

Are Traditional Performance Reviews Outdated? An Empirical Analysis on Continuous, Real-Time Feedback in the Workplace

Online Appendix A

Table A.1 Selective Review of Performance Feedback Studies and Positioning of the Current Study

Study	Methodology	Units of Analysis	Independent variables	Dependent variables	Moderating variables	Feedback application-based?	Feedback real-time?
Current Study	Observational study in several organizational settings using real-time feedback application	Actual feedback instances	Anonymity, relationship	Rating score (0-5)	Previous scores received and given, tit-for-tat strategy, gender	Yes	Yes
Belschak & Den Hartog (2009)	Scenario studies of public or private feedback; subsequent survey	Individual survey responses	Positive or negative valence	Work attitudes and behavioral intentions	Emotional reaction to feedback	No	No
Britt et al. (2016)	Survey of employed persons	Individual survey responses	Gender, training	Feedback behaviors	Training impacted giving but not receiving feedback	No	No
Cooper (2006)	Observational study in paper mill	Supervisors' daily safety observations	Observation frequency, goal-setting, feedback meetings, feedback data	Safety performance	Visible ongoing support	No	No

Festing et al. (2014)	Survey of top management in five countries	Individual survey response	Gender	Satisfaction with performance management systems	None	N/A	N/A
Hekman et al. (2017)	Field and lab test of 350 executives rated by peers and bosses	Actual feedback instances	Gender, race	Performance rating score	Leader competence, diversity-valuing behavior	No	Yes
Hsia et al. (2016)	Online peer feedback system and interviews with 100 dance students	Actual feedback instances	Peer ratings vs peer ratings plus feedback	Peer and group dance performance improvement	Intrinsic motivation, self-efficacy	No	No
Huang et al. (2019)	Online performance feedback interactions to motivate user-generated content	Actual feedback instances	Message framing (competitive, cooperative, or individualistic)	Motivation	Gender	Yes	No
Ilies et al. (2007)	Survey and lab test of 197 undergraduates completing task and receiving feedback	Individual survey response	Performance feedback valence (positive vs negative)	Individual affect response	Self-esteem	No	Yes
Jung & Sosik (2003)	31 student work groups performed decision-making tasks over 15 weeks	Performance accuracy	Individual and group performance feedback	Subsequent performance improvement	Group potency, collective efficacy	No	No

Kurtzberg et al. (2005)	Peer appraisal in pen-and-paper and online conditions	Actual feedback instances	Paper vs e-mail feedback	Positive or negative valence	Feelings of social obligation	No	No
Lefroy et al. (2017)	Focus groups and interviews with feedback application users	Individual survey response	Use of feedback application	Rating of application's utility and usability	Comfort and familiarity with application	Yes	No
Lizzio et al. (2003)	Management and undergraduate survey	Individual survey response	Gender	Assessment of features of effective feedback	None	No	No
Motro & Ellis (2017)	Lab study assessed participants' responses to video of feedback recipients	Individual survey response	Employee reaction to negative feedback	Manager appraisal of employee reaction	Gender, crying, perceived typicality	No	No
Pampino et al. (2004)	Two-part study in a retail setting	Task completion	Task clarification, goal setting, access to rewards, and graphic feedback	Completion of routine maintenance tasks	Feedback frequency	No	No
Steelman & Rutkowski (2004)	Survey of 405 employees	Individual survey response	Feedback valence (positive or negative)	Employee motivation to improve	Source credibility, feedback quality, feedback delivery	No	No

Online Appendix B

We take a log form of our dependent variable (the score rating) and rerun our model. The results are robust and are presented in Table B.1.

Table B.1 The Impact of Rating Relationship on Feedback Scores: Log Model

VARIABLES	Full Sample
Subordinate	0.0454*** [3.512]
Peer	0.0499*** [4.754]
Anon	-0.0910*** [-7.532]
AveRecipientGotRated	-0.00493 [-0.461]
AveGiverRatedOthers	-0.0249* [-1.742]
AveGiverGotRated	0.111*** [5.581]
AveRecipientRatedOthers	0.0503*** [4.563]
RecipientTime	0.000151 [1.552]
Organization dummies	Yes
Group dummies	Yes
Competency dummies	Yes
Observations	4,975

Robust t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

Online Appendix C

To help us identify the latent topics of raters’ comments, we use latent Dirichlet allocation (LDA), which is a Bayesian statistical and information retrieval technique (Blei et al. 2003; Lee et al. 2016). The input to the LDA topic model is a set of documents, which in our study is the text of raters’ comments. The output of LDA is a predefined number (K) of topics from these posts. In addition, LDA provides a posterior topic distribution over all K topics for each post. In other words, for each comment m , the LDA model outputs a K -vector, $\langle T_{m,1}, T_{m,2}, \dots, T_{m,K} \rangle$, where $T_{m,k}$ represents the weight of topic k associated with comment m . Also, for each comment m and topic k , $T_{m,k} \geq 0$, and $\sum_{k=1}^K T_{m,k} = 1$.

We run the LDA model with $K = 3, 5$, and 10 topics, and find that 3 is the optimal number of latent topics in terms of the posterior log-marginal likelihood (Nam et al. 2017). A list of the most frequent words associated with each latent topic is shown in Table C.1. We can see that topics 2 and 3 are more likely to be associated with personal opinions. In contrast, topic 1 is more likely to be associated with giving specific comments. For each comment m , if the weight of topic 1 is greater than the sum of the weights of topics 2 and 3, then we say this rater receives event-based feedback. Then, we construct two dummy variables: (i) *InitialEvent_j*, which indicates whether a recipient whose initial score is less than 4 and receives event-based feedback, and (ii) *InitialNotEvent_j*, which indicates whether a recipient whose initial score is less than 4 and does not receive event-based feedback.

Table C.1 Keywords Associated with Each Topic

Topic #	Keywords
1	work, thanks, job, team, your, problem, improve, change, issue, help, always, appreciate, feedback, keep, helpful, our, support, make, professional, care, leader, beyond, training, hard, think, advice, continue, grow, provide, presentation, learning
2	great, very, best, really, good, much, pleasure, positive, awesome, love, well, excellent, truly, nice, wonderful, most, amazing, greatly, dedication, outstanding, happy, friendly, fantastic, like, exceptional, impressive
3	not, difficult, especially, stressful, negative, critical, complex, haven’t, probably, confusing, easily, less, long, harder, limited, lost, properly, lack, doesn’t, frustrated, overall, lose, snappy, hazard

We estimate the following regression equation:

$$\begin{aligned}
 Score_{ijt} = & a_i + \beta_0 + \beta_1 Subordinate_{ij} + \beta_2 Peer_{ij} + \beta_3 Anon_{ij} + \beta_4 AveRecipientGotRated_{jt} + \\
 & \beta_5 AveGiverRatedOthers_{it} + \beta_6 AveGiverGotRated_{it} + \beta_7 AveRecipientRatedOthers_{jt} + \\
 & \beta_8 RecipientTime_{jt} + \beta_9 RecipientTime_{jt} * (InitialEvent_j) + \beta_{10} RecipientTime_{jt} * \\
 & (InitialNotEvent_j) + \beta_{11} InitialEvent_j + \beta_{12} InitialNotEvent_j + controls + \varepsilon_{ijt}.
 \end{aligned}$$

Table C.2 The Impact of Rating Relationship on Feedback Scores: Event-Based Feedback

VARIABLES	(2) Full Sample
Subordinate	0.188*** [4.418]
Peer	0.204*** [5.732]
Anon	-0.251*** [-6.217]
AveRecipientGotRated	-0.000869 [-0.0132]
AveGiverRatedOthers	-0.0737 [-1.431]
AveGiverGotRated	0.337*** [4.841]
AveRecipientRatedOthers	0.153*** [4.322]
RecipientTime	0.000492 [1.448]
RecipientTime*InitialEvent	0.00481** [2.187]
RecipientTime*InitialNotEvent	0.00136 [1.625]
InitialEvent	0.0116** [2.032]
InitialNotEvent	-0.0134 [-0.547]
Organization dummies	Yes
Group dummies	Yes
Competency dummies	Yes
Observations	4,975

Robust t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

The estimation results are presented in Table C.2. We find that the coefficient on $RecipientTime_{jt} * (InitialEvent_j)$ is positive and significant, and the coefficient on $RecipientTime_{jt} * (InitialNotEvent_j)$

is not significant. It suggests that (i) the scores of recipients who receive initial low scores and event-based feedback increase more significantly than those of recipients with initial high scores; and (ii) the scores of recipients who receive initial low scores and do not receive event-based feedback are not statistically different from those of recipients with initial high scores. Recall that our previous result is that scores of recipients with initial low scores increase more significantly than those of other recipients. Our analysis demonstrates that this result is mainly driven by employees who received event-based feedback rather than employees who received comments about personal opinions.