Asia Pacific Journal of Information Systems
Vol. 33 No. 4 (December 2023), 1118-1134

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2023.33.4.1118

Validity of Language-Based Algorithms Trained on Supervisor Feedback Language for Predicting Interpersonal Fairness in Performance Feedback

Jisoo Ock[a,*], Joyce S. Pang[b]

[a] *Associate Professor, Department of Business Administration, Pusan National University, Korea*
[b] *Associate Professor, Division of Psychology, Nanyang Technological University, Singapore*

A B S T R A C T

Previous research has shown that employees tend to react more positively to corrective feedback from supervisors to the extent they perceive that they were treated with empathy, respect, and concern towards fair interpersonal treatment in receiving the feedback information. Then, to facilitate effective supervisory feedback and coaching, it would be useful for organizations to monitor the contents of feedback exchanges between supervisors and employees to make sure that supervisors are providing performance feedback using languages that are more likely to be perceived as interpersonally fair. Computer-aided text analysis holds potential as a useful tool that organizations can use to efficiently monitor the quality of the feedback messages that supervisors provide to their employees. In the current study, we applied computer-aided text analysis (using closed-vocabulary text analysis) and machine learning to examine the validity of language-based algorithms trained on supervisor language in performance feedback situations for predicting human ratings of feedback interpersonal fairness. Results showed that language-based algorithms predicted feedback interpersonal fairness with reasonable level of accuracy. Our findings provide supportive evidence for the promise of using employee language data for managing (and improving) performance management in organizations.

Keywords: Feedback Delivery, Performance Management, Interpersonal Fairness, Computer-Aided Text Analysis, Machine Learning

Ⅰ. Introduction

Supervisors are increasingly being held accountable for employee development through on-going feedback and coaching (e.g., communicating on-going behavioral expectations, regular informal feedback exchanges on specific tasks, social recognition of performance, short-term goal-setting). An important advantage of continuous performance management through on-going feedback and coaching

*Corresponding Author. E-mail: jisoo.ock@pusan.ac.kr

is its utility for effective goal setting. In most organizations, performance appraisals occur infrequently. Although organizations vary on how often they evaluate their employees' job performance, the gap between performance appraisals is typically as long as a year and sometimes even longer (Meyer, 1991; Spence and Keeping, 2011). As a result, performance goals that are set in a performance evaluation session tend to be on a long-term basis (e.g., achieve certain performance goal for the next performance appraisal cycle; Murphy and Cleveland, 1995). However, as jobs become increasingly complex, it is often difficult to set specific performance goals that will retain their relevance over an extended period of time because the circumstances surrounding a job and performance on that job are constantly changing (Pulakos and O'Leary, 2011). Thus, for many jobs, continuously providing short-term goals that are applicable to each changing work situations are likely to be much more useful than providing broad, long-term goals that may become obsolete or irrelevant with changing work circumstances.

Even for jobs that are relatively constant and predictable, continuously communicating how employees are doing respect to their goals, what they should do to help meet those goals, and continually revising performance goals with respect to individual employee's progress towards his/her goals are useful managerial interventions for driving performance and engagement, because they allow employees to develop concrete understanding of what the organizations expect from them, how they are doing with respect to those expectations, and specific steps they need to take in order to meet them (London, 2003; Yukl, 2002). Additionally, when corrective feedback is provided immediately following a negative performance episode, it is more likely to produce changes in job behavior than when feedback is given later, when both feedback providers and feedback recipients are less likely to have accurate recollections about the performance episode and the context in which the performance episode took place (Gregory et al., 2008). Taken together, there are strong theoretical and empirical reasons to believe that regular informal feedback is a useful managerial intervention for employee development.

However, previous models of feedback seeking behavior have suggested that the effectiveness of performance feedback is dependent on the degree to which supervisors help create an environment that facilitates and promotes feedback seeking and the use of performance feedback (Steelman et al., 2004; Whitaker et al., 2007). Among a number of supervisor factors, feedback delivery, which refers to employees' perceptions about the intentions of the supervisor in providing feedback, can affect how employees react and respond to the feedback information (Fedor et al., 1989; Steelman et al., 2004). Specifically, employees are more likely to react positively to feedback (e.g., stronger motivation towards task-related goals, stronger perceived helpfulness of feedback information) to the extent that they perceive that they were treated with empathy, respect, and concern about their well-being in receiving the feedback information (Leung et al., 2001; O'Malley and Gregory, 2011; Steelman et al., 2004).

Several empirical studies have provided support for the effect of positive supervisor feedback delivery on employee reactions to feedback. For example, Young et al. (2017) conducted a video-based online experiment and found that participants who were exposed to supervisor empathic concern (observable manifestation of concern towards others; Batson, 2011) in negative feedback reported greater increase in positive affect and higher effectiveness evaluation of the supervisor's feedback behavior compared to participants in the control condition. Also, in a field study, Young et al. (2017) found that supervisors who provided high quality negative feedback (as rated by the supervisors' subordinate employees) were rated as being more promotable (as rated by the supervisors' bosses) and that this relationship was especially stronger for supervisors with high subordinate perceptions of empathic concern. Similarly, to the extent that employees perceived that their supervisors provided performance feedback with a higher level of interpersonal consideration, employees had lower perceived cost and higher perceived value of feedback seeking (VandeWalle et al., 2000), and reported higher satisfaction with feedback and stronger motivation to use feedback information to improve job performance (Steelman and Rutkowski, 2004).

Research also suggests that one of the factors that feedback recipients might consider in determining the feedback provider's feedback delivery is the language that feedback providers use in communicating feedback information. For example, Nguyen et al. (2017) found that critical feedback with positive affective language (e.g., including positive statements like "good work so far") was associated with increased positive emotions, decreased annoyance and frustration, and increased work quality compared to critical feedback that did not contain positive affective language. This suggests that one of the ways that supervisors can improve employees' perceptions of feedback delivery is by using language that is more likely to be perceived as being supportive, encouraging, and empathetic.

In the current study, we refer to the degree of consideration, empathy, and respect that supervisors show towards their employees in providing performance feedback as *feedback interpersonal fairness*. The positive effect that interpersonally fair feedback delivery has on feedback outcomes is consistent with findings from the literature on organizational justice, which has shown that employees evaluate the fairness of organizational decisions based on the perceived fairness of the interpersonal treatment they receive from the organizational decision makers (Cropanzano et al., 2007).

The positive effect that interpersonally fair feedback delivery has on feedback outcomes suggests that organizations can facilitate the effectiveness of continuous performance management practices by monitoring the interpersonal attitude that supervisors demonstrate in providing performance feedback to their employees and engaging in appropriate interventions (e.g., training; Gallo and Steelman, 2019) to address ineffective feedback delivery behavior. Although there are traditional methods that organizations can pursue for monitoring supervisor feedback effectiveness (e.g., interviewing, or administering a retrospective survey to employees about the quality of their supervisors' feedback), technology is enabling ways for organizations to make such evaluations more quickly, efficiently, and on a much broader scale (Sheets et al., 2019). Namely, recent advances in computer-aided text analysis have opened the door for researchers to capture behavior-related variance from various types of language data in organizational contexts (Speer, 2021).

In the current study, we applied computer-aided text analysis (CATA) on supervisor language data in performance feedback situations to develop a language-based machine learning (ML) algorithm for predicting human ratings of feedback interpersonal fairness. We propose that validity evidence for such language-based model can have important practical value for managing and improving continuous performance practices in organizations.

Ⅱ. Method

2.1. Sample and Procedure

Data for this study was obtained from previous research that examined the relationship between supervisor personality and feedback interpersonal fairness (Ock and Pang, 2022). Specifically, participants were recruited from a panel maintained by a crowdsourcing behavioral research platform (Prolific). Participants were required to have supervisory experience as well as prior experience conducting performance appraisal and providing performance feedback to participate in the study.

Participants consisted of 246 supervisors in the U.S. and the U.K. working in a variety of industries (e.g., healthcare, retail, finance). The sample consisted of more males (n = 158, 64.2%) than females (n = 88, 35.8%), with a mean age of 36.1 years (SD = 11.2) and a mean working experience of 15.5 years (SD = 10.6). There was a large gender disparity in our sample. However, the gender ratio in our sample closely approximates the gender ratio of managers in organizations. According to Chilazi et al. (2021), females make up only 38% of managers. The gender disparity in the distribution of managers is even worse in Asia-Pacific, with a report indicating that the proportion of females who hold managerial positions in private organizations was just 20% (UN Women, 2021).

Participants were provided with three critical incidents describing a situation in which a supervisor witnessed an ineffective performance episode of his/her employee (see <Appendix A> for the specific critical incidents used). Then, assuming the role of the supervisor described in the critical incident, participants were asked to write what they would say to the employee in that situation, in verbatim, using at least 50 words. The study was conducted online. Thus, participants entered the feedback into a text box that was provided.

2.2. Measurement of Feedback Interpersonal Fairness

Two undergraduate research assistants who were blind to the purpose of the study rated each feedback comment using three items that measure the feedback delivery dimension of the Feedback Environment Scale (Steelman et al., 2004). The items captured the degree to which the rater perceived that the feedback was provided in an interpersonally just manner (i.e., prosocial feedback delivery; see <Appendix B> for list of the items). The items were on a 7-point Likert scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The interrater reliability of the ratings was .92. Because the level of agreement between raters was high, the average ratings were used to indicate participants' feedback interpersonal fairness scores.

2.3. Analysis of Performance Feedback Language Using Computer-Aided Text Analysis

We employed closed-vocabulary text analysis to analyze the performance feedback language data. Closed-vocabulary text analysis uses the word frequency procedure, which quantitatively characterizes a given text by examining the number of times individual words of a text is categorized into pre-defined categories of related words and phrases, called *dictionaries*, relative to the total number of words in the text. The relative frequency scores that are calculated for each dictionary can be used as outputs that are used in subsequent statistical analyses with other variables (Eichstaedt et al., 2021; McKenny et al., 2018).

Each dictionary contains a pre-determined list of words and phrases that are theoretically believed to represent a particular category, much like how measures of psychological constructs contain a list of items that are thought to capture the underlying latent construct being measured (Eichstaedt et al., 2021). Different dictionaries are available to measure different aspects of the text like parts of speech, informal speech, and punctuations. Dictionaries that measure words that reflect psychological dimensions (e.g., anger, sadness, affiliations) are also available. The words in a dictionary may not necessarily be similar in meaning or frequently co-occur. Rather, different words of a dictionary reflect different aspects of a common construct that holistically define the construct when they are measured together (Eichstaedt et al., 2021). For example, a dictionary that Speer et al. (2019) developed for capturing *leading and deciding* dimension of the Great Eight job performance factors (Kurz and Bartram, 2002) includes words and phrases like "authority," "delegates," "take responsibility," and "took over," each of which reflects the different ways that behaviors related to leading others and making decisions in an organizational environment might be described in the English language.

We used a closed-vocabulary text analysis software called Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015). In addition to counting and scoring the proportion of text corresponding to different psychological categories in its default dictionary (e.g., *negative emotion*, *positive emotion*, *anger*, *sadness*), it employs algorithms to derive summary variables that evaluate the degree to which the text is formal and logical vs. informal and personal (*analytical thinking*); being spoken from the perspective of confidence vs. tentativeness (*clout*); honest and disclosing vs. guarded and distanced (*authentic*); and positive and upbeat vs. anxious and hostile (*emotional*). Also, it analyzes various linguistic features of the text like parts of speech (e.g., *adjectives*, *conjunctions*, *prepositions*, *verbs*), informal language use (e.g., *swear words*, *fillers*), time orientations (e.g., *past focus*, *present focus*), personal drives (e.g., *affiliation*, *power*), and so forth. LIWC has been used in previous research to operationalize language data for making predictions about HR-relevant factors, such as job satisfaction (based on online company reviews; Jung and Suh, 2019), performance appraisal ratings (based on performance narrative comments; Speer et al., 2019), and gender biases in reference letters (based on reference letter contents; Madera et al., 2019).

2.4. Machine Learning Prediction

Predictive Modeling. We entered LIWC category scores as predictors and human ratings of feedback interpersonal fairness as an outcome variable in a ML prediction model. The model was trained and tested using the *caret* package in R (Kuhn, 2008). Specifically, we used elastic net regression and *k*-fold cross-validation with *k* being set to five, splitting the sample data into random 80% training samples and 20% test samples five times. Each test sample was independent such that a participant was included in the test sample only once. For each iteration, predictive model algorithm was built on the training sample then cross-validated on the test sample to test how well the algorithm performed in an independent sample. We quantified prediction accuracy as the correlation between the actual and predicted scores that was calculated using the optimal tuning parameters determined by elastic net regression.

Elastic Net Regression. Elastic net regression is a type of penalized regression that adds a constraint to the linear regression model by imposing a penalty that shrinks the regression coefficients for predictor variables that do not contribute to prediction. Penalized regression is a useful method for dealing with overfitting by decreasing the complexity of a model when there is a large multivariate dataset (Chapman et al., 2016). Elastic net produces a regression model that shrinks the regression coefficients for non-contributing predictor variables by imposing L1-norm (penalty term imposed in LASSO regression) and L2-norm (penalty term imposed in ridge regression) penalty terms. These penalty terms add an important extension to the linear regression analysis such that the sum of the absolute values of regression coefficients (L1-norm) or the sum of the squared values of regression coefficients (L2-norm) are constrained to not exceed a specified value called *tuning parameter*, symbolized as lambda (λ). Penalized regression adjusts the tuning parameter value to obtain the optimal regression weights that reduces model complexity and overfitting and increases cross-validation accuracy. As lambda reaches zero, the penalty terms have a weaker effect, and the regression coefficient estimates approach the values found with linear regression. Conversely, as lambda increases, the penalty terms have a stronger effect and shrink the regression coefficients for non-contributing predictor variables to zero (L1-norm) or close to zero (L2-norm). In elastic net regression, there is an additional tuning parameter, symbolized as alpha (α), that assigns the weight given to L1-norm and L2-norm penalties. Alpha ranges between zero and 1.0. As alpha reaches 1.0, more weight is given to L1-norm penalty (i.e., penalty terms approach LASSO regression), and as alpha reaches zero, more weight is given to L2-norm penalty (i.e., penalty terms approach ridge regression).

In training predictive models, *caret* systematically varies the alpha and lambda values according to a specified parameter. Then, it provides the alpha and lambda value combination that provides the highest cross-validated accuracy for a given outcome. In the current study, we tried five values each for alpha and lambda (using default values in caret). Model accuracy was quantified as the correlation between the actual and predicted scores on human ratings of feedback interpersonal fairness. Specifically, we calculated model accuracy in the test sample for each of the five folds for each set of tuning parameters, then reported the correlations for the optimal tuning parameters.

We present correlation coefficients instead of regression coefficients because of algorithmic uncertainties in elastic net regression that allow the specific weights and rankings of predictors to vary (Hickman et al., 2021). Namely, because different data is used to train each model across the five-fold cross-validated models, some level of variability in regression weights and rank-order of those weights are expected due to sampling error. Moreover, because elastic net regression minimizes or excludes the effect of predictors that do not contribute to meaningful prediction, the cross-validated models for each dependent variable may include somewhat different sets of LIWC categories. Thus, following the procedures in Hickman et al. (2021), we used bivariate correlations that use full information in the sample to examine the relationships between LIWC categories and the outcome variable.

Ⅲ. Results

<Table 1> Descriptive Statistics and Zero-Order Correlations between Demographic Variables and Feedback Interpersonal Fairness Rating

Variable	*M*	*SD*	1.	2.	3.	4.
1. Gender	.6	.5	-			
2. Age	36.1	11.2	-.07	-		
3. Tenure	15.5	10.6	-.05	.93**	-	
4. Interpersonal fairness	4.7	1.6	.22**	-.08	-.04	*.92*

Note: M = mean; SD = standard deviation. Gender was coded as 0 = male; 1 = female. Feedback interpersonal fairness was rated on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Reliability for feedback interpersonal fairness is interrater reliability. ** $p < .01$.

The descriptive statistics of participant gender, age, tenure (years worked), and feedback interpersonal fairness ratings are provided in <Table 1>. The average feedback interpersonal fairness rating was high (M = 4.7, SD = 1.6), indicating that participants generally provided feedback in a way that showed support and consideration towards the feedback recipient. The correlations indicated that female participants showed higher feedback interpersonal fairness ($r = .22$, $p < .01$). Participants' age or tenure was not correlated with feedback interpersonal fairness ratings.

<Table 2> provides the statistically significant correlations between LIWC categories and feedback interpersonal fairness. The correlations showed that 22 LIWC categories were significantly correlated with feedback interpersonal fairness (out of 93 total categories). We found that *word count* ($r = .47$, $p < .01$) and *words per sentence* ($r = .24$, $p < .01$) was positively correlated with feedback interpersonal fairness ($r = .47$, $p < .01$). Similarly, *authentic* showed reliable positive correlation with feedback interpersonal fairness ($r = .19$, $p < .01$). Results also showed that *negative emotion* and *informal* categories were negatively correlated with feedback interpersonal fairness ($r = -.24$, $p < .01$ and -.15, $p < .01$, respectively). Although all correlations with $p < .05$ were reported

<Table 2> Statistically Significant Correlations between LIWC Categories and Human Ratings of Feedback Interpersonal Fairness

LIWC Variable	Feedback Interpersonal Fairness
Adverb	.15
Affect	-.20
Authentic	.19
Biological Processes	-.21
Cognitive Processes	.18
Comma	.17
Conjunction	.23*
Feeling	.21
Focus Present	-.15
Function	.17
Informal	-.15
Leisure	-.14
Negative Emotion	-.20
Perceptual Processes	.15
Relativity	.25*
Space	.22
Tentative	.21
Time	.21
Word Count	.47*
Work	-.19
Words Per Sentence	.24*
2nd Person Pronouns	-.18

Note: All correlations are significant at $p < .05$. * denotes correlations significant after Bonferroni correction ($p < .0005$).

<Table 3> Five-Fold Cross-Validated Accuracy for Predicting Human Ratings of Feedback Interpersonal Fairness

	Elastic Net Parameters					
Outcome	Alpha	Lambda	$\bar{r}$	$r_{\min}$	$r_{\max}$	r_{SD}
Interpersonal Fairness	0.9	2.487249	.43	.31	.61	.34

Note: Tuning parameters reported for the optimal models. $\bar{r}$ was calculated by correlating predicted and human ratings of feedback interpersonal fairness. $r_{\min}$ = minimum correlation; $r_{\max}$ = maximum correlation; r_{SD} = standard deviation of the correlations.

in <Table 2>, it should be noted that a smaller number of correlations (four per model) remained significant after applying Bonferroni correction (i.e., $p \leq .0005$; significance denoted by *).

The prediction accuracy results for the predictive algorithm are provided in <Table 3>. The optimal tuning parameters was .9 for alpha and 2.49 for lambda. The alpha value indicates that the penalty in elastic net regressions was close to LASSO penalty (as it was close to 1.0), and the large lambda value indicates that strong penalty was imposed on the regression weights because many LIWC categories did not contribute to meaningful prediction of the outcome variable. The average correlation between the predicted and reported scores for feedback interpersonal fairness across the five test folds was $\bar{r}$ = .43 (r_{SD} = .34).

Ⅳ. Discussion

The results showed that language-based algorithms trained on supervisor language in performance feedback situations predicted human ratings of feedback interpersonal fairness with a reasonable level of accuracy. Additionally, we identified several specific feedback language features that are associated with feedback interpersonal fairness, allowing us to develop a more concrete understanding of specific language features that predict feedback interpersonal fairness.

Specifically, we found that language features like *negative emotions* (e.g., words that express anger) and *informal language* (e.g., swear words, fillers, shortened text) were associated with lower feedback interpersonal fairness. The negative relationship between expression of negative emotions and perceived interpersonal fairness of feedback indicates that it is important for supervisors to control their emotions when interacting with their employees regarding their job performance. This may be difficult to do when supervisors are reacting to employee behaviors that seriously undermine the functioning of the organization or when they are dealing with employees who are being rude and disrespectful. Controlling emotions may be even more difficult when reactions to such employee behaviors have to be immediate. In those situations, it may be advisable for supervisors to find ways that would allow them to be calmer and objective in providing feedback (e.g., speaking to the employee in private after taking some time to recede angry reaction to the negative episode).

Similarly, our results suggest that the use of informal language in performance feedback is likely to be perceived as inconsiderate, which is expected to have negative effect on the acceptability of feedback information. Informal language includes the use of shortened texts that are commonly used on the Internet (e.g., btw, lol, thx). Shortened words are usually less applicable in spoken feedback (although the use of abbreviated words and phrases are becoming increasingly common in spoken Korean language, especially among younger generations), they can certainly be used in performance comments that are digitally provided.

Also, we found that expression of *authenticity* in feedback language, which indicates an honest, personal, and disclosing form of communication, is associated with higher feedback interpersonal fairness rating. This finding suggests that being open and candid in communicating feedback information is likely to be perceived positively by feedback recipients. This relationship makes theoretical sense because we expect that feedback language that involves showing empathy and consideration towards the feedback recipient to require honesty and communication on a personal level rather than being distant and guarded.

We found that objective feedback language features like *word count* and *words per sentence* were significant predictors of feedback interpersonal fairness. Although objective language features like *word count* are seemingly unrelated to interpersonal fairness, it is reasonable to suggest, based on the analysis of the contents of feedback with high vs. low word count, that these language features indirectly capture the level of effort that the supervisors put into providing feedback, which in turn may be related to feedback interpersonal fairness. For example, supervisors who are less concerned about providing feedback in a fair, interpersonal just manner might be less concerned about being perceived as terse or snappish in their interaction with employees, which in many cases, likely involves using fewer words.

It was difficult to explain some of the significant relationships that feedback language features had with feedback interpersonal fairness. For example, feedback language features like *conjunctions* (use of conjunctions like and, but, whereas) and *relativity* (LIWC category that contains words related to details of position, time, or action) showed significant positive correlations with feedback interpersonal fairness. Given the inductive data-driven nature of the current predictive models, we do not expect all (if any) of these difficult-to-explain predictive relationships to withstand the rigor of replication and cross-validation. Future work that replicates, cross-validates, and expands the limited context in which performance feedback occurred in the current study is crucial for building greater confidence in the stability of the predictive accuracy of feedback language-based algorithms for predicting feedback interpersonal fairness.

4.1. Contributions of the Current Study

The current study makes unique contributions to the performance management practice and literature in two important ways. First, as technologies for facilitating real-time digital feedback exchanges become increasingly prevalent in organizational settings (Ewenstein et al., 2016; Petryk et al., 2022; Rivera et al., 2021), the current study provides useful and timely illustration as to how organizations can take advantage of those tools for improving performance management. Namely, the current study results provide empirical evidence that CATA and ML can be used to train algorithms using supervisor feedback language data to make inferences regarding the level of feedback interpersonal fairness that supervisors show when they provide performance feedback to their employees. This suggests that organizations can significantly increase the efficiency of evaluating the quality of feedback that supervisors provide to their employees. Specifically, the current study results indicate that organizations can evaluate and monitor the interpersonal fairness of performance feedback that supervisors provide to their employees in real-time, as opposed to traditional methods (e.g., retrospective survey) that tend to be much less efficient and less comprehensive, if such measures are taken to manage the quality of supervisor feedback at all.

Additionally, the availability of information regarding supervisors' feedback quality means that organizations can use that information to identify and reward supervisors who are providing performance feedback in a considerate and interpersonally fair manner. In doing so, organizations can send a clear signal to supervisors that fair interpersonal treatment of employees in providing performance feedback is important and valued. This is likely to contribute to facilitating performance feedback exchanges in organizations. Previous models of performance feedback have indicated that supervisors, and more specifically, the level of empathy and concern towards the employee that supervisors show during performance feedback, is an important contextual factor that contributes to the overall organizational support for day-to-day feedback processes (Steelman et al., 2004; Whitaker et al., 2007). The contextual aspects of feedback processes in an organization, which is referred to as *feedback environment* (Steelman et al., 2004), is said to predict various outcomes of feedback (e.g., satisfaction with feedback, motivation to use feedback information) as well as the frequency with which employees will seek feedback from their supervisors. Thus, to the extent that CATA can facilitate organizational interventions (i.e., efficient and effective monitoring of supervisor feedback language that allows organizations to identify and reward appropriate feedback language) that can encourage supervisors to provide performance feedback in a more interpersonally fair manner, CATA is expected to help improve performance management practices in organizations.

Second, the current study results provide more concrete understanding of the specific features of feedback language that make feedback more (or less) likely to be perceived as being interpersonally fair. In addition to examining the validity of language-based models for predicting feedback interpersonal fairness, we identified specific features of feedback language that are predictive of perceived interpersonal fairness of feedback (e.g., negative emotions, informal language, authenticity, word count). Thus, our findings provide useful guidance regarding *how* performance feedback language can be designed to improve employee reactions. This has important implications for automation of performance feedback in organizations. That is, it may be possible to train large language models to learn such specific feedback language features enhance employee reactions. Then, supervisors may enter simple keywords or phrases that describe their employees' performance, then ask language models to generate feedback in a way that are likely to be perceived as being considerate and interpersonally fair.

4.2. Limitations and Future Research Directions

Findings from the current study should be considered with a few limitations in mind. First, the contexts in which the participants provided feedback were limited to employees engaging in ineffective behaviors on the job that are not task-related. The purpose of this approach was to make sure that the feedback situations were general enough so that participants could provide reasonable feedback even if they do not have managerial experience in the settings that were described in the critical incidents that were provided. However, in many organizations, it is more common for supervisors to address employees' task-related behaviors through feedback. To the extent that there is a meaningful difference in the variability in the language features between feedback about task-related behaviors and non-task-related behaviors, we can expect there to be a difference in the validity of the models that are based on those feedback. For example, supervisors might be more formal and logical when providing feedback about task-related behaviors than when providing feedback about non-task-related behaviors, in which case the predictive validity of the language feature associated with formal vs. informal nature of language (*analytical thinking*) might be lower in task-related feedback contexts. Future research that examines the validity of language models that are based on task-related feedback language would be a useful extension to the current study.

Second, the method that we used to collect feedback comments from the participants may have affected their feedback language. Namely, participants were asked to provide performance feedback through text rather than through speech in a face-to-face conversation. As a result, participants are likely to have been able to better control their feedback comments. For example, in providing performance feedback through text, participants may have written their initial comments, reviewed them, and edited them to enhance the clarity, interpretability, and fairness of the comments. Although we can assume that supervisors would also prepare and think about what they would say to their employees in providing performance feedback in a face-to-face manner, the process is likely to be less thorough or rigorous as the amount of preparation that supervisors might put into providing performance feedback through text. Moreover, when supervisors provide feedback through text, they might be less concerned about how employees might react to the feedback comments because such reactions would not be immediately available to them, which in turn, might affect how they approach feedback delivery. For example, in a face-to-face feedback conversation, some supervisors might plan on being very critical of employee before the feedback session, but eventually change their approach if it seems that such feedback delivery style is ineffective (e.g., employee seems to be visibly angry and unaccepting of the feedback comments). Although text-based feedback is not uncommon and may become more prevalent with the increasing use of technology that take advantage of the efficiency and timeliness of text-based feedback, it would be useful for future research to examine the validity of language-based models developed on supervisors' verbal feedback (which can be transcribed to text using artificial intelligence-powered software) for predicting feedback interpersonal fairness. Specifically, such research should entail not only the level of interpersonal fairness expressed in verbal language, but also the level of interpersonal fairness expressed in non-verbal behavior in feedback delivery.

Third, we used closed-vocabulary method for analyzing participants' feedback comments, which has several important limitations. Namely, because words are counted without consideration of context or order, it is often difficult to accurately interpret different types of lexical ambiguities (e.g., same word having more than one meaning, irony, sarcasm) using a closed-vocabulary method (Schwartz et al., 2013). Although dictionaries can be developed to analyze documents more accurately in specific contexts, their accuracies are likely to suffer when they are applied to documents that are outside of those contexts. Additionally, closed-vocabulary methods typically rely on a smaller number of words in analyzing texts, which limits the level of accuracy that can be achieved, especially when they are applied to analyze large amounts of text written in general, non-specific contexts (e.g., social media; Kern et al., 2016). However, closed-vocabulary method has several useful properties. Namely, as mentioned, dictionaries are grounded in theory. Thus, the language variables that are extracted from closed-vocabulary methods are easily interpretable. Similarly, dictionaries can be developed for texts written in specific contexts, which allows for more accurate analysis of texts written in those contexts (Eichstaedt et al., 2021). Nonetheless, future research could consider using a more powerful, modern computer-based language models or using dictionaries that are specifically tailored to languages that are used in professional business contexts for examining the validity of language models for predicting feedback effectiveness.

Finally, there is an important need to examine the effect that cultural context may have on the study results, particularly in cultural contexts in which inequality in the distribution of power between groups, including in organizational contexts, is accepted (i.e., power distance). For example, in many Asian countries that have high power distance (e.g., China, Malaysia, Singapore, South Korea), power difference between people higher vs. lower in the organizational hierarchy is often accepted and considered the norm. In these contexts, supervisors may be more critical or react more negatively to ineffective employee performance, especially if it entails behaviors that are against the socially accepted norm regarding power distance (e.g., insubordination, public display of disrespect against the supervisor). Future research that cross-validates the validity of language-based models for predicting feedback interpersonal fairness in additional samples, especially in Asian contexts, would be a useful extension to the current study (see Pang and Ock, 2023, for a review of application of theoretical models of feedback delivery behavior in Asian organizational contexts).

4.3. Concluding Comments

The lack of strong system for monitoring feedback exchanges between supervisors and employees poses a challenge to successful adaptation of continuous performance management practices in organizations. Findings from the current study suggest that organizations can partially automatize the monitoring of continuous performance management practices by applying CATA and ML on supervisor feedback language, which is becoming more readily available and accessible with the advent of technology that allows organizations to easily collect, store, and analyze various types of employee language data. However, our ML predictive model results are invariably limited by sampling error and the limited context in which feedback was delivered in the current study. Also, given that there were a few inexplicable feedback language features that contributed to the prediction of feedback interpersonal fairness, not all inferences about feedback (in)effectiveness that may be derived from the analysis of supervisor feedback language are likely to be useful. Thus, there is an important need for future research to replicate and cross-validate our findings to build greater confidence in the validity of such system for monitoring continuous performance management practices.

Acknowledgements

This work was supported by a 2-Year Research Grant of Pusan National University.

<References>

[1] Batson, C. D. (2011). *Altruism in Humans.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195341065.001.0001

[2] Chapman, B. P., Weiss, A., and Duberstein, P. R. (2016). Statistical learning theory for high dimensional prediction: Application to criterion-keyed scale development. *Psychological Methods, 21*(4), 603-620. https://doi.org/10.1037/met0000088

[3] Chilazi, S., Bohnet, I., and Hauser, O. (2021). Achieving gender balance at all levels of your company. *Harvard Business Review.* Retrieved from https://hbr.org/2021/11/achieving-gender-balance-at-all-levels-of-your-company

[4] Cropanzano, R., Bowen, D. E., and Gilliland, S. W. (2007). The management of organizational justice. *Academy of Management Perspectives, 21*(4), 34-48. https://doi.org/10.5465/amp.2007.27895338

[5] Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V. A., Smith, L. K. Buffone, A., Iwry, J., Seligman, M. E. P., and Ungar, L. H. (2021). Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods, 26*(4), 398-427. https://doi.org/10.1037/met0000349

[6] Ewenstein, B., Hancock, B., and Komm, A. (2016). Ahead of the curve: The future of performance management. *McKinsey Quarterly.* Retrieved from www.mckinsey.com/business-functions/organization/our-insights/ahead-of-the-curve-the-future-of-performance-management

[7] Fedor, D. B., Eder, R. W., and Buckley, M. R. (1989). The contributory effects of supervisor intentions on subordinate feedback responses. *Organizational Behavior and Human Deicision Process, 44*(3), 396-414. https://doi.org/10.1016/0749-5978(89)90016-2

[8] Gallo, J. R., and Steelman, L. A. (2019). Using a training intervention to improve the feedback environment. In L. A. Steelman and J. R. Williams (Eds.), *Feedback at work* (pp. 163-174). Springer Nature. https://doi.org/10.1007/978-3-030-30915-2_9

[9] Gregory, J. B., Levy, P. E., and Jeffers, M. (2008). Development of a model of the feedback process within executive coaching. *Consulting Psychology Journal: Practice and Research, 60*(1), 42-56. https://doi.org/10.1037/1065-9293.60.1.42

[10] Hickman, L., Saef, R., Ng, V., Woo, S. E., Tay, L., and Bosch, N. (2021). Developing and evaluating language-based machine learning algorithms for inferring applicant personality in video interviews. *Human Resource Management Journal.* https://doi.org/10.1111/1748-8583.12356

[11] Jung, Y., and Suh, Y. (2019). Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. *Decision Support Systems, 123,* 113074. https://doi.org/10.1016/j.dss.2019.113074

[12] Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., and Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods, 21*(4), 507-525. https://doi.org/10.1037/met0000091

[13] Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1-26. https://doi.org/10.18637/jss.v028.i05

[14] Kurz, R., and Bartram, D. (2002). Competency and individual performance: Modeling the world of work. In I. T. Robertson, M. Callinan, and D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 227-255). Wiley.

[15] London, M. (2003). *Job feedback: Giving, seeking, and using feedback for performance improvement.* Erlbaum.

[16] Leung, K., Su, S., and Morris, M. W. (2001). When is criticism not constructive? The roles of fairness perceptions and dispositional attributions in employee acceptance of critical supervisory feedback.

Human Relations, 54(9), 1155-1187. https://doi.org/10.1177/0018726701549002
[17] Madera, J. M., Hebl, M. R., Dial, H., Martin, R., and Valian, V. (2019). Raising doubt in letters of recommendation for academia: Gender differences and their impact. *Journal of Business and Psychology, 34*(3), 287-303. https://doi.org/10.1007/s10869-018-9541-1
[18] Meyer, H. H. (1991). A solution to the performance appraisal feedback enigma. *Academic of Management Executive, 5*(1), 68-76. https://doi.org/10.5465/AME.1991.4274724
[19] McKenny, A. F., Aguinis, H., Short, J. C., and Anglin, A. H. (2018). What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management, 44*(7), 2909-2933. https://doi.org/10.1177/0149206316657594
[20] Murphy, K. R., and Cleveland, J. N. (1995). *Understanding Performance Appraisal: Social, Organizational and Goal-Oriented Perspectives.* Sage.
[21] Nguyen, D. T., Gancarz, T., Ng, F., Dabbish, L. A., and Dow, S. P. (2017). Fruitful feedback: Positive affective language and source anonymity improve critique reception and work outcomes. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1024-1034). ACM. https://doi.org/10.1145/2998181.2998319
[22] Ock, J., and Pang, J. S. (2022, April). *Effect of Personality and Knowledge about Professional Conduct on Feedback Delivery* [Poster]. Society for Industrial and Organizational Psychology Annual Conference, Seattle, WA, United States.
[23] O'Malley, A. L., and Gregory, J. B. (2011). Don't be such a downer: Using positive psychology to enhance the value of negative feedback. *The Psychologist-Manager Journal, 14*(4), 247-264. https://doi.org/10.1080/10887156.2011.621776
[24] Pang, J. S., and Ock, J. (2023). Supervisors' prosocial feedback delivery: Dispositional trait and motivational concerns. In J. Wood, J. Ramsay, K. Thirumaran, and E. Ng (Eds.), *Managing People across Asia-Pacific: An Organizational Psychology Approach.* Edward Elger Publishing.
[25] Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015.* Austin, TX: University of Texas at Austin.
[26] Petryk, M., Rivera, M., Bhattacharya, S., Qiu, L., and Kumar, S. (2022). How network embeddedness affects real-time performance feedback: An empirical investigation. *Information Systems Research, 33*(4), 1467-1489. https://doi.org/10.1287/isre.2022.1110
[27] Pulakos, E. D., and O'Leary, R. S. (2011). Why is performance management broken? *Industrial and Organizational Psychology, 4*(2), 146-164. https://doi.org/10.1111/j.1754-9434.2011.01315.x
[28] Rivera, M., Qiu, L., Kumar, S., and Petrucci, T. (2021). Are traditional performance reviews outdated? An empirical analysis on continuous, real-time feedback in the workplace. *Information Systems Research, 32*(2), 517-540. https://doi.org/10.1287/isre.2020.0979
[29] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., and Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE, 8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791
[30] Sheets, T. L., Belwalkar, B. B., Toaddy, S. R., and McClure, T. K. (2019). Filling the I-O/technology void. In R. N. Landers (Ed.), *The Cambridge Handbook of Technology and Employee Behavior* (pp. 22-37). Cambridge University Press. https://doi.org/10.1017/9781108649636.003
[31] Speer, A. B. (2021). Scoring dimension-level job performance from narrative comments: Validity and generalizability when using natural language processing. *Organizational Research Methods, 24*(3), 572-594. https://doi.org/10.1177/1094428120930815
[32] Speer, A. B., Schwendeman, M. G., Reich, C. C., Tenbrink, A. P., and Siver, S. R. (2019). Investigating

the construct validity of performance comments: Creation of the Great Eight narrative dictionary. *Journal of Business and Psychology, 34*(6), 747-767. https://doi.org/10.1007/s10869-018-9599-9

[33] Spence, J. R., and Keeping, L. (2011). Conscious rating distortion in performance appraisal: A review, commentary, and proposed framework for research. *Human Resource Management Review, 21*(2), 85-95. https://doi.org/10.1016/j.hrmr.2010.09.013

[34] Steelman, L. A., Levy, P. E., and Snell, A. F. (2004). The feedback environment scale: Construct definition, measurement, and validation. *Educational and Psychological Measurement, 64*(1), 165-184. https://doi.org/10.1177/0013164403258440

[35] Steelman, L. A., and Rutkowski, K. A. (2004). Moderators of employee reactions to negative feedback. *Journal of Managerial Psychology, 19*(1), 6-18. https://doi.org/10.1108/02683940410520637

[36] UN Women (2021). Snapshot of women's leadership in Asia and the Pacific. Retrieved from https://asiapacific.unwomen.org/en/news-and-events/in-focus/csw/snapshot-of-womens-leadership-in-asia-and-the-pacific#_ftn17

[37] VandeWalle, D., Ganesan, S., Challagalla, G. N., and Brown, S. P. (2000). An integrated model of feedback-seeking behavior: Disposition, context, and cognition. *Journal of Applied Psychology, 85*(6), 996-1003. https://doi.org/10.1037/0021-9010.85.6.996

[38] Whitaker, B. G., Dahling, J. J., and Levy, P. (2007). The development of a feedback environment and role clarity model of job performance. *Journal of Management, 33*(4), 570-591. https://doi.org/10.1177/0149206306297581

[39] Young, S. F., Richard, E. M., Moukarzel, R. G., Steelman, L. A., and Gentry, W. A. (2017). How empathetic concern helps leaders in providing negative feedback: A two-study examination. *Journal of Occupational and Organizational Psychology, 90*(4), 535-558. https://doi.org/10.1111/joop.12184

[40] Yukl, G. (2002). *Leadership in Organizations.* Prentice Hall.

<Appendix A>

Please read the descriptions about employee behavior on the job below. Read each description carefully and assuming that you are the supervisor of the employee in the description, write what you would say to the employee (in verbatim) in these situations. For each of your response, use at least 50 words.

1. You are a manager of a busy restaurant. Because you are a bit short on staff, you ask one of your servers to work tomorrow night. S/he scoffs at you and says, "I'd rather not, thanks." What would you say to the server?
2. You are a manager of a clothing store. One day, you overhear Angie, a veteran employee, telling a new clerk that because employees are paid minimum wage, most of them sometimes take home clothes for themselves. At closing time, you call Angie to your office to discuss this issue. What would you say?
3. You are the manager of a small factory. You walk past one of your employees working on a dangerous machine and smell beer on his/her breath. You call him/her over and say…

<Appendix B>

Feedback Environment Scale

Please read the participants' written feedback carefully and indicate how accurately the following statements describe his/her feedback on a scale ranging from 1 to 7, where…

1 = very inaccurate
2 = somewhat inaccurate
3 = slightly inaccurate
4 = neither accurate nor inaccurate
5 = slightly accurate
6 = somewhat accurate
7 = very accurate

Feedback Delivery

1. The participant was supportive when giving feedback.
2. The participant was considerate of the feedback recipient's feelings.
3. The participant did not treat the feedback recipient very well when providing performance feedback.

◆ About the Authors ◆

Jisoo Ock

Jisoo Ock is an associate professor of human resource management in the Department of Business Administration at Pusan National University. Dr. Ock's research focuses on measurement of psychological constructs and its application for predicting people's behavior in organizational settings.

Joyce S. Pang

Joyce S. Pang is an associate professor of psychology in the Division of Psychology at Nanyang Technological University. Dr. Pang is a personality psychologist, with a broad interest in the influences of multiple social identities and individual differences on well-being, performance, and health psychology outcomes within different social contexts.

Submitted: May 26, 2023; 1st Revision: November 9, 2023; Accepted: December 7, 2023