively related to commenting behavior (providing support for H1d). Conscientiousness was not significantly related to commenting behavior and therefore H1c was not supported.

Exploratory multiple regression analyses were conducted to examine the unique influence of each of these predictor variables in the presence of the others. Logistic regression is traditionally used to answer questions related to the prediction of group membership (in this case, whether or not the participant commented) and is the appropriate choice when some of the predictors of interest are discrete (i.e., education) and others are continuous (e.g., satisfaction). For this analysis, all four predictors (education, conscientiousness, interest in the topic, and satisfaction) were entered into the equation at the same time. This allowed evaluation of the relative (i.e., independent) contribution of each variable in predicting group membership (Tabachnick & Fidell, 2001).

The significant omnibus chi-square test \( \chi^2(8)=30.33, p<.001 \) resulting from the logistic regression described above showed that one level of education, interest in the topic, satisfaction with the instructor, and satisfaction with the utility of the training were significantly related to commenting behavior, while conscientiousness was not (see Table 6). In terms of education, the odds of providing a comment were 1.64 times greater for a person with some college education but no four-year degree or higher than a person with a high
school education only (\(B=.50, p=.05\)). This finding is consistent with the bivariate results related to education, which also showed significant differences between these two levels of education. Interest in the topic was positively related to commenting behavior (\(B=.32, p<.001\)), such that individuals who self-reported an additional unit (i.e., scale point) of interest in the topic were 1.37 times more likely to provide a comment than those who indicated less interest in the topic. The lack of significance for conscientiousness was consistent with the bivariate results. Satisfaction with the instructor was positively related to commenting behavior (\(B=.66, p=.01\)), while satisfaction with the utility of training was negatively related to commenting behavior (\(B=-.33, p<.001\)). Individuals who self-reported an additional unit of satisfaction with the instructor were 1.94 times more likely to comment than not comment, while individuals who self-reported an additional unit of satisfaction with the utility of training were .72 times less likely to comment than not.

Comment tone. Pearson product-moment correlations were used to address H2a-2c (see Table 4). For these hypotheses, the predictors of interest were satisfaction, agreeableness, and emotional stability and the criterion was comment tone. There were four significant correlations between the predictor variables and comment tone. All four dimensions of satisfaction were positively related to comment tone, providing support for H2a. H2b and H2c were not supported, as indicated by the nonsignificant correlation coefficients shown in Table 4.

An exploratory standard multiple regression analysis was conducted to determine the impact of these predictors when the others were considered (see Table 7). The overall regression was significant [\(R^2=.20, F(6,210)=8.90, p<.001\)]. In contrast with the bivariate
analysis, only one dimension of satisfaction was significant. Specifically, the only significant predictor in this regression equation was satisfaction with the utility of training, which was positively related to comment tone ($B=.25$, $p=.01$).

**Comment scope.** Pearson product-moment correlations were used to address H3a-3b (see Table 4). For these hypotheses, the predictors of interest were satisfaction and conscientiousness and the criterion was comment specificity. There were four significant correlations between the predictor variables and comment scope. Conscientiousness, openness to experience, emotional stability, and satisfaction with course materials and learning environment were positively related to comment scope. The satisfaction finding opposes H3a which predicted a negative relationship between satisfaction and comment scope. H3b, which predicted a positive relationship between conscientiousness and comment scope, was supported.

The exploratory multiple regression analysis revealed a significant overall regression model [$R^2=.08$, $F(5,211)=3.49$, $p=.01$; see Table 8]. Satisfaction with the course materials and learning environment ($B=.20$, $p=.03$) and conscientiousness ($B=.19$, $p=.01$) were both positively related to comment specificity; these relationships are consistent with the bivariate findings.

**Comment purpose.** Pearson product-moment correlations were used to address H4a-4b (see Table 4). For these hypotheses, the predictors of interest were conscientiousness and extraversion and the criterion was comment prescriptiveness. None of the hypothesized relationships between the predictors and comment purpose were supported. Unexpectedly, openness to experience was negatively related to comment purpose while satisfaction with
the course materials and learning environment was positively related to comment purpose. The exploratory regression analysis was not significant \( R^2 = .00, F(2,219) = .41, p = .66; \) See Table 9.

Discussion

In this study, approximately half of the respondents (53%) provided comments in response to a single open-ended question on a training evaluation survey. This response rate is higher than the response rates (ranging from 34-40%) that have been reported in previous commenting behavior research (Borg, 2005; Poncheri et al., 2008; Siem, 2005). This higher response rate may be a function of the population being studied (adult, military sample), the type of survey administered (training evaluation), or the administration context (i.e., required participation) in this study.

The first set of hypotheses in this study focused on the question of “who” comments and explored the individual differences (i.e., education, interest in the topic, conscientiousness, and satisfaction) expected to predict commenting behavior. Bivariate analyses showed support for a positive relationship between interest in the topic and commenting behavior. While this hypothesis has not been explored in the open-ended comments literature, this finding is consistent with past research and theory in the area of general survey nonresponse. Although the operationalization of interest used in the current study differed from the types of measures used in the survey nonresponse literature, the findings emerged as expected, supporting the alternative operationalization chosen for this research.
The relationship between satisfaction and commenting behavior has been explored in the open-ended comments literature, but has not been examined in the particular context of training evaluation. Consistent with previous research a negative relationship was found between two dimensions of satisfaction (i.e., satisfaction with the utility of training and satisfaction with language training in the unit) and commenting behavior. The replication of this finding in the context of training evaluation extends the “bad is stronger than good” finding beyond employee attitude surveys.

Although previous empirical work in the area of unit and item nonresponse has provided support for a positive relationship between education and response, the two studies which have explored this relationship in the commenting behavior literature (Clayton et al., 1999; McNeely, 1990) have failed to find significant results. In the present study, bivariate analyses indicated partial support for the relationship between education and commenting behavior. Although the omnibus chi-square test did not show a significant difference, the follow-up 2x2 chi-square tests indicated that individuals with some college education but no four-year degree or higher were more likely to comment than those without college experience. However, there was no difference between those with and without a four-year degree or higher, regardless of whether those without a degree had college experience. The non-significant difference between those with a four-year degree or higher and those with no college experience seems especially surprising. There is no straightforward explanation for this pattern of results; thus, the nature of education’s seemingly complex effect on commenting behavior is presently undetermined and not altogether understood.
The second, third, and fourth hypotheses focused on the question of “how” respondents comment and explored the individual differences that impacted the tendency to provide comments that varied in tone, scope, and purpose. None of these hypotheses had been previously explored and therefore all findings pertaining to the second, third, and fourth hypotheses are new to this area of research.

The average comment tone rating was “somewhat negative” which corroborates previous research demonstrating a negativity bias in comments (Poncheri et al., 2008). The bivariate analyses showed that all four dimensions of satisfaction were positively correlated with comment tone. The positive relationship between these variables indicates that more satisfied individuals are more likely to provide comments that are positive in tone, while more dissatisfied individuals are more likely to provide comments that are negative in tone.

The average comment scope rating was “somewhat specific.” As hypothesized, conscientious individuals were more likely to provide specific comments. The relationship between conscientiousness and comment specificity makes theoretical sense when considering the definition of conscientiousness as a tendency to be detail-oriented.

Contrary to the proposed negative relationship between satisfaction and comment specificity, individuals who were satisfied with the course materials and learning environment were more likely to provide specific comments. The original hypothesis was proposed based on the idea that negative events tend to lead to more cognitive processing than positive events (Taylor, 1991) and that those who are satisfied may not be motivated to provide a lot of details about the source of their satisfaction. However, the nature of this relationship may be dependent on the type of survey being administered. In this study, the
survey was an evaluation for a one-time training event. Dissatisfied respondents may not have been motivated to provide specific comments since they were finished with the course and therefore there was no personal benefit from doing so. On the contrary, those who were dissatisfied may have been frustrated and therefore unmotivated to provide specific details. Meanwhile, in another type of survey, such as a climate survey, perhaps dissatisfied individuals would be particularly motivated to offer specific details if they believe that the information they provide could prompt action that may benefit their work lives.

The average comment purpose rating was between “not at all prescriptive” and “somewhat prescriptive.” This is unsurprising because the type of question asked in this study can be classified as descriptive – no specific suggestions were directly solicited. None of the hypothesized relationships between the predictors and comment prescriptiveness were supported in the bivariate analyses.

Coding Implications

The type of qualitative coding employed in this study has not been widely used in previous literature. Smither and Walker’s (2004) study was used to develop the basic coding protocol, which included discrete comment coding followed by an evaluation of comment characteristics using Likert-type scales. There are several implications of using this coding scheme that merit discussion. For discrete comment coding, a decision was made to define the unit of analysis as topic areas/themes and a decision rule was employed to keep sentences complete (i.e., phrases were not separated from one another within a sentence). The decision to keep sentences intact was based on the notion that some parts of sentences may become uninterpretable without the complete sentence. However, there are instances where this does
artificially prevent the separation of distinct ideas. This decision rule makes sense from a content analysis perspective, but may lead to complications for coders when rating comment characteristics. For this study, one of the main goals of breaking comments into discrete components was to ease the burden of coding. In general, the discrete commenting portion of the coding did lead to easier coding of comment characteristics even though there may have been occasional instances of mixed comments as a result of the decision to keep sentences whole.

Another related issue is the interplay between mixed comments and the use of the middle categories on the Likert-type response scales (i.e., “neither negative nor positive” for comment tone, “somewhat specific” for comment scope, and “somewhat prescriptive” for comment purpose). When a discrete comment contained two opposing elements (e.g., “I liked training, but it was too long”), coders were instructed to do mental averaging to assign a rating. So, in the case of comment tone, a somewhat negative and somewhat positive element of the comment could have cancelled one another out, leading to a “neither negative nor positive” rating. Theoretically, the same ratings would have been assigned had these elements of the comment been separated. However, there is a potential for the decision rules related to discrete comment coding to complicate other coding that builds on it. For example, if a sentence remains intact and there is one somewhat negative and one very positive phrase, the tone rating assigned is a judgment call between “neither negative nor positive” and “somewhat positive.” The comment tone scale does not allow for raters to assign decimal values. A word of advice to other researchers working in this area is to consider the
importance of identifying truly discrete comments and balance the concern of rater burden with concerns for interpretability.

Limitations

Three noteworthy limitations of this research are related to measurement, internal, and external validity. From a measurement perspective, it is possible that mono method bias could account for some of the findings in this study, since all data were collected via questionnaires. This is particularly relevant in instances in which both the predictor and criterion were measured on the same survey instrument (e.g., interest in the topic was measured on the post-training questionnaire, which also included the comment block). However, this limitation does not apply to all of the relationships explored in this study. One important strength of this study design was that the data were not all collected at a single point in time, but were gathered at two different time points. The personality variables explored were collected months prior to the commenting opportunity and therefore mono method bias is likely less of a concern for findings involving these variables.

Another limitation is related to internal validity. The nature of the analyses in this study as correlational prevents unequivocal statements about the direction of the significant relationships. However, it is less intuitive that commenting behavior and characteristics would impact variables such as interest in the topic and satisfaction than the reverse. Furthermore, the predictors measured on the post-training questionnaire appeared before the comment block. It is, however, possible that unmeasured “third” variables may have accounted for covariation between the predictors and the criteria explored in this study.
The final limitation is related to generalizability of these findings to other populations, administration contexts, survey types, and question types. The participants in this study were predominantly male and members of the military. The norms for commenting behavior in this population may vary significantly from those found in non-military and more gender-balanced populations. There is a need to explore predictors of commenting behavior and comment characteristics in other populations.

The administration context of this questionnaire should also be considered. This study explored a required survey administered before and after a lengthy foreign language training program. This is unlike many other survey contexts in which responding is left to the discretion of the survey recipient. Moreover, other training surveys may be shorter in length and focused on a different topic area. At present, there are important unknowns regarding whether this study’s findings extend to alternative (e.g., shorter duration of the training program, shorter length of the questionnaire) administration contexts.

It was initially argued that both survey type and question type should be considered when investigating commenting behavior. Most of the research in this area has been conducted using organizational climate surveys or 360-degree feedback surveys, but even these studies have not explored many of the individual differences investigated in this research. Therefore, it is unclear whether this study’s answers to the questions of “who” comments and “how” respondents comment generalize to organizational climate surveys, 360-surveys, or surveys using different open-ended comment prompts. Perhaps, for example, surveys including open-ended questions that solicit specific suggestions do not encourage input from the same types of individuals who were especially inclined to respond to this
study’s general, descriptive question. In short, it is important to acknowledge that this study’s findings could be moderated by survey type and/or question type.

**Future Research**

There are several avenues for future research in this largely unexplored topic area of commenting behavior. The coding implications suggest the need for additional research using this coding protocol and also for additional theoretical development in the area of quantifying qualitative data. Content analysis is a more commonly used technique for analyzing qualitative data; however, the potential associated with training raters to quantify qualitative responses can increase the utility of this type of data and expand our knowledge base, not only in the survey literature but in any research areas where open-ended comments are used.

Overall, the limitations in this study suggest two broad areas for future research – replication and expansion. There is a strong need for replication in other populations using different types of surveys and open-ended questions. There is also a need to consider and explore other theoretically relevant individual difference and situational characteristics that predict the tendency to comment and the characteristics of comments provided.

Variations in how often and how open-ended comments are solicited are worthy of investigation. In this study, only one open-ended item was provided, and it was both general and descriptive. It would be informative to investigate commenting behavior by varying the number of comment prompts provided. For example, researchers should study patterns of commenting behavior across multiple questions on the same questionnaire and also longitudinally across different survey administrations. Researchers should also examine how altering the item wording in terms of scope and purpose impacts the quantity and nature of
comments provided. The type of question asked may have a large impact on who comments and what type of comments are provided. A question seeking a very specific suggestion for improvement to a training program may encourage comments from a different type of responder than asking a general question about attitudes toward the training. Understanding what types of questions are likely to elicit more valued responses may increase the utility of open-ended comments for organizational decision-makers.

Research in this area would also benefit from further exploration of personality, attitudinal, and situational variables that influence commenting behavior and characteristics of comments. The results of this study revealed significant findings which were not anticipated (e.g., positive relationship between emotional stability and comment specificity). Although there is no theoretical explanation for these relationships, a reexamination of these variables in future research can help to determine if the unanticipated relationships are spurious or robust.

Additionally, it should be noted that the current study included “state-like” predictors (i.e., interest in the topic, satisfaction) – that is, attitudes which may fluctuate over time. Also included were predictors that are more “trait-like” in that they are somewhat more stable over time (i.e., education, personality). Findings revealed that state-like variables were more influential than variables which are traditionally considered more trait-like. Future research should examine whether the state/trait distinction matters in this line of research and explore additional “state” predictors, such as mood, in an attempt to gain a great understanding of commenting behavior.
The stronger relationship between state predictors and the outcomes in this study also suggests that situational factors may play an important role in commenting behavior. Although this study focused on individual differences, it is unwise to ignore the impact of the situation. Future studies should explore situational features, such as time constraints and the presence of others who have shared a similar experience (i.e., members of the same training class filling out the evaluation together) on commenting behavior and comment characteristics.

Another area in need of additional attention lies at the intersection of training research and commenting research. As previously discussed, trainee satisfaction is frequently measured; however, there are many unanswered questions about the dimensionality of this construct (Alvarez et al., 2004; Bassi et al., 1996; Goldstein & Ford, 2002; Saari et al., 1988). It is important for researchers to take a fine-grained look at satisfaction, as was done in this study, because studies that assume a unitary construct may mask important nuances in the relationship between satisfaction and commenting behavior (or other criteria). The exploratory multivariate analyses conducted in this study suggest that the dimensionality of satisfaction may indeed be relevant when examining commenting behavior. The findings showed that individuals who were satisfied with their instructor were more likely to comment, while individuals who were dissatisfied with the utility of training were more likely to comment. Together, these findings suggest a potential difference in the way individuals react when evaluating a person (e.g., instructor) as opposed to an object or entity (e.g., training utility). Specifically, it is possible that respondents are more likely to provide negative comments when rating an object or entity than when evaluating a person.
To this end, the majority of past research has shown that most comments tend to be negative in tone (e.g., Borg, 2005; Garcia et al. 2004, Poncheri et al., 2008). However, a notable exception is a study conducted by Smither and Walker (2004) in the context of multirater (or 360-degree) feedback. In their study, the majority of comments were positive in tone. In addition, in Poncheri et al.’s (2008) study, Immediate Supervision was the only dimension on the climate survey where there was an even number of negative, neutral, and positive statements, unlike the other dimensions (e.g., Training and Development) where there was a higher occurrence of negative comments. Combined with the findings from the current study, this pattern of results suggests that perhaps questions in which an individual (particularly one close in relationship to the evaluator) is being assessed lead to a positive bias in response. Future research should test this proposition, exploring the effect of question target (i.e., please provide comments about a person v. please provide comments about a thing) on comment tone, specifically, but also on commenting behavior and other outcomes.

Practical Implications

There are many ways that the coding protocol and findings from this study can be put to practical use in organizations. The method for coding comment tone, scope, and purpose developed in this study can be used to help practitioners organize and present information to clients. Comments can be sorted into categories and associated with quantitative findings to bring those findings to life. For example, if the protocol developed for this study were used to code comments, practitioners responsible for presenting results could quickly identify comments that exemplify or illustrate closed-ended rating trends. This would help clients grasp the meaning and sentiment underlying the quantitative results. This would also allow
clients to request specific types of comments (i.e., very specific and very prescriptive) and avoid the burden of “weeding through” lists of verbatim comments. Although the coding can be somewhat time intensive, computer coding software may become sophisticated enough in the future so that many of these comment characteristics could be coded using automated systems.

The positive relationship between satisfaction and comment tone is an important replication of previous research findings (e.g., Poncheri et al., 2008). Results from this study showed that comment tone was consistent with quantitative ratings of satisfaction. This provides insights into how respondents view and use commenting opportunities. It implies that the comments participants offered were generally used to expound upon their level of satisfaction in words. That is, respondents typically used the comment block as an opportunity to expand and elaborate, not to point out exceptions to their quantitatively reported satisfaction. This finding is important to communicate to training evaluators so that they can appropriately interpret both the quantitative and the qualitative information provided by respondents.

The implications of understanding “who” comments and “how” respondents comment can lead to a better understanding of the information obtained from these questions. Researchers who study survey response behavior (e.g., Rogelberg et al., 2003) have warned against the dangers of nonresponse bias in the broader context of survey response. Item nonresponse bias, particularly in the case of open-ended items, is also an important concern for recipients of survey data. Managers and training designers should care about the difference between commenting responders and non-responders because there is evidence to
suggest that those who comment are different from those who do not. In this study, those who commented differed from those who did not in terms of interest in the topic, satisfaction, and education providing evidence of item nonresponse bias. Furthermore, open-ended comments have characteristics that may be particularly appealing to recipients of survey feedback. The vividness and language used in verbatim comments may overwhelm the sometimes bland and difficult to understand quantitative ratings. Managers and those evaluating training may inadvertently make decisions based on comments, while ignoring quantitative data as a result of this unbalanced focus. As the findings from this study show, if managers and training designers do pay more attention to the comments than the quantitative results, they may be making decisions based only on the highly interested and dissatisfied respondents. By doing this, these feedback recipients may develop a skewed perception of how their trainees really feel about the training experience.

This study also investigated what types of input commenters are likely to provide. Managers and training designers should be interested in this information because it can help gain an even deeper understanding of comment nonresponse bias. As previously discussed, comments tend to be more appealing than quantitative data. Additionally, certain types of comments are arguably more memorable than others. For example, highly negative comments may stand out to managers and training designers to the extent that strong and explicit language is used. Additionally, very specific comments may be more memorable since this type of feedback is highly desired, especially when it translates into easy to implement, actionable improvements to the training program.
As the results from this study show, certain types of commenters are more likely to make certain types of comments. Dissatisfied commenters are more likely to make comments that are negative in tone. Conscientious commenters are more likely to provide specific comments. Managers and training designers should be aware of the dangers of making decisions based on this subset of commenters. This group may only be a small subset of the population. Making changes based on this small group may be a poor use of time, money, and resources to benefit only a few; or worse, this practice may result in interventions that work against the needs of the majority.

Identifying and educating managers and training designers about the potential for item nonresponse bias in open-ended comments is a noteworthy challenge for practitioners. At a minimum, it is important for decision makers to be aware of the potential for comment nonresponse bias. This awareness may discourage an overreliance on qualitative results or at least lead to a more careful consideration of decisions made exclusively on the basis of comments. Practitioners can also consider implementing interventions (providing prompts or coaching to obtain more specific comments from respondents known or suspected to be low in conscientiousness, for example) to increase the likelihood of hearing from the voices that often go unheard.
References


Marcus, B., & Shutz, 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.


Siem (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, California.


Table 1

*Interrater Reliability and Agreement*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Interrater Reliability (ICC)</th>
<th>Interrater Agreement (Median $r_{wg}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete Comment Coding</td>
<td>221</td>
<td>.96</td>
<td>--</td>
</tr>
<tr>
<td>Comment Tone</td>
<td>641</td>
<td>.94</td>
<td>1.00</td>
</tr>
<tr>
<td>Comment Scope</td>
<td>641</td>
<td>.84</td>
<td>.83</td>
</tr>
<tr>
<td>Comment Purpose</td>
<td>641</td>
<td>.95</td>
<td>1.00</td>
</tr>
</tbody>
</table>

1 In this table, \(N\) represents the number of comments, not the number of participants.
2 One rater inadvertently skipped a comment when coding for discrete comments. The arbitrating rater agreed with the first rater.
3 Since it is inappropriate to use $r_{wg}$ to assess the interrater agreement for a count variable, the percentage of absolute agreement and agreement within one level were calculated. Absolute agreement was 68%, while agreement within one level was 94%.
Table 2

*Exploratory Factor Analysis: Interest in the Topic*

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would like to learn as many languages as possible.</td>
<td>.72</td>
</tr>
<tr>
<td>I don't particularly like the process of language learning and I do it only because I am required to participate in language training.</td>
<td>.80</td>
</tr>
<tr>
<td>I would rather spend the time dedicated to language training on training other skills.</td>
<td>.53</td>
</tr>
<tr>
<td>I am motivated to perform well in language training so that I will be able to perform my missions more effectively.</td>
<td>.53</td>
</tr>
<tr>
<td>If I learn to speak this language well, I will have many opportunities to use it.</td>
<td>.49</td>
</tr>
</tbody>
</table>

| Eigenvalues | 2.53 |
| % of Variance | 50.5 |

*These items were reverse-scored prior to conducting the exploratory factor analysis.*
Table 3

**Exploratory Factor Analysis: Satisfaction**

<table>
<thead>
<tr>
<th>Items: Indicate your level of satisfaction with…</th>
<th>Factor Loadings</th>
<th>Eigenvalues</th>
<th>% of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>The instructor's knowledge of course material and subject matter.</td>
<td>.90 - .12 .09</td>
<td>8.52</td>
<td>47.40</td>
</tr>
<tr>
<td>The instructor's ability to keep the interest of the class.</td>
<td>.93 .05 .02</td>
<td>2.50</td>
<td>13.91</td>
</tr>
<tr>
<td>The instructor's ability to teach the course content.</td>
<td>1.01 .04 .02</td>
<td>1.16</td>
<td>6.45</td>
</tr>
<tr>
<td>The instructor's responsiveness to student questions and problems.</td>
<td>.93 .03 .03 .07</td>
<td>1.23</td>
<td>6.83</td>
</tr>
<tr>
<td>The instructor's overall effectiveness.</td>
<td>.94 .02 .02</td>
<td>1.02</td>
<td>6.83</td>
</tr>
<tr>
<td>The match of course objectives with your idea of what would be taught.</td>
<td>.05 .87 - .11 .00</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The relevance of training you received for your specific job tasks or functions.</td>
<td>.00 .97 - .10 .03</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The extent to which the course prepared you to perform on your missions.</td>
<td>-.03 .88 -.01 .05</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The quality of this training course overall.</td>
<td>.17 .49 .28 -.02</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The quality of course materials.</td>
<td>-.14 .38 .45 -.02</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The quality of audio and visual aids used by the instructor.</td>
<td>.43 -.05 .45 - .04</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The computer resources available for this course.</td>
<td>-.16 -.09 .70 -.07</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The classroom's physical environment.</td>
<td>.07 -.15 .57 .11</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The length of the training course.</td>
<td>-.04 .12 .42 .13</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The pace of the course material presented.</td>
<td>.17 .16 .40 .10</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The availability of language training courses for individuals in your unit.</td>
<td>.00 -.05 .03 .89</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The communication of language training information to members of your unit.</td>
<td>-.03 -.01 -.05 1.02</td>
<td>47.40</td>
<td>66.44</td>
</tr>
<tr>
<td>The quality of language training in your unit.</td>
<td>.02 .11 .05 .76</td>
<td>47.40</td>
<td>66.44</td>
</tr>
</tbody>
</table>

Since this item had a substantial cross-loading on the factor labeled “Instructor,” the item was dropped and not used to compute the composite for satisfaction with course materials and learning environment.
### Table 4

**Descriptive Statistics and Correlation Matrix**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Education</td>
<td>419</td>
<td>2.02</td>
<td>.73</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Interest in the Topic</td>
<td>419</td>
<td>5.03</td>
<td>1.07</td>
<td>.13** (.75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Conscientiousness</td>
<td>419</td>
<td>3.97</td>
<td>.53</td>
<td>.05</td>
<td>.17** (.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Openness to Experience</td>
<td>419</td>
<td>3.82</td>
<td>.50</td>
<td>.17**</td>
<td>.29** (.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Emotional Stability</td>
<td>419</td>
<td>3.91</td>
<td>.60</td>
<td>.07</td>
<td>.18** (.97)</td>
<td>.46**</td>
<td>.14**</td>
<td>.30**</td>
<td>.30**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Agreeableness</td>
<td>419</td>
<td>3.68</td>
<td>.55</td>
<td>.10*</td>
<td>.31**</td>
<td>.21**</td>
<td>.39**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
<td>.37**</td>
</tr>
<tr>
<td>8. Sat with instructor</td>
<td>419</td>
<td>5.37</td>
<td>1.79</td>
<td>.04</td>
<td>.21**</td>
<td>.06</td>
<td>.01</td>
<td>.06</td>
<td>.13*</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>9. Sat with utility of training</td>
<td>419</td>
<td>4.10</td>
<td>1.56</td>
<td>.04</td>
<td>.33**</td>
<td>.10*</td>
<td>.07</td>
<td>.05</td>
<td>.05</td>
<td>.08</td>
<td>.08</td>
<td>.08</td>
<td>.08</td>
<td>.08</td>
<td>.08</td>
<td>.08</td>
<td>.08</td>
<td>.08</td>
</tr>
<tr>
<td>10. Sat with language training in the unit</td>
<td>414</td>
<td>4.35</td>
<td>1.19</td>
<td>.03</td>
<td>.23**</td>
<td>.09</td>
<td>.10*</td>
<td>.04</td>
<td>.09</td>
<td>.12*</td>
<td>.38**</td>
<td>.57**</td>
<td>(.93)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Sat with course materials and learning environment</td>
<td>419</td>
<td>4.07</td>
<td>1.22</td>
<td>.06</td>
<td>.24**</td>
<td>.09</td>
<td>.06</td>
<td>.02</td>
<td>.06</td>
<td>.08</td>
<td>.37**</td>
<td>.60**</td>
<td>.61**</td>
<td>(.71)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Comment behavior</td>
<td>419</td>
<td>--</td>
<td>--</td>
<td>.08</td>
<td>.10*</td>
<td>.01</td>
<td>.06</td>
<td>.03</td>
<td>.03</td>
<td>.05</td>
<td>.04</td>
<td>.13**</td>
<td>.11*</td>
<td>.07</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>13. Comment tone</td>
<td>222</td>
<td>2.12</td>
<td>.60</td>
<td>.10</td>
<td>.05</td>
<td>.00</td>
<td>.01</td>
<td>.13</td>
<td>.01</td>
<td>.07</td>
<td>.31**</td>
<td>.42**</td>
<td>.31**</td>
<td>.36**</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>14. Comment scope</td>
<td>222</td>
<td>1.96</td>
<td>.47</td>
<td>.09</td>
<td>.13</td>
<td>.20**</td>
<td>.02</td>
<td>.17*</td>
<td>.23**</td>
<td>.06</td>
<td>.10</td>
<td>.11</td>
<td>.09</td>
<td>.17*</td>
<td>--</td>
<td>.05</td>
<td>--</td>
<td>.05</td>
</tr>
<tr>
<td>15. Comment purpose</td>
<td>222</td>
<td>1.60</td>
<td>.63</td>
<td>.03</td>
<td>.01</td>
<td>.06</td>
<td>.03</td>
<td>.15*</td>
<td>.06</td>
<td>.02</td>
<td>.10</td>
<td>.09</td>
<td>.18**</td>
<td>--</td>
<td>.17**</td>
<td>.20**</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

**Notes:**
- Coefficient alpha is provided on the diagonal for items where it is relevant.
- Education scale: 1=High school, 2=Some college, 3=Four-year degree or higher (BA/BS, MA/MS, PhD/Ed.D.). Personality scales: 1=very inaccurate, 5=very accurate.
- Satisfaction scales: 1=extremely dissatisfied, 5=very satisfied.
- Comment behavior: 0=no comment made, 1=comment made.
- 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.
- *p < .05; **p < .01; two-tailed.

1 Openness to Experience was not collected as part of this study, but is included in this table as a point of interest for the reader.
2 This variable was transformed prior to all bivariate and multivariate analyses to reduce negative skewness. The untransformed mean and standard deviation is presented to ease comparison with other satisfaction variables.
3 Five respondents failed to complete items assessing satisfaction with language training in the unit. Therefore, analyses involving this variable include five fewer cases than the rest of the analyses.
4 Interrater reliability and agreement statistics for these variables are presented in Table 3.
### Table 5

*Percentage of Respondents who Commented by Education Level*

<table>
<thead>
<tr>
<th>Education</th>
<th>No Comment</th>
<th>Made Comment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>1. High School</td>
<td>60</td>
<td>56.6</td>
<td>46</td>
</tr>
<tr>
<td>2. Some College</td>
<td>84</td>
<td>42.6</td>
<td>113</td>
</tr>
<tr>
<td>3. Four-year Degree or Higher</td>
<td>53</td>
<td>45.7</td>
<td>63</td>
</tr>
<tr>
<td>Total</td>
<td>197</td>
<td>47.0</td>
<td>222</td>
</tr>
</tbody>
</table>

*Note.* The percentages in this table are based on row totals.
Table 6

*Logistic Regression Analysis Predicting Commenting Behavior*

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$SE$</th>
<th>Wald $\chi^2$</th>
<th>df</th>
<th>Sig.</th>
<th>Exp($B$)</th>
<th>95% CI</th>
<th>Fit Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some College</td>
<td>.50</td>
<td>.25</td>
<td>3.80*</td>
<td>1</td>
<td>.05</td>
<td>1.64</td>
<td>1.00</td>
<td>2.70</td>
</tr>
<tr>
<td>Four-year Degree or Higher</td>
<td>.40</td>
<td>.28</td>
<td>2.02</td>
<td>1</td>
<td>.16</td>
<td>1.50</td>
<td>.86</td>
<td>2.61</td>
</tr>
<tr>
<td>(BA/BS, MA/MS,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph.D./Ed.D.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-.06</td>
<td>.20</td>
<td>.10</td>
<td>1</td>
<td>.75</td>
<td>.94</td>
<td>.64</td>
<td>1.38</td>
</tr>
<tr>
<td>Interest in the Topic</td>
<td>.32</td>
<td>.11</td>
<td>8.85**</td>
<td>1</td>
<td>.00</td>
<td>1.37</td>
<td>1.11</td>
<td>1.69</td>
</tr>
<tr>
<td>Sat with instructor</td>
<td>.66</td>
<td>.25</td>
<td>7.20**</td>
<td>1</td>
<td>.01</td>
<td>1.94</td>
<td>1.20</td>
<td>3.16</td>
</tr>
<tr>
<td>Sat with utility of training</td>
<td>-.33</td>
<td>.10</td>
<td>10.84**</td>
<td>1</td>
<td>.00</td>
<td>.72</td>
<td>.60</td>
<td>.87</td>
</tr>
<tr>
<td>Sat with language training</td>
<td>-.13</td>
<td>.12</td>
<td>1.21</td>
<td>1</td>
<td>.27</td>
<td>.88</td>
<td>.70</td>
<td>1.10</td>
</tr>
<tr>
<td>in the unit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sat with course materials</td>
<td>.04</td>
<td>.12</td>
<td>.14</td>
<td>1</td>
<td>.71</td>
<td>1.04</td>
<td>.83</td>
<td>1.31</td>
</tr>
<tr>
<td>and learning environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cox &amp; Snell $R^2$</strong></td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nagelkerke $R^2$</strong></td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model $\chi^2$</strong></td>
<td>30.33**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*N = 414; *$p < .05$, **$p < .01$

1For this analysis high school is the referent category. The additional education categories are interpreted in comparison to high school.
Table 7

*Multiple Regression Predicting Comment Tone*

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>Model Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sat with instructor</td>
<td>.11</td>
<td>.09</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Sat with utility of training</td>
<td>.09</td>
<td>.04</td>
<td>.25**</td>
<td></td>
</tr>
<tr>
<td>Sat with language training in the unit</td>
<td>.03</td>
<td>.04</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Sat with course materials and learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>environment</td>
<td>.07</td>
<td>.04</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>-.02</td>
<td>.07</td>
<td>-.02</td>
<td></td>
</tr>
<tr>
<td>Emotional stability</td>
<td>-.04</td>
<td>.06</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>$F$</td>
<td></td>
<td></td>
<td></td>
<td>8.90**</td>
</tr>
</tbody>
</table>

*N = 217; **p < .01.*
Table 8

*Multiple Regression Predicting Comment Scope*

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$SE_B$</th>
<th>$\beta$</th>
<th>Model Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sat with instructor</td>
<td>.07</td>
<td>.07</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Sat with utility of training</td>
<td>-.02</td>
<td>.03</td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>Sat with language training in the unit</td>
<td>-.02</td>
<td>.03</td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>Sat with course materials and learning environment</td>
<td>.08</td>
<td>.04</td>
<td>.20*</td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.18</td>
<td>.06</td>
<td>.19**</td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ = .08  
$F$ = 3.49**  

$N = 217$; *$p < .05$, **$p < .01$
Table 9

Multiple Regression Predicting Comment Purpose

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>Model Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>.07</td>
<td>.08</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>.03</td>
<td>.06</td>
<td>.03</td>
<td>.00</td>
</tr>
</tbody>
</table>

$R^2$                | .00  |       |       |           |

$F$                  | .41  |       |       |           |

*N = 222.*
Appendices
Appendix A. Interest in the Topic Items

*Scale: 1 = strongly disagree, 2 = disagree, 3 = somewhat disagree, 4 = neither agree nor disagree, 5 = somewhat agree, 6 = agree, 7 = strongly agree.*

1. I would like to learn as many languages as possible.

2. I don't particularly like the process of language learning and I do it only because I am required to participate in language training. (RS)

3. I would rather spend the time dedicated to language training on training other skills. (RS)

4. I am motivated to perform well in language training so that I will be able to perform my missions more effectively.

5. If I learn to speak this language well, I will have many opportunities to use it.

*Note. RS = reverse-scored.*
Appendix B. Personality Items: IPIP (Goldberg, 1999)

Scale: 1 = very inaccurate, 2 = moderately inaccurate, 3 = neither inaccurate nor accurate, 4 = moderately accurate, and 5 = very accurate.

Extraversion

1. Am the life of the party.
2. Don't talk a lot. (RS)
3. Feel comfortable around people.
4. Keep in the background. (RS)
5. Start conversations.
6. Have little to say. (RS)
7. Talk to a lot of different people at parties.
8. Don’t like to draw attention to myself. (RS)
9. Don’t mind being the center of attention.
10. Am quiet around strangers. (RS)

Agreeableness

1. Feel little concern for others. (RS)
2. Am interested in people.
3. Insult people. (RS)
4. Sympathize with others' feelings.
5. Am not interested in other people's problems. (RS)
6. Have a soft heart.
7. Am not really interested in others. (RS)
8. Take time out for others.
9. Feel others' emotions.
10. Make people feel at ease.

**Conscientiousness**

1. Am always prepared.
2. Leave my belongings around. (RS)
3. Pay attention to details.
4. Make a mess of things. (RS)
5. Get chores done right away.
6. Often forget to put things back in their proper place. (RS)
7. Like order.
8. Shirk my duties. (RS)
9. Follow a schedule.
10. Am exacting in my work.

**Emotional Stability**

1. Get stressed out easily. (RS)
2. Am relaxed most of the time.
3. Worry about things. (RS)
4. Seldom feel blue.
5. Am easily disturbed. (RS)
6. Get upset easily. (RS)
7. Change my mood a lot. (RS)
8. Have frequent mood swings. (RS)
9. Get irritated easily. (RS)
10. Often feel blue. (RS)

Openness to Experience

1. Have a rich vocabulary.
2. Have difficulty understanding abstract ideas. (RS)
3. Have a vivid imagination.
4. Am not interested in abstract ideas. (RS)
5. Have excellent ideas.
6. Do not have a good imagination. (RS)
7. Am quick to understand things.
8. Use difficult words.
9. Spend time reflecting on things.
10. Am full of ideas.

Note. RS = reverse scored.
Appendix C. Satisfaction Items

Scale: 1 = extremely dissatisfied, 2 = dissatisfied, 3 = somewhat dissatisfied, 4 = neither satisfied nor dissatisfied, 5 = somewhat satisfied, 6 = satisfied, 7 = extremely satisfied.

Indicate your level of satisfaction with:

1. The instructor's knowledge of course material and subject matter.
2. The instructor's ability to keep the interest of the class.
3. The instructor's ability to teach the course content.
4. The instructor's responsiveness to student questions and problems.
5. The instructor's overall effectiveness.
6. The match of course objectives with your idea of what would be taught.
7. The relevance of training you received for your specific job tasks or functions.
8. The extent to which the course prepared you to perform on your missions.
10. The quality of audio and visual aids used by the instructor.
11. The computer resources available for this course.
12. The classroom's physical environment.
13. The length of the training course.
14. The pace of the course material presented.
15. The quality of this training course overall.
16. The availability of language training courses for individuals in your unit.
17. The communication of language training information to members of your unit.
18. The quality of language training in your unit.
Appendix D. Rater Training Protocol for Identifying Discrete Comments

Step 1: Determine the unit of analysis: Topic areas/themes.

- Think about the following when determining whether a comment is discrete:
  - If you were going to break the comments into a bulleted list, how many bullets would there be and what would they be?
  - Discrete comments have different target topics and cover different content areas.
  - If the comment can be understood on its own, then it is discrete. Key words like “it” and “this” in a subsequent sentence referring to a topic or content area in a previous sentence are examples of this situation.
    - An example comment was provided for illustration purposes.

Step 2: Rater training.

- Review six examples of classifying comments as discrete.
- Together, we will code a sub-sample of 20 comments (randomly chosen).
- Rules of thumb:
  - Keep sentences together.
  - Keep generic evaluation sentences separate from more specific comments.
  - Break numbered lists apart as separate comments.

Step 3: Identify the number of discrete comments.

- Two coders will read each comment and enter in a cell adjacent to the comment the number of discrete comments provided.
- In addition, each coder will place brackets around each discrete comment.
Step 4: Assessment of interrater reliability and interrater agreement.

Step 5: Review and separate.

• The third rater will review discrete comment codes, make a final decision related to the number of discrete comments, and separate the discrete comments into separate columns in the Excel file for subsequent coding.
Appendix E. Definitions of Comment Tone, Scope, and Purpose

Comment Tone

- Comment tone is often rated on a continuum of negative-positive.
- In this study, comment tone will be rated on the following scale: 1 = *very negative*, 2 = *somewhat negative*, 3 = *neither negative nor positive*, 4 = *somewhat positive*, 5 = *very positive*.
- Consider the following prompt when rating comment tone: What is the tone of this comment?

*Comment Tone Definitions:*

**Very Negative** – Comments that contain opinions that are negative in content and use strong words (e.g., hated, wasted, horrible) and symbols (e.g., !) that reflect the respondent’s high level of dissatisfaction.

**Somewhat Negative** – Comments that contain opinions that are negative in content, but the level of dissatisfaction is not as severe as with the *very negative* category. The words are not as strong as those that would be used in a *very negative* comment (e.g., dislike, not pleased).

**Neither Negative nor Positive** – Comments that do not provide a clear indication of the person’s attitude toward the topic. These comments read like factual statements and contain very little evidence of attitude towards the topic – either positive or negative. In addition, these are comments that provide a somewhat negative component and a somewhat positive component that balance one another (i.e., x was good, but y was bad).
**Somewhat Positive** – Comments that contain opinions that are positive in content, but the level of satisfaction is not as evident as with the *very positive* category. The words are not as strong as those that would be used in a *very positive* comment (e.g., ok, good).

**Very Positive** – Comments that contain opinions that are positive in content and use strong words (e.g., excellent, very good) and symbols (e.g., !) that reflect the respondent’s high level of satisfaction.

**Comment Scope**

- Comment scope refers to whether a comment is general or specific. In this study, the focus will be on the degree of comment specificity.

- In this study, comment specificity will be rated on the following scale: 1 = *not at all specific*, 2 = *somewhat specific*, 3 = *very specific*.

- Consider the following prompt when rating comment specificity: How specific is this comment?

**Comment Scope Definitions:**

**Not at all Specific** – Comments that are very general and contain very few or vague details. These comments provide only general information.

**Somewhat Specific** – Comments that provide some specific information, but are not highly detailed.

**Very Specific** – Comments that provide a lot of details and specific information. These comments are very targeted.
**NOTE: It is easy to confuse wordiness with specificity. Please focus on the amount of detail provided when rating specificity as opposed to the number of words. It is possible to have a wordy comment that is not at all specific or a short comment that is very specific.**

**Comment Purpose**

- Comment purpose refers to whether a comment is descriptive (provides statements about a particular issue) or prescriptive (provides suggestions or recommendations for change). In this study the focus will be on the degree to which comments provide suggestions or recommendations.

- In this study, comment purpose will be rated on the following scale: 1 = not at all prescriptive, 2 = somewhat prescriptive, 3 = very prescriptive.

- Consider the following prompt when rating comment prescriptiveness: How prescriptive is this comment (i.e., to what extent does this comment provide recommendations or suggestions)?

**Comment Purpose Definitions:**

**Not at all Prescriptive** – Comments that do not contain suggestions for improvement. These comments simply state the way things are, not the way things ought to be.

**Somewhat Prescriptive** – Comments that contain some suggestions for improvement, but the suggestions are not strong ones. The language used hints at a suggestion, but is not direct.

**Very Prescriptive** – Comments that provide direct suggestions for improvement – they contain unambiguous statements about the way things ought to be.
Appendix F. Rater Training Protocol for Coding Comment Tone, Scope, and Purpose

Step 1: Overview of definitions and coding protocol.

- Lead researcher will review definition of comment tone, scope, and purpose and key questions.
- Lead researcher will review six examples for each response category of comment tone, scope, and purpose.

Step 2: Practice round 1.¹

- Three raters will practice coding a sample of 20 comments from a training file. Raters will code each comment independently, discuss coding decisions, and develop rules of thumb for subsequent coding.

Step 3: Practice round 2.¹

- Practice coding a sample of 100 comments from the training file. Raters will code each comment independently, but will not discuss. Preliminary interrater reliability and agreement statistics will be assessed after this stage of practice coding to determine whether additional training is needed.

Step 4: Comment coding.¹

- Three raters will read each comment in the data file and code for tone, scope, and purpose.

Step 5: Assessment of interrater reliability and agreement.

Step 6: Aggregation

¹The definitions, key questions, and six examples for each response category will be available to raters during the practice sessions and during actual coding.
• Mean comment ratings for comment tone, scope, and purpose ratings will be calculated across raters for each discrete comment.

• Discrete comment ratings for tone, scope, and purpose will be aggregated back to the comment level (i.e., the mean rating for each discrete comment will be calculated for all comments provided by a participant to determine the mean tone, scope, and purpose).