It Takes Just 120 Seconds: Predicting Satisfaction in Technical Support Calls

Judith A. Hall
Northeastern University

Phil Verghis
Klever

William Stockton
Klever

Jin X. Goh
Northeastern University

ABSTRACT

This study investigated whether information gleaned from the first two minutes of technical support telephone conversations could predict the callers' satisfaction with the technical support person. The first two minutes of 84 calls from employees of a company to their help desk (47 technical support persons) were measured using (1) new participants who listened to or read the conversations and rated their impressions of the technical support person and their satisfaction while playing the caller role (proxy callers), and (2) other raters who rated the caller and aspects of the interaction as a whole. A word count software program (Linguistic Inquiry and Word Count, LIWC) was also used to examine the support persons' communication style. Proxy callers' satisfaction, their ratings of the support persons' behavior (particularly on items indicative of a positive, caller-centered behavior style), and the support persons' use of the first person singular category of the LIWC were all significant predictors of the original callers' satisfaction. These findings have implications for companies' selection and training of customer support employees. © 2014 Wiley Periodicals, Inc.

In the technical support domain, as in customer support and social life more generally, communication quality is important. In particular, when a caller—customer or client—asks for help with a technical problem, time is often of the essence, the need is great, and frustration comes easily. Although companies around the world record technical support conversations and administer satisfaction surveys, the present authors were unable to locate empirical research that measured actual behavior in these interactions. The goal of the present research was to determine whether dimensions of the technical support person's and caller's communication during a recorded telephone conversation would predict the caller's satisfaction with the interaction.

The present study employed methods from the social psychology and health communication fields to study communication processes and callers' satisfaction in 84 calls made to 47 technical support persons in a large company. The present study used the "thin slice" methodology, that is, only a small portion of the total conversation was analyzed—specifically, two minutes taken from conversations that averaged over nine minutes in duration. The thin slice methodology has steadily gained ground in multiple fields, including consumer psychology (Ambady, Bernieri, & Richeson, 2000; Ambady & Rosenthal, 1992; Peracchio & Luna, 2006). Numerous studies have shown that thin slices provide information that is reliable in terms of inter-judge agreement (Murphy, 2005), and representative in terms of correlations with measurements based on longer excerpts of behavior from the same interactions (Murphy, 2005; Roter, Hall, Blanch-Hartigan, Larson, & Frankel, 2011). Thin slices also permit accurate judgments of many kinds of cue contents (e.g., personality, Carney, Colvin, & Hall, 2007; sexual orientation, Rule & Ambady, 2008a; emotion, Matsumoto et al., 2000), and ratings made of thin slices have predictive validity (e.g., ratings of CEO's faces predict company profits, Rule & Ambady, 2008b; ratings of doctors' nonverbal voice tone predict patients' satisfaction in medical encounters, Haskard, Williams, DiMatteo, Heritage, & Rosenthal, 2008).

From the health communication field, the present authors adopted the proxy patient methodology by

employing proxy callers—that is, participants who made ratings of the technical support person and of their own satisfaction while imagining they were the original caller. In research on physician–patient communication, new viewers of recorded medical conversations may be asked to take on the role of the patient and make evaluations of the physician or rate their satisfaction while in that role. The proxy patient has been validated as a useful stand-in when the desired information is not, or cannot be, gathered directly from the patient or when the researcher wishes to present experimentally manipulated stimuli to patients (Blanch-Hartigan, Hall, Krupat, & Irish, 2013; van Vliet et al., 2012).

In the present study, some of the participants were proxy callers who responded to the technical support person and reported their own satisfaction as they thought the original caller would have. If validated against the original callers' satisfaction, the proxy caller (or proxy customer) approach could prove extremely useful in communication research in many business-related areas. In the present study, untrained undergraduate raters served as proxy callers who either listened to or read transcripts of technical support conversations. In addition, other student listeners (third-party raters) made ratings of the caller in each conversation, and made ratings of the conversation as a whole. Because the proxy callers rated both the technical support persons and their satisfaction as they thought the original (actual) callers would have rated it, it was possible to assess the validity of the proxy callers' satisfaction ratings, as well as the predictive value of their ratings of the technical support persons' behavior.

Because the goal was to understand the impact of the communication process itself, it was important that the raters be blind to the factual outcome of the call, lest that knowledge influence their ratings. If proxy callers' perceptions of the support person's communication could predict the original callers' satisfaction without knowledge of problem resolution, this would strongly implicate communication as a determinant of the original callers' satisfaction. Therefore, the two-minute thin slices that were analyzed were always the very first two minutes of conversations that were always longer than two minutes (often considerably longer).

An automated method of analyzing word usage was also used with the technical support providers' part of the dialogue to explore the relationship of selected linguistic categories to the original callers' satisfaction (Linguistic Inquiry and Word Count, LIWC; Pennebaker, Booth, & Francis, 2007). The LIWC software was originally developed for the purpose of analyzing emotional writing samples such as those concerning traumatic experiences, and it has been widely used in social and health psychology to understand various phenomena such as personality, emotions, social relationships, honesty, and thinking styles (Pennebaker, Booth, & Francis, 2007; Pennebaker, Mehl, & Nieder-hoffer, 2003; Tausczik & Pennebaker, 2010). The LIWC calculates the percentages of words that fall within certain meaning or function categories. For instance, words such as *I, me, mine* would fall under the category of "I" words in the LIWC dictionary.

Evidence from survey research, customer service, and sales suggests that communication style does predict outcomes of telephone interactions in business-related contexts. Oksenberg, Coleman, and Cannell (1986) obtained listeners' impressions of the voices of 12 telephone interviewers who were known to have high or low rates of success in obtaining the cooperation of callees to be interviewed. Listeners heard the first 30 seconds of actual attempted telephone survey contacts. Interviewers with low refusal rates had voices that were perceived to be more expressive, faster, more attractive, more positive, more competent, clearer, and more distinct than those with high refusal rates.

Hecht and LaFrance (1995) obtained listeners' impressions of 20 telephone operators' voices based on 5-second clips of their actual calls. When the operator's voice was perceived to be more enthusiastic, sympathetic, professional, expressive, and clear in pronunciation, the calls were shorter in duration. The authors speculated that calls go more smoothly and efficiently, and therefore more quickly, when the operator has favorable vocal qualities.

Peterson, Cannito, and Brown (1995) studied the voices of 21 door-to-door salespeople while they read a scripted overture such as they might make to an actual potential customer. As in the preceding studies, naïve listeners made ratings of the voices and, in addition, acoustic variables were objectively measured. Salesmen who were more successful at completing sales had faster voices, due in part to shorter pause durations, and voices with more falling pitch contour (perceptually, their sentences sounded more declarative and less questioning).

Ambady, Krabbenhoft, and Hogan (2006) also used short excerpts to examine the voices of 12 regional sales managers in an interview with a researcher. One-minute excerpts were electronically filtered so that the linguistic content was unintelligible while the nonverbal vocal characteristics (e.g., speed, contour, and loudness) were still perceptible. Naïve raters' impressions of the interpersonal quality of these excerpts (e.g., cooperative, enthusiastic, and supportive) predicted ratings of performance given to the managers by their own superiors. This result was replicated when the raters heard the unfiltered (normal speech) excerpts.

These studies, in which listeners' impressions predicted the outcomes of social influence (gaining compliance with doing an interview, selling more of a product, being efficient when that is considered a good thing, and being a successful manager) strongly support the argument that the vocal style and/or content of telephone conversations is an important contributor to outcomes that are valued in the business domain. However, none of these studies was on technical support and none used client satisfaction as the outcome.

METHOD

Database of Conversations and Original Caller Satisfaction

The study protocol was approved by the Committee on Human Subject Protection of Northeastern University. A major, tier 1 investment bank provided a convenience sample of digital audio recordings of their employees who called their helpdesk with technical support problems. Many routine problems were included, such as problems with login, password reset, request for Lotus Notes configuration, and many more ranging from simple to complex issues. Surveys are sent randomly to callers to the helpdesk. Only those calls for which the caller subsequently returned a satisfaction survey were used in this study. No specific criteria were used in call selection by the company other than that the calls had resulted in resolution, i.e., resolved by the first responder at the helpdesk and that the original caller had completed a satisfaction survey. The original callers indicated their satisfaction on two items: Courtesy and professionalism of the service desk support person and overall satisfaction with the service desk (both on a 1–10 scale, very low to very high).

Calls shorter than approximately 2-½ minutes in length were not used. The 84 calls that were analyzed averaged 9.31 minutes in duration, with a range from 2.48 minutes to 30.70 minutes. These 84 calls were made to 47 different technical support persons (17% female), with the number of calls per support person ranging from one to seven, as follows: 28 support persons had one call, 11 had two calls, three had three calls, three had four calls, one had five calls, and one had seven calls.

Raters

Raters were 57 Northeastern University undergraduates who received partial course credit in Introductory Psychology or $10.00 for participation. No demographic data were collected, but in this population the majority group is White, average age is about 20, gender is approximately equally distributed between male and female, and many academic majors are represented. In addition, a research assistant coded the gender of the technical support person and his/her accent on a three-point scale (no accent to strong accent).

Materials and Procedure

Thirty-six of the raters were assigned to take the role of proxy callers. Of these, 21 listened to the two-minute conversation excerpts (proxy callers, listening condition) and 15 read a typed transcript of the same two-minute excerpts (proxy callers, transcript condition). Twenty-one additional raters listened and made third-party ratings of the caller's behavior and the over-

all interaction (third-party raters, listening condition). There were no third-party ratings made in the transcript condition. Transcript ratings were obtained in order to understand how much the results would depend on the nonverbal vocal cues available in the listening condition, versus the words by themselves.

All ratings were done in groups of one to three raters, with all the raters in a given group assigned to the same rater role. Because of the length of the rating tasks, rater groups were presented with only as many calls as could be rated in an experimental hour. Each of the 84 call excerpts was listened to by three proxy callers and three (different) third-party raters, and each of the 84 transcript excerpts was read by five (again, different) proxy callers. The choice of how many raters to use per call was based on past experience with similar rating studies.

The experimenter explained that the study was on communication in the customer support context, obtained informed consent, and then described the task requirements. Raters listened to digital audio of the calls using headphones or read a typed transcript booklet. After listening to or reading each excerpt, raters completed either the proxy caller ratings or the third-party ratings (depending on their assignment), before moving on to the next call.

The satisfaction items used by the proxy callers were identical to those used by the original callers and are shown in Table 1. The other rating items are also shown in Table 1. The ratings were selected to capture the dimensions of positive-negative valence, arousal, and dominance and are similar to global ratings used in the physician-patient interaction literature (Roter & Larson, 2002) as well as representing the three basic dimensions of the semantic differential (Osgood, Suci, & Tannenbaum, 1957). In the table, items that were used in the transcript condition are indicated with a superscript; only proxy caller ratings were collected in the transcript condition and only items that made sense in the text-reading context.

Analysis

The multiple raters who rated each conversation were averaged for each call, and then ratings of multiple calls to the same technical support person were averaged within support person, so that all analyses were based on 47 technical support persons and all correlations had 45 degrees of freedom.

Word Usage. Using just the technical support person's part of the call transcripts as input, the LIWC software (Pennebaker, Booth, & Francis, 2007) was used to calculate percentages of different kinds of word usage. Within each of the more than 70 LIWC categories, numerous words and word stems are identified by the LIWC dictionary to be tallied. Particular interest was in pronoun use (first and second person) because they might reflect a more personal tone in the conversation.

Table 1. Rating Scales.

Principal Components Factors and Items

Proxy callers' satisfaction
  Satisfaction with courtesy and professionalism
    (very low . . . very high)
  Overall satisfaction (very low . . . very high)
Proxy callers' ratings of the technical support person
  "Caller-centered support person" factor ($\alpha = 0.94$)
    Impatient, abrupt . . . patient, listened well[a]
    Irritated, rude . . . warm, friendly[a]
    Ignorant, not knowledgeable . . . well informed,
      knowledgeable[a]
    Not respectful . . . very respectful[a]
    Uncaring . . . sympathetic[a]
    Not a likeable person . . . a very likeable person[a]
    Easy going, relaxed . . . aroused, upset[a] (−)
    Interested, attentive . . . bored, uninvolved[a] (−)
    Unattractive voice/speech style . . . attractive
      voice/speech style[a]
    Hard to make out words . . . easy to make out
      words
    Spontaneous, natural . . . mechanical, following
      script[b] (−)
  "Dominant support person" factor (just one item)
    Meek, submissive . . . dominant, pushy[a]
Third-party raters
  "Difficult caller" factor ($\alpha = 0.84$)
    Caller was calm, relaxed . . . aroused, upset
    Caller was courteous, appreciative . . . demanding,
      dominant
    Caller was accommodating, cooperative . . .
      resistant, argumentative
    Caller was irritated, rude . . . warm, friendly (−)
    Caller was ignorant, not knowledgeable . . . well
      informed, knowledgeable (−)
  "Good interaction" factor ($\alpha = 0.94$)
    Rapport (low . . . high)
    Smoothness (low . . . high)
    Efficiency (low . . . high)

Note: All ratings were made on a 1–10 scale. (−) means the item was reversed when added to the composite factor. Alphas refer to the listening condition.
[a]Also rated in transcript condition.
[b]Loaded on its own factor in transcript condition.

Also investigated were the LIWC categories for negation, affect, positive emotion, negative emotion, sadness, tentativeness, certainty, assent, and nonfluency.

RESULTS

Preliminary Analyses

The proxy callers' and third-party raters' impressions (leaving out the proxy callers' satisfaction ratings) for the listening condition were subjected to principal components analysis. A clear four-factor solution was obtained using Varimax rotation, with the factors labeled "caller-centered support person" (11 items, Cronbach's $\alpha = 0.94$), "difficult caller" (five items, $\alpha = 0.84$), "good interaction" (three items, $\alpha = 0.94$), and "dominant sup-

port person" (one item). The items defining each factor are shown in Table 1. Factors were created by averaging the relevant items after appropriate reversals of scale polarity. For the transcript condition, the "caller-centered support person" factor also emerged with the exception that the item on mechanical style loaded on its own, separate factor. Also, as in the listening condition, the dominance item loaded separately on its own factor.

Gender and accent were not significantly related to the original callers' satisfaction and are not discussed further.

Descriptive Data for Satisfaction Variables

Table 2 presents descriptive data for satisfaction. Satisfaction was highest and most variable in the original callers, with progressively lower means and less variability in the two groups of proxy callers. Satisfaction with courtesy/professionalism and overall satisfaction were strongly correlated with each other in each caller group: For original callers, $r = 0.72$; for listening proxy callers, $r = 0.90$; for transcript proxy callers, $r = 0.90$ (all $p < 0.001$).

Correlations between Listening and Transcript Conditions

Listening and transcript proxy callers agreed significantly with each other in their satisfaction. For courtesy/professionalism the correlation was 0.32, $p < 0.05$, and for overall satisfaction it was 0.40, $p < 0.01$. In terms of impressions of the support person's behavior, the two groups of proxy callers had significant agreement on a number of individual items, with their respective factors for "caller-centered support person" being correlated significantly, $r = 0.34, p < 0.05$. Therefore, there was significant overlap between the two rating conditions both on satisfaction and on the impressions made by the speakers.

Is Proxy Callers' Satisfaction Valid?

Table 3 shows that in both groups of proxy callers (listening and transcript), the proxy callers' satisfaction was significantly correlated with the original callers' satisfaction. These correlations were roughly equivalent in magnitude for the two groups of proxy callers. Whether listening or reading, proxy callers could significantly predict the original callers' satisfaction based on exposure to only the first two minutes of the conversation.

Do Impressions of Behavior Predict the Original Callers' Satisfaction?

Table 4 shows the correlations between raters' impressions and the original callers' satisfaction, based on the

Table 2. Descriptive Statistics for Satisfaction Ratings.

Source	Courtesy/Professionalism			Overall Satisfaction		
	Mean	Range	*SD*	Mean	Range	*SD*
Original callers	8.89	3.00–10.00	1.52	8.24	2.00–10.00	2.00
Proxy callers (listening)	7.18	4.00–9.00	1.31	6.77	3.67–9.00	1.38
Proxy callers (transcript)	6.39	4.20–7.80	0.83	6.47	4.00–7.83	0.92

Note: All ratings were made on the same 1–10 scale.

Table 3. Validity of the Proxy Callers' Satisfaction Ratings for Listening and Transcript Conditions: Correlations between Proxy Callers' and Original Callers' Corresponding Scales.

Rating Condition	Correlation
Listening condition	
Courtesy and professionalism	0.42**
Overall satisfaction	0.32*
Transcript condition	
Courtesy and professionalism	0.38**
Overall satisfaction	0.37**

*$p < 0.05$, **$p < 0.01$ (all two-tail).

Table 4. Predictors of Original Callers' Satisfaction.

Condition	Courtesy and Professionalism	Overall Satisfaction
Listening condition		
Proxy callers		
"Caller-centered support person"	0.42**	0.27+
"Dominant support person"	0.16	0.09
Third-party raters		
"Difficult caller"	−0.24	−0.21
"Good interaction"	0.21	0.19
Transcript condition		
Proxy callers		
"Caller-centered support person"	0.38**	0.32*
"Dominant support person"	−0.09	0.08
"Mechanical support person"	−0.13	0.04

Note: Correlations are for the factors based on the principal components analysis.

+$p < 0.10$, *$p < 0.05$, **$p < 0.01$ (all two-tail).

factors derived from the principal components analysis. The Appendix shows the correlations for all individual items. Most notably, the "caller-centered support person" factor for both the listening and transcript groups of proxy callers was a significant predictor of the original callers' satisfaction (as were many of the individual items making up that factor; see Appendix). In the listening condition, the "difficult caller" factor was not significantly correlated with the original callers' satisfaction with courtesy/professionalism, though the correlation was negative as one might expect. Furthermore, the "argumentative" item within that factor was a strong negative predictor, $r = -0.41$, $p < 0.01$, mean-ing that original callers who were rated as more argumentative reported lower satisfaction. Perceived dominance of the support person and ratings of the overall interaction were not predictors of the original callers' satisfaction.

Thus, the most potent source of variance in predicting the original callers' satisfaction was the proxy callers' impressions of a positive and supportive style on the part of the support person. Overall, this held for both the listening and transcript conditions, but there were a few interesting differences for specific items (see Appendix). In the listening modality, the qualities of warmth, sympathy, and (lack of) boredom were more predictive of the original callers' satisfaction than was the case when proxy callers just read the words. These are qualities likely tied to nonverbal cues in the voice. On the other hand, the quality of being well informed and seeming (not) aroused were better predictors of satisfaction in the transcript modality than in the listening modality.

Because listeners heard both the technical support person and the original caller, the possibility was considered that the raters based their impressions on the caller's behavior even when they were told to rate the support person. To investigate this, the original caller's satisfaction with courtesy/professionalism was regressed onto the "caller-centered support person" factor and the "difficult caller" factor simultaneously in a multiple regression analysis of the listening ratings. Both of these variables were independent predictors of satisfaction when each was controlled for the other: For "caller-centered support person," β (standardized regression coefficient) = 0.43, $p < 0.01$; for "difficult caller," $\beta = -0.25$, $p = 0.06$. This shows that proxy caller ratings of the support person were an independent predictor of the original callers' satisfaction.

Word Usage

The LIWC (word-usage) categories of negation, affect, positive emotion, negative emotion, sadness, tentativeness, certainty, assent, and nonfluency were examined with respect to the original callers' satisfaction. The support person's use of first person singular forms (references to I, me, myself, etc.; hereafter called I-words) was significantly positively correlated with the original callers' satisfaction with both courtesy/professionalism and overall satisfaction, $r = 0.43$ and 0.38, respectively,

both $p < 0.01$. I-words were significantly negatively correlated with the "difficult caller" factor based on listening ratings, $r = -0.38$, $p < 0.01$, meaning that more I-words were used with callers who were less argumentative, dominant, and so forth. When both I-words and the "difficult caller" factor were put simultaneously into a regression analysis to predict the original callers' satisfaction with courtesy/professionalism, "difficult caller" was not a significant predictor ($\beta = -0.08$, $p = 0.57$)—reduced from its zero-order correlation of -0.24 (Table 4)—while I-words remained significant ($\beta = 0.40$, $p = 0.01$). Thus, the relation of I-words to the original caller's satisfaction was not due to the caller's behavior that was reflected in the "difficult caller" factor.

The LIWC category called "assent" (use of words such as uh-huh, okay, and yes) was negatively predictive of the original callers' satisfaction (for courtesy/professionalism), $r = -0.32$, $p < 0.05$. When technical support persons made more use of assent words, the original callers were less satisfied with them. Assent was also marginally significantly correlated with the "difficult caller" factor, $r = 0.27$, $p < 0.07$, meaning that more assent was associated with the caller being rated as more argumentative, dominant, and so forth; furthermore, assent was negatively correlated with the support person's rated respectfulness, $r = -0.29$, $p = 0.05$, in both listening and transcript conditions, suggesting that more assent was perceived as less respectful. High usage of assent words may therefore be a marker for an interaction that is not going well. Interestingly, though assent was not related to the overall "caller-centered support person" factor in the listening condition, $r = -0.09$, it was negatively related to that factor in the transcript condition, $r = -0.33$, $p < 0.05$, suggesting that proxy callers did not like reading many occurrences of assenting words. Proxy callers did not appear to mind *hearing* those assent words, except for the apparent negative impact on their estimation of respectfulness as noted above.

To untangle the various associations involving assent in the listening condition, a regression analysis was performed to predict original callers' satisfaction with courtesy/professionalism from assent, the "difficult caller" factor, and the support person's respectfulness. The "difficult caller" and support person's respectfulness remained as predictors of satisfaction ($\beta = -0.25$, $p < 0.07$, and 0.48, $p < 0.001$, respectively), while assent was not an independent predictor of satisfaction, $\beta = -0.11$, $p = 0.42$.

Word Usage and Proxy Caller Ratings Considered Together

In a multiple regression, the original callers' satisfaction with courtesy/professionalism was regressed onto the caller's use of I-words and the proxy callers' ratings on the "caller-centered support person" factor in the listening condition. Together, these two predictors accounted for 40% of variance in the original callers' sat-

isfaction, and both were significant independent predictors: I-words, $\beta = 0.48$, $p < 0.001$, and "caller-centered support person," $\beta = 0.47$, $p < 0.001$. The corresponding analysis for the proxy callers in the transcript condition also showed highly significant regression coefficients for both predictors which together explained 32% of the variance.

DISCUSSION

Improving the quality of communication in customer service can benefit companies, customer service employees of those companies, and customers. Companies may benefit because of increased customer loyalty and reduced costs resulting from fewer escalations and call-backs. Customer service employees may benefit because of reduced stress and burnout, and customers may benefit because of a reduction in hassles (stresses of daily life) that would otherwise diminish their general well-being.

The present study on communication in technical support produced findings that could be important for all of these stakeholders as well as for researchers. Proxy callers—stand-ins for the original callers—were able to significantly predict the original callers' satisfaction, based on hearing or reading just the first two minutes (comprising less than 25% of the average call duration) and without knowing how the caller's problem was resolved. This alone implicates the communication process, not just problem resolution, in satisfaction, a finding that fits with the extensive literature on physician-patient communication (Roter & Hall, 2006). If proxy callers were exposed to more of each call, one can assume their validity for predicting satisfaction would be even higher. One reason why this finding is important stems from the low response rates to customer satisfaction surveys; if proxy callers are acceptable substitutes for the actual caller, many research avenues are opened up, as is the case in the physician-patient communication domain where proxy patients are increasingly used (van Vliet et al., 2012).

Prediction was generally very similar whether the ratings were made by listeners or transcript raters, though some different patterns were noted for specific rating items. Thus, it appears that in this study there was not a great deal of extra information gained by including nonverbal cues in the voice, such as pitch, loudness, rate, variation and so forth that one might think would be strong predictors of satisfaction. Perhaps in this context, satisfaction is equally predicted from both modalities.

Proxy callers' impressions of the technical support person, the use of first person pronouns by the support person, and raters' impressions of the behavior of the caller also significantly predicted the original callers' satisfaction, thus providing insight into what communication qualities are important to satisfaction (and which are not, as some proxy caller ratings of the support persons were not predictive of the original callers'

satisfaction—dominance, for example). The most predictive cluster of support person behaviors, in both the listening and transcript modalities, was what the present authors called the "caller-centered support person," consisting of items such as warm, respectful, sympathetic, not bored, likeable, and informed. This cluster bears a strong resemblance to what is called patient-centeredness in the medical communication field, which is widely demonstrated to predict the satisfaction of medical patients as well as clinical outcomes (Mead & Bower, 2000; Roter & Hall, 2006; Stewart et al., 1995). Analysis that took ratings of the caller's behavior into account confirmed that the association of a positive style in the support person with the original callers' satisfaction was independent of the association of caller behavior with the original callers' satisfaction.

It should be remembered, however, that analysis at the level of the support person meant that ratings of the individual callers were combined within each support person and therefore speak to how the support person's callers tended to behave, not how individual callers behaved. There was no correlation, however, between the "caller-centered support person factor" and the "difficult caller" factor ($r = 0.02$), meaning that there was no mutual influence at the support person level between the behavior of support persons and callers.

The association of first person pronouns to satisfaction can speculatively be interpreted as reflecting a more personal exchange between support person and caller. However, first person plurals ("we" words) were not predictive of satisfaction, though that would be predicted by the physician-patient communication literature in which "we" words are considered a marker for good partnership building (Roter & Hall, 2006). Perhaps the failure of "we" words to predict satisfaction in the present study was due to the very low percentage of usage (less than 1% compared to over 5% for first person singular usage).

Another intriguing interpretation is based on a qualitative difference between physician-patient conversations and technical support calls. One could assume that customers call the support person because they cannot solve their own problem and would like the support person to do so. This would require the support person to be a single agent of change; hence, using the term "I" might signal a desirable attitude on the part of the support person. However, in medical interactions there is a stronger need for collaboration, and patients often expect to participate in decision making and to share a therapeutic alliance with the physician. Therefore, the use of "we" words would be a marker of a patient-centered attitude in the physician and would likely be a stronger predictor of patient satisfaction.[1]

Slatcher, Vazire, and Pennebaker (2008) found, in an analysis of instant messaging exchanges in romantic couples, that I-word usage by the woman predicted her male partner's satisfaction with the relationship, as well as the prospectively measured longevity of the relationship, while We-words did not. The similarity of these findings to our own is intriguing and suggests that linguistic usage may be a fruitful avenue in customer service research, especially if longer-term outcomes (e.g., customer loyalty, analogous to relationship longevity in Slatcher, Vazire, & Pennebaker) can be predicted. In future research, it would be elucidating to examine the conversational context in which the I-words were used.

There are several limitations to the present study. Because it was correlational, inferences about paths of causation can only be speculative. Furthermore, only fairly routine technical support problems were studied, and the research was done in only one company in one industry (financial services). Furthermore, the participants who served as proxy callers were undergraduates, quite dissimilar in age and life experience to the original callers. Nevertheless there was extensive evidence that their ratings were valid, in spite of this. Moreover, previous studies involving ratings of conversations in the business context have also used student raters with success (Ambady, Krabbenhoft, & Hogan, 2006; Hecht & LaFrance, 1995; Oksenberg et al., 1986).

Future directions for research using the proxy caller paradigm are both methodological and substantive. Methodological questions include determining the optimal duration of thin slice excerpts for predicting satisfaction, the optimal number of proxy callers (or other raters) per call, and the optimal number of original callers per support person. Clearly there are trade-offs between reducing measurement error by using more in all of these categories (longer durations, more proxy callers, and more original callers per support person) versus how much investment future researchers or companies wish to make in a rather labor-intensive research methodology. Other measurement methods could be used, and other behaviors measured, as well as other outcomes besides satisfaction. On the substantive side, research in this vein could contribute to improvements in employee selection and training.

---

[1] We thank an anonymous reviewer for this interpretation.

REFERENCES

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychological Bulletin, 111, 256–274.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), Advances in experimental social psychology, 32 (pp. 201–272). San Diego, CA: Academic Press.

Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. Journal of Consumer Psychology, 16, 4–13.

Blanch-Hartigan, D., Hall, J. A., Krupat, E., & Irish, J. T. (2013). Can naïve viewers put themselves in the patients' shoes? Reliability and validity of the analogue patient methodology. Medical Care, 51, 16–21.

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin-slice approach to accuracy of first impressions. Journal of Research in Personality, 41, 1054–1072.

Haskard, K. B., Williams, S. L., DiMatteo, M. R., Heritage, J., & Rosenthal, R. (2008). The provider's voice: Patient satisfaction and the content-filtered speech of nurses and physicians in primary medical care. Journal of Nonverbal Behavior, 32, 1–20.

Hecht, M. A., & LaFrance, M. (1995). How (fast) can I help you? Tone of voice and telephone operator efficiency in interactions. Journal of Applied Social Psychology, 25, 2086–2098.

Matsumoto, D., LeRoux, J, Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., et al. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian brief affect recognition test (JACBART). Journal of Nonverbal Behavior, 24, 179–209.

Mead, N., & Bower, P. (2000). Patient-centredness: A conceptual framework and review of the empirical literature. Social Science & Medicine, 51, 1087–1110.

Murphy, N. A. (2005). Using thin slices for behavioral coding. Journal of Nonverbal Behavior, 29, 235–246.

Oksenberg, L., Coleman, L., & Cannell, C. F. (1986). Interviewers' voices and refusal rates in telephone surveys. Public Opinion Quarterly, 50, 97–111.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning. Oxford, UK: University of Illinois Press.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. Annual Review of Psychology, 54, 547–577.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic inquiry and word count: LIWC 2007 [computer software]. Austin, TX: LIWC.net.

Peracchio, L. A., & Luna, D. (2006). The role of thin-slice judgments in consumer psychology. Journal of Consumer Psychology, 16, 25–32.

Peterson, R. A., Cannito, M. P., & Brown, S. P. (1995). An exploratory investigation of voice characteristics and selling effectiveness. Journal of Personal Selling & Sales Management, 15, 1–15.

Roter, D. L., & Hall, J. A. (2006). Doctors talking with patients/patients talking with doctors: Improving communication in medical visits, 2nd ed. Westport, CT: Praeger.

Roter, D. L., & Larson, S. (2002). The roter interaction analysis system (RIAS): Utility and flexibility for analysis of medical interactions. Patient Education and Counseling, 46, 243–251.

Roter, D. L., Hall, J. A., Blanch-Hartigan, D., Larson, S., & Frankel, R. M. (2011). Slicing it thin: New methods for brief sampling analysis using RIAS-coded medical dialogue. Patient Education and Counseling, 82, 410–419.

Rule, N. O., & Ambady, N. (2008a). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. Journal of Experimental Social Psychology, 44, 1100–1105.

Rule, N. O., & Ambady, N. (2008b). The face of success: Inferences from chief executive officers' appearance predict company profits. Psychological Science, 19, 109–111.

Slatcher, R. B., Vazire, S., & Pennebaker, J. W. (2008). Am 'I' more important than 'we'? Couples' word use in instant messages. Personal Relationships, 15, 407–424.

Stewart, M., Brown, J. B., Weston, W. W., McWhinney, I. R., McWilliam, C. L., & Freeman, T. R. (1995). Patient-centered medicine: Transforming the clinical method. Thousand Oaks, CA: Sage Publications.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29, 24–54.

van Vliet, L. M., van der Wall, E., Albada, A., Spreeuwenberg, P. M., Verheul, W., & Bensing, J. M. (2012). The validity of using analogue patients in practitioner-patient communication research: Systematic review and meta-analysis. Journal of General Internal Medicine, 27, 1528–1543.

Correspondence regarding this article should be sent to: Judith A. Hall, Department of Psychology, Northeastern University, Boston, MA 02115 (j.hall@neu.edu).

APPENDIX

Correlations between Individual Ratings and Original Callers' Satisfaction

Condition and Rating	Courtesy and Professionalism	Overall Satisfaction
Transcript condition		
Proxy callers' ratings of the technical support person		
Impatient, abrupt . . . patient, listened well	0.31*	0.23
Irritated, rude . . . warm, friendly	0.47***	0.29*
Ignorant, not knowledgeable . . . well informed, knowledgeable	0.26+	0.25+
Not respectful . . . very respectful	0.49***	0.33*
Uncaring . . . sympathetic	0.51***	0.36**
Not a likeable person . . . a very likeable person	0.44**	0.28+
Unattractive voice/speech style . . . attractive voice/speech style	0.36**	0.18
Hard to make out words . . . easy to make out words	0.13	0.08
Easy going, relaxed . . . aroused, upset	−0.05	−0.04
Spontaneous, natural . . . mechanical, following script	−0.27+	−0.17
Interested, attentive . . . bored, uninvolved	−0.44**	−0.23
Meek, submissive . . . dominant, pushy	0.16	0.09
Third-party raters		
Caller was calm, relaxed . . . aroused, upset	−0.04	−0.07
Caller was courteous, appreciative . . . demanding, dominant	−0.25+	−0.21
Caller was accommodating, cooperative . . . resistant, argumentative	−0.41**	−0.26+
Caller was irritated, rude . . . warm, friendly	0.28+	0.21
Caller was ignorant, not knowledgeable . . . well informed, knowledgeable	0.00	0.08
Rapport (low . . . high)	0.30*	0.21
Smoothness (low . . . high)	0.19	0.20
Efficiency (low . . . high)	0.12	0.13
Transcript condition		
Proxy callers' ratings of the technical support person		
Impatient, abrupt . . . patient, listened well	0.25+	0.27+
Irritated, rude . . . warm, friendly	0.29*	0.20
Ignorant, not knowledgeable . . . well informed, knowledgeable	0.45**	0.41**
Not respectful . . . very respectful	0.41**	0.34*
Uncaring . . . sympathetic	0.27+	0.13
Not a likeable person . . . a very likeable person	0.35*	0.29*
Unattractive speech style . . . attractive speech style	0.28+	0.25+
Easy going, relaxed . . . aroused, upset	−0.36**	−0.33*
Interested, attentive . . . bored, uninvolved	−0.15	−0.09
Meek, submissive . . . dominant, pushy	−0.09	0.08
Spontaneous, natural . . . mechanical, following script	−0.13	0.04

Note: All ratings were made on a 1–10 scale.

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ (all two-tail).