7-13-2018 2:00 PM

Examining Rating Source Differences in Narrative Performance Feedback

Kevin Doyle
The University of Western Ontario

Supervisor
Dr. Richard D. Goffin
The University of Western Ontario

Graduate Program in Psychology
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Abstract

The purpose of the present study was to examine the narrative feedback quality and content of comments from supervisors, peers, and subordinates in a multisource performance feedback context. Research on performance management interventions tends to focus on issues such as rater training, scale development, scale formats, and reducing test and rater bias. However, other components in performance management interventions have received little attention, including narrative feedback. Narrative feedback takes the form of written comments describing the ratee's performance on different dimensions. The narrative feedback quality variables included favorability, specificity, goal content, and feedback length. Predictor variables of narrative feedback quality including rater familiarity, rater acquaintanceship time, and ratee position tenure were also investigated. The narrative feedback content variables included the amount of relative content, absolute content, task content and trait content.

The data were collected using a commercial multi-source feedback instrument which included numeric ratings and narrative feedback from the perspectives of the ratees' supervisors, peers and subordinates. A sample of 200 ratees with manager or director in their title were selected. Each of the 8,967 comments were coded by four trained research assistants. The results indicated that supervisors provided the highest quality narrative feedback, peers and subordinates were comparable. Rater familiarity tended to be positively related to narrative feedback quality, and, interestingly, acquaintanceship time tended to be negatively related to narrative feedback quality, suggesting that acquaintanceship time should not be used as a proxy for familiarity. Ratee position tenure was negatively related to narrative feedback quality, however the

relationship was smallest for peers suggesting the use of peer raters for longer-tenured ratees. The rating source comparisons of the narrative feedback content variables suggested that all sources used about the same amount of each content type, and that the relationships between the content variables and narrative feedback quality were comparable across rating sources. The overall results for relative, absolute, and trait feedback content suggested that they were related to positive description and included little actionable content. Task content had the largest positive relationships with narrative feedback quality, indicating that future rater training should focus on the provision of task content.

Keywords: narrative feedback; 360 degree feedback; multisource feedback; familiarity; acquaintanceship time; position tenure; feedback content.

Acknowledgements

I would like to start by recognizing all the people who put so much time and effort into making this project succeed. My supervisor, Dr. Rick Goffin, went above and beyond mentoring and supporting me throughout this project, from idea conception to reporting the insights. My lab mates Justin Feeney, Kabir Daljeet, and Rebecca Factor also supported this project by assisting with managing the dataset, developing training for the coders, and being subject matter experts for scale development.

I would like to thank Dr. Paul Tremblay. His guidance during the dissertation proposal regarding structuring the dataset and statistical analyses elevated the project. Dr. Mitch Rothstein's subject matter expertise was also helpful at the proposal stage of this project. I would like to thank my departmental defence committee of Dr. Natalie Allen, Dr. Joan Finnegan, and Dr. Alex Benson. Their input and questions helped make the dissertation more defensible and prepared me for the senate defence. I would also like to thank my senate defence committee or Dr. Chet Robbie, Dr. Ann Peng, Dr. Natalie Allen, and Dr. Joan Finnegan. I really enjoyed speaking with each of you about the project and appreciated the insight you provided me.

I would like to thank all the faculty that have helped shape my undergraduate and graduate school experience. Dr. John Meyer was an excellent instructor and mentor both in the classroom and through RUWP projects. Dr. Susan Pepper introduced me to Industrial/Organizational Psychology in her undergraduate leadership class in 2008 and introduced me to my Honours Thesis supervisors, Dr. Tom O'Neill and Dr. Natalie Allen. Dr. O'Neill provided me with my first experience with psychological research

which helped move me toward graduate studies in Industrial/Organizational Psychology. Dr. Allen played a very influential role in my education. She taught me during my undergraduate studies, welcomed me into her teamwork lab, and hired me as a research assistant while I applied to graduate school. The time and energy she invested really pushed me to pursue my doctorate. My supervisor, Dr. Rick Goffin played an enormous role in my development as a researcher and instructor. He pushed me to round out my skill-set by encouraging me to teach lab courses, act as recruitment lead for the Industrial/Organizational Psychology area, attend conferences, and get involved in consulting projects through the RUWP.

I would like to thank my friends in the Industrial/Organizational Psychology program. Graduate school can be difficult to navigate without a strong support network, and I had one of the best. Finally, I'd like to thank my family for providing me support throughout my post-secondary career. Thank you for listening to the ups and downs. Thank you for being there to listen and celebrate the small wins. I truly appreciate the love and support.

Table of Contents

List of Tables

List of Appendices

Performance management can be defined as a "continuous process of identifying, measuring, and developing the performance of individuals and teams and aligning performance with the strategic goals of the organization" (Aguinis, 2009, p. 2). The development of performance management systems begins by identifying key tasks and skills that are necessary to be effective at a particular job, generally through a process called job analysis (see Cascio & Aguinas, 2011). The next undertaking is to develop a set of scales that adequately represent the tasks and skills identified in the job analysis and to link these with the organizational goals and vision. The set of scales comprises the annual performance appraisal or performance evaluation. Performance evaluation has two main purposes, to support administrative functions and to assist in employee development. Administrative functions include decisions regarding pay raises, promotion, termination, allocation of rewards, etc. Although administrative functions are central to the performance management process, the present research looks specifically at the employee development component. It is believed that by providing an employee with feedback regarding their performance that they will have a better understanding of their strengths and weaknesses and change their behavior accordingly.

With employee development being a central tenet in the performance management literature, it is surprising that one-third of performance feedback interventions result in decreased performance (Kluger & DeNisi, 1996). The meta-analysis by Kluger and Denisi (1996) has become highly cited largely because researchers have yet to fully understand why so many feedback interventions result in decreased performance. One possible reason is researchers' fixation on the numeric or measurement component of performance management (i.e., Murphy & Cleveland, 1995).

Research on performance management interventions tends to focus on issues such as scale development, scale formats, and reducing test and rater bias (i.e. Austin & Villanova, 1992; Landy & Farr, 1980; Nowack & Mashihi, 2012). However, other components common in performance management interventions have received little attention, including narrative feedback. Narrative feedback generally takes the form of written comments that explain why particular ratings were given. A recent benchmarking study indicated that 85% of multisource feedback instruments contain narrative feedback items (3D Group, 2013), underscoring just how common narrative performance feedback is (Brutus, 2010). Despite the prevalence of narrative feedback, the overwhelming majority of performance feedback research has focused on numeric performance ratings (i.e., Ilgen & Moore, 1987; Ludwig & Goomas, 2009; Vigoda-Gadot & Angert, 2007).

The purpose of the present study was to examine the narrative feedback quality and content of comments from supervisors, peers, and subordinates in a multisource feedback context. We were only able to find two studies that have examined narrative feedback in the context of performance evaluation (David, 2013; Wilson, 2010). Both of these studies focus solely on narrative performance feedback provided by the supervisor, leaving important feedback provided by other rating sources unexamined (i.e., peers and subordinates). Because very little research has focused on narrative performance feedback, and there are no published findings for peer and subordinate raters, we took an inductive approach to uncover effects which would in turn lead to theory development. The approach used, as suggested by Hambrick (2007), involved the generation of results from a large sample analysis that informed researchers what we need theory development for. Our investigation solely utilized research questions in order to better understand and

report the observed effects. Thus, we see the present study as the groundwork for many studies to come. The data were collected using a commercial multi-source feedback instrument which included numeric ratings and narrative feedback from the perspectives of the ratees' supervisors, peers and subordinates.

We next discuss the background of narrative feedback research, followed by a discussion of multisource feedback systems and rating sources (i.e., supervisors, peers, and subordinates), and narrative feedback quality. The remainder of the introduction will address three lines of investigation. The first is the overall differences in the quality of narrative feedback across the ratings sources. The second is the introduction of predictor variables and how these might impact the quality of the narrative feedback provided. The predictor variables included the rater's familiarity with the ratee's work behavior, the acquaintanceship time of the rater with the ratee, and the position tenure of the ratee. The third is the content of the narrative feedback provided and how the content differentially related to narrative feedback quality.

Narrative Feedback Background

As previously mentioned, in response to the findings of Kluger and Denisi (1996) several researchers have turned toward narrative feedback as a means to investigate why some performance evaluations result in decreased performance. Wilson (2010) was the first to investigate narrative feedback with regard to performance evaluation. Wilson's (2010) study investigated different performance descriptors supervisors utilized when providing feedback to their subordinates, and potential differences based on ratee ethnicity. Two researchers developed the dictionary for performance descriptors. Additionally, two researchers coded the first 60 performance appraisals to calculate inter-

rater reliability. The remaining 607 performance appraisals were coded by a single researcher. Wilson's (2010) findings indicate that supervisors provided overwhelmingly positive comments. Further, the positivity of the comments often contradicted the associated numeric ratings. Finally, Wilson's (2010) findings suggest that supervisors emphasized a different set of factors across ethnic groups in arriving at an overall evaluation.

Following Wilson's (2010) work, David (2013) set to develop and test a theory of quality narrative feedback. David (2013) suggested that supervisor feedback that is both directive (lengthy, specific, and includes goals) and motivational (favorable and high in interactional justice) would be related to year-lagged performance. David's (2013) investigation followed the performance of 1,019 nurses. The data were collected from the organization's automated performance appraisal system. Similar to Wilson (2010), David (2013) had 5 researchers code the first 100 performance appraisals to demonstrate inter-rater reliability. Following this, David (2013) coded the remaining data herself. David (2013) found that both favorability and interactional justice had direct and indirect effects on year-lagged employee performance.

The present study makes new contributions over what has already been investigated in the domain of narrative performance feedback in a number of ways. First, we extended our focus beyond the supervisor to also investigate effects for peer raters and subordinate raters. Second, we adapted David's (2013) measures of narrative feedback quality in order to address a number of issues. David's (2013) scales had an inconsistent number of scale points and the scale point labels were often categorical rather than continuous. Third, both Wilson (2010) and David (2013) involved other

researchers to code an initial subset of data in order to calculate inter-rater reliability statistics, following which the data were coded by a single researcher. We sought a more rigorous approach, however. We developed a Frame of Reference (FOR; Bernardin, 1979) training program to train our research assistants. Further, each comment was coded by four trained research assistants, and the author did not code any of the data to remove the possibility of bias.

Multisource Feedback Systems

Multisource feedback systems are tools that gather information about a target employee from two or more rating sources (Balzer, Greguras & Raymark, 2005). These sources may include an employee's supervisors, peers, subordinates, customers, etc. Generally, multisource feedback systems lend themselves to management positions that can take advantage of multiple rating sources and perspectives. Multisource feedback systems are considered to be primarily developmental tools. In a recent survey, 98 percent of organizations cited employee development as one of the uses of their multisource feedback system (3D Group, 2013). It has been suggested that ratings from different sources provide different perspectives on the performance of any given employee, which can help guide the development and improvement process (i.e., Balzer et al., 2005). The present study made use of ratings from supervisors, peers, and subordinates.

Narrative Feedback Quality

Performance feedback that is predominantly numeric provides insufficient context (David, 2013). Hence, it can be unclear to employees why they received a particular rating. This becomes more apparent as job complexity increases (David, 2013). For

instance, managerial roles tend to be multi-faceted and involve tasks and skills that are difficult to quantify. As an example, a manager receives a score of 2 on a 5-point scale measuring facilitating teamwork, where a score of 1 is well below expectations and 5 is well above expectations. This manager is working to better herself as a leader and wants to improve her performance, but she is unsure what component of facilitating teamwork she needs to improve upon. She may interpret her score as an indication that she needs to conduct team building in order to build comradery. However, the rater may have supplied the lower score because they find that team tasks are poorly structured and expectations are not being made clear. The context provided in narrative feedback is necessary for developing precise goals that drive the development process. Furthermore, evidence shows that employees pay attention to narrative feedback (Bracken & Rose, 2011), more than they do the numeric ratings (Ferstl & Bruskiewicz, 2000). Thus, not only do employees receive important context when they are provided with narrative feedback, but they are likely to attend to it and internalize it.

David (2013) contends that narrative feedback should be both motivational and directional to be high quality. Thus, high quality narrative feedback should not only provide vital information with regard to how the employee should improve their current performance but also provide the support and encouragement necessary to meet their improvement goals.

T**he motivational component of narrative feedback quality**. Narrative feedback tends to be more motivational when it avoids harsh criticism and provides ample support (DeNisi & Pritchard, 2006). Accordingly, the favorability of narrative feedback is likely to influence how readily the ratee accepts and acts on their goals. In

support of this, David (2013) found that favorability demonstrated significant direct and indirect effects on year-lagged performance. Favorability is judged by the degree to which the feedback is positive rather than negative. David (2013) also included interactional justice in the motivational component of narrative feedback quality which captured how the rater treated the ratee with dignity, respect, kindness and consideration in the feedback provided. During a pre-screening study examining the adapted scales from David (2013), the interactional justice component and favorability component were highly correlated ($r = .949$, $p < .001$). Therefore, only favorability was retained to reduce redundancy. The prescreening is further discussed in the methods section.

The directive component of narrative feedback quality. The directive component assists employee development by affecting the ease to which the ratee can glean important information regarding their performance and set relevant and specific goals. Locke and Latham (1984) proposed that specific, detailed, and accepted goals work best to motivate behavior. As such, the directive component of narrative feedback quality includes three indices (David, 2013). The first is specificity, defined as the degree to which the feedback is detailed and supported with behavioral examples. The second is goal content, defined as the degree to which the rater provides actionable steps to improve performance. The third is simply the narrative feedback's length – longer narrative comments are generally presumed to contribute to higher quality feedback.

Section 1: Overall Differences in Narrative Feedback Quality by Rating Source

We were only able to find two studies that have examined narrative feedback in the context of performance evaluation (David, 2013; Wilson, 2010). Both of these studies focus solely on narrative performance feedback provided by the supervisor, leaving

important feedback provided by other rating sources unexamined (i.e., peers and subordinates). As mentioned, based on David's (2013) conceptions of narrative feedback quality, our first line of investigation was to examine possible differences in narrative quality between rating sources. This is pertinent for two main reasons. First, identifying the rating sources that provide higher quality narrative feedback would allow researchers and practitioners to sample more heavily from these sources to ensure that the ratee is receiving the best information on which to base their professional development. Second, should we find differences in narrative feedback quality between rating sources, the results will provide researchers with a base from which to explore why this occurred. For these reasons we propose:

Research Question 1: Does feedback from different rating sources (supervisors, peers, and subordinates) vary on the indices of narrative feedback quality (RQ1a: favorability; RQ1b: specificity; RQ1c: goal content; and RQ1d: feedback length)?

Section 2: Predictor Variables and Narrative Feedback Quality

Understanding the contextual factors related to narrative feedback quality is also an important endeavor. In particular, by discovering the characteristics of the rater and ratee that are associated with higher quality narrative feedback, practitioners may be able to maximize the usefulness of the narrative feedback that is provided to the ratee.

Familiarity with the ratee's work behavior. In accordance with the Realistic Accuracy Model (RAM), Funder (1995) suggested that those who are more familiar with the ratee are more likely to be exposed to relevant cues, detect those cues, and refer to them when providing ratings. While the RAM model is intended to describe how people rate others' personality, we see it as a good framework for understanding the behavior of

providing narrative feedback. Raters familiar with the ratee's work will likely be able to recall specific behavioral instances to support their feedback, to create goals that have relevance to the ratee, and to tailor their feedback so that the ratee is likely to accept and act on it. Therefore, it is likely that raters more familiar with the ratee's work behavior are better equipped to provide high quality narrative feedback. Rater selection is very important in multisource feedback systems. There are often many peers and subordinates from whom to choose potential raters. Self-reported familiarity with the ratee's work behavior could be a simple and cost-effective criterion for selecting raters to help ensure that the ratee receives high quality narrative feedback. Based on Funder's propositions and in accordance with the RAM (Funder, 1995), we expect that raters who report being more familiar with the ratee will provide higher quality narrative feedback. Therefore, we propose the following:

Research Question 2: Will the quality of the narrative feedback (RQ2a: favorability; RQ2b: specificity; RQ2c: goal content; RQ2d: feedback length) vary as a function of rater familiarity with the ratee's work behavior?

It is likely that the level of familiarity with the ratee's work behavior differs across rating sources. Supervisors and subordinates are often working with the ratee on a daily basis and may report higher familiarity than peers. This prompted the following research question.

Research Question 3: Do the different rating sources (supervisors, peers, and subordinates) vary on their reported level of familiarity with the ratee's work behavior?

The relationship between rater familiarity with the ratee's work behavior and narrative feedback quality may not be the same for each rating source. Raters prefer to provide feedback anonymously largely due to decreased fear of possible retribution once the feedback has been delivered (e.g., Bracken & Rose, 2011; Nowack & Mashihi, 2012). Thus, anonymous raters may be less afraid to provide constructive criticism or negative feedback should it be warranted. This is especially important for subordinate raters whose outcomes may be dependent on the ratee, their supervisor. Several studies have reported that when subordinates are not assured of anonymity, ratings are more lenient and the raters report that they rated differently than they would have if anonymity was ensured (e.g., Bracken & Rose, 2011; Nowack & Mashihi, 2012). Anonymity is likely to be a less precious commodity for supervisors because they are less vulnerable to revenge by the ratee. Moreover, anonymity is often not feasible in the typical situation of a sole primary supervisor per ratee. Peer raters are likely somewhere in between subordinates and supervisors with regard to their need of anonymity.

However, the nature of narrative feedback may jeopardize the anonymity generally provided in multisource feedback systems. High quality narrative feedback is thought to include specific behavioral examples, which may inadvertently identify the rater to the ratee. Therefore, it is likely that the higher the narrative feedback quality, the more identifiable the rater becomes. As a result, subordinate raters may choose to be less specific and provide less feedback in an attempt to remain anonymous. This would reduce the variability of the indices of narrative feedback quality, resulting in a smaller relationship between familiarity and narrative feedback quality for subordinates. Similarly, peer raters may be affected by the reduction of anonymity that may be

associated with quality narrative feedback. This might be to a lesser extent because peers are less vulnerable than subordinates to the "ratee revenge", however a disgruntled peer has greater potential to influence others who are more powerful than would generally be the case with disgruntled subordinates. Therefore, there is reason to believe that familiarity with the ratee's work behavior may be differently related to the quality of the narrative feedback provided by alternate rating sources, potentially as a function of desire for anonymity and/or fear of reprisals. This is an important avenue for research because methods to select raters who are likely to provide high quality narrative feedback may not be effective for all rating sources, and may actually result in lower narrative feedback quality. Thus, we ask the following question:

> *Research Question 4*: Does the relationship between rater's reported familiarity with the ratee's work behavior, and quality of narrative feedback (RQ4a: favorability; RQ4b: specificity; RQ 4c: goal content; and RQ 4d: feedback length), differ between the rating sources (supervisor, peer, and subordinate)?

Acquaintanceship time. Acquaintanceship time is the amount of time the rater has known the ratee in their current capacity. Whereas acquaintanceship time is likely related to rater familiarity with the ratee's work behavior, it is distinct in that it does not ask specifically about how familiar the rater is with the ratee's work behavior. Similar to familiarity, Funder's (1995) propositions suggest the notion that the longer the rater has been acquainted with the ratee in the rater's current role, the more accurate narrative feedback they should be able to provide. As with the rater's familiarity with the ratee's work behavior, acquaintanceship time could be used to select raters to ensure the ratees are receiving quality narrative feedback. Objectively, acquaintanceship time is easier to

assess than a rater's familiarity with the ratee's work behavior, and may prove to be an expedient proxy for rater familiarity. Therefore, we propose the following:

Research Question 5: Will the quality of the narrative feedback (RQ5a: favorability; RQ5b: specificity; RQ5c: goal content; RQ5d: feedback length) vary as a function of acquaintanceship time?

It is likely that the acquaintanceship time of the rater with the ratee differs across rating sources. On average, supervisors and peer-raters are likely to have known the ratee in a working capacity for longer are therefore likely to have knowledge of more instances of behavior from which to provide feedback than would subordinate raters. This prompted the following research question.

Research Question 6: Do the different rating sources (supervisors, peers, and subordinates) vary on their reported acquaintanceship time with the ratee?

Funder's (1995) propositions regarding the RAM suggest that the more familiar the rater is with the ratee, the more opportunity the rater has likely had to observe the ratee's behavior. Certain rating sources likely have acquaintanceship time and their familiarity with the ratee's work behavior inextricably tied together. For instance, subordinates and supervisors are likely to interact with the ratee regularly. However, peer raters may not interact with the ratee on a regular basis and may interact with them on only a small range of tasks. It is for this reason we ask the following question:

Research Question 7: Does the relationship between acquaintanceship time and quality of feedback (RQ7a: favorability; RQ7b: specificity; RQ7c: goal content; and RQ7d: feedback length) differ between the rating sources (supervisor, peer, and subordinate)?

Position tenure. Whereas the purpose of investigating the relationship between the previous variables (familiarity and acquaintanceship time) and narrative feedback quality was to assist in the selection of raters, the purpose of investigating the relationship between ratee position tenure and narrative feedback quality was to assist in the selection of ratees who are likely to receive high quality narrative feedback. Multisource feedback systems are time consuming and expensive to administer. As a practical concern, it is important to understand which ratees are likely to receive high quality narrative feedback to help reduce lost time and money on uninformative reports.

There is a widespread assumption that employees who have been working in the same position and/or organization for longer are generally better performers than those who have been in the position and/or organization for less amount of time (i.e., Ng & Feldman, 2010). The reason this pervasive assumption persists is twofold. First, employees who have been in the same position for longer amounts of time know how to do their jobs better than those with less experience (Wagner, Ferris, Fandt, & Wayne, 1987). Second, poorer performing employees are likely to experience voluntary or involuntary turnover before they spend longer amounts of time in the position (Schneider, Goldstein, & Smith, 1995). These claims are supported by two theories. Human Capital Theory suggests that long-tenured workers are better performers because they have accumulated more job related knowledge over the course of their careers which is likely to make them better performers (Becker, 1964). Attraction-Selection-Attrition (ASA) theory suggests that person-organization fit increases with tenure (Schneider, et al., 1995). Employees who experience high levels of person-organization fit are likely to perform better because their values match with those of the company's culture and their

skills are a good match to the position's demands (Kristof-Brown, Zimmerman, & Johnson, 2005). ASA theory also suggests that the selection processes operating in the development of employee-organization relationships is mutual. Employees are generally attracted to organizations and positions that reflect their interests. Similarly, organizations tend to hire only those applicants who fit with their conceptualizations of high performers (Bretz, Ash, & Dreher, 1989). Additionally, ASA theory suggests that employees will voluntarily turnover should they perceive a lack of fit, just as organizations will eventually remove employees who do not have the right set of characteristics and skills. Therefore, raters may provide less critical feedback to long tenured ratees under the assumption that those who have spent more time in their current position have garnered the skills and proficiency to do their jobs well. Therefore, we ask the following:

Research Question 8: Will the quality of the narrative feedback (RQ8a: favorability; RQ8b: specificity; RQ8c: goal content; RQ8d: feedback length) vary as a function of ratee position tenure?

Because position tenure is solely a function of the ratee, mean differences between rating sources were not investigated. Position tenure may impact performance behaviors in different ways. As one example, accumulating more experience with a specific role may increase task proficiency on a fairly narrow set of tasks (McEnrue, 1988) and limit the employee's exposure to different and novel methods being used elsewhere. Assuming that employees who have been in a position for a longer period of time are likely higher performers, supervisors will likely be content with their performance and provide less detailed feedback (i.e., ASA theory; Schneider et al., 1995). Subordinate raters, however, are likely newer to the organization. They may be exposed

to novel methods of completing tasks which may be in conflict with the more traditional methods of their supervisor. Therefore, subordinate raters may be best situated to provide high quality narrative feedback because their perspective has been influenced less by organizational norms. It is likely that peer raters fall somewhere in between supervisor and subordinate raters.

Alternatively, subordinate raters may view a long-tenured supervisor in high regard and may be less inclined to provide high quality narrative feedback. Along the same lines, supervisors may see the long-tenured ratee as someone who has become comfortable in their current position and provide high quality narrative feedback to encourage them to develop professionally. Differences in the relationship between ratee position tenure and narrative feedback quality across rating sources may suggest that certain rating sources are better at providing narrative feedback to ratees of different position tenure. This information would allow practitioners to reduce wasted time and money collecting information from sources that are not likely to provide high quality narrative feedback. Therefore, we ask the following question:

> *Research Question 9*: Does the relationship between ratee position tenure and the quality of narrative feedback (RQ9a: favorability; RQ9b: specificity; RQ9c: goal content; and RQ9d: length) differ between the rating sources (supervisor, peer, and subordinate)?

Section 3: Narrative Performance Feedback Content and Feedback Quality

The third line of investigation shifts focus to examine what content is associated with narrative feedback quality. As mentioned, much of the research in the area of performance evaluation focuses on the rating scales and less on the narrative component

of the process (Murphy & Cleveland, 1995). Therefore, the present study draws from the rating scale line of research to ask questions regarding the content of the narrative performance feedback. The first distinction is whether the narrative feedback uses relative or absolute metrics for comparison. Relative feedback content makes use of social comparison, while absolute feedback content makes use of standards and anchors prescribed by the organization to describe the level of performance. We will first discuss relative feedback content, followed by absolute feedback content. The second distinction is whether the narrative feedback content draws the ratee's attention to their behavior, task feedback content, or to their personal characteristics, trait feedback content. Thus, we will discuss task feedback content followed by trait feedback content.

Relative feedback content. One of the most recent developments in the area of performance evaluation is the introduction of relative performance scales (i.e., Goffin, Gellatly, Paunonen, Jackson, & Meyer, 1996). These scales make use of social comparison by evaluating the ratee against a large referent group, likely employees with comparable roles and jobs (Goffin, Jelley, Powell, & Johnston, 2009; Kruglanski & Mayseless, 1990). For instance, the Relative Percentile Method (RPM), asks the raters to provide percentile ratings of the ratee compared to all others in that position (Goffin et al., 1996). The scale may ask the rater to evaluate a fast-food cashier by indicating the percentage of all fast food workers that the employee being rated performs better than. Relative performance scales have been shown to increase the validity and reduce the leniency of ratings (Freund & Kasten, 2012). Extending this line of research to narrative performance feedback, raters may feel inclined to provide feedback that makes use of social comparison to help describe the ratee's level of performance and motivate them to

act on the areas of improvement. An example might include, “Helen is the best leader I have ever had”. Another is “Compared to the managers here at the plant, Joel’s approach to team building could use some work”. In the first example, Helen’s performance was compared to all the leaders the rater has worked with. In the second example, Joel’s performance was compared to all managers at the plant.

Social comparison theory (SCT; e.g., Festinger, 1954; Kruglanksi & Mayseless, 1990) suggests that people continually evaluate themselves because there is value in having accurate assessments of one’s own attributes. Furthermore, when suitable objective criteria for self-evaluation are not available, evaluation takes place through comparisons with others. For instance, if an employee is wondering how well they are performing their job, they are likely to compare their perception of their performance against those with whom they work. Goffin and Olson (2011) suggest that comparative judgements occur naturally and constantly in our day to day lives. Because people are constantly comparing themselves to others, relative performance feedback is likely to be internalized easily and thus acted upon. Therefore, we ask the following question.

Research Question 10: Will the quality of the narrative feedback (RQ10a: favorability; RQ10b: specificity; RQ10c: goal content; RQ10d: feedback length) vary as a function of the amount of relative content?

Should relative feedback content prove to be beneficial with regard to the narrative feedback quality, it will be important to know which sources provide more of it. For this reason we asked the following question.

Research Question 11: Do the different rating sources (supervisors, peers, and subordinates) vary on the amount of relative content in the feedback that they provide?

One of the cornerstones of relative performance methodology is a good representation of the group being used for comparison. Consider the example of a manager of line workers at a manufacturing plant who is having their performance rated and is being compared to all other line worker managers. The line worker manager's supervisor will likely have more exposure to the performance of other line worker managers and should be able to effectively use social comparison in their narrative feedback. Subordinates, however, likely have little exposure to many line worker managers and may use social comparison less frequently and less effectively than their managers. Alternatively, subordinates are likely less familiar with organizational policy regarding performance levels than the ratee's supervisors. Because they may be unsure if the ratee's performance is meeting expectations, subordinates may use social comparison in lieu of understanding organizational performance benchmarks as suggested by SCT. Peer raters are likely in the same or similar position as the ratee. Therefore, the peer rater's performance is also included in the comparison group when providing relative narrative feedback. Peer raters may not provide much relative feedback to keep their own performance from influencing the narrative feedback they provide to the ratee. It is for these reasons we ask the following questions:

Research Question 12: Does the relationship between relative feedback content and narrative feedback quality (RQ12a: favorability, RQ12b: specificity, RQ12c:

goal content, and RQ12d: feedback length) differ between the rating sources (supervisor, peer, and subordinate)?

Absolute Feedback Content. Whereas relative rating scales require that raters compare the ratee to a group using social comparison, absolute methods require the rater to compare the ratee's performance to standards set by the organization. Examples of these standards include anchors such as poor, excellent, satisfactory, meeting expectations, exceeding expectations, etc. When providing narrative feedback to the ratee, raters may feel inclined to use absolute language to describe the level of performance and motivate the ratee to improve. Therefore, we ask the following question.

Research Question 13: Will the quality of the narrative feedback (RQ13a: favorability; RQ13b: specificity; RQ13c: goal content; RQ13d: feedback length) vary as a function of the amount of absolute content?

Should absolute feedback prove to be beneficial with regard to the narrative feedback quality, it will be important to know which sources provide more of it in order to ensure the ratee is receiving the maximal amount of useful feedback. For this reason we asked the following question.

Research Question 14: Does narrative feedback from different rating sources (supervisors, peers, and subordinates) vary on the amount of absolute content provided?

In order to make effective use of absolute feedback content, the rater must be aware of organizational standards and policy. Understanding what the organization deems to be effective or ineffective performance in a specific domain will enable the rater to successfully describe and evaluate the ratee's performance. The level of exposure to

organizational standards and policy is likely to differ between rating sources with supervisors being more exposed, followed by peers and subordinates respectively. Therefore, we ask the following:

Research Question 15: Does the relationship between absolute feedback content and narrative feedback quality (RQ15a: favorability, RQ15b: specificity, RQ15c: goal content, and RQ15d: feedback length) differ between the rating sources (supervisor, peer, and subordinate)?

Task Feedback Content. Task feedback content focuses the ratee's attention on specific behaviors or tasks (Smither & Walker, 2004). Task feedback can be useful to the employee because it directly addresses the behavior that they exhibit. From this information, the employee can adjust their behavior in adherence to the narrative feedback provided which makes it useful in goal setting. An example might be, "Lloyd should develop agendas and disseminate them prior to team meetings". This statement draws the employee's attention to a behavior that he can do to improve his performance. Another example is "Candace does a very good job at managing the patient database". This statement indicates what Candace is doing well so she can continue performing this behavior. Therefore, we ask the following question.

Research Question 16: Will the quality of the narrative feedback (RQ16a: favorability; RQ16b: specificity; RQ16c: goal content; RQ16d: feedback length) vary as a function of the amount of task feedback?

Should task feedback prove to be beneficial with regard to the narrative feedback quality, it will be important to know which sources provide more of it. This led to the following question:

Research Question 17: Does narrative feedback from different rating sources (supervisors, peers, and subordinates) vary on the amount of task content provided?

The effective use of task feedback content is likely to differ between rating sources. A rater's ability to effectively address an employee's behavior and provide specific examples likely depends on a number of factors. The first is the exposure to the relevant behavior that is being addressed. In order to provide specific and detailed task feedback, the rater must be able to draw upon instances of the behavior in question (Funder, 1995). Therefore, rating sources that work more closely with the ratee will likely be able to provide more task feedback. Thus, subordinates, as the recipients of leadership behavior, may be in a good position to provide task feedback. Peer raters may not be as exposed to the ratee's leadership behavior, and therefore less able to provide effective task feedback. The second factor is the rater's behavioral representation of what is good and bad performance which may affect their ability to make effective comparisons (Bernardin, 1979). Supervisors are likely in the best position as they are probably more exposed to organizational performance standards. Along this line of reasoning, peer raters likely have less exposure to organizational performance standards than supervisors, and subordinate raters are likely least exposed. Therefore, based on exposure to ratee behavior and exposure to organizational performance standards, we expected differences between the rating sources on the amount of task feedback provided, which prompted the following questions.

Research Question 18: Does the relationship between task feedback content and narrative feedback quality (RQ18a: favorability, RQ18b: specificity, RQ18c: goal

content, and RQ18d: feedback length) differ between the rating sources (supervisor, peer, and subordinate)?

Trait Feedback Content. Narrative feedback can also bring the ratee's personal traits or characteristics into focus (Smither & Walker, 2004). Trait feedback addresses stable characteristics in the employee and is often perceived as less actionable. For instance, "Susan does not have the leadership ability to properly manage this team". In this scenario, the attention was on Susan's leadership ability and it did not specifically address a task or behavior. Another example could be, "Barry has the intelligence needed to succeed in this position". In this example, Barry was told that he had the capability to succeed in his position due to his intelligence. Because trait feedback content highlights personal characteristics of the employee, which are less actionable, it may be associated with lower narrative feedback quality, however this has yet to be investigated. Therefore, we asked the following:

Research Question 19: Will the quality of the narrative feedback (RQ19a: favorability; RQ19b: specificity; RQ19c: goal content; RQ19d: feedback length) vary as a function of the amount of trait content?

Should trait feedback prove to be beneficial with regard to the narrative feedback quality, it will be important to know which sources provided more of it. Should trait feedback prove not to be beneficial with regard to narrative feedback quality it can be addressed through rater training. This led to the following question:

Research Question 20: Does narrative feedback from different rating sources (supervisors, peers, and subordinates) vary on the amount of trait content provided?

The use of trait feedback is also likely to differ across rating sources. In a similar vein to task feedback content, it could be that those closest to the employee being rated will provide more trait content. Raters close to the ratee are likely to know the ratee personally and thus should be able to provide more nuanced information regarding their characteristics and disposition. However, it could also be that less familiar ratees use trait feedback content to describe the employee because they are less familiar with the specific behaviors they exhibit. In other words, because they cannot comment on specific behaviors of the employee, they rely on global comments regarding their personality or ability. Therefore, it is unclear which rating source will provide more trait feedback content and how trait feedback content relates to indices of feedback quality for the different rating sources.

Research Question 21: Does the relationship between trait feedback content and narrative feedback quality (RQ21a: favorability, RQ21b: specificity, RQ21c: goal content, and RQ21d: feedback length) differ between the rating sources (supervisor, peer, and subordinate)?

Method

Participants

An archival database of responses to a commercial multisource performance rating instrument was used for this study. This instrument asked raters to provide numeric performance feedback on 51 leadership behaviors which are grouped into four factors: cognitive managerial skills; interpersonal managerial skills; personal managerial skills; and teamwork, supervision, planning, and productivity (i.e., O'Neill, McLarnon, & Carswell, 2015). Factor descriptions and select leadership behavior descriptions can be

found in Appendix A. Ratees were leaders and managers from a range of industries including manufacturing, healthcare, finance, and information technology. Raters included the ratees' supervisors, peers and subordinates. The data were collected as part of development and succession planning initiatives (i.e., not for formal administrative decision making). Aside from providing numeric ratings, the raters were also asked to provide narrative feedback for the leadership behaviors as they saw fit. Therefore, it was not necessary for raters to provide narrative feedback for any or all leadership behaviors. The database contained 171,531 narrative comments for 4,385 ratees.

We chose to narrow the scope of our investigation to managers and directors. Attempts were made to code all the comments provided to the 2,123 ratees who had director or manager in their job title using the Linguistic Inquiry and Word Count software (LIWC2015; Pennebaker, Booth, Boyd, & Francis, 2015). Due to the complexity of the pertinent variables and the inconsistent and often short length of the comments provided, the use of the software was not successful. Therefore, we chose to code the narrative comments manually. We selected a random sample of 200 ratees from the 2,123 ratees who had manager or director in their job title. The sample of 200 ratees seemed appropriate because the cost associated with hiring and training additional research assistants would have been excessive. Of the 200 ratees, 111 were male, 65 were female, and 24 did not provide information regarding their gender. Due to confidentiality concerns, no information on age or ethnicity was collected. The final sample consisted of 63,423 ratings with 8,967 associated narrative comments. Thus, the narrative feedback response rate was 12.4 percent. Broken down by rating source, supervisors had a response rate of 23.7 percent (7,653 ratings with 2,377 comments), peers had a response

rate of 10.5 percent (26,924 ratings with 3,159 comments), and subordinates had a response rate of 10.6 percent (28,846 ratings with 3,431 comments).

Of the 8,967 narrative comments, 1,085 were removed because they did not contain feedback regarding the ratee's performance. This included statements such as "not applicable" or "I am not in a good position to be providing feedback on this dimension". Other comments were removed because the rater made a mistake. For instance, some wrote the numeric rating in the narrative feedback field during the assessment. Each of the narrative comments was coded by four research assistants. Comments were removed if two or more of the four research assistants coding the comment agreed that it was not a comment. If only one research assistant thought it was not a comment, that research assistant's ratings were removed and that comment was judged based on the remaining three research assistants. In summary, we studied a total of 7,882 comments. Each comment was associated with one of 200 ratees and each comment was coded by 4 research assistants.

Narrative Feedback Quality Measures

The indices of narrative feedback quality (favorability, specificity, goal content, and length) were adapted from David's (2013) measure of narrative feedback quality. These scales were coded by trained research assistants using the procedure described later in this section.

Favorability. Favorability was defined as the degree to which the feedback was positive and reflected well on the ratee, or negative and tended to focus on the shortcomings of the ratee (David, 2013). It was measured using a 5-point Likert scale

where 1 was extremely unfavorable, 3 was neutral, and 5 was extremely favorable (see Appendix B).

Specificity. Specificity was defined as the degree to which the feedback provided was detailed and supported by behavioral examples (David, 2013). It was measured using a 5-point Likert scale where 1 was nonspecific, 3 was moderately specific, and 5 was extremely specific (See Appendix B).

Goal Content. Goal content was defined as the degree to which the rater provided the ratee with actionable steps to improve performance (David, 2013). It was measured using a 5-point Likert scale where 1 was no goal content, 3 was a moderate amount of goal content, and 5 was a large amount of goal content (See Appendix B).

Length. Narrative feedback length was operationalized as the total number of words in the narrative feedback and was measured electronically.

Rater-Ratee Relationship Variables

These rater-ratee relationship variables were a part of the commercial multisource feedback tool used in the present study. As such, the rater completed these variables when they completed the instrument itself.

Familiarity with the Ratee's Work Behavior. Rater familiarity was measured using a single 7-point Likert scale item asking the rater to indicate "How well are you acquainted with the work behavior of the target?" where 1 was not at all, 4 was moderately well, and 7 was extremely well. This information was collected from the raters at the time they completed the leadership assessment.

Acquaintanceship Time. Acquaintanceship time was measured using a single item asking the rater "Please indicate how long you've known the target in your current

capacity". A 6-point scale with the following response options was used: 1 (less than 6 months), 2 (6 months to less than 1 year), 3 (1 year to less than 2 years), 4 (2 years to less than 5 years), 5 (5 years to less than 10 years), and 6 (10 years or more). This information was collected from the raters at the time they completed the leadership assessment.

Ratee-Reported Variable

The ratee-reported variable was a part of the commercial multisource feedback tool used in the present study. As such, the ratee completed this variable when they completed the instrument itself.

Position Tenure. Position tenure was measured using a single item asking the ratee "Please indicate how long you've been in your current position". A 6-point scale with the following response options was used: 1 (less than 6 months), 2 (6 months to less than 1 year), 3 (1 year to less than 2 years), 4 (2 years to less than 5 years), 5 (5 years to less than 10 years), and 6 (10 years or more). This information was collected from the ratees at the time they completed the leadership assessment.

Narrative Feedback Content Variables

The narrative feedback content variables were also coded by trained research assistants. The coding and training procedures are addressed next.

Relative feedback content. Relative feedback content was defined as the extent to which the rater provided feedback that made use of social comparison. It was measured using a 5-point Likert scale where 1 was no relative content, 3 was a moderate amount of relative content, and 5 was a large amount of relative content (See Appendix B).

Absolute feedback content. Absolute feedback content was defined as the extent to which the rater provided feedback that made use of adjective descriptors to indicate performance level. It was measured using a 5-point Likert scale where 1 was no absolute content, 3 was a moderate amount of absolute content, and 5 was a large amount of absolute content (See Appendix B).

Task feedback content. Task feedback content was defined as the extent to which the rater provided feedback that made reference to specific behaviors and tasks. It was measured using a 5-point Likert scale where 1 was no task content, 3 was a moderate amount of task content, and 5 was a large amount of task content (See Appendix B).

Trait feedback content. Trait feedback content was defined as the extent to which the rater provided feedback that made reference to personal qualities of the ratee. It was measured using a 5-point Likert scale where 1 was no trait content, 3 was a moderate amount of trait content, and 5 was a large amount of trait content (See Appendix B).

Narrative Feedback Quality Coding Procedure

In the present study we used a deductive approach to qualitative analysis as suggested by Elo and Kyngas (2008). First, we identified the item-level comment as the unit of analysis. An item-level comment is the narrative feedback that an individual rater provided based on one of the 51 dimensions of leadership performance. Examples of individual narrative comments can be found in Appendix C. Raters were not required to provide comments for any or all of the performance dimensions.

The second step, as outlined by Elo and Kyngas (2008), was to code the narrative comments according to the categories and codes. Prior to coding the entire data set, the four graduate students coded a sample of 100 item-level comments to ensure an adequate

level of inter-rater reliability. The graduate students all had research experience in the area of performance evaluation and were familiar with the variables being coded. The graduate students were trained by acquainting them with the narrative feedback quality variables and scales. The graduate students went through the same sample of 100 narrative comments and coded them for narrative feedback quality. Once this was completed, inter-rater reliability statistics were calculated and differences in ratings were discussed. The Cronbach's alpha for each of the variables were as follows: favorability was .897, specificity was .839, goal content was .901, relative content was .933, and absolute content was .814. The interactional justice scale had a Cronbach's alpha of .842, however it was highly correlated with favorability (r = .949, p <.001). Thus, it was removed to reduce redundancy. Favorability was retained because it demonstrated the largest effects in David's (2013) work. These findings suggested that we should continue with the coding procedure which will be discussed next.

We used paid research assistants to code the data for the present study. We hired ten third and fourth year students in linguistics as research assistants. We believed that their knowledge of language was an asset in rating the narrative performance feedback. We had four research assistants coding each item-level comment. With 8,967 comments in total, this was a large endeavor. Research assistants were brought in for 3 hour sessions which occurred four times a week. The number of sessions each coder attended per week varied according to the research assistants' schedules and availability. The coding process lasted 12 weeks, and totaled 376 research assistant hours. Coding took place in a private room and was supervised by the author. The author was present during each session to organize and manage the data set, and to answer any questions the research assistants had

during the session. The research assistants were provided with a dataset at the beginning of each session, and gave that data set back to the author at the end of each session in order to maintain security over the data.

The research assistants received rater training following the principles of Frame of Reference training (FOR; Bernardin, 1979). The purpose of FOR training was to help coders adopt the same metric when it came to providing ratings by reducing idiosyncrasies in raters' conceptualization and operationalization of the constructs being measured. This will be further discussed below.

Research Assistant Training

As mentioned, the research assistant training was based on the principles of FOR Training (Bernardin, 1979). Two training sessions were offered to a total of 13 research assistant applicants, of which 10 were retained. Each session lasted two hours, and each applicant could choose which session worked according to their schedule. The research assistants were given a training package which included the training slides so they could follow along and use as a reference when coding. Each session began with a description of the study and why it was important. Following the introduction, the research assistants were introduced to the seven variables they would be coding: favorability, specificity, goal content, relative content, absolute content, task content, and trait content. Each variable was then expanded upon individually. This involved a definition of the variable, an introduction to the scale used to code it, and examples of narrative feedback to work through as a group. When it came to the examples, a written comment taken from the data set was put up on the projector along with the rating scale used to measure the specific variable in question. Research assistants were asked to record how they would

code that comment. Once everyone had coded the comment, they were encouraged to share their rating and how they decided on that rating with the group at large. Once the discussion was over, the expert ratings of each example, as provided by the author and three other graduate students with knowledge of the area, were shared and words deemed important to the variable of interest were highlighted in the example. The variables favorability, specificity, and goal content each had two examples. Relative and absolute content were introduced and discussed together, as were task and trait content. These variables were more complex and more difficult to code, thus more examples were provided. Relative and absolute content had nine examples, while task and trait content had five.

Following the training, research assistants were asked to code a sample of 100 item-level narrative comments. The 100-item measure was used to show inter-rater reliability of the research assistants. The Cronbach's alpha for each of the variables were as follows: favorability was .966, specificity was .928, goal content was .950, relative content was .965, absolute content was .889, task content was .879, and trait content was .919. This was used as an indication that the training had been adopted, that the research assistants had a similar approach to rating the variables in question, and that we should continue coding the dataset in its entirety.

Statistical Analyses

All analyses were run using random intercepts mixed models in SPSS. Because the item-level comment was the unit of interest, this statistical procedure seemed ideal as it allowed us to control for the ratee while investigating the qualities of individual comments. We could not control for the rater due the necessary anonymity associated

with multisource ratings which made it impossible to track raters across ratees. Thus, this analysis accounted for the dependencies associated with the ratee in the dataset. The only exception to this was the ratee's position tenure. As this is a ratee level variable, the effect would not be detected if we also controlled for the ratee.

As mentioned, our analyses clustered at the level of the ratee. We did this because the item-level comment was the unit of interest and the ratees received an inconsistent number of comments. In order to justify clustering at the ratee level, several models were estimated in order to demonstrate the amount of variance accounted for by the ratee for each of the outcome variables. The intraclass correlation (ICC)(1) value for favorability was .187 suggesting that the ratee accounted for 18.7 percent of the variance. The ICC(1) value for specificity was .147 suggesting that the ratee accounted for 14.7 percent of the variance. The ICC(1) for goal content was .155 suggesting that the ratee accounted for 15.5 percent of the variance. Finally, the ICC(1) value for feedback length was .156 suggesting that the ratee accounted for 15.6 percent of the variance. These values serve as indications that clustering at the ratee level, and thus controlling for this shared variance, was justified.

The predictor variables were assessed through simple slopes analyses with the predictor as the fixed effect for the overall models (familiarity, acquaintanceship time, position tenure, relative content, absolute content, task content, and trait content), and the predictor and rating source interaction term as the fixed effect for the moderation models (i.e. predictor*rating source). The variables were standardized to assist with interpretability of the results. Therefore, each slope can be interpreted similar to a partial

correlation. The intercept indicates the value of the dependent variable when the independent variable is zero.

Rather than using effect coding for the interaction effects, the rating source variable was added to the model as a categorical factor with three levels. The three levels being supervisors, peers, and subordinates. The rating source and predictor interaction term was then added to the model as a fixed effect. This allowed for the estimation of a common intercept and unique slopes for each of the three rating sources. The interaction effects have an associated test of significance which is also reported.

Effect Size Interpretation

In order to assist in the interpretability of the results, each of the relationships investigated have the associated measure of effect size reported. There has been a recent push in the literature for the inclusion of effect size metrics (Aguinis, Werner, Abbott, Angert, Park, & Kohlhausen, 2010). Furthermore, researchers have recognized that small effect sizes may be of practical significance, and that the cut-offs used to categorize effect size metrics are subjective (i.e., Aguinis & Harden, 2009; Cortina & Landis, 2009). In order to address these concerns we used the following guidelines to ensure that small effect sizes were not discarded, and that our categories of effect size strength differentiated between the observed relationships. With regard to research questions involving a comparison of means we used the guidelines for Cohen's d outlined by Cohen (1988). Therefore, a Cohen's d value of .2 was a *small effect*, a value of .5 was a *medium effect*, and a value of .8 was a *large effect*. Additionally, we classified significant mean comparisons with a Cohen's d value with a value of .1 as an *approaching-small effect*.

Similarly, we adapted Cohen's (1988) guidelines for the effect size of correlations for the research questions involving an investigation of relationship strength. We believed that Cohen's guidelines were not specific enough to differentiate between the observed effect sizes. Therefore, we supplemented Cohen's (1988) guidelines with three additional categories. The guidelines for effect size of the observed partial correlations are as follows: .05 was an *approaching-small effect,* .1 was a *small effect,* .2 was an *approaching-medium effect,* .3 was a *medium effect,* .4 was an *approaching-large effect,* and a .5 was a *large effect.*

Results

Section 1: Overall Narrative Feedback Outcome Comparisons

The means and standard errors relevant to Research Question 1 can be found in Table 1. Research Question 1a asked whether there were differences in the mean level of favorability of the narrative feedback provided by the different rating sources. The results indicated that supervisors provided more favorable feedback than subordinates ($t(30864) = 3.448$, $p = .001$, $d = .0393$) but the comparison of supervisors and peers did not reach significance ($t(31181) = 1.480$, *ns*, $d = .0168$). The comparison between peer and subordinate raters indicated that peers provided more favorable feedback ($t(31305) = 2.155$, $p = .031$, $d = .0244$). The effect sizes of the comparisons in Research Question 1a were very small indicating consistency in the favorability of the narrative feedback provided across rating sources.

Research Question 1b asked whether there were differences in the mean level of specificity of the feedback provided by the different rating sources. The results indicated that supervisors provided more specific feedback than peers ($t(30719) = 3.263$, $p = .001$,

Table 1. Means and Standard Errors of Outcome Variables by Rating Source

	Supervisor				Peer				Subordinate			
	Raw		Standardized		Raw		Standardized		Raw		Standardized	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Favorability	3.635	0.0316	0.0292	0.0341	3.614	0.0314	0.00653	0.0339	3.585	0.0315	-0.025	0.034
Specificity	2.507	0.0338	-0.0314	0.0311	2.452	0.0335	-0.0808	0.0309	2.497	0.0336	-0.0406	0.0297
Goal Content	1.475	0.0264	0.0552	0.0311	1.359	0.0262	-0.0813	0.0309	1.35	0.0263	-0.0927	0.031
Length	15.368	0.44	-0.0508	0.03	13.194	0.437	-0.199	0.0298	14.724	0.438	-0.0947	0.0299

Note: All scales ranged from 1-5 with the exception of feedback length which was the number of words in the narrative feedback provided

d = .0372) but the comparison of supervisors and subordinates did not reach significance (*t*(30194) = .591, *ns*, *d* = .00680). The comparison between peer and subordinate raters indicated that subordinates provided more specific narrative feedback (*t*(30881) = 2.789, *p* = .005, *d* = .0159). The effect sizes of the comparisons in Research Question 1b were very small indicating consistency in the specificity of the narrative feedback provided across rating sources.

Research Question 1c asked whether there were differences in the mean level of goal content in the feedback provided by the different rating sources. The results indicated that supervisors provided more goal content than peers (*t*(30756) = 8.827, *p* < .001, *d* = .101) and subordinates (*t*(30243) = 9.334, *p* < .001, *d* = .107). The comparison between peer and subordinate raters did not reach significance (*t*(30922) = 0.772, *ns*, *d* = .00878). The effects sizes for Research Question 1c indicated that the comparison of goal content of supervisors and peers, as well as supervisors and subordinates resulted in approaching-small effects.

Research Question 1d asked whether there were differences in the mean length of the feedback provided by the different rating sources as indicated by word count. The results indicated that supervisors provided longer feedback than peers (*t*(30827) = 9.964, *p* < .001, *d* = .114) and subordinates (*t*(30350) = 2.883, *p* = .004, *d* = .0331). The comparison between peer and subordinate raters indicated that subordinates provided longer feedback (*t*(30982) = 7.355, *p* < .001, *d* = .0836). The effects sizes for Research Question 1d indicated that the comparison of feedback length for supervisors and peers resulted in an approaching-small effect.

Section 2: Predictor Variables and Narrative Feedback Quality

The research questions involving the predictor variables assessed through simple slopes analyses. As mentioned, the variables were standardized. Therefore each slope can be interpreted similar to a partial correlation. The intercept indicates the value of the dependent variable when the independent variable is zero. Because we are testing many models we are reporting only the slopes and the slopes' significance in the results section as they are the most pertinent to the research questions. Furthermore, a summary table outlining the largest effects for each predictor variable is included for interpretability.

Familiarity. Research Question 2 asked whether the narrative feedback quality would vary as a function of rater familiarity with the ratee's work behavior. A summary of the largest effects for familiarity can be found in Table 2. The models used to investigate this research question used data from all rating sources. The intercepts and slopes pertaining to Research Question 2 can be found in Table 3. Research Question 2a asked whether the favorability of the associated feedback would vary as a function of rater familiarity. The slope of favorability on rater familiarity was -.0331 ($p < .001$) indicating that as rater familiarity increased, the favorability of the narrative feedback decreased. Research Question 2b asked whether the specificity of the feedback would vary as a function of rater familiarity. The slope of specificity on rater familiarity was .0236 ($p < .001$) indicating that as rater familiarity increased, the specificity of the narrative feedback increased. Research Question 2c asked whether the amount of goal content in the feedback would vary as a function of rater familiarity. The slope of goal content on rater familiarity was .0619 ($p < .001$), indicating that as rater familiarity increased, the amount of goal content in the narrative feedback increased. Finally, Research Question 2d asked whether the feedback length would vary as a function of

Table 2. Summary of Effect Sizes of Relationships for Familiarity

		Narrative Feedback Quality Variables		
Group	Favorability	Specificity	Goal Content	Feedback Length
Overall			Approaching-small positive	
Supervisors	Small negative		Approaching-medium positive	
Peers		Approaching-small positive		Small positive
Subordinates				

Table 3. Slopes and Standard Errors for Familiarity and Outcome Variables

	Intercept	SE	Sig.	Slope	SE	Sig.
Favorability	.00380	.0330	*ns*	-.0331	.00623	<.001
Specificity	-.0511	.0284	*ns*	.0236	.00614	<.001
Goal Content	-.0397	.0301	*ns*	.0619	.00628	<.001
Length	-.115	.0289	<.001	.0437	.00604	<.001

rater familiarity. The slope of feedback length on rater familiarity was .0437 ($p < .001$) indicating that as rater familiarity increased, the length of the narrative feedback increased. Overall, the results for Research Question 2 suggested that although the relationships were very small, rater familiarity with the ratee's work behavior was associated with narrative feedback that was less favorable, but more specific, contained more goal content, and lengthier. The strongest relationship was found for goal content which met the criteria for an approaching-small effect.

Research Question 3 asked whether there were differences in the mean familiarity with the ratee's work behavior between the different rating sources. The means and standard errors can be found in Table 4. The results indicated that supervisors were more familiar than peers ($t(31689) = 49.662$, $p < .001$, $d = .558$) and subordinates ($t(31695) = 6.407$, $p < .001$, $d = .0720$). Furthermore, subordinate raters were more familiar than peers ($t(31680) = 45.247$, $p < .001$, $d = .508$). Therefore, supervisors were the most familiar with the work behavior of the ratee, followed by subordinates and peers respectively. The comparison of supervisors and peers, and the comparison of subordinates and peers resulted in medium effect sizes.

Research Question 4 asked whether the relationship between familiarity with the ratee's work behavior and narrative feedback quality would vary between rating sources. This was assessed by estimating a model with a common intercept, but different slopes for each of the rating sources. As mentioned, a summary of the largest effects for familiarity can be found in Table 2. The intercepts and slopes pertaining to this research question can be found in Table 5. Research Question 4a asked whether the relationship of rater familiarity and the favorability of the narrative feedback would vary between rating

Table 4. Means and Standard Errors of Additional Variables by Rating Source

	Supervisor				Peer				Subordinate			
	Raw		Standardized		Raw		Standardized		Raw		Standardized	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Familiarity	5.73	0.0692	0.247	0.0518	4.856	0.0691	-0.407	0.0517	5.615	0.0691	0.161	0.0517
Acquaintanceship	3.97	0.0695	0.0699	0.0593	3.92	0.0695	0.0293	0.0593	3.63	0.0695	-0.219	0.0593
Relative	1.119	0.0146	0.042	0.0317	1.099	0.0146	-0.0115	0.0374	1.107	0.146	0.0103	0.0375
Absolute	1.845	0.0243	0.0187	0.0278	1.777	0.0241	-0.0588	0.0276	1.814	0.0242	-0.0162	0.0277
Task	2.07	0.0256	-0.0428	0.0243	2.072	0.0254	-0.0900	0.024	2.021	0.0253	-0.0409	0.0241
Trait	1.464	0.0209	-0.0547	0.0233	1.529	0.0206	0.0171	0.0229	1.495	0.0207	-0.0205	0.0231

Note: The scale for Familiarity ranged from 1–7. The scale for Acquaintanceship Time ranged from 1-6. The scales of Relative, Absolute, Task, and Trait ranged from 1–5.

Table 5. Slopes and Standard Errors for Familiarity and Outcome Variables by Rating Source

				Supervisor			Peer			Subordinate		
	Intercept	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.
Favorability	.0152	.0330	*ns*	-.111	.0130	<.001	-.0226	.0102	.027	-.0008	.101	*ns*
Specificity	-.0474	.0285	*ns*	.0238	.0128	*ns*	.0615	.0101	<.001	-.0134	.00994	*ns*
Goal Content	-.0707	.0298	.019	.254	.0130	<.001	.00808	.0103	*ns*	.00837	.0101	*ns*
Length	-.106	.0291	<.001	.0283	.0126	.024	.112	.00993	<.001	-.0139	.00977	*ns*

sources. The intercepts and slopes pertaining to this research question can be found in Table 5. A test of the interaction between rating source and rater familiarity was significant, $F(3, 30577) = 25.727$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors (-.111) was significant ($p < .001$), as was that for peers (-.0226; $p = .027$). However, the slope for subordinates was not found to be significant (-.0008; *ns*). Overall, it appears that as rater familiarity increased, the favorability of the narrative feedback decreased for supervisors and peers, but not for subordinates. Furthermore, the relationship for supervisors reached a small effect size.

Research Question 4b asked whether the relationship between familiarity with the ratee's work behavior and specificity would vary between the rating sources. A test of the interaction between rating source and rater familiarity was significant, $F(3, 29798) = 13.749$, $p = .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors (.0238) was not found to be significant (*ns*), neither was the slope for subordinates (-.0134; *ns*). The slope for peers was found to be significant, however (.0615; p <.001) Overall, it appears that as rater familiarity increased, the specificity of the narrative feedback increased (approaching-small effect) for peers, but not for supervisors or subordinates.

Research Question 4c asked whether the relationship between familiarity with the ratee's work behavior and goal content would vary between the rating sources. A test of the interaction between rating source and rater familiarity was significant, $F(3, 29881) = 127.845$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors (.254) was significant ($p < .001$). However, the slopes

were not found to be significant for peers (.00808; *ns*) nor subordinates (.00837; *ns*). Overall, it appears that as rater familiarity increased, the goal content of the narrative feedback increased for supervisors with an approaching-medium effect size.

Research Question 4d asked whether the relationship between familiarity with the ratee's work behavior and feedback length would differ between rating sources. A test of the interaction between rating source and rater familiarity was significant, $F(3, 30047) = 43.789$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors (.0283) was significant ($p = .024$), as was the slope for peers (.112; $p < .001$). However, the slope for subordinates was not found to be significant (-.0139; *ns*). Overall, it appears that as rater familiarity increased, the length of the narrative feedback increased most noticeably for peers (small effect size).

Taken together, the results for Research Question 4 suggest that the relationship between familiarity and the quality of narrative feedback was generally positive for supervisors and peers with the exception of favorability. The most notable findings for supervisors included a small negative effect for familiarity and an approaching-medium positive effect for goal content. The most notable relationships for peers included an approaching-small positive effect for specificity and a small positive effect for feedback length. The results did not indicate any notable relationships for subordinates.

Acquaintanceship Time. Research Question 5 asked whether the quality of the narrative feedback would vary as a function of acquaintanceship time. A summary of the largest effects for acquaintanceship time can be found in Table 6. The intercepts and slopes pertaining to this research question can be found in Table 7. Research Question 5a asked whether the favorability of the feedback would vary as a function of

Table 6. Summary of Effect Sizes of Relationships for Acquaintanceship Time

		Narrative Feedback Quality Variables		
Group	Favorability	Specificity	Goal Content	Feedback Length
Overall				Approaching-small negative
Supervisors		Approaching-small negative		
Peers				Approaching-small negative
Subordinates		Approaching-small negative		Approaching-small negative

Table 7. Slopes and Standard Errors for Acquaintanceship Time and Outcome Variables

	Intercept	SE	Sig.	Slope	SE	Sig.
Favorability	.00403	.0328	*ns*	.00631	.00684	*ns*
Specificity	-.0526	.0285	*ns*	-.0414	.00673	<.001
Goal Content	-.0394	..0300	*ns*	.00303	.00689	*ns*
Length	-.117	.0287	<.001	-.0513	.00662	<.001

acquaintanceship time. The slope of favorability on acquaintanceship time was not found to be significant (.00631; *ns*). Research Question 5b asked whether the specificity of the feedback would vary as a function of acquaintanceship time. The slope of specificity on acquaintanceship time was -.0414 ($p < .001$) indicating that as acquaintanceship time increased, the specificity of the narrative feedback decreased. Research Question 5c asked whether the amount of goal content provided would vary as a function of acquaintanceship time. The slope of goal content on acquaintanceship time was not significant (.00303; *ns*). Research Question 5d asked whether feedback length would vary as a function of acquaintanceship time. The slope of feedback length on acquaintanceship time was significant (-.0513; $p < .001$) indicating that as acquaintanceship time increased, the length of the narrative feedback decreased. Taken together, these results indicated that acquaintanceship time is either unrelated or negatively related to narrative feedback quality. Only feedback length resulted in an effect size that was approaching-small.

Research Question 6 asked whether there were differences in the reported acquaintanceship time between the different rating sources. The means and standard errors can be found in Table 4. The results indicate that supervisors had more acquaintanceship time with the ratee than peers ($t(31657) = 3.345$, $p = .001$, $d = .0376$) and subordinates ($t(31671) = 22.685$, $p < .001$, $d = .255$). Furthermore, peers had more acquaintanceship time with the ratees than subordinates ($t(31642) = 20.929$, $p < .001$, $d = .235$). Therefore, supervisors had the longest acquaintanceship time with the ratee, followed by peers and subordinates. The comparison of supervisors and subordinates, and the comparison of peers and subordinates resulted in small effect sizes.

Research Question 7 asked whether the relationship between acquaintanceship time and narrative feedback quality differed between rating sources. This was assessed by estimating a model with a common intercept, but different slopes for each of the rating sources. As mentioned, a summary of the largest effects for acquaintanceship time can be found in Table 6. The intercepts and slopes pertaining to this research question can be found in Table 8. Research Question 7a asked whether the relationship of acquaintanceship time and favorability would differ between the rating sources. A test of the interaction between rating source and acquaintanceship time was not significant, $F(3, 29513) = 1.221$, $p = .30$. Furthermore, the slope for supervisors was not significant (-.00331; *ns*), neither were those for peers (-.00401; *ns*) nor subordinates (.0167; *ns*). Overall, the relationship between acquaintanceship time and favorability was not found to be significant across rating sources, suggesting consistency.

Research Question 7b asked whether the relationship of acquaintanceship time and specificity would differ between the rating sources. A test of the interaction between rating source and acquaintanceship time was significant, $F(3, 28341) = 14.993$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was significant (-.0522; $p < .001$), as was that for subordinates (-.0524; $p < .001$). However, the slope for peers was not found to be significant (-.0149; *ns*). Overall, as acquaintanceship time increased, the specificity of the narrative feedback decreased for both supervisors and subordinates, both of which had an approaching-small effect size.

Research Question 7c asked whether the relationship of acquaintanceship time and the amount of goal content would differ between the rating sources. A test of the

Table 8. Slopes and Standard Errors for Acquaintanceship Time and Outcome Variables by Rating Source

				Supervisor			Peer			Subordinate		
	Intercept	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.
Favorability	.00523	.0329	*ns*	-.00331	.0130	*ns*	-.00401	.0121	*ns*	.0167	.00925	*ns*
Specificity	-.0540	.0285	*ns*	-.0522	.0127	<.001	-.0149	.0120	*ns*	-.0524	.00911	<.001
Goal Content	-.0369	.0301	*ns*	-.0308	.0131	.018	-.00776	.0122	*ns*	.0244	.00934	.009
Length	-.118	.0289	<.001	-.0153	.0125	*ns*	-.0606	.0118	<.001	-.0616	.00897	<.001

interaction between rating source and acquaintanceship time was significant, $F(3, 28480) = 4.920$, $p = .002$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was significant and negative (-.0308; $p = .018$). The slope for subordinates was also found to be significant but positive (.0244; $p = .009$). However, the slope for peers was not found to be significant (-.00776; *ns*). Therefore, a significant negative relationship was found for supervisors and a significant positive relationship was found for subordinates. This may account for why the overall relationship in Research Question 5c was not found to be significant. However, neither of the effect sizes for these relationships were particularly noteworthy.

Research Question 7d asked whether the relationship of acquaintanceship time and feedback length would differ between the rating sources. A test of the interaction between rating source and acquaintanceship time was significant, $F(3, 28671) = 23.856$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was not found to be significant (-.0153; *ns*). However, the slopes were found to be significant for both peers (-.0606; $p < .001$) and subordinates (-.0616; $p < .001$). Overall, the results demonstrate that as acquaintanceship time increased, the length of the narrative feedback decreased for peers and subordinates. The effect size for these relationships was approaching-small.

Taken together, the results of Research Question 7 suggest that the relationship between acquaintanceship time and narrative feedback quality was negative in nature for all rating sources. Supervisors had an approaching-small negative effect for specificity, peers had an approaching-small negative relationship for feedback length, and

subordinates had an approaching-small negative relationship for both specificity and feedback length.

Position Tenure. Research Question 8 asked whether the quality of the narrative feedback provided would vary as a function of ratee position tenure. A summary of the largest effects for position tenure can be found in Table 9. The intercepts and slopes pertaining to this research question can be found in Table 10. Research Question 8a asked whether the favorability of the feedback would vary as a function of ratee position tenure. The slope of favorability on ratee position tenure was found to be significant (-.0586; p <.001), indicating that as ratee position tenure increased, the favorability of the narrative feedback received decreased. The effect size of position tenure and favorability was approaching-small. Research Question 8b asked whether the specificity of the feedback would vary as a function of ratee position tenure. The slope of specificity on ratee position tenure was also found to be significant (-.111; $p < .001$) indicating that as ratee position tenure increased, the specificity of the narrative feedback decreased. Furthermore, the effect size of the relationship was small. Research Question 8c asked whether the amount of goal content provided would vary as a function of ratee position tenure. The slope of goal content on ratee position tenure was also found to be significant (-.0222; p <.001). Therefore, the amount of goal content decreased as ratee position tenure increased. Research Question 8d asked whether the feedback length would vary as a function of ratee position tenure. The slope of feedback length on ratee position tenure was found to be significant as well (-.117; $p < .001$), another small effect size. The results of Research Question 8d indicated that as position tenure increased, the length of the narrative feedback decreased. Taken together, the narrative feedback quality decreased as

Table 9. Summary of Effect Sizes of Relationships for Position Tenure

	Narrative Feedback Quality Variables			
Group	Favorability	Specificity	Goal Content	Feedback Length
Overall	Approaching-small negative	Small negative		Small negative
Supervisors		Approaching-small negative	Small negative	Small negative
Peers	Small negative	Small negative	Approaching-small positive	Small negative
Subordinates		Small negative		Small negative

Table 10. Slopes and Standard Errors for Tenure in Position and Outcome Variables

	Intercept	SE	Sig.	Slope	SE	Sig.
Favorability	.000	.00561	*ns*	-.0586	.00561	<.001
Specificity	.000	.00558	*ns*	-.111	.00558	<.001
Goal Content	.000	.00562	*ns*	-.0222	.00562	<.001
Length	.000	.00558	*ns*	-.117	.00558	<.001

ratee position tenure increased. The most notable relationships included small negative effects for specificity and length, and an approaching-small negative effect for favorability.

Research Question 9 asked whether the relationship between ratee position tenure and narrative feedback quality differed between the different rating sources. This was assessed by estimating a model with a common intercept, but different slopes for each of the rating sources. As mentioned, a summary of the largest effects for position tenure can be found in Table 9. The intercepts and slopes pertaining to this research question can be found in Table 11. Research Question 9a asked whether the relationship of ratee position tenure and favorability would differ between the rating sources. A test of the interaction of rating source and ratee position tenure was significant, $F(3, 31694) = 47.519$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (-.0420; $p < .001$), as were those for peers (-.101; $p < .001$) and subordinates (-.0294; $p < .001$). Overall, as ratee position tenure increased, the favorability of the narrative feedback decreased. This effect was the largest for peers with a small effect size.

Research Question 9b asked whether the relationship of ratee position tenure and specificity differed between the rating sources. A test of the interaction between rating source and ratee position tenure was significant, $F(3, 31694) = 141.311$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (-.0643; $p < .001$), as were those for peers (-.125; $p < .001$) and subordinates (-.132; $p = .001$). Overall, as ratee position tenure increased, the specificity of the narrative feedback decreased. The effect size of negative

Table 11. Slopes and Standard Errors for Tenure in Position and Outcome Variables by Rating Source

				Supervisor			Peer			Subordinate		
	Intercept	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.
Favorability	-.00019	.00561	*ns*	-.0420	.0109	<.001	-.101	.00933	<.001	-.0294	.00918	.001
Specificity	.00109	.00558	*ns*	-.0643	.0108	<.001	-.125	.00928	<.001	-.132	.00913	<.001
Goal Content	-.00142	.00561	*ns*	-.114	..0109	<.001	.0560	.00932	<.001	-.0327	.00917	<.001
Length	-.00053	.00558	*ns*	-.135	.0108	<.001	-.118	.00928	<.001	-.102	.00913	<.001

relationships for peers and subordinates were small, and the effect size for the negative relationship for supervisors was approaching-small.

Research Question 9c asked whether the relationship of ratee position tenure and goal content differed between the rating sources. A test of the interaction between rating source and ratee position tenure was significant, $F(3, 31694) = 52.679$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors (-.114; $p < .001$) was found to be significant and negative, as was the slope for subordinates (-.0327, $p < .001$). The slope for peers was found to be significant and positive, however (.0560; $p < .001$). Therefore, the relationship between ratee position tenure and goal content was negative for supervisors and subordinates, and positive for peers. The effect size for supervisors was small and the effect size for peers was approaching-small.

Research Question 9d asked whether the relationship of ratee position tenure and feedback length differed between the rating sources. A test of the interaction between rating source and ratee position tenure was significant, $F(3, 31694) = 147.415$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (-.135; $p < .001$), as were those for peers (-.118; $p < .001$) and subordinates (-.102; $p < .001$). Overall, as ratee position tenure increased, the length of the narrative feedback decreased for all three sources fairly consistently, with all three demonstrating small effect sizes.

Taken together, these results for position tenure indicate that as ratee position tenure increased, the quality of the narrative feedback decreased. Notable relationships for supervisors included small negative effects for goal content and feedback length, as

well as an approaching-small negative effect for specificity. Notable relationships for peers included small negative effects for favorability, specificity, and feedback length, as well as an approaching-small positive effect for goal content. Notable relationships for subordinates included small negative effects for specificity and feedback length.

Section 3: Narrative Feedback Content and Narrative Feedback Quality

The narrative feedback content variables were judged by the amount present. For example, the scale ranges for absolute feedback content ranged from "no absolute content" to "large amount of absolute content" (See Appendix B). Therefore, the narrative feedback quality variables of specificity and feedback length were expected to have large relationships with the narrative feedback content variables; more content will be longer and likely perceived as more specific. However, the narrative feedback quality variables of specify and feedback length were still useful for the comparisons between rating sources and are therefore still reported.

Relative Content. Research Question 10 asked whether the quality of the narrative feedback would vary as a function of the amount of relative content present. A summary of the largest effects for relative content can be found in Table 12. The intercepts and slopes pertaining to this research question can be found in Table 13. Research Question 10a asked whether the favorability of the feedback would vary as a function of the amount of relative content. The slope of favorability on relative content was found to be significant (.0716; p <.001), an approaching-small effect. Therefore, as relative content increased, so too did the favorability of the narrative feedback. Research Question 10b asked whether the specificity of the narrative feedback would vary as a function of the amount of relative content. The slope of specificity on relative content

Table 12. Summary of Effect Sizes of Relationships for Relative Feedback Content

	Narrative Feedback Quality Variables			
Group	Favorability	Specificity	Goal Content	Feedback Length
Overall	Approaching-small positive	Small positive		Approaching-small positive
Supervisors	Small positive	Small positive		Small positive
Peers	Small positive	Approaching-small positive		Approaching-small positive
Subordinates	Small positive	Small positive		Approaching-small positive

Table 13. Slopes and Standard Errors for Relative Content and Outcome Variables

	Intercept	SE	Sig.	Slope	SE	Sig.
Favorability	.00282	.0328	*ns*	.0716	.00571	<.001
Specificity	-.0528	.0293	*ns*	.114	.00562	<.001
Goal Content	-.0394	.0299	*ns*	-.00773	.00579	*ns*
Length	-.116	.0292	<.001	.0851	.00554	<.001

was also found to be significant (.114; $p < .001$) indicating that as relative content increased, the specificity of the narrative feedback increased, a small effect. Research Question 10c asked whether the amount of goal content would vary as a function of the amount of relative content. However, the slope of goal content on relative content was not found to be significant (-.00773; *ns*). Research Question 10d asked whether feedback length would vary as a function of the amount of relative content. The slope of feedback length on relative content was found to be significant (.0851; $p < .001$) indicating that as relative content increased, the length of the narrative feedback also increased, an approaching-small effect. Taken together, the results for Research Question 10 indicate that the amount of relative content was associated with more favorable, more specific, and longer narrative feedback. However, the relationship between relative feedback content and goal content was not found to be significant. The most notable relationships included a small positive effect for specificity, and approaching-small positive effects for favorability and feedback length.

Research Question 11 asked whether there were differences in the mean amount of relative content in the narrative feedback across rating sources. The means and standard errors can be found in Table 4. The results suggest that supervisors provided narrative feedback with more relative content than peers ($t(31477) = 3.552$, $p < .001$, $d = .0400$) and subordinates ($t(31312) = 2.053$, $p = 0.040$, $d = .0232$). The comparison between peer and subordinate raters did not reach significance ($t(31550) = 1.519$, *ns*, $d = .0171$). Therefore, supervisors provided more relative content in their narrative feedback than did peers and subordinates. However, the effect sizes for these comparisons were

very small suggesting consistency in the amount of relative feedback provided across rating sources.

Research Question 12 asked whether the relationship of the amount of relative content and narrative feedback quality differed between the rating sources. This was assessed by estimating a model with a common intercept, but different slopes for each of the rating sources. As mentioned, a summary of the largest effects for relative content can be found in Table 12. The intercepts and slopes pertaining to this research question can be found in Table 14. Research Question 12a asked whether the relationship of relative content and favorability differed between the rating sources. A test of the interaction between rating source and relative content was significant, $F(3, 31671) = 52.792$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was significant (.0641; $p < .001$), as were those for peers (.0780; $p < .001$) and subordinates (.0716; $p < .001$). Overall, as relative content increased, the favorability of the narrative feedback also increased resulting in consistent approaching-small effect sizes for all rating sources.

Research Question 12b asked whether the relationship of relative content and specificity differed between the rating sources. A test of the interaction between rating source and relative content was significant, $F(3, 31665) = 139.808$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (.129; $p < .001$), as were those for peers (.0926; $p < .001$) and subordinates (.122; $p < .001$). Overall, as relative content increased, the specificity of the narrative feedback increased as well. The relationship for supervisors and subordinates

Table 14. Slopes and Standard Errors for Relative Content and Outcome Variables by Rating Source

				Supervisor			Peer			Subordinate		
	Intercept	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.
Favorability	.00297	.0328	*ns*	.0641	.000993	<.001	.0780	.00911	<.001	.0716	.00944	<.001
Specificity	-.0531	.0293	*ns*	.129	.00976	<.001	.0926	.00897	<.001	.122	.00929	<.001
Goal Content	-.0393	.0299	*ns*	-.00985	.0101	*ns*	-.00069	.00924	*ns*	-.0132	.00957	*ns*
Length	-.117	.0292	<.001	.113	.00963	<.001	.0892	.00884	<.001	.0555	.00916	<.001

resulted in small positive effect sizes, and the relationship for peers resulted in an approaching-small effect size.

Research Question 12c asked whether the relationship of relative content and goal content differed between the rating sources. A test of the interaction between rating source and relative content was not found to be significant, $F(3, 31662) = .935$. Furthermore, the slope for supervisors was not found to be significant (-.00985; *ns*), neither were those for peers (-.00069; *ns*) nor subordinates (-.0132; *ns*).

Research Question 12d asked whether the relationship of relative content and feedback length differed between the rating sources. A test of the interaction between rating source and relative content was significant, $F(3, 31667) = 85.314$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (.113; $p < .001$), as were those for peers (.0892; $p < .001$) and subordinates (.0555; $p < .001$). Overall, as relative content increased, the length of the narrative feedback increased. The increase was most notable for supervisors (a small effect), followed by peers and subordinates respectively (approaching-small effects).

Taken together, increases in relative feedback content were associated with increases in narrative feedback quality, which was fairly consistent across rating sources. Notable relationships for supervisors included small positive effects for specificity and feedback length, as well as an approaching-small positive effect for favorability. Notable relationships for peers include approaching-small effects for favorability and specificity. Notable relationships for subordinates include a small positive effect for specificity, as well as approaching-small positive effects for both favorability and feedback length. Also

noteworthy were the relationships for goal content and relative content which were found not to be significant for all rating sources.

Absolute Content. Research Question 13 asked whether narrative feedback quality varied as a function of the amount of absolute content present. A summary of the largest effects for absolute content can be found in Table 15. The intercepts and slopes pertaining to this research question can be found in Table 16. Research Question 13a asked whether the favorability of the narrative feedback would vary as a function of the amount of absolute content present. The slope of favorability on absolute content was found to be significant (.187; p <.001) and indicated a small effect size. Therefore, as absolute content increased, the favorability of the associated feedback increased as well. Research Question 13b asked whether the specificity of the narrative feedback would vary as a function of the amount of absolute content. The slope of specificity on absolute content was found to be significant (.105; p < .001) indicating that as absolute content increased, the specificity of the narrative feedback increased. The results for specificity also indicated a small effect size. Research Question 13c asked whether the amount of goal content would vary as a function of the amount of absolute content. The slope of goal content on absolute content was found to be significant (-.0750; p < .001) and indicated an approaching-small effect size. Thus, as the amount of absolute content increased, the amount of goal content decreased. Research Question 13d asked whether feedback length would vary as a function of the amount of absolute content. The slope of feedback length on absolute content was found to be significant (.0936; p < .001) indicating that as absolute content increased, the length of the narrative feedback also increased. The results for feedback length also indicated an approaching-small effect size.

Table 15. Summary of Effect Sizes of Relationships for Absolute Feedback Content

	Narrative Feedback Quality Variables			
Group	Favorability	Specificity	Goal Content	Feedback Length
Overall	Small positive	Small positive	Approaching-small negative	Approaching-small positive
Supervisors	Small positive	Approaching-small positive	Approaching-small negative	Approaching-small positive
Peers	Small positive	Small positive	Approaching-small negative	Small positive
Subordinates	Approaching-medium positive	Small positive	Approaching-small negative	Approaching-small positive

Table 16. Slopes and Standard Errors for Absolute Content and Outcome Variables

	Intercept	SE	Sig.	Slope	SE	Sig.
Favorability	.00866	.0313	*ns*	.187	.00531	<.001
Specificity	-.0488	.0282	*ns*	.105	.00531	<.001
Goal Content	-.0412	.0298	*ns*	-.0750	.00545	<.001
Length	-.113	.0288	<.001	.0936	.00523	<.001

Taken together, larger amounts of absolute content were related to narrative feedback that was more favorable, more specific, and longer. However, larger amounts of absolute content were also related to less goal content. The results indicated small positive effects for favorability and specificity, as well as an approaching-small positive effect for feedback length. The effect for goal content was approaching-small and negative.

Research Question 14 asked whether there were differences in the mean amount of absolute content in the narrative feedback between rating sources. The means and standard errors can be found in Table 4. The results suggest that supervisors provided narrative feedback with more absolute feedback than peers ($t(29766) = 4.881$, $p < .001$, $d = .0566$) and subordinates ($t(28841) = 2.146$, $p = 0.032$, $d = .0253$). The comparison between peer and subordinate raters indicated that subordinates provided more absolute feedback ($t(29948) = 2.815$, $p = .005$, $d = .0325$). Therefore, supervisors provided more absolute content in their narrative feedback followed by subordinates and peers respectively. However, the effect sizes for these comparisons were very small, suggesting consistency in the amount of absolute content provided across rating sources.

Research Question 15 asked whether the relationship between the amount of absolute content and narrative feedback quality would differ between rating sources. This was assessed by estimating a model with a common intercept, but different slopes for each of the rating sources. As, mentioned, a summary of the largest effects for absolute content can be found in Table 15. The intercepts and slopes pertaining to this Research Question can be found in Table 17. Research Question 15a asked whether the relationship of absolute content and favorability would differ between the rating sources. A test of the interaction between rating source and absolute content was significant, $F(3, 31632) =$

Table 17. Slopes and Standard Errors for Absolute Content and Outcome Variables by Rating Source

				Supervisor			Peer			Subordinate		
	Intercept	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.
Favorability	.00918	.0314	*ns*	.173	.0102	<.001	.172	.00894	<.001	.209	.00831	<.001
Specificity	-.0487	.0282	*ns*	.0995	.0102	<.001	.104	.00896	<.001	.108	.00832	<.001
Goal Content	-.0411	.0298	*ns*	-.0765	.0105	<.001	-.0761	.00919	<.001	-.0732	.00854	<.001
Length	-.114	.0288	<.001	.0982	.0100	<.001	.118	.00882	<.001	.0692	..00819	<.001

417.077, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (.173; $p < .001$), as were those for peers (.172; $p < .001$) and subordinates (.209; $p < .001$). Overall, as absolute content increased, the favorability of the narrative feedback also increased. The effect sizes for supervisors and peers were small, and the effect size for subordinates was approaching-medium.

Research Question 15b asked whether the relationship of absolute content and specificity would differ between the rating sources. A test of the interaction between rating source and absolute content was significant, $F(3, 31653) = 129.407$, $p < .001$, warranting a more in-depth look at the slopes for each rating source. The slope for supervisors was found to be significant (.0995; $p < .001$), as were those for peers (.104; $p < .001$) and subordinates (.108; $p < .001$). Overall, as absolute content increased, the specificity of the narrative feedback increased consistently across the rating sources. The effect sizes for peers and subordinates were small, and the effect size for supervisors was approaching-small.

Research Question 15c asked whether the relationship of absolute content and goal content would differ between the rating sources. A test of the interaction between rating source and absolute content was significant, $F(3, 31645) = 63.172$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (-.0765; $p < .001$), as were those for peers (-.0761; $p < .001$) and subordinates (-.0732; $p < .001$). Overall, as absolute content increased, the amount of goal content provided decreased consistently for all rating sources. The effect size was approaching-small for all rating sources.

Research Question 15d asked whether the relationship of absolute content and feedback length would differ between the rating sources. A test of the interaction between rating source and absolute content was significant, $F(3, 31647) = 112.577$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (0982; $p < .001$), as were those for peers (.118; $p < .001$) and subordinates (.0692; $p < .001$). Overall, as absolute content increased, the length of the narrative feedback increased as well. The effect size was small for peers and approaching-small for supervisors and subordinates.

Taken together, as absolute content increased, the narrative feedback quality did as well. The exception to this finding pertained to goal content. Across rating sources, increased absolute content was related to decreased goal content. Notable relationships for supervisors included a small positive effect for favorability, approaching-small positive effects for the narrative feedback quality variables of specificity and comment length, and an approaching-small negative effect for goal content. Notable relationships for peers included small positive effects for favorability, specificity, and feedback length, as well as an approaching-small negative effect for goal content. Notable relationships for subordinates included an approaching-medium effect for favorability, a small positive effect for specificity, and approaching-small positive effect for feedback length, and an approaching-small negative effect for goal content.

Task Content. Research Question 16 asked whether narrative feedback quality would vary as a function of the amount of task content present. A summary of the largest effects for task content can be found in Table 18. The intercepts and slopes pertaining to this research question can be found in Table 19. Research Question 16a asked whether

Table 18. Summary of Effect Sizes of Relationships for Task Feedback Content

		Narrative Feedback Quality Variables		
Group	Favorability	Specificity	Goal Content	Feedback Length
Overall		Approaching-large positive	Approaching-medium positive	Medium positive
Supervisors		Approaching-large positive	Medium positive	Approaching-large positive
Peers		Approaching-large positive	Approaching-medium positive	Medium positive
Subordinates		Approaching-large positive	Small positive	Medium positive

Table 19. Slopes and Standard Errors for Task Content and Outcome Variables

	Intercept	SE	Sig.	Slope	SE	Sig.
Favorability	.00304	.0328	*ns*	-.0129	.00545	.018
Specificity	-.0251	.0228	*ns*	.427	.00482	<.001
Goal Content	.0265	.0282	*ns*	.218	.00537	<.001
Length	-.0924	.0240	<.001	.378	.00485	<.001

the favorability of the narrative feedback would vary as a function of the amount of task content present. The slope of favorability on task content was found to be significant (-.0129; $p = .018$). Therefore, as task content increased the favorability of the narrative feedback decreased. However, the effect for the relationship between task content and favorability was very small. Research Question 16b asked whether the specificity of the narrative feedback would vary as a function of the amount of task content present. The slope of specificity on task content was also found to be significant (.427; $p < .001$) indicating that as task content increased, the specificity of the narrative feedback increased. The effect size for specificity on task content was approaching-large. Research Question 16c asked whether the amount of goal content in the narrative feedback would vary as a function of the amount of task feedback. The slope of goal content on task content was also found to be significant (.218; $p < .001$) indicating that as task content increased, the amount of goal content increased. The effect size for goal content on task feedback was approaching-medium. Research Question 16d asked whether the feedback length of the narrative feedback would vary as a function of the amount of absolute feedback. The slope of feedback length on task content was also found to be significant (.378; $p < .001$) indicating that as task content increased, the length of the narrative feedback also increased. The effect size of feedback length on task content was medium. Taken together, the results indicate that narrative feedback with more task content tended to be more specific, contain more goal content, and lengthier. However, increased task content was also related to less favorable narrative feedback. Notable relationships include an approaching-large positive effect for specificity, a medium positive effect for feedback length, and an approaching-medium positive effect for goal content.

Research Question 17 asked whether there were differences in the mean amount of task content in the narrative feedback between rating sources. The means and standard errors can be found in Table 4. The results suggest that supervisors provided narrative feedback with more task content than peers ($t(28347) = 3.009$, $p = .003$, $d = .0357$) but the comparison of supervisors and subordinates did not reach significance ($t(26898) = -0.118$, *ns*, $d = .00144$). The comparison between peer and subordinate raters indicated that subordinates provided more task feedback ($t(28413) = 3.283$, $p = .001$, $d = .0390$). Therefore, supervisors and subordinates provided more task content in their narrative feedback than peers. However, the effect sizes for these comparisons were very small, again indicating consistency in the amount of task content provided across rating sources.

Research Question 18 asked whether the relationship between task content and narrative feedback quality would differ between the different rating sources. This was assessed by estimating a model with a common intercept, but different slopes for each of the rating sources. As mentioned, a summary of the largest effects for task content can be found in Table 18. The intercepts and slopes pertaining to this research question can be found in Table 20. Research Question 18a asked whether the relationship of task content and favorability would differ between the rating sources. A test of the interaction between rating source and task content was significant ($F(3, 31618) = 3.816$, $p = .01$), warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (-.0256; $p = .013$), as was that for subordinates (-.0194; $p = .024$). However, the slope for peers was not found to be significant (.00468; *ns*). Overall, the relationship between task content and favorability was significant and

Table 20. Slopes and Standard Errors for Task Content and Outcome Variables by Rating Source

				Supervisor			Peer			Subordinate		
	Intercept	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.
Favorability	.00341	.0328	*ns*	-.0256	.0103	.013	.00468	.00917	*ns*	-.0194	.00862	.024
Specificity	-.0257	.0228	*ns*	.431	.00915	<.001	.404	.00812	<.001	.444	.00764	<.001
Goal Content	-.0267	.0282	*ns*	.303	.0102	<.001	.200	.00903	<.001	.175	.00849	<.001
Length	-.0926	.0241	<.001	.410	.00921	<.001	.365	.0817	<.001	.367	.00768	<.001

negative for both supervisors and peers. However, the effects for all relationships were very small, suggesting consistency across rating sources.

Research Question 18b asked whether the relationship of task content and specificity would differ between the rating sources. A test of the interaction between rating source and task content was significant, $F(3, 31665) = 2612.548$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (.431; $p < .001$), as were those for peers (.404; $p < .001$) and subordinates (.444; $p < .001$). Overall, as task content increased, the specificity of the narrative feedback increased as well. The relationship was consistent across rating sources, all of which demonstrated approaching-large effect sizes.

Research Question 18c asked whether the relationship of task content and goal content would differ between the rating sources. A test of the interaction between rating source and task content was significant, $F(3, 31642) = 585.201$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (.303; $p < .001$), as were those for peers (.200; $p < .001$) and subordinates (.175; $p < .001$). Overall, as task content increased, the amount of goal content provided also increased. The results indicated a medium effect size for supervisors, an approaching-medium effect size for peers, and a small effect size for subordinates.

Research Question 18d asked whether relationship of task content and feedback length would differ between the rating sources. A test of the interaction between rating source and task content was significant, $F(3, 31655) = 2031.336$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors

was found to be significant (.410; $p < .001$), as were those for peers (.365; $p < .001$) and subordinates (.367; $p < .001$). Overall, as task content increased, the length of the narrative feedback increased as well. The results indicated an approaching-large effect for supervisors and medium effects for peers and subordinates.

Taken together, as task content increased, the narrative feedback quality did as well. The different rating sources demonstrated consistently positive findings across the indices narrative feedback quality, with the exception of favorability. Notable relationships for supervisors included approaching-large positive effects for specificity and feedback length, as well as a medium positive effect for goal content. Notable relationships for peers include an approaching-large positive effect for specificity, a medium positive effect for feedback length, and an approaching-medium positive relationship for goal content. Notable relationships for subordinates include an approaching-large positive effect for specificity, a medium positive effect for feedback length, and a small positive effect for goal content.

Trait Content. Research Question 19 asked whether narrative feedback quality would vary as a function of the amount of trait content present. A summary of the largest effects for trait content can be found in Table 21. The intercepts and slopes pertaining to this research question can be found in Table 22. Research Question 19a asked whether the favorability of the narrative feedback would vary as a function of the amount of trait content. The slope of favorability on trait content was found to be significant (.0665; $p <.001$) indicating an approaching-small effect. Therefore, increased trait content was associated with increased favorability. Research Question 19b asked whether the specificity of the narrative feedback would vary as a function of the amount of trait

Table 21. Summary of Effect Sizes of Relationships for Trait Feedback Content

	Narrative Feedback Quality Variables			
Group	Favorability	Specificity	Goal Content	Feedback Length
Overall	Approaching-small positive	Small positive		Small positive
Supervisors		Small positive		Small positive
Peers	Approaching-small positive	Small positive		Approaching-small positive
Subordinates	Approaching-small positive	Small positive		Approaching-small positive

Table 22. Slopes and Standard Errors for Trait Content and Outcome Variables

	Intercept	SE	Sig.	Slope	SE	Sig.
Favorability	.00510	.0327	*ns*	.0665	.00534	<.001
Specificity	-.0491	.0283	*ns*	.102	.00526	<.001
Goal Content	-.0401	.0299	*ns*	-.0319	.00541	<.001
Length	-.113	.0287	<.001	.0985	.00518	<.001

content. The slope of specificity on trait content was found to be significant (.102; $p <$ .001) indicating that as trait content increased, the specificity of the narrative feedback increased. The effect size for specificity on trait content was small. Research Question 19c asked whether the amount of goal content would vary as a function of the amount of trait content. The slope of goal content on trait content was found to be significant (-.0319; $p < .001$) and negative. Therefore, increased trait content was associated with decreased goal content. Research Question 19d asked whether the feedback length would vary as a function of the amount of trait content. The slope of feedback length on trait content was also found to be significant (.0985; $p < .001$) indicating that as trait content increased, the length of the narrative feedback also increased. The effect size for feedback length on trait content was approaching-small. Taken together, the results indicate that more trait content was related to more favorable, more specific, and longer narrative feedback. However, the relationship between trait feedback content and goal content was not found to be significant. Notable relationships include a small positive effect for specificity, as well as approaching-small positive effects for favorability and feedback length.

Research Question 20 asked whether there were differences in the mean amount of trait content in the narrative feedback across rating sources. The means and standard errors can be found in Table 4. The results suggest that supervisors provided narrative feedback with less trait content than peers ($t(27163) = 4.506$, $p < .001$, $d = .05468$) and subordinates ($t(25320) = 2.099$, $p = .036$, $d = .0264$). The comparison between peer and subordinate raters indicated that subordinates provided less trait feedback ($t(27089) = 2.476$, $p = .013$, $d = .0301$). Therefore, supervisors provided the least amount of trait

feedback, followed by subordinates and peers respectively. However, the effect sizes for these comparisons were very small for all rating sources, suggesting consistency in the amount of trait content provided.

Research Question 21 asked whether the relationship between trait content and narrative feedback quality would differ between the rating sources. This was assessed by estimating a model with a common intercept, but different slopes for each of the rating sources. As mentioned, a summary of the largest effects for trait content can be found in Table 21. The intercepts and slopes pertaining to this research question can be found in Table 23. Research Question 21a asked whether the relationship of trait content and favorability would differ between the rating sources. A test of the interaction between rating source and trait content was significant, $F(3, 31602) = 53.160$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (.0486; $p < .001$), as were those for peers (.0776; $p < .001$) and subordinates (.0680; $p < .001$). Overall, as trait content increased, the favorability of the narrative feedback also increased. The relationships were fairly consistent across rating sources with approaching-small effects for peers and subordinates.

Research Question 21b asked whether the relationship of trait content and specificity would differ between the rating sources. A test of the interaction between rating source and trait content was significant, $F(3, 31628) = 130.286$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (.117; $p < .001$), as were those for peers (.0741; $p < .001$) and subordinates (.118; $p < .001$). Overall, as trait content increased, the

Table 23. Slopes and Standard Errors for Trait Content and Outcome Variables by Rating Source

				Supervisor			Peer			Subordinate		
	Intercept	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.	Slope	SE	Sig.
Favorability	.00485	.0327	*ns*	.0486	.0105	<.001	.0776	.00881	<.001	.0680	.00854	<.001
Specificity	-.0487	.0283	*ns*	.117	.0103	<.001	.0741	.00868	<.001	.118	.00841	<001
Goal Content	-.0400	.0299	*ns*	-.0125	.0106	*ns*	-.0340	.00893	<.001	-.0426	.00865	<.001
Length	-.113	.0287	<.001	.114	..0287	<.001	.0872	.00854	<.001	.0989	.00828	<.001

specificity of the narrative feedback increased as well. The relationships were again fairly consistent across rating sources with small effects for supervisors and subordinates, as well as an approaching-small effect for peers.

Research Question 21c asked whether the relationship of trait content and goal content would differ between the rating sources. A test of the interaction between rating source and trait content was significant, $F(3, 31620) = 13.207$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was not found to be significant (-.0125; *ns*). However, the slopes were found to be significant for both peers (-.0340; $p < .001$) and subordinates (-.0426; $p < .001$). Therefore, increases in trait content were related to fairly consistent decreases in goal content for both subordinates and peers. However, the effect sizes for these relationships were very small.

Research Question 21d asked whether the relationship of trait content and feedback length would differ between the rating sources. A test of the interaction between rating source and trait content was significant, $F(3, 31623) = 121.934$, $p < .001$, warranting a more in-depth investigation of the slopes for each rating source. The slope for supervisors was found to be significant (.114; $p < .001$), as were those for peers (.0872; $p < .001$) and subordinates (.0989; $p < .001$). Overall, as trait content increased, the length of the narrative feedback increased as well. The relationships were again very consistent with a small effect for supervisors, as well as approaching-small effects for both peers and subordinates.

Taken together, as trait content increased, the narrative feedback quality did as well. Notable relationships for supervisors included small positive effects for specificity

and feedback length. Notable relationships for peers included approaching-small positive effects for favorability, specificity, and feedback length. Notable relationships for subordinates included a small positive relationship for specificity, as well as approaching-small positive relationships for favorability and feedback length.

Discussion

As evidenced by the very limited amount of extant research on this topic, little attention has been given to the narrative component of performance evaluation, despite being an important piece of many performance evaluation interventions. As mentioned, performance feedback that is predominantly numeric provides insufficient context (David, 2013). Hence, it can be unclear to employees why they received a particular rating. The context provided in narrative feedback is necessary for developing precise goals that drive the development process. The present study builds on, and extends, what little is known about narrative feedback quality (e.g., David, 2013).

Section 1: Overall Differences in Narrative Feedback Quality by Rating Source

The first purpose of the present study was to investigate which rating source provided higher quality narrative feedback based on the quality variables of favorability, specificity, goal content, and length. This was especially pertinent because previous studies that have investigated narrative feedback quality have focused solely on supervisory narrative feedback (David, 2013; Wilson, 2010). This is the first investigation of peer and subordinate narrative feedback quality. The largest effects for Research Question 1 were found for the supervisor-peer and supervisor-subordinate comparisons of goal-content, as well as the supervisor-peer comparison of feedback length. These effects were approaching-small suggesting supervisors provided slightly

higher quality narrative feedback over the other sources. However, most of the observed effects for Research Question 1 were very small, indicating consistency in the feedback provided to ratees regardless of who is providing it.

Section 2: Predictor Variables and Narrative Feedback Quality

The second purpose of the present study was to investigate the relationship between familiarity with the ratee's work behavior, rater acquaintanceship time, and ratee position tenure and the quality of the narrative feedback provided. This was an important endeavor as it allowed us to investigate possible mechanisms that would support the selection of raters who are likely to provide high quality narrative feedback, and ratees who are likely to receive high quality narrative feedback.

Familiarity with the ratee's work behavior. The research questions related to familiarity with the ratee's work behavior largely draw from the Realistic Accuracy Model (RAM; Funder, 1995) and Funder's propositions which suggest that those who are more familiar with the ratee are more likely to be exposed to relevant cues, detect those cues, and refer to them when providing ratings. In line with this model, Research Question 2 asked whether narrative feedback quality varied as a function of the rater's level of familiarity with the ratee's work behavior. Overall, the results of Research Question 2 indicated that higher familiarity was related to narrative feedback that was more specific, contained more goal content, and was lengthier. The results for the relationship between familiarity and favorability, however, were in the opposite direction, indicating that more familiar raters provided narrative feedback that was less favorable than less familiar raters. Although somewhat surprising, the favorability results are in line with Bernardin and Villanova's (2005) findings suggesting that raters may not feel

efficacious in providing negative feedback if they are not familiar with the ratee's work behavior. Overall, the effects for Research Question 2 were very small. The only relationship to reach an approaching-small effect size was between familiarity and goal content. Thus, the very small observed effects indicated that the relationship between familiarity with the ratee's work behavior and narrative feedback quality, although generally positive, may not be of large consequence.

Research Question 3 asked whether there were mean differences in the reported familiarity with the ratee's work behavior between rating sources. The results of Research Question 3 indicated that supervisors reported the greatest familiarity with the ratee's work behavior, followed by subordinate raters, and finally peer raters. The effect sizes of these tests indicate that supervisors and subordinates reported much higher familiarity than peer raters. These findings suggest that when looking for alternate sources to supervisory narrative feedback, subordinate raters may be in a better position than peer raters.

Research Question 4 investigated the relationship between familiarity with the ratee's work behavior and narrative feedback quality for each of the rating sources. The most notable findings for supervisors included a small negative effect for the relationship between familiarity and favorability, and an approaching-medium positive effect for the relationship between familiarity and goal content. The most notable relationships for peers included approaching-small positive effects for the relationships between familiarity and the narrative feedback quality variables of specificity and feedback length. The results did not indicate any notable relationships for subordinates.

The findings for Research Question 4 indicate that supervisors provided more actionable content as their familiarity with the ratee's work behavior increased as evidenced by an approaching-medium positive effect for goal content. This increase in goal content was likely perceived as less favorable as indicated by the associated small negative effect between familiarity and favorability for supervisors. Peers provided more specific, an approaching-small positive effect, and longer feedback, a small positive effect, as familiarity with the ratee's work behavior increased. This suggests that although peers provided more specific and longer feedback to the ratee, they did not provide more actionable content. The findings for supervisors and peers were generally in line with Funder's (1995) RAM. However, none of the effects for subordinates were found to be significant. Therefore, the results suggest that the relationship between familiarity and narrative feedback quality differed across rating sources.

The differences in the effects for each of the rating sources and how they align with Funder's (1995) RAM suggest that another variable is likely affecting the relationship between familiarity and narrative feedback quality. Because the relationships for the subordinates were the smallest, and supervisors the largest, these findings may indicate that subordinate raters may not provide higher quality narrative feedback due to the desire for anonymity and/or fear of reprisals. This may also explain why peer rates provided more specific and longer feedback to more familiar ratees but did not provide more goal content. Rather than provide more goal content which may be misconstrued as harsh, peer raters provided more description.

Acquaintanceship time. Similar to familiarity, the research questions concerning the amount of time the rater has known the ratee in their current capacity,

acquaintanceship time, used Funder's (1995) RAM as a framework. Research Question 5 asked whether the quality of the narrative feedback provided would vary as a function of rater acquaintanceship time with the ratee. The results for the relationships between acquaintanceship time and the narrative feedback quality variables of favorability and goal content were not found to be significant. Furthermore, the results for the relationships between acquaintanceship time and the narrative feedback quality variables of specificity and feedback length were significant, but negative. The only noteworthy effect was for the relationship between acquaintanceship time and feedback length, which was approaching-small. The negative and not significant findings for acquaintanceship time and narrative feedback quality are important because familiarity with the ratee's work behavior and the amount of time the rater has known the ratee in their current capacity appear to be similar variables but have opposing relationships with narrative feedback quality. Similar to the results for familiarity, the effects for acquaintanceship time were very small suggesting somewhat limited utility.

The above findings regarding acquaintanceship time were for all rating sources. Research Questions 6 and 7 investigated differences between the different rating sources. The results of Research Question 6 indicated that supervisors had the most acquaintanceship time with the ratee, followed by peer raters and subordinate raters respectively. The comparison of supervisors and subordinates, and the comparison of peers and subordinates resulted in small effect sizes. As found above, acquaintanceship time appeared to be related to lower narrative feedback quality, suggesting that subordinate raters may be a good source of narrative feedback. These findings are further investigated below as the relationships are tested for each rating source.

Research Question 7 investigated the relationship between acquaintanceship time and narrative feedback quality for each of the rating sources. The results indicated that supervisors, peers, and subordinates all had relationships between acquaintanceship time and favorability that were not found to be significant. Supervisors had significant negative relationships between acquaintanceship time and specificity as well as between acquaintanceship time and goal content. Peers had a significant negative relationship between acquaintanceship time and feedback length. Subordinates had significant negative relationships between acquaintanceship time and the narrative feedback quality variables of specificity and feedback length, and a positive relationship between acquaintanceship time and goal content. The only notable relationship for supervisors was an approaching-small negative effect between acquaintanceship time and specificity. Peers also had only one notable relationship which was an approaching-small negative effect between acquaintanceship time and feedback length. Subordinates had two approaching-small negative effects which were between acquaintanceship time and the narrative feedback quality variables of specificity and feedback length. The results for acquaintanceship time indicate that as acquaintanceship time increased, the quality of the associated narrative feedback decreased fairly consistently for each rating source. The decrease in narrative feedback quality mostly concerned the specificity and length of the feedback.

The opposing findings of rater familiarity and acquaintanceship time with narrative feedback quality were interesting because of the logical relationship between the two predictor variables. Reasonably, raters who have known the ratee for longer in their current position should be exposed to more instances of their work behavior. This

would imply that these two variables should be highly related. Originally, we proposed that acquaintanceship time might be used as a proxy for familiarity as it is easier to assess and the two variables were likely related. As it turns out, this was a dangerous assumption. Practitioners and researchers who use acquaintanceship time as a method of rater selection may be doing more harm than good.

Position Tenure. The purpose of investigating ratee position tenure with regard to narrative feedback quality was to better understand who is likely to receive high quality narrative feedback. Research Question 8 asked whether narrative feedback quality would vary as a function of ratee position tenure. The findings for Research Question 8 indicated that as position tenure increased, the favorability, specificity, goal content, and length of the narrative feedback decreased. The most notable relationships included small negative effects between position tenure and the narrative feedback quality variables of specificity and feedback length, and an approaching-small negative effect between position tenure and favorability. The findings are in line with Human Capital Theory (Becker, 1964) and ASA theory (Schneider et al., 1995) suggesting that those who have been in their position for longer have likely learned the requisite skills and are a good fit for the position, and therefore require less constructive feedback. Similarly, raters may view long-tenured employees in high regard and may be less inclined to provide high quality narrative feedback. Thus, should a longer tenured employee require development, rater training may be necessary in order to ensure they are receiving the feedback they need.

Research Question 9 broke down the relationships between position tenure and narrative feedback quality by rating source. Notable relationships for supervisors

included small negative effects between position tenure and the narrative feedback quality variables of goal content and feedback length, as well as an approaching-small negative effect between position tenure and specificity. Notable relationships for peers included small negative effects between position tenure and the narrative feedback quality variables of favorability, specificity, and feedback length. Peers also had an approaching-small positive effect between position tenure and goal content. Notable relationships for subordinates included small negative effects between position tenure and the narrative feedback quality variables of specificity and feedback length. Of the three predictor variables in Section 2 (familiarity, acquaintanceship time, and ratee position tenure), ratee position tenure demonstrated some of the largest relationships.

The results for position tenure seem to support the notion that subordinate raters are likely to hold long-tenured employees in high regard and provide feedback with lower narrative feedback quality. The negative results for supervisors might indicate that they perceive longer-tenured employees as having garnered the requisite skills for their work, explaining the negative relationship between ratee position tenure and narrative feedback quality (ASA theory; Schneider et al., 1995). Peer raters had the only positive relationship between ratee position tenure and goal content, suggesting that they might be a good source of constructive feedback for longer tenured employees. Because peer raters are likely in a similar position to the ratee, minute differences in the ratee's performance might be made more salient due to social comparison (Festinger, 1954). The positive relationship between position tenure and goal content for peers was accompanied by the strongest negative relationship between position tenure and favorability. A similar pattern

emerged for supervisors when investigating familiarity in Research Question 4, providing additional support for the apparent tradeoff between goal content and favorability.

The results for ratee position tenure suggest that practitioners may find that supplementing supervisory ratings with peer ratings to be more effective than supplementing with subordinate ratings when it comes to longer-tenured employees. This should be investigated further because the overall differences in narrative feedback quality indicate that, generally, subordinate raters are in a slightly better position to provide high quality narrative feedback over peer raters. Long-tenured ratees may be the exception to this finding.

Section 3: Narrative Feedback Content and Narrative Feedback Quality

The third purpose of the present study was to investigate the relationship between the feedback content and narrative feedback quality. The feedback content variables included relative content, absolute content, task content, and trait content. This was an important endeavor as it allowed us to investigate what content was associated with higher quality narrative feedback. The implications include the development of rater training to provide appropriate content to the ratee. As previously mentioned, the narrative feedback quality variables of specificity and length demonstrated larger relationships with the feedback content variables due to the nature of the variables. The content variables were judged by the amount present, therefore, it is not surprising that more content was associated with longer feedback, and that the feedback was perceived as more specific. However, the narrative feedback quality variables of specificity and length are still useful for the comparisons between rating sources.

Relative feedback content. The purpose of investigating relative feedback content was to determine whether feedback based in social comparison was related to the indices of narrative feedback quality. Research Question 10 asked whether narrative feedback quality would vary as a function of the amount relative feedback content present. The results generally indicate that increased relative feedback content was associated with increased narrative feedback quality. The relationship between the amount of relative feedback content and the narrative feedback quality indices of favorability, specificity and feedback length were all positive and significant. However, the relationship between relative feedback content and goal content was not found to be significant. Notable relationships included a small positive effect for the relationship between relative feedback content and specificity, as well as approaching-small positive effects for the relationships between relative feedback content and the narrative feedback quality variables of favorability and feedback length.

Research Question 11 investigated the mean differences in the amount of relative feedback content provided by the different rating sources. The effect sizes for these comparisons in Research Question 11 were very small suggesting consistency in the amount of relative feedback provided across rating sources.

Research Question 12 investigated the relationship between relative feedback content and the indices of narrative feedback quality for each rating source. The results for Research Question 12 suggest consistency in the relationships between relative feedback content and narrative feedback quality across rating sources. All three rating sources had small positive effects between relative feedback content and favorability, and relationships between relative feedback content and goal content that were not found to

be significant. The results for the relationships between relative feedback content and specificity indicated small positive effects for supervisors and subordinates, as well as an approaching-small positive effect for peers. Further, the results for the relationship between relative feedback content and feedback length indicated a small positive effect for supervisors, as well as approaching-small positive effects for peers and subordinates. Interestingly, based on the positive relationship between relative feedback content and specificity and feedback length, as well as a relationship with goal content that as not found to be significant, relative feedback content was likely used primarily for behavior description rather than providing actionable content. Furthermore, based on the positive relationship of relative feedback content and favorability, and the relationship between relative content and goal content found not to be significant, it is likely that relative feedback content may have been used for ingratiation. In other words, as the amount of relative content increased, the favorability of the feedback increased without providing more actionable content for the ratee. These are interesting findings and should continue to be investigated.

Absolute feedback content. The purpose of investigating the amount of absolute feedback content was to determine whether feedback using adjective-based performance descriptors would be related to the indices of narrative feedback quality. Research Question 13 asked whether the narrative feedback quality would vary as a function of the amount of absolute content provided. The results generally indicate that the amount of absolute content was associated with narrative feedback quality. The results indicated small positive effects for the relationships between absolute feedback content and the narrative feedback quality variables of favorability and specificity, as well as an

approaching-small positive effect between absolute feedback content and feedback length. The effect for the relationship between absolute feedback content and goal content was approaching-small and negative. This suggests that more absolute feedback content, while more specific, more favorable, and longer, was associated with less actionable content for the ratee. Again, the relationships for absolute feedback content indicate a trade-off between favorability and goal content.

Research Question 14 investigated the mean differences in the amount of absolute feedback content provided by the three rating sources. The effect sizes for these comparisons were very small, suggesting consistency in the amount of absolute content provided across rating sources.

Research Question 15 investigated the relationship between absolute feedback content and the indices of narrative feedback quality for each rating source. Similar to what was found for relative feedback content, the relationships for absolute content and the narrative feedback quality variables suggested consistency across rating sources. Notable relationships for supervisors included a small positive effect for the relationship between absolute feedback content and favorability, approaching-small positive effects between absolute feedback content and the narrative feedback quality variables of specificity and comment length, and an approaching-small negative effect between absolute feedback content and goal content. Notable relationships for peers include small positive effects for the relationships between absolute feedback content and the narrative feedback quality variables of favorability, specificity, and feedback length, as well as an approaching-small negative effect between absolute feedback content and goal content. Notable relationships for subordinates include an approaching-medium effect between

absolute feedback content and favorability, a small positive effect between absolute feedback content and specificity, an approaching-small positive effect between absolute feedback content and feedback length, and an approaching-small negative effect between absolute feedback content and goal content. Similar to the results of Research Question 13c, all three rating sources had significant negative relationships between the amount of absolute feedback content and goal content. This suggests that across all ratings sources, more absolute feedback content was associated with less goal content. This is an indication that absolute feedback content might be used universally for performance level description and ingratiation as it was related to less goal content for all rating sources, but more specific and longer narrative feedback. As such, rater training should be implemented to ensure that raters are providing useful narrative to the ratee. Multisource feedback systems are developmental in nature and those who utilize them are generally expecting information on how to improve their performance. It could be frustrating for a leader who is expecting feedback on how to improve to receive overly positive description of how they are currently performing with little constructive criticism or comments regarding future performance. This issue should be more thoroughly investigated. Further, the relationships for absolute feedback content indicate a trade-off between favorability and goal content, with more absolute feedback content related to less goal content but more favorable narrative feedback.

Task Feedback Content. As previously mentioned, task feedback content focuses the ratee's attention on specific behaviors or tasks making it beneficial when it comes to developing goals to improve performance. Notable relationships for Research Question 16 include an approaching-large positive effect between task feedback content

and specificity, a medium positive effect between task feedback content and feedback length, and an approaching-medium positive effect between task feedback content and goal content. The relationship between task feedback content and favorability was significant, but the effect size was very small. This may be an indication that task feedback content is not associated with the tradeoff between favorability and goal content that the other feedback content variables demonstrated. The effect sizes found for task feedback content are some of the largest in the present study.

Research Question 17 indicated that the effect sizes for the comparisons across rating sources were very small, indicating consistency in the amount of task content provided across rating sources.

Research Question 18 asked whether there were differences between rating sources in the relationship between task feedback content and narrative feedback quality. The different rating sources demonstrated consistently positive findings between task feedback content and the indices of narrative feedback quality, with the exception of favorability. Notable relationships for supervisors included approaching-large positive effects between task feedback content and the narrative feedback quality variables of specificity and feedback length, as well as a medium positive effect between task feedback content and goal content. Notable relationships for peers include an approaching-large positive effect between task feedback content and specificity, a medium positive effect between task feedback content and feedback length, and an approaching-medium positive effect between task feedback content and goal content. Notable relationships for subordinates include an approaching-large positive effect between task feedback content and specificity, a medium positive effect between task

feedback content and feedback length, and a small positive effect between task feedback content and goal content. The relationship between task feedback content and favorability was very small for all rating sources.

Trait Feedback Content. Narrative feedback can also bring the ratee's personal traits or characteristics into focus (Smither & Walker, 2004). Trait feedback content addresses stable characteristics in the employee and is often perceived as less actionable. The results for Research Question 19 indicated a small positive effect between trait feedback content and specificity, as well as approaching-small positive effects between trait feedback content and the narrative feedback quality variables of favorability and feedback length. The relationship between trait feedback content and goal content was significant, however the effect was very small.

Research Question 20 indicated consistency in the amount of trait feedback content provided across rating sources. The comparisons between rating sources indicated very small effect sizes.

Research Question 21 asked whether there were differences between rating source in the relationship between the amount of trait feedback content and narrative feedback quality. Similar to the results of the other feedback content variables, the results for trait feedback content suggest consistency across rating sources. Notable relationships for supervisors included small positive effects between trait feedback content and the narrative feedback content variables of specificity and feedback length. Notable relationships for peers included approaching-small positive effects between trait feedback content and the narrative feedback quality variables of favorability, specificity, and feedback length. Notable relationships for subordinates included a small positive effect

between trait feedback content and specificity, as well as approaching-small positive effects for the narrative feedback content variables of favorability and feedback length. Similar to the overall effect found in Research Question 19c, the effects for the relationship between trait feedback content and goal content were very small for all rating sources.

Implications

The present study suggests a number of implications that should be taken into account by researchers and practitioners. The first implication is that supervisors provided higher quality narrative feedback than peers and subordinates. Further, the comparisons between peer and subordinate raters produced mixed results, however none of the comparisons resulted in encouraging effect sizes. These findings suggest that narrative feedback provided by supervisors should be given precedence over the other two sources. There is no evidence to suggest that peer or subordinate raters should be considered over the other based on the overall comparisons.

The second implication is that familiarity with the ratee's work behavior and the amount of time the rater has known the ratee in their current position, acquaintanceship time, are not similar variables. Reasonably, acquaintanceship time is a much easier method of rater selection for practitioners and researchers than is asking raters how familiar they are with the ratee's work behavior. However, as indicated in the present study, these two variables had very different relationships with narrative feedback quality. Familiarity was positively related to narrative feedback quality and acquaintanceship time was negatively related to narrative feedback quality. Therefore, acquaintanceship time should not be used as a proxy for familiarity.

A third implication relates to the question of who should provide narrative feedback to long-tenured employees. The results indicated that narrative feedback quality tended to decrease as ratee position tenure increased for all rating sources. The exception to this findings was found for peer raters and indicated that the amount of goal content provided increased with ratee position tenure. Therefore, peer raters may be best situated to provide narrative feedback to long-tenured employees over the other rating sources.

The results for relative and absolute feedback indicated that relative and absolute feedback content had fairly similar relationships with narrative feedback quality. The biggest difference was that increases in the amount of absolute content were more favorable and provided less goal content. As mentioned, this may be an indication that absolute feedback content is being used for ingratiation which may be addressed through rater training.

The results for task and trait feedback indicated that task feedback was associated with higher narrative feedback quality than trait feedback. Task feedback demonstrated the strongest relationships with narrative feedback quality across all of the content variables studied. Additionally, task feedback content was the only narrative feedback content variable to have a positive relationship with the amount of goal content provided. Further, the approaching-medium positive effect between task feedback content and goal content does not have the associated trade-off with favorability that other predictor variables had in the present study. Therefore, task feedback content was used to provide actionable content to the ratee without being perceived as harsh or negative. Rater training should focus on increasing the amount of task feedback provided to ratees.

The final implication is that, supervisors, peers, and subordinates did not differ on the amount of the various types of narrative feedback content provided, nor did the relationships with narrative feedback quality differ across rating sources. Therefore, rating sources utilized the different types of narrative feedback similarly and provided similar amounts of each type of narrative feedback content. Because the relationships were found to be consistent across rating sources, future training interventions focused on feedback content likely do not need to be tailored to each rating source.

In summary, the findings of the present study suggest a few general principles to be followed when collecting narrating performance feedback. The first is to select raters who meaningfully interact with the ratee on a regular basis or at least with regard to the behavior being addressed. There is a distinction between knowing the behavior of the ratee and simply knowing who the ratee is, as exemplified by the differences between familiarity and acquaintanceship time. The second is to train the raters so they are confident in providing constructive feedback. The majority of the findings indicate that peer and subordinate raters provided lower quality narrative feedback. Further, peer and subordinate raters had a much lower narrative feedback response rate than supervisors. There are a number of reasons for why this might occur, but we see rater training, and communication regarding anonymity and the goals of the multisource feedback tool, to be essential in addressing these issues. The need for rater training is exemplified in the familiarity results. Subordinates indicated comparable levels of familiarity with the ratee's performance as supervisors, however the effects of familiarity on narrative feedback quality were not significant for subordinate raters. The link between observing the relevant behavior and putting it down on paper is obscured and needs to be addressed.

Finally, the overwhelming results of the content variables suggest that task feedback content is perceived as the highest quality, and is the only type to show an overall effect for goal content. Rater training programs should capitalize on this and encourage the use of this type of feedback when implementing developmental multisource feedback systems.

Limitations

The first limitation of the present study was the inability to draw causal inferences. As mentioned the data in the present study were obtained from a commercial instrument currently being used for employee development. The data were archival, and therefore the investigators were unable to exert control over how and when the data were collected. While causal inferences may not be able to be drawn, the data does provide interesting relationships for this commercial multisource instrument.

A second limitation is the use of only one instrument. The observed effects are only for managers and directors who underwent development using this specific instrument. Other instruments likely vary in their implementation, and may demonstrate differences in their effects based on the variables studied. Previously, we mentioned that narrative feedback quality has rarely been studied, and we could find no current publications that investigated peer and subordinate narrative feedback. Thus, we see the present study as a starting point for future researchers to build upon, utilizing alternate instruments and incorporating more rating sources.

We acknowledge the fact that we conducted a large number of statistical tests and that there might be concern regarding the capitalization on chance contributing to finding significant relationships. However, we had such small p-values that it seems unlikely that

the results were found due to chance. Thus, the very small p-values suggest a low study-wise error rate. Replication of the present study's findings is encouraged.

Study Strengths

There are a number of strengths of the present study that separate it from the work that has previously been done in the area of narrative performance feedback. First, we utilized a large sample of industry data which assists in the generalizability of the findings. The data were archival and therefore the researchers had no influence in the collection of the data. Second, our variables were coded by research assistants who received training based on the principles of Frame of Reference (FOR) training (Bernardin, 1979). This ensured that the research assistants adopted a similar metric when coding the data. Third, previous studies (David, 2013; Wilson, 2010) utilized multiple coders for a small portion of the data to demonstrate inter-rater reliability, however the authors coded the majority of the data on their own. We had four research assistants code each comment to ensure the quality of the ratings for the entire dataset. Further, the author of the present study recused themselves from coding any of the data to avoid potential influence. Finally, the variables were provided from two sources: the coders hired and trained for the study, as well as the raters and ratees of the multisource performance intervention. The variables of familiarity, acquaintanceship time, and position tenure were collected from the rater and/or ratee at the time of the multisource feedback intervention.

Future Research

An important area of research is the apparent trade off of favorability and goal content. Many of the relationships observed throughout the present study indicated that as

goal content increased, the favorability of the narrative feedback decreased and vice versa. This is important because the purpose of developmental feedback interventions, such as the one used in the present study, are to provide the ratee with feedback to help them improve their performance. David (2013) suggested that there are two components to narrative feedback quality, the motivational component and the directional component. The results of the present study indicate that these two components may be in conflict with one another. Future research should address how favorable the feedback needs to be in order for the ratee to want to act on it, and how much goal content is optimal for feedback acceptance and implementation. It is likely that some sort of balance needs to be found between these two variables and this may be influenced by ratee individual differences.

Future research should also investigate the outcomes of narrative feedback quality. David (2013) found that favorability had direct and indirect effects on year-lagged employee performance. This investigation included supervisory narrative feedback only. We believe it would beneficial to extend this investigation to peer and subordinate narrative feedback as well. As mentioned, evidence shows that employees pay attention to narrative feedback (Antonioni, 1996), often more than they do the numeric ratings (Ferstl & Bruskiewicz, 2000). Therefore, we believe that narrative feedback quality is likely to predict performance outcomes.

An additional concern regarding narrative performance feedback is the potential for differences to occur based on ratee demographics. One such variable is ratee gender. Studies concerning numeric performance ratings suggest that there are situations of gender inequality in performance ratings (Bowen, Swim, & Jacobs, 2015). We believe

that investigating the narrative performance feedback quality of the comments provided to men and women would be a worthwhile and novel perspective for investigating potential gender biases in performance evaluation. We have the capability to look at gender differences in the present study, however it would have doubled the number of research questions, making the project much too large. Additional demographic variables would also be interesting to evaluate, however the present dataset is somewhat limited with what can be investigated due to the overwhelming concern for rater and ratee anonymity through the multisource feedback process.

Research on narrative performance feedback should also look to adopt models that would help structure research moving forward. For instance, Murphy & Cleveland (1995) propose a four-component model of performance evaluation which contains the elements of the rating context, the performance judgment, the performance rating, and the evaluation of the performance appraisal system. This model makes the distinction between the rater's judgements which are the private evaluations of the ratee's performance, and ratings which represent the public statements about the ratee's performance. Similarly, it is quite likely that the narrative feedback provided by raters and their actual judgments regarding the ratee's performance differ. By conducting research to investigate performance judgements and narrative performance feedback, researchers may begin to understand why raters provide the narrative that they do. For instance, in the present data subordinates had comparable familiarity with the ratee's work behavior as supervisors, and significantly higher familiarity than peers. However, none of the relationships between familiarity and narrative feedback quality were significant for subordinates. This may be an indication of a gap between the judgements

and narrative feedback provided by subordinates. This could be investigated using think-aloud methods.

Both absolute feedback content and trait feedback content demonstrated positive relationships with the narrative feedback quality variable of favorability without also demonstrating an effect for goal content. The ratee may be providing overly positive absolute and trait narrative content in an effort to ingratiate the ratee. Thus, the intentions of the ratee should be investigated as a predictor of narrative feedback quality. Spence and Keeping (2013) propose a framework for understanding managers' intentions when rating employee performance. We believe this model could be extended to the narrative feedback domain, as well as to other rating sources. The model investigates a number of rater intentions including the intention to be accurate, the attention to avoid conflict, the intention to be benevolent, and the intention to impression manage. It is likely that the highly favorable nature of absolute and trait feedback content could be explained by the raters' intention to avoid conflict. In this way, raters are likely inclined to provide positive feedback to avoid hurting or angering the ratee. Alternatively, the ratee is likely in a position of power for subordinates and peers which could indicate the intention to impression manage. Raters in this case might be motivated to provide highly positive narrative feedback in order to put themselves in an advantageous position with the ratee. This model should be studied more closely to better understand the different rater intentions with regard to providing narrative performance feedback.

Conclusion

This study's findings have important implications for the collection of narrative feedback in a multisource context. Supervisors provided the highest quality narrative

feedback. Peers and subordinates were comparable with regard to narrative feedback quality. This suggests that when looking for additional narrative feedback, researchers and practitioners should match the additional rating sources to the rating context. The lower narrative feedback quality for peer raters might be partially explained by familiarity, as they reported the lowest familiarity with the ratee's work behavior across all sources. However, familiarity appears to be a good indicator of narrative feedback quality for supervisors and peers. Therefore selecting highly familiar raters may result in higher quality narrative feedback, although this has yet to be tested empirically. Acquaintanceship time tended to be related negatively with narrative feedback quality, suggesting that it should not be used as a proxy for familiarity with the ratees work behavior. When collecting narrative feedback for longer-tenured ratees, peers are likely to provide higher quality feedback. However, all rating sources' narrative feedback quality decreased as ratee position tenure increased. When it comes to the content of the narrative feedback, the results for relative and absolute feedback content suggested that both were related to positive description and little actionable content. This finding was more apparent for absolute feedback content. Additionally, task feedback content was associated with the greatest increases in narrative feedback quality. This suggests that future rater training should focus on how to provide task content feedback to the ratee.

References

Aguinis, H. (2009). An expanded view of performance management. In J. W. Smither, & M. London (Eds.), *Performance management: Putting research into action; performance management: Putting research into action* (pp. 1-43) Jossey-Bass, San Francisco, CA.

Aguinis, H., & Harden, E. E. (2009). In Lance C. E., Vandenberg R. J. (Eds.), *Sample size rules of thumb: Evaluating three common practices* Routledge/Taylor & Francis Group, New York, NY.

Aguinis, H., Werner, S., Abbott, J., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods, 13*, 515-539. doi:http://dx.doi.org/10.1177/1094428109333339

Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. *Organizational Dynamics, 25*, 24-38. doi:http://dx.doi.org/10.1016/S0090-2616(96)90023-6

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. *Journal of Applied Psychology, 77*, 836-874. doi:http://dx.doi.org/10.1037/0021-9010.77.6.836

Balzer, W. K., Greguras, G. J., & Raymark, P. H. (2004). Multisource feedback. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment, vol. 4: Industrial and organizational assessment; comprehensive handbook of psychological assessment, vol. 4: Industrial and organizational assessment* (pp. 390-411) John Wiley & Sons Inc, Hoboken, NJ.

Becker, G. (1964). Human capital: A theoretical and empirical analysis with special reference to education. New York: Columbia University Press.

Bernardin, H.J. (1979) Rater training: A critique and reconceptualization. Proceedings of the Academy of Management, 131-135.

Bernardin, H. J., & Villanova, P. (2005). Research streams in rater self-efficacy. *Group & Organization Management, 30,* 61-88. doi:http://dx.doi.org/10.1177/1059601104267675

Bowen, C., Swim, J. K., & Jacobs, R. R. (2000). Evaluating gender biases on actual job performance of real people: A meta-analysis. *Journal of Applied Social Psychology, 30,* 2194-2215. doi:http://dx.doi.org/10.1111/j.1559-1816.2000.tb02432.x

Bracken, D. W., & Rose, D. S. (2011). When does 360-degree feedback create behavior change? and how would we know it when it does? *Journal of Business and Psychology, 26,* 183-192. doi:http://dx.doi.org/10.1007/s10869-011-9218-5

Bretz, R. D., Ash, R. A., & Dreher, G. F. (1989). Do people make the place? an examination of the attraction-selection-attrition hypothesis. *Personnel Psychology, 42,* 561-581. doi:http://dx.doi.org/10.1111/j.1744-6570.1989.tb00669.x

Brutus, S. (2010). Words versus numbers: A theoretical exploration of giving and receiving narrative comments in performance appraisal. *Human Resource Management Review, 20,* 144-157. doi:http://dx.doi.org/10.1016/j.hrmr.2009.06.003

Cascio, W.F. & Aguinis, H (2011*). Applied psychology in human resource management* (7th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc. Chapter 5.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.) Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Cortina, J. M., & Landis, R. S. (2009). In Lance C. E., Vandenberg R. J. (Eds.), *When small effect sizes tell a big story, and when large effect sizes don't.* Routledge/Taylor & Francis Group, New York, NY.

David, E. M. (2013). Examining the role of narrative performance appraisal comments on performance. *Human Performance, 26,* 430-450. doi:http://dx.doi.org/10.1080/08959285.2013.836197

DeNisi, A. S., & Pritchard, R. D. (2006). Performance appraisal, performance management and improving individual performance: A motivational framework. *Management and Organization Review, 2,* 253-277. doi:http://dx.doi.org/10.1111/j.1740-8784.2006.00042.x

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62,* 107-115. doi:http://dx.doi.org/10.1111/j.1365-2648.2007.04569.x

Ferstl, K. L., & Bruskiewicz, K. T. (2000). *Self–other agreement and cognitive reactions to multirater feedback*. Paper presented at the 15th annual conference of the Society of Industrial and Organizational Psychology, New Orleans, LA.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117-140. doi: https://doi.org/10.1177/001872675400700202

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin, 138,* 296-321. doi:http://dx.doi.org/10.1037/a0026556

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102,* 652-670. doi:http://dx.doi.org/10.1037/0033-295X.102.4.652

Goffin, R. D., Gellatly, I. R., Paunonen, S. V., Jackson, D. N., & Meyer, J. P. (1996). Criterion validation of two approaches to performance appraisal: The behavioral observation scale and the relative percentile method. *Journal of Business and Psychology, 11*, 23-33. doi:http://dx.doi.org/10.1007/BF02278252

Goffin, R. D., Jelley, R. B., Powell, D. M., & Johnston, N. G. (2009). Taking advantage of social comparisons in performance appraisal: The relative percentile method. *Human Resource Management, 48*, 251-268. doi:http://dx.doi.org /10.1002/hrm.20278

Goffin, R. D., & Olson, J. M. (2011). Is it all relative? comparative judgments and the possible improvement of self-ratings and ratings of others. *Perspectives on Psychological Science, 6*, 48-60. doi:http://dx.doi.org/10.1177/1745691610393521

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal, 50*, 1348-1352. doi:http://dx.doi.org/10.5465/AMJ.2007.28166119

Ilgen, D. R., & Moore, C. F. (1987). Types and choices of performance feedback. *Journal of Applied Psychology, 72*, 401-406. doi:http://dx.doi.org/10.1037/0021-9010.72.3.401

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback

intervention theory. *Psychological Bulletin, 119*, 254-284. doi:http://dx.doi.org/10.1037/0033-2909.119.2.254

Kristof-Brown, A., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individual's fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58*, 281-342. doi:http://dx.doi.org/10.1111/j.1744-6570.2005.00672.x

Kruglanski, A. W., & Mayseless, O. (1990). Classic and current social comparison research: Expanding the perspective. *Psychological Bulletin, 108*, 195-208. doi:http://dx.doi.org/10.1037/0033-2909.108.2.195

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*, 72-107. doi:http://dx.doi.org/10.1037/0033-2909.87.1.72

Locke, E. A., & Latham, G. P. (1984). *Goal setting: A motivational technique that works*. Englewood Cliffs, NJ: Prentice Hall.

Ludwig, T. D., & Goomas, D. T. (2009). Real-time performance monitoring, goal-setting, and feedback for forklift drivers in a distribution centre. *Journal of Occupational and Organizational Psychology, 82*, 391-403. doi:http://dx.doi.org/10.1348/096317908X314036

McEnrue, M. P. (1988). Length of experience and the performance of managers in the establishment phase of their careers. *Academy of Management Journal, 31*, 175-185. doi:http://dx.doi.org/10.2307/256504

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Sage Publications, Inc, Thousand Oaks, CA.

Ng, T. W. H., & Feldman, D. C. (2010). Organizational tenure and job performance. *Journal of Management, 36,* 1220-1250. doi:http://dx.doi.org/10.1177/0149206309359809

Nowack, K. M., & Mashihi, S. (2012). Evidence-based answers to 15 questions about leveraging 360-degree feedback. *Consulting Psychology Journal: Practice and Research, 64,* 157-182. doi:http://dx.doi.org/10.1037/a0030011

O'Neill, T. A., McLarnon, M. J. W., & Carswell, J. J. (2015). Variance components of job performance ratings. *Human Performance, 28,* 66-91. doi:http://dx.doi.org/10.1080/08959285.2014.974756

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Schneider, B., Goldstein, H. W., & Smith, D. B. (1995). The ASA framework: An update. *Personnel Psychology, 48,* 747-773. doi:http://dx.doi.org/10.1111/j.1744-6570.1995.tb01780.x

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

Spence, J. R., & Keeping, L. M. (2013). The road to performance ratings is paved with intentions: A framework for understanding managers' intentions when rating employee performance. *Organizational Psychology Review, 3,* 360-383. doi:http://dx.doi.org/10.1177/2041386613485969.

3D Group. (2013). *The 10 Most Common 360-Degree Feedback Practices in 2009. 3D Group Technical Report #8326E*. Berkley, CA: Data Driven Decisions, Inc.

Vigoda-Gadot, E., & Angert, L. (2007). Goal setting theory, job feedback, and OCB: Lessons from a longitudinal study. *Basic and Applied Social Psychology, 29*, 119-128. doi:http://dx.doi.org/10.1080/01973530701331536

Wagner, J. A., Ferris, G. R., Fandt, P. M., & Wayne, S. J. (1987). The organizational tenure—job involvement relationship: A job-career experience explanation. *Journal of Occupational Behaviour, 8*, 63-70. doi:http://dx.doi.org/10.1002/job.4030080108

Wilson, K. Y. (2010). An analysis of bias in supervisor narrative comments in performance appraisal. *Human Relations, 63*, 1903-1933. doi:http://dx.doi.org/10.1177/0018726710369396

Appendix A

Leadership Development Instrument Factors and Sample Behaviors

Factor	Behavior	Definition
Cognitive Managerial Skills		Characterized by decision making, problem solving, analytical skills, technical proficiency, and the ability to demonstrate creativity and objectivity in working through problems, decisions, and risks.
	Decisiveness	The ability to make clear-cut and timely decisions with the appropriate amount of information.
	Analytical Orientation	Demonstrating a preferences for problems requiring precise, logical reasoning, and showing an ability to dissect and understand complex, multifaceted problems.
	Creativity	Demonstrating the ability to initiate original and innovative ideas, products, and approaches.
Interpersonal Managerial Skills		Working effectively and cooperatively with people, and maintaining positive interpersonal relationships.
	Social Astuteness	The ability to accurately read and respond diplomatically to organizational trends and norms, as well as effectively deal with organizational politics.
	Conflict Management	The ability to mediate and resolve conflicts and disagreements in a manner best for all parties involved.
	Listening	A willingness to take the time to listen to others' questions and concerns, and to hear their points of view on workplace issues.
Personal Managerial Skills		The ability to self-manage, remain focused, and encourage subordinates through support and understanding.

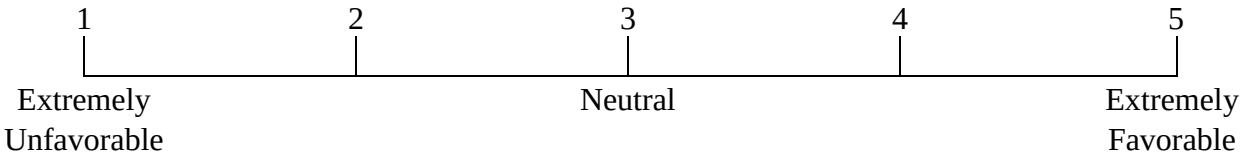
	General Leadership Effectiveness	Influencing and guiding the behavior of others in a certain direction by providing motivation, coaching, and support.
	Self-Discipline	The ability to resist impulse, remain focused, and see a project through to completion.
	Dependability	The ability to be counted on to meet commitments and deadlines.
Teamwork, Supervision, Planning, & Productivity		Capabilities involving setting clear and inspirational objectives, planning and initiating structure, communicating performance expectations and priorities, and monitoring employee and team progress toward long-term goals.
	Inspirational Role Model	The ability to set a positive and inspirational example for subordinates to follow.
	Motivating Others	Showing enthusiasm and providing encouragement, recognition, constructive criticism, and coaching to subordinates.
	Organizing the Work of Others	Clearly defining roles and responsibilities for subordinates, and letting them know exactly what tasks should be done and what results are expected.

Appendix B

Narrative Feedback Quality Scales

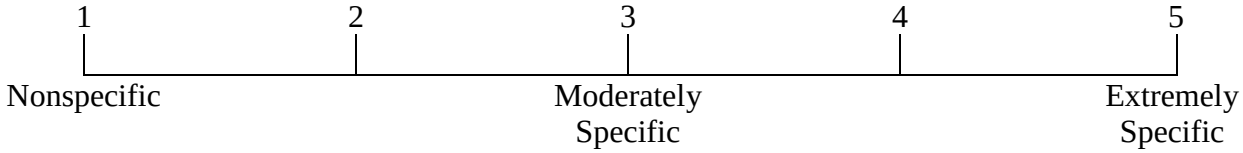
Favorability

The degree to which the feedback is positive or negative.



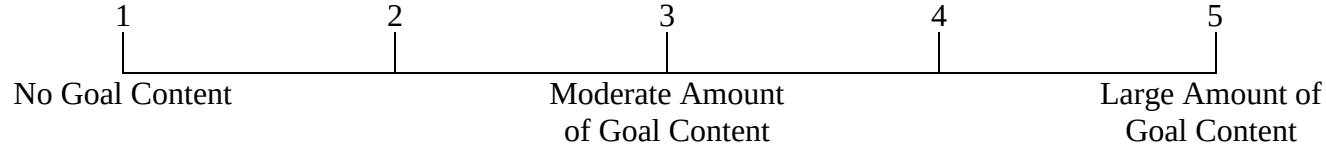
Specificity

The degree to which the feedback provided is detailed and supported by behavioral examples.



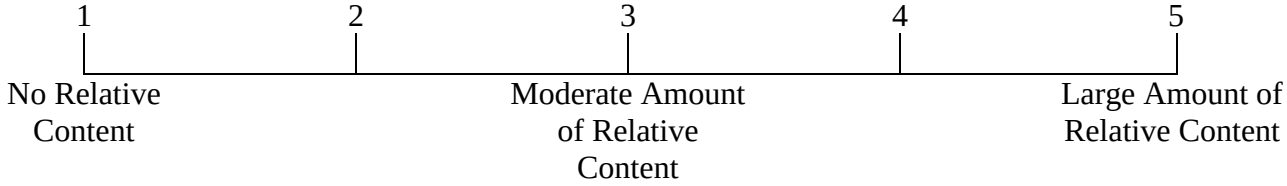
Goal content

The degree to which the rater provides the ratee with actionable steps to improve performance.



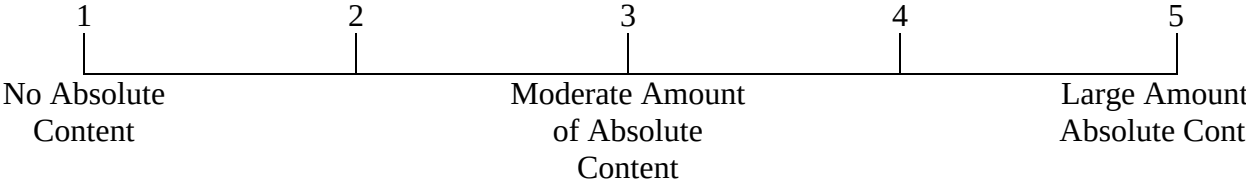
Relative content

The degree to which comparative language is used to describe the performance rating.



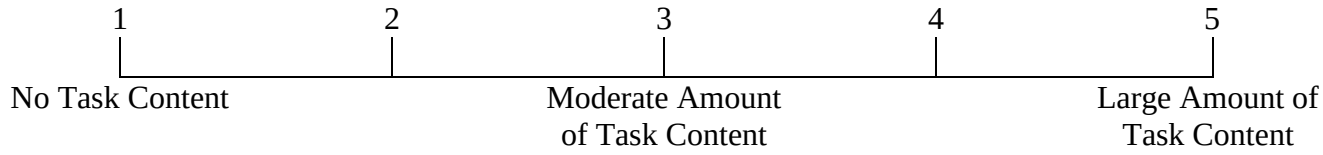
Absolute content

The degree to which non-comparative language is used to describe the performance rating.



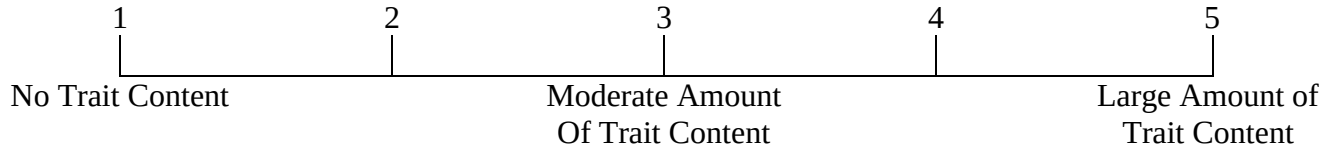
Task content

The degree to which the target's behavior on a task are the focus of the performance rating.



Trait content

The degree to which the target's personality traits or attributes are the focus of the performance rating.



Appendix C

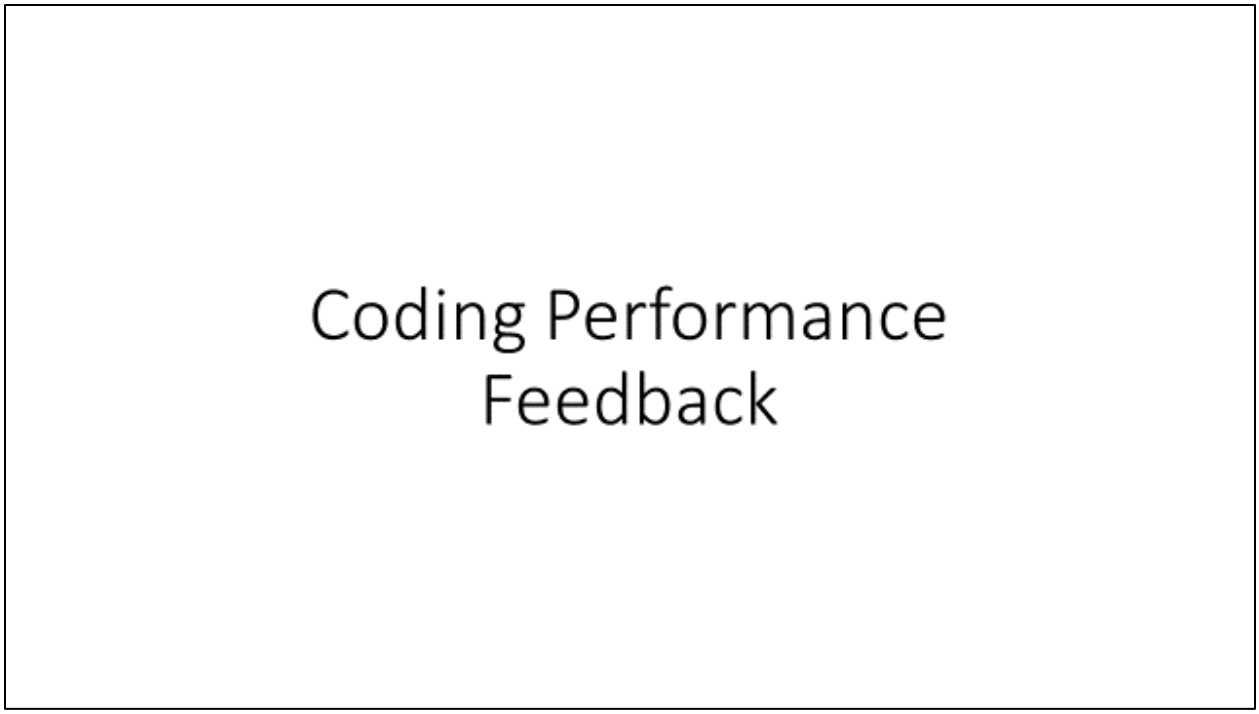
Sample Rater Comments

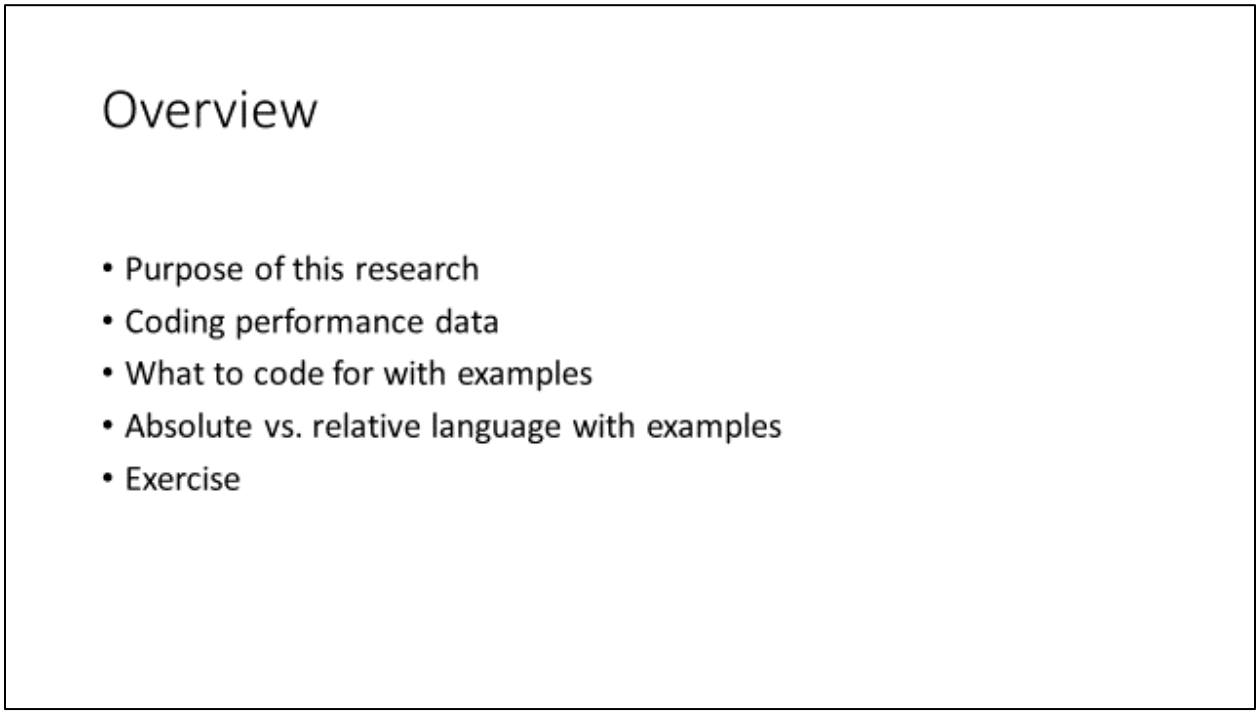
Item	Comment
Presenting a positive role model for other people at work, demonstrating by example how to achieve organizational objectives.	I do not know how to advise the step required to move to "inspirational" however, as a role model she exudes professionalism and a solid commitment to the company, the employees and the members overall.
Involving subordinates in the formulation, evaluation, and implementation of business decisions and work projects.	TARGET is highly effective in this behaviour. Examples include: effective delegation of meaningful tasks to subordinates; involving subordinates (specifically the Director of Finance) in senior management decision making forums.
Cultivating a sense of teamwork and cohesion; acting to increase the effectiveness of the group as a whole.	TARGET is one of our best practitioners of teamwork.
Persuading people to adopt particular courses of action.	TARGET should be more forceful at times when attempting to influence others on a particular point of view.
Places a high value on interpersonal relationships and continuously promotes the development of these relations.	Not only promoting such but also trying to recover those relationships that have slipped.
Readily approaching and conversing with others on the job.	She demonstrates her interest in others and provides a warm reception for everyone.
Directing others to carry out work responsibilities on one's behalf.	TARGET, you are very aware of work and other pressures on your subordinates and you make every effort to ensure their workload is balanced, however, you sometimes do that at your own expense. You may try to delegate even more to them - while monitoring the impact. You may be pleasantly surprised!
Helping to retain the best workers in the organization.	TARGET has had no negative staff turnover to my knowledge. Staff appear to be very competent.

Making clear and timely decisions in the face of competing priorities or ideas.	TARGET is constantly improving in this area. I have been critical of her need to gather more and more facts when often additional facts do not alter the decision but may delay it unnecessarily. She is doing much better in decision making.
Expecting and communicating high standards of performance for both oneself and for others.	Has personal high standards. Not always expected of or to subordinates.
Demonstrating an ability to influence, direct, assist, train and motivate others' work.	I am confident that the leadership she provides to her areas is 100% on all these factors.
Gathering and analyzing information, and evaluating the performance of others to determine if progress is on track. Exercising legitimate control over the organization and its members.	My only concern is that the volume of activity by subordinates is sometimes skewing the results rather than the right activities skewing the results. She and her area are the primary monitors and controllers of the entire organization and it is carried out quite well especially as the focus is more on the remedial rather than the history.
Creating a work environment that attracts people that fit with the organization and the job and selecting those likely to be effective.	can only provide moderate as organization allows
Defining precisely the work roles and tasks of others, including the relative importance of the tasks.	I have less observation on this one and am providing perception more than fact. I have seen alignment through the new scorecards but this is just recent and requires one year to see if it fulfills the focus on priorities.
Helps others to energize, direct, and maintain high levels of appropriate work behavior.	I believe she does this quite well within her units and amongst her peer group.
Keeping leaders and people in authority well informed about key issues.	TARGET is diligent on keeping those that need to know in the know in a timely manner even if it is devastating news.
Keeping direct and insubordinates well informed on key issues.	From what I have witnessed or received feedback she is timely and informative with subordinates and allows ample opportunity for them to provide feedback.

Appendix D

Rater Training Slides





Why is this research important?

- Providing feedback is an essential part of evaluating workplace performance
- Formal feedback processes can affect the development, productivity and future of the employee
- Almost no research has been done assessing feedback employees provide during performance evaluations
- We believe that certain characteristics of the feedback itself may be associated with the numerical ratings being provided to employees

Coding Feedback

- In order to examine relationships between performance feedback and numeric evaluations we need to code the performance feedback
  - First, read the definition of the variable
  - Then, read performance data and decide the extent to which the variable is being communicated
  - Find key ideas, points or words in order to come up with your ratings

Variables We Need to Code

- There are 3 variables that will be coded
  - Favorability
  - Specificity
  - Goal Content

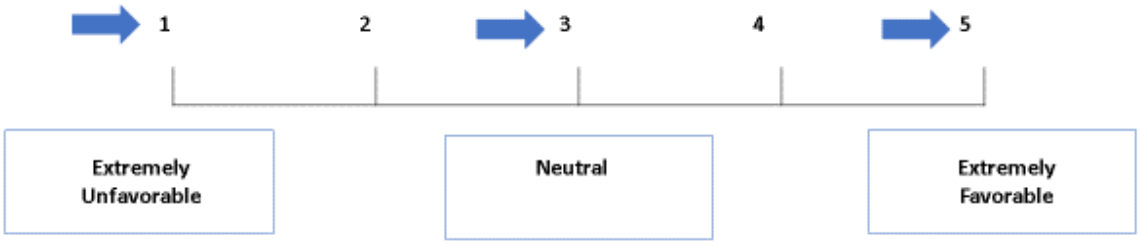
The Process

- In the next several slides, we will define each variable
- A sample comment from the data set will be displayed
- Decide what you think the rating should be (1-5)
- The ratings will be discussed and explained

Favorability

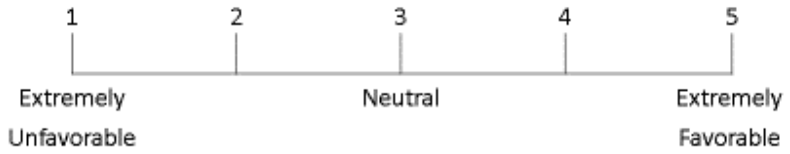
- The degree to which the feedback is positive or negative.

Favorability



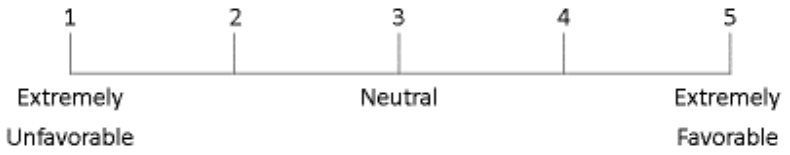
Example

- "TARGET has ambition but needs to nail the tasks that have been presented in her role at the moment. TARGET needs to embody what it is to be a leader and with this increased influence and respect will come. This comes down to very simple elements in her day to day work."
  - Favorability rating: 2



Example

- "Overall, TARGET does this well. I think, and as we have discussed, TARGET needs to think about how he can better leverage his administrative resources"
  - Favorability rating: 4



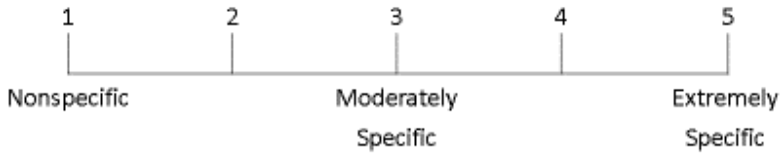
Specificity

- The degree to which the feedback provided is detailed and supported by behavioral examples.

Specificity

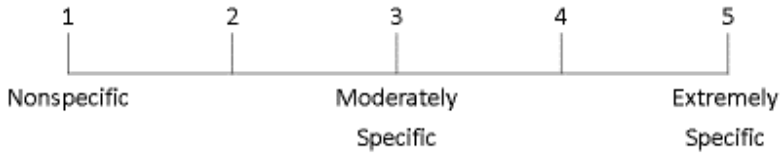
Example

- "TARGET needs to understand the company big picture in relation to cost and margin."
  - Specificity rating: 2



Example

- "TARGET sets tasks for people that report to her - she will check on the progress at regular times. I don't think TARGET utilizes the up and coming teams to help and train (due to time restraints)"
  - Specificity rating: 5



Goal Content

- The degree to which the rater provides the ratee with actionable steps to improve performance.

Goal Content

Example

- "This is the biggest area of opportunity for TARGET."
  - Goal content rating: 2

Example

- "In general, TARGET's ability to prioritize is good, but when there are a very high number of competing demands on his time, his prioritization skills are minimized. Some suggestions: keep notes, make better use of whiteboards in office to keep track of projects and assignments to staff (and keep them updated), get regular updates on large projects from staff, not just through email but in person so that questions can be asked/answered."
  - Goal content rating: 5

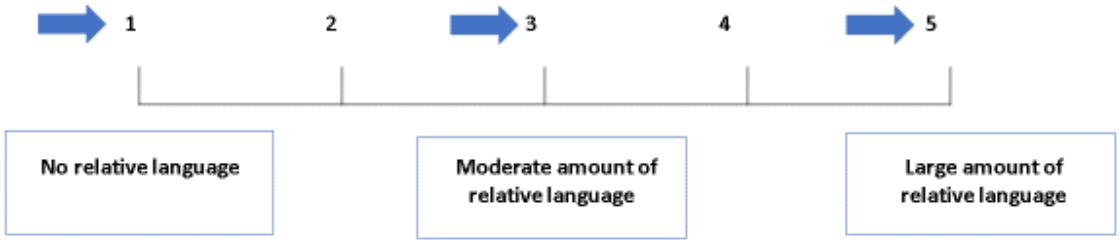
Relative and Absolute Language

- **Absolute language** compares the behavior of the target to standards of performance, using terms such as favorable, unfavorable, excellent, poor, great, etc.
  - No reference point
- **Relative language** compares the behavior of the target against a "pool" of people, using terms such as average, above average, below average, better than most, etc.
  - Uses a reference point

Relative Language

- The extent to which comparative language is used to describe the performance rating.

Relative Rating



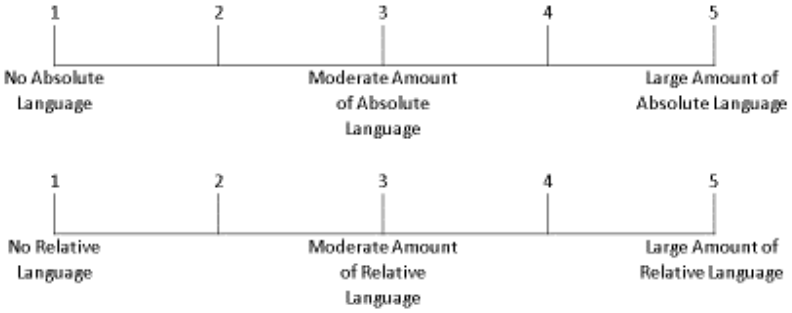
Absolute Language

- The extent to which non-comparative language is used to describe the performance rating.

Absolute Rating

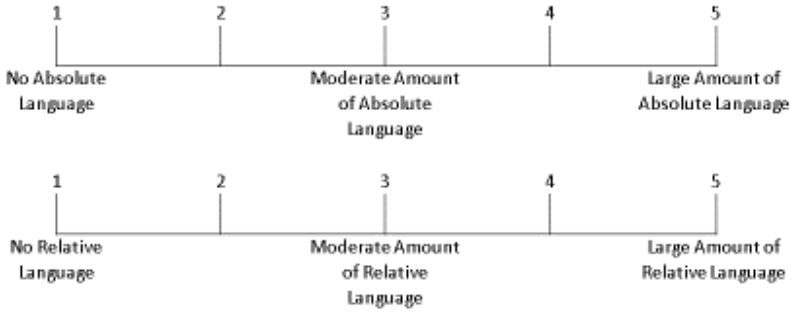
Example 7

- "TARGET works hard and is task driven. Successful when it comes to execution."



Example 8

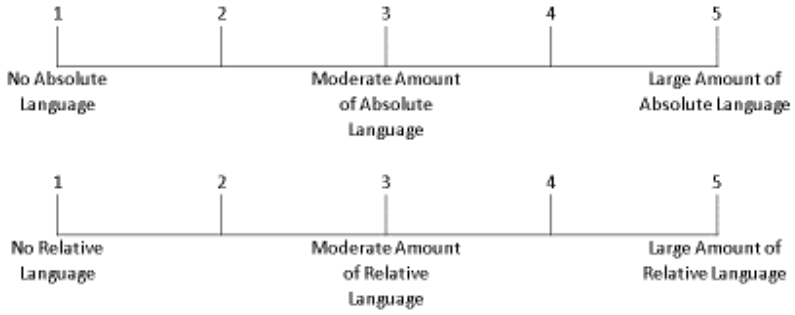
- "The agile methodology has helped TARGET facilitate this behavior and he has done well with embracing it."



+

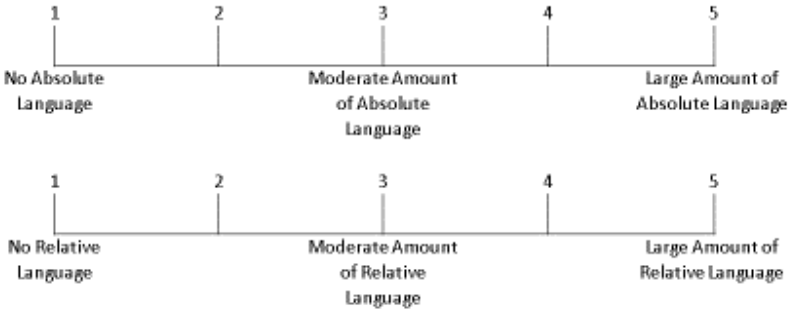
Example 9

- "the achievement of your goals will be dependant on how you respect the goals of others"



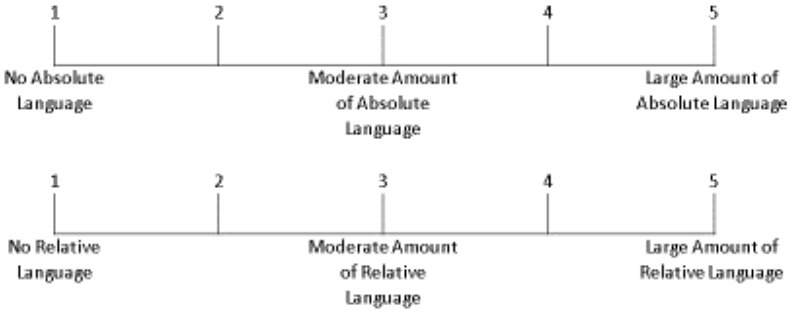
Example 10

- "The below average rating may only reflect the reality of organizational priorities and constraints"



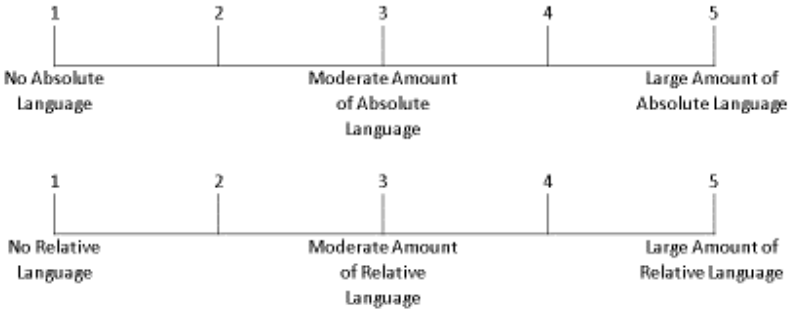
Example 11

- "TARGET works hard and is an above average employee. His work is impeccable and precise."



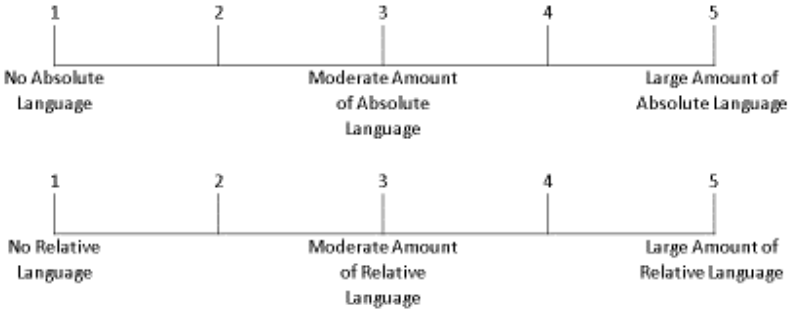
Example 12

- "TARGET works hard and for an employee in his position he definitely accomplishes an above average amount of quantity and quality of work. He has a vested interest in all areas of the branch."



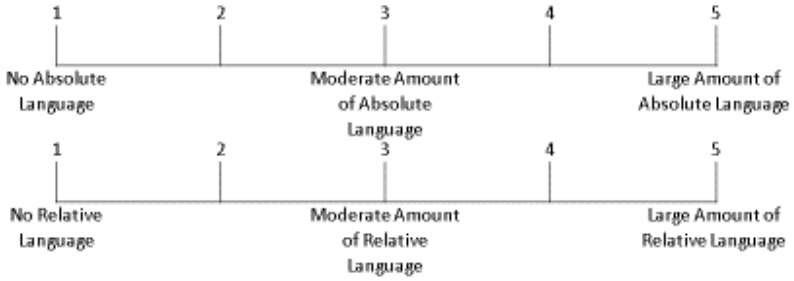
Example 13

- "The appearance of this question caused me to have an epiphany. TARGET is one of the most objective people with whom I have interacted in my career."



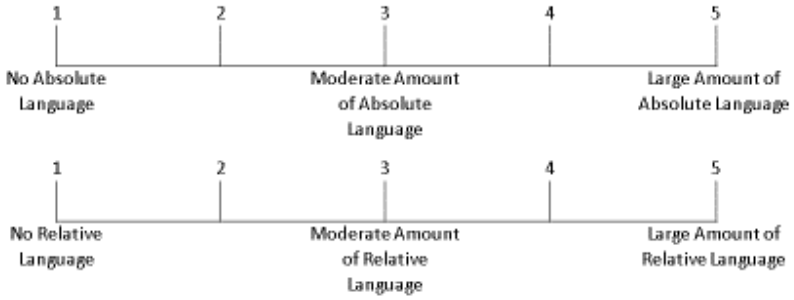
Example 14

- "TARGET motivates the team effectively. TARGET gives informative feedback to ensure all members meet potential and does so in such a way that employees do not feel intimidated. This is not something that I have seen with other leaders. TARGET is above average in inspiring the team to work hard and to meet potential. Definitely the most inspiring leader I've worked with."



Example 15

- "Not the best at managing time. Seems overwhelmed at times. Looks flustered more so than previous leader and many last minute projects compared to the last manager. Does not seem like time management is a strength due to her below average performance."



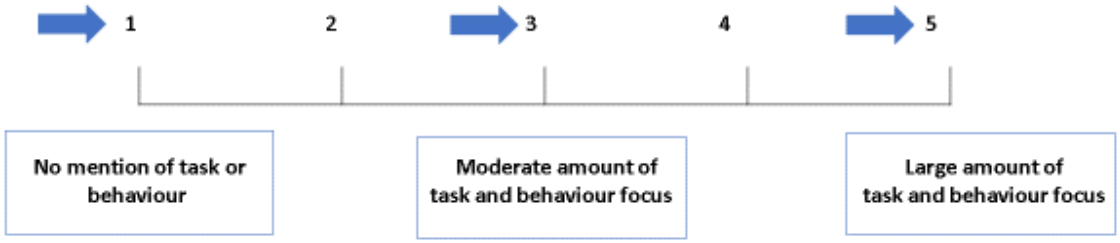
Trait & Task Based Comments

- Trait-based comments are those that focus on the person's attributes.
- Task-based comments are those that focus on the person's ability and behaviour when carrying out certain tasks.

Task Rating

- The extent to which the person's ability and behaviour on a task are the focus on the performance rating.

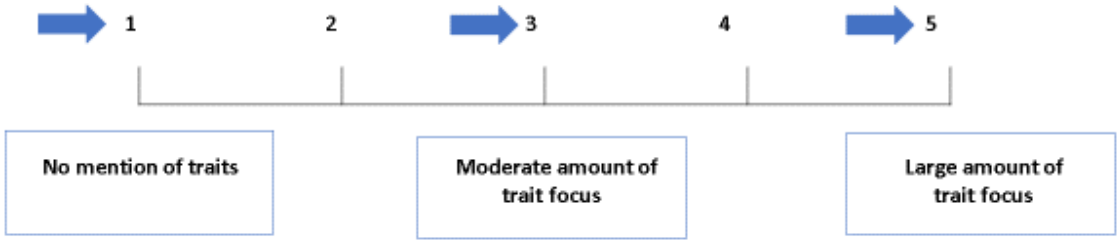
Task Rating



Trait Rating

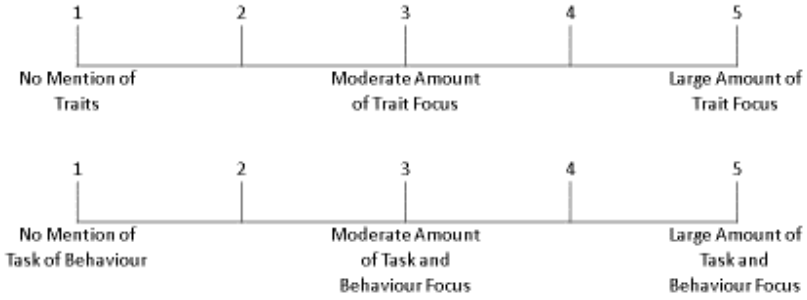
- The extent to which the person's personality traits and attributes are the focus of the performance rating

Trait Rating



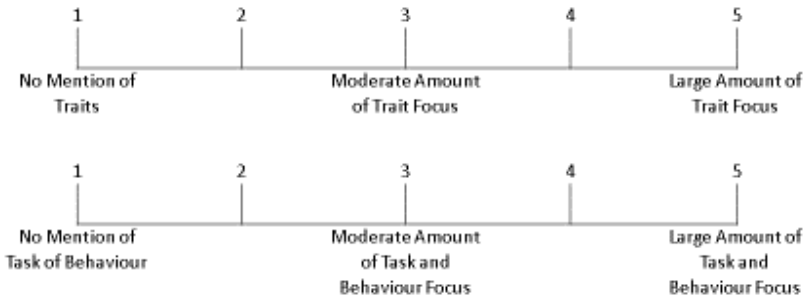
Example 16

- "Generally socially aware but can be quite insensitive on occasions. I have seen situations where he was downright rude to a customer."



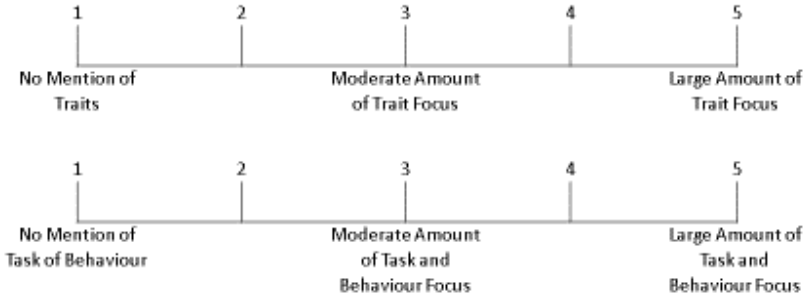
Example 17

- Be open to Staff that have the abilities you are looking for; not just a title.



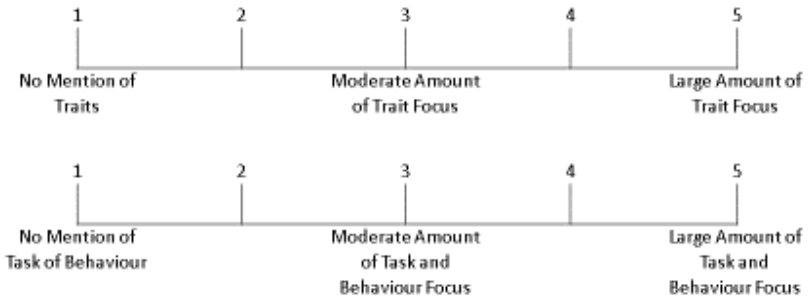
Example 18

- Very intelligent and can work through to a solution.



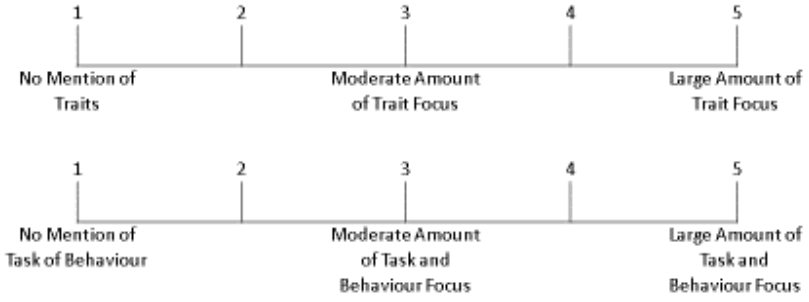
Example 19

- TARGET tends to be a bit risk adverse and will proceed cautiously only once risks have been identified and quantified when ever possible.



Example 20

- Highly efficient presenter. Always pitches well.



Curriculum Vitae

Name: Kevin M. Doyle

Post-secondary Education and Degrees:

University of Western Ontario
London, Ontario, Canada
2013-2018 Ph.D.
Dissertation: "Examining Rating Source Differences in Narrative Performance Feedback"
Supervisor: Dr. Richard Goffin

University of Western Ontario
London, Ontario, Canada
2011-2013 M.Sc.
Thesis: "Subordinate Ratings of Supervisor Performance: Balancing Accountability and Anonymity"
Supervisor: Dr. Richard Goffin

University of Western Ontario
London, Ontario, Canada
2005-2010 HB.A.
Thesis: "Facet-level Personality Predictors of Perceptions of Group Processes"
Supervisor: Dr. Thomas O'Neill

Honors and Awards:

Council of Canadian Departments of Psychology
Certificate of Teaching Excellence
2017

Human Resources Professionals Association
London & District Chapter Scholarship
2015-2016

Province of Ontario Graduate Scholarship (OGS)
Doctoral Scholarship
2014-2015

Province of Ontario Graduate Scholarship (OGS)
Doctoral Scholarship
2013-2014

Social Science and Humanities Research Council (SSHRC)
Masters Scholarship
2012-2013

Related Work Experience:

Talent Management Consultant, Assessment Practice Lead
Carswell Partners Inc.
2014-2018

Teaching Assistant
University of Western Ontario
2011-2017

Talent Management Consultant
Jackson Leadership
2015

Research Assistant
University of Western Ontario
2010-2011

Service:

Coordinator of Graduate Student Recruitment
University of Western Ontario, I/O Psychology
2014-2017

Conference Co-organizer
South Western Ontario Industrial/Organizational Student Conference
2012, 2016

Executive Member
Psychology Graduate Student Association
2012-2013

Journal Articles:

Doyle, K. M., & Goffin, R. D. (in press). Accountability and Accuracy in Subordinates' Ratings of Supervisors' Performance. *Social Behavior and Personality, 46*. doi:http://dx.doi.org/10.2224/sbp.6967

Non-Refereed Contributions:

Doyle, K. M. (2013) How to Capitalize on Subordinate and Peer Raters. *Public Sector Digest*.

Manuscripts under Review:

Doyle, K. M., Goffin, R. D., & Woycheshin, D. E. Peer-Rated Organizational Citizenship Behavior: Does Familiarity Improve Rating Quality? Manuscript submitted to *Journal of Personnel Psychology*. Revise & resubmit stage.

Feeney, J. R., Goffin, R. D., Daljeet, K. N., Schneider, T. J., & Doyle, K. M. The Soft Underbelly of Socio-comparative Ratings? Accuracy, Leniency, and Reactions from Ratees and Raters. Manuscript submitted to *International Journal of Selection and Assessment*.

Refereed Conference Publications:

Doyle, K. M., Goffin, R. D., Factor, R., Daljeet, K. N., Feeney, J. R., & Carswell, J. J. (2018). *Examining Rating Source Difference in Multisource Narrative Performance Feedback.* Poster presented at the 33rd annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Feeney, J. R., Goffin, R. D., Daljeet, K. N., Factor, R. J., & Doyle, K. M. (2018). *Multisource Performance Management: Improving Rater Agreement and Reducing Leniency.* Poster presented at the 33rd annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Doyle, K. M., Goffin, R. D., & Daljeet, K. N. (2016). *Increasing Acceptance of Work Performance feedback through Social Comparison.* Poster presented at the 28thAssociation for Psychological Science annual convention, Chicago, IL.

Doyle, K. M., Goffin, R. D., & Woycheshin, D. E. (2016). *Peer-Rated Contextual/Citizenship Performance: Matching Construct to Rating Source.* Poster presented at the 31st annual conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.

Doyle, K. M., Goffin, R. D., & Woycheshin, D. E. (2015). *The Discrepancy Between Self and Others' Ratings of Contextual/Citizenship Behavior as a Function of Impression Management and Self-Deceptive Enhancement.* Poster presented at the meeting of The International Society for the Study of Individual Differences Conference, London, Ontario.

Doyle, K. M., & Goffin, R. D. (2015). *Subordinate Ratings of Supervisor Performance: Balancing Accountability and Anonymity*. Poster presented at the annual meeting of the Canadian Psychological Association, Ottawa, ON.

Doyle, K. M., Goffin, R. D., & Woycheshin, D. E. (2015). *Contextual/Citizenship Performance: Do Peer-ratings Reflect Rater or Rated?* Poster presented at the 30th annual conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Doyle, K. M., O'Neill, T. A., Allen, N. A. (2011). *Personality and the prediction of group processes and emergent states*. Poster presented at the annual meeting of the Canadian Psychological Association, Toronto, ON.