

American Sociological Review

Visibility and Gender Disparity in Low-Promotability Work

Journal:	<i>American Sociological Review</i>
Manuscript ID	Draft
Manuscript Type:	Original Article
Keywords:	Evaluation research, Gender, Organizations, Low-Promotability Work, Task Design
Abstract:	Providing evaluative feedback is an example of a task with low promotability: critical to organizational functioning but offering little opportunity for career advancement. Men consistently invest less effort in such work than women, a gap often treated as fixed. We argue instead that gendered engagement depends partly on task design, and that visibility, meaning who can observe the work, can narrow gender disparities by encouraging men to engage more. We test this with a quasi-natural experiment at a multinational pharmaceutical company, whose 360-degree feedback platform inadvertently lost the option for givers to hide feedback from recipients. Applying a difference-in-differences framework to linguistic measures of engagement within matched giver-recipient dyads, we find that when feedback was made visible to its recipient, men increased their engagement more than women, in both effort and recipient-orientation, narrowing the pre-existing gender gap. Feedback content remained largely unchanged. The engagement increases were strongest among male managers giving downward feedback to subordinates, where pressure to account for their feedback was greatest. The recipient's gender did not moderate these patterns. Thus, expanding who observes low-promotability work can narrow gendered engagement gaps, primarily by increasing engagement from men in positions of authority.



Low-promotability work, meaning work that is necessary for organizational functioning but does not contribute to an individual's career advancement, is inherent in many organizational contexts. It is also persistently gendered. Widely shared cultural beliefs route women toward supportive, less visible contributions (Ridgeway 2011), with women accordingly taking on more low-promotability work than men (Babcock et al. 2017a; Kaplan 2022; Misra, Lundquist, and Templer 2012). In large part, this gender disparity in contributing what turns out to be low-promotability work is a feature of the segregation of men and women into different types of occupations or job titles (Benschop and Doorewaard 1998; Weeden 2002; Reskin and Roos 1990). However, even within the same occupation, employers disproportionately assign these discounted, low-promotability tasks to women (Chan and Anteby 2016).

Furthermore, recent evidence suggests that even when men and women are assigned an equal number of low-promotability tasks, they often engage with and perform these tasks in systematically different ways (Conzon 2023; Feldberg 2022; Ranganathan and Shivaram 2021; Brewer et al. 2020). In particular, the amount of time and cognitive effort invested tends to be less for men and greater for women, especially when it comes to engagement, which includes both the effort invested and the orientation toward the recipient. For instance, in the context of workplace assessment, female attending physicians provide longer, more detailed, and more job specific feedback to residents compared to men, and are more likely to offer helpful and supportive evaluations, especially when working with poor performers (Nelson et al. 2023).

The less time and effort a professional devotes to important but low-promotability tasks, the more effort they can devote to tasks likely to be recognized for promotion, thus gaining a career advantage. As with other forms of low-promotability work, feedback work is typically performed by workers whose primary responsibilities lie outside evaluative labor: physicians

1
2
3 whose core work centers on patient care, lawyers whose performance is assessed through billable
4
5 hours, faculty whose promotion depends primarily on research output, and managers who are
6
7 evaluated on operational or financial performance rather than the quality of evaluations they
8
9 provide. In this way, a gender difference in engagement and performance can contribute to
10
11 gendered workplace outcomes (Babcock et al. 2022). It also increases pressures and elevates the
12
13 risk of burnout for those performing more low-promotability work (Vial and Cowgill 2022;
14
15 Wharton 2009; Eagan and Garvey 2015; Hart and Cress 2008). Finding ways of reducing this
16
17 inequality therefore clearly seems a worthwhile direction for research.
18
19

20
21 We propose that the design of tasks can reshape how women and men engage in them
22
23 and alter patterns of unequal participation in devalued work, and we focus on one particular
24
25 design aspect of a task: visibility. By visibility, we mean the extent to which others in the
26
27 organization, such as supervisors, peers, as well as subordinates, can directly observe, recognize,
28
29 or otherwise become aware of the performance of these tasks. To illuminate this design aspect
30
31 we connect a classical sociological insight—namely, that conduct is shaped by who is positioned
32
33 to observe it—with recent organizational research on workplace visibility. Bringing the two
34
35 together lets us ask a question neither addresses on its own: how does the observability of an
36
37 evaluative act change an employee's engagement when they perform that act?
38
39

40
41 By changing the nature of the audience (Simmel 1950; Goffman 1959), visibility can
42
43 encourage men's engagement with the work as compared with women's. Visibility is likely to
44
45 matter most when it exposes work to audiences whose judgments carry relational and
46
47 reputational weight (Simmel 1950; Goffman 1959), raising the prospect of social evaluation in a
48
49 way that observation by indifferent parties would not (Lerner and Tetlock 1999). Giving an
50
51 evaluation is an exercise of authority over the person evaluated (Bourdieu 1984, 1991), and
52
53
54
55
56
57
58
59
60

visibility can in turn change whether that authority must be sustained before the person it concerns. In this way visibility activates a state of accountability. To be accountable, in this sense, is to be positioned as potentially needing to render an account of one’s conduct to an audience entitled to judge it (Garfinkel 1967). Visibility may thus narrow gender differences in engagement, in both the effort and the recipient-orientation it reflects, by increasing accountability (Castilla 2008, 2015; Elvira and Graham 2002) and heightening pressures to conform to performance expectations (Botelho 2024; Connell and Wood 2005; Kelly et al. 2010; Williams 2001).

To examine how visibility shapes gendered engagement in low-promotability work, we analyze data from a quasi-natural experiment within a large pharmaceutical company’s 360-degree feedback platform. This context provides a rare opportunity to observe how the visibility of feedback reshapes patterns of engagement in a real organizational setting. During the study period, an inadvertent platform change removed the option for feedback givers to “Hide the Feedback from the Employee”. Beforehand, feedback givers had the option to share the feedback only privately with the recipient’s manager, but once the “hide feedback” option was removed, feedback recipients were guaranteed to see the comments relating to them. This discontinuity allows us to isolate how expanding the audience affects men’s and women’s engagement with the low-promotability task of providing evaluative feedback to others.

Because feedback is addressed to a particular audience, the act of giving it takes its character from that address, i.e. from the relational situation rather than from the giver alone (Emirbayer 1997). We treat engagement as this addressed act and distinguish two dimensions of it: first, the effort a giver invests in writing the text of the feedback, and second, the recipient-orientation of the text, i.e. the extent to which the giver’s writing choices are oriented toward

1
2
3 how the audience will receive it. This second dimension follows from the relational character of
4 evaluation itself, in which an assessment is shaped by and toward the party who receives it (Blau
5 1964). A design change that expands the audience to include a new class of readers bears directly
6 on this second dimension (Simmel 1950), which makes it a sharp test of how visibility reshapes
7 the work.
8
9

10
11
12 Using a difference-in-differences framework applied to linguistic measures of
13 engagement, we estimate how men and women differentially adjusted their feedback behavior
14 following the platform change. We find that the increase in visibility prompted a greater rise in
15 feedback engagement among men than among women, narrowing the pre-existing gender gap in
16 feedback engagement. These effects were especially evident among male managers providing
17 downward feedback to subordinates. Importantly, the gender of the feedback recipient did not
18 influence these effects, nor did we find a material change in the substantive utility of the
19 feedback, suggesting that the observed shifts in language stem from heightened accountability
20 activated by the implication of an expanded audience rather than by gender-conscious responses
21 or by changes in feedback content.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 Together, these findings demonstrate how the visibility of low-promotability work can
38 alter gendered engagement patterns by amplifying accountability pressures, particularly among
39 those in positions of authority. By tracing these dynamics within an ongoing organizational
40 process, this study extends theories of task design and gendered engagement, highlighting how
41 visibility operates as a structural lever that can reduce inequality in everyday organizational
42 practices.
43
44
45
46
47
48
49
50

51 **LOW-PROMOTABILITY WORK AND GENDER INEQUALITY**

52 *Evaluative Feedback System Design and Inequality*

53
54
55
56
57
58
59
60

1
2
3 Giving feedback is a common form of low-promotability organizational work performed by and
4
5 for both men and women. It is a routine but often underrecognized aspect of organizational life,
6
7 embedded in formal practices such as performance reviews and evaluations as well as informal
8
9 processes such as mentoring and day-to-day coaching. As a core element of the human resource
10
11 systems used to motivate employees and improve performance, feedback is essential to
12
13 organizational functioning and employee development (Baron and Kreps 1999). Yet the act of
14
15 providing feedback is time-consuming (Buckingham and Goodall 2015) and widely perceived as
16
17 both burdensome and aversive by both givers and recipients (Pfeffer 1994; Cleveland, Lim, and
18
19 Murphy 2007). Providing feedback offers little direct material or career reward to those who
20
21 provide it, with employees frequently perceiving it as low-return, symbolic, relational labor that
22
23 is unlikely to translate into career advancement (Bourdieu 1991; Babcock et al. 2022).
24
25
26 Employees who prioritize career advancement might therefore be expected to minimize their
27
28 engagement with feedback tasks were it not for systemic task design elements capable of
29
30 constraining their behavior. The design of feedback systems is therefore not neutral. It organizes
31
32 the distribution of undervalued labor and, in doing so, may shape the extent to which inequality
33
34 is reproduced or reversed.
35
36
37
38
39

40 Thus, the design of feedback and performance assessment systems remains an active site
41
42 of experimentation and innovation. For instance, many organizations now emphasize frequent,
43
44 conversational, real-time feedback processes, often in place of traditional appraisal systems
45
46 centered on annual merit-based ratings by managers (Cappelli and Travis 2016; Murphy 2020;
47
48 Rivera et al. 2021). Other organizations have moved toward more socially embedded feedback
49
50 models, in which employees receive input from a broader set of colleagues through peer- and
51
52 network-based assessment systems (Fiore and Souza 2021). One of the most common feedback
53
54
55
56
57
58
59
60

mechanisms is the 360-degree evaluation, used by over 85% of Fortune 500 companies (Zenger 2016). Rather than relying solely on formal assessments from direct managers, 360-degree feedback incorporates input from colleagues, supervisors, and subordinates. These systems vary in whether feedback is anonymous or fully identifiable. While many multisource feedback systems are designed to be anonymous in order to encourage honesty and reduce bias (Atwater, Waldman, and Brett 2002), some organizations have moved toward fully transparent (visible to anyone in the organization) 360-degree feedback systems (Bernstein and Li 2017).

Previous studies have examined how feedback design shapes the gender inequality of outcomes for those who are recipients of the feedback. Although formalized evaluation structures such as standardized evaluation criteria and performance rubrics are found to influence evaluators' judgments in ways that can reduce gender disparities (Elvira and Graham 2002), evaluators nevertheless are still observed to draw on gender frames and schemas of competences (Ridgeway and Correll 2004; Ridgeway 2011), resulting in biased assessments of those being evaluated. In one study of formalized performance reviews, Correll et al. (2020) show that gendered language in the review text influences career trajectories of the receivers and reinforces gender bias in managerial assignments. Relatedly, Rivera and Tilcsik (2019) show that changes in seemingly technical design elements, such as reducing the number and range of rating scales (e.g. shifting from a 10-point to a 6-point scale) narrow gender gaps by limiting evaluators' subjective judgments. A recent review suggests that simplicity in the design of evaluation systems helps to curtail the operation of implicit biases (Abraham, Botelho, and Lamont-Dobbin 2024).

By contrast, far less is understood about the work of providing feedback itself: specifically, how the task of giving evaluative feedback is engaged in and how that impacts

1
2
3 evaluators. This gap is significant, as feedback is inherently relational: the evaluation of one
4
5 worker depends on the often-invisible labor of another, and the act of giving feedback
6
7 simultaneously shapes not only the assessment of the subject of review but how the giver of the
8
9 evaluative feedback is perceived (Blau 1964; Goffman 1959). Yet prior research has rarely
10
11 examined how particular features of these systems shape the act of giving feedback in real
12
13 organizational settings.
14
15

16
17 Our study addresses this gap. As a design feature, visibility expands who can observe and
18
19 assess feedback, transforming what was once a confidential evaluative act into a more observable
20
21 social process (Goffman 1959). We submit, therefore, that visibility can influence how feedback
22
23 work gets done, with implications for gender inequality in low-promotability work.
24
25

26
27 **VISIBILITY AND GENDER DISPARITY IN LOW PROMOTABILITY WORK**

28
29 Low-promotability work such as giving evaluative feedback is typically characterized by its
30
31 invisibility (Fletcher 1999; Kaplan 2022; Babcock et al. 2022). Because these tasks are typically
32
33 sequestered, informal, and organizationally undervalued (Cech 2013), invisibility not only
34
35 obscures those who perform them but also allows unequal participation to persist unchallenged.
36
37 This invisibility, moreover, is not gender-neutral: it affords men greater latitude to disengage
38
39 from such work unobserved (Simpson and Lewis 2005; Lewis and Simpson 2010). Men are
40
41 accordingly less likely than women to take it on and face fewer social or occupational penalties
42
43 for opting out, especially when the work is feminized or relational in nature, as men face weaker
44
45 expectations to engage in such tasks and incur fewer consequences for disengagement (Heilman
46
47 2001; Eagly and Karau 2002). Invisibility, combined with the low-promotability of this work,
48
49 then creates conditions under which men and women respond differently.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 For women, the invisibility of low-promotability work operates primarily to erode
4 recognition of meritorious performance. Women are both disproportionately more likely to be
5 assigned to low-promotability tasks (Chan and Anteby 2016) and to volunteer and accept
6 requests to do such work (Babcock et al. 2017a), even if there is no penalty for refusing
7 (Babcock et al. 2017b). Keeping this work invisible therefore renders women's contributions
8 especially vulnerable to being overlooked and undervalued (Tyler and Cohen 2010). This
9 dynamic is further reinforced by differences in socialized behavior and gendered expectations
10 that create a participation imbalance in low-promotability work in the first place (Correll 2001;
11 El-Alayli, Hansen-Brown, and Ceynar 2018). Socialization steers women, more than men,
12 toward emotional labor (Hochschild 1983; Brody 1993; Hibbard and Buhrmester 1998; Chaplin,
13 Cole, and Zahn-Waxler 2005), contributing, for example, to heavier service burdens among
14 female academics (Bellas 1999) and gendered expectations that women provide "nurturing" to
15 students (El-Alayli et al. 2018).

16
17 These differences allow men to withdraw from undervalued work with fewer
18 consequences, while women continue to engage with greater effort (Tyler and Cohen 2010).
19 Greater visibility, however, removes this cover.

20 ***Relational Work and Visibility***

21
22 Visibility matters to gendered engagement in giving evaluative feedback because it can change
23 the social nature of the act. Garfinkel (1967) argues that everyday conduct is "accountable" in
24 the sense of being observable and reportable: people produce actions in ways that allow others to
25 recognize and describe them as orderly and appropriate, or not. Accountability, at its core, means
26 rendering an account of one's conduct to an audience positioned to judge it (Scott and Lyman
27 1968; Huising and Silbey 2021). Once an audience can observe the act, in other words, the act is
28 positioned to be assessed against shared expectations for what counts as appropriate. Evaluation,
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

moreover, is not a neutral transmission of judgment but a form of social exchange: in assessing another, a person also confers or withholds approval, and so adjusts the standing of both parties to the relationship (Blau 1964), which means that the evaluator may incur risk to their own status. Thus, the evaluator, or feedback-giver, may be more or less accountable to the degree that the tasks are visible or reportable. In relational work, such as giving feedback, performers may update or adjust their approach depending on the observability of their work, managing how they appear once an audience is present to observe them (Goffman 1959).

In such relational work, therefore, the management of the interactional surface is often not an irrelevant packaging but rather the very substance of what is being exchanged (Goffman 1967). What an audience credits, in other words, is set not just by the content of an act but by the manner of its delivery. The value of action is fixed by the audience that receives it (Bourdieu 1991), and that audience judges appropriateness through the way the act is conducted as much as through its propositional content (Bourdieu 1984). Moreover, in communication whose work is relational, a communicative act can establish and sustain a bond between parties largely independently of the information it conveys (Malinowski 1923), so that how something is said can perform the relational work of an exchange even as its substance holds constant (Weeks 2004). The standing a giver draws from the act therefore depends on the standards of the audience positioned to judge it, and on the manner through which the act meets them.

When visibility increases, this accreditation process updates accordingly. A newly present audience brings its own set of expectations to the exchange, which the giver must now take account of if the act is to meet those expectations (Goffman 1959). In particular, moving from a pairing to a group, or expanding exchanges between two parties to three, “completely changes them” in a way unmatched by further group expansion (Simmel 1950: 96).

Gendered Norms and the Display of Authority

The norms to which people are held accountable are not uniform. Visibility may activate a general orientation toward audience-facing self-presentation, but the forms that self-presentation can take, and the standards by which it is judged appropriate, differ for men and women (Bourdieu 1984; Heilman 1983). When the norms with which conduct is assessed are themselves gendered, identical behavior is read and valued differently depending on the gender of the actor performing it and the audience observing it (West and Zimmerman 1987). An enlarged audience, then, does not bring a single standard to bear on every giver. Gender operates as a diffuse status characteristic, attaching widely held expectations of competence and worthiness to the categories “man” and “woman” independent of any individual’s demonstrated capacity (Ridgeway 2001, 2011; Ridgeway and Correll 2004).

The audience does not weigh the act alone but the act against the gendered expectation associated with the author, judging whether a given display is fitting for the one performing it. A manner that conveys authority and command is congruent with the higher status culturally accorded to men, and reads from a man as competence; the same manner from a woman runs against the lower status her category is assigned, and reads instead as a claim to standing she is not granted or seen as deserving. Because speaking or commanding with authority (Wrong 1979) is congruent with the behavior generally expected of men and incongruent with that expected of women, the two do not face the same audience assessment/terrain. Women who command or request deference to their statements or requests violate the prescriptions attached to their category and draw evaluative penalties for the violation, a pattern documented across role-congruity and backlash research (Eagly and Karau 2002; Rudman and Glick 2001). The penalty is not abstract: holding a position of authority raises men’s ability to command attention before others but leaves women’s power relatively less changed, because women who attempt to

exercise power often anticipate and incur backlash from observers of both genders (Brescoll 2011). Visibility enlarges the audience before which an authoritative woman can be penalized. It also adds to the work she must do, since women in positions of authority carry an additional burden of managing and leveling their status so that their standing is read as legitimate, a demand men in the same positions are spared (Cardador, Hill, and Salles 2022).

Because an audience judges the manner of an act against standards that are themselves gendered, a wider audience does not move all givers in one direction. Where the operative standard is shared across gender, for instance with the effort a task receives or the warmth with which it is pitched to the party who receives it, greater visibility should pull men and women toward a common level. Where the standard is gendered, as with the display of authority, a man receives license to project more of it while for a woman the identical projection carries costs he does not bear. Convergence and divergence are therefore not separate effects but two expressions of one status logic: the level of authoritative self-presentation an audience will credit is conditioned on the gender and standing of the one who performs it. A performer must therefore calibrate their presentation carefully, and this is especially true for women.

VISIBILITY AND INEQUALITY IN EVALUATIVE FEEDBACK

Transparency and Narrowing Gender Gaps

Research on evaluation systems lets us anticipate when an enlarged audience narrows a gender gap and when it widens that gap. Evaluations are, of course, themselves social products. Managers’ assessments of those below them do not emerge in a vacuum but are shaped by the social context in which they occur and by the presence or absence of constraints on managerial discretion. Where there are few constraints, performance ratings and the rewards based on them may display demographic imbalance (Castilla 2011). In contrast, Castilla’s (2015) study of

merit-based pay shows that when evaluative processes and outcomes were made transparent and managers were required to justify their decisions, gender disparities in rewards narrowed.

Prior work has established that making evaluation visible to overseeing authorities reshapes how evaluators behave: when managers must account for their decisions to Human Resources or senior leadership, discretion is constrained and disparities in outcomes narrow (Castilla 2008, 2015). But this work traces the evaluator's behavior for the sake of its downstream consequences for the recipient (the fairness of the rating, the size of the pay gap) rather than treating the evaluator's own engagement in the act as the outcome of interest. Other accounts share this orientation from different angles: observation by superiors can lead the observed to manage the surface of their conduct rather than its substance (Bernstein 2012), and visibility to peers can raise effort through social pressure (Mas and Moretti 2009). What none addresses is what happens to the evaluator's conduct when the evaluative act is made visible downward, to the person being evaluated. The subordinate, the person over whom the evaluator holds authority, has not been studied as an audience who might influence the evaluative act, including its performance as a function of gender.

Downward Exposure and Authority

The pressures of this downward exposure should be strongest when participation becomes observable to an audience for whom the outcome is consequential. Relevant to the sociological framework we have outlined, work in social psychology has treated accountability as a mechanism of norm enforcement (Tetlock 1985). When individuals anticipate that their conduct will be evaluated by others, they modify their behavior to align with shared expectations and standards (Tetlock 1983; Lerner and Tetlock 1999). Anticipation of external evaluation triggers what Tetlock (1983, 1992) terms "preemptive self-criticism," in which individuals simulate the

audience’s reactions in advance and adjust their behavior to appear reasonable and defensible. The most familiar such audience is the one above: being observed by a superior who controls one’s rewards predictably leads people to sharpen their conduct and put a defensible face on their work (Goffman 1959: 106–140; Bernstein 2012).

Visibility to those below carries anticipatory pressure of its own, governed by a different logic of what counts as reasonable and defensible. Authority is not a one-way imposition but a relation that draws on recognition from those subject to it; even within a hierarchy the superordinate remains in part answerable to the response of those below (Simmel 1950: 185–189, 190–223). Being observed by subordinates therefore activates an accountability latent in the relation itself: a party in a position of authority must now expect their actions to be read by the very people over whom that authority is held, and to have its exercise weighed by them.

Hence, feedback as a form of social exchange is not only an evaluative act but a relational one (Blau 1964; Tomaskovic-Devey and Avent-Holt 2019; Bandelj 2020). Visibility may thus draw men toward increased engagement by pulling their behavior upward toward the standard the newly added audience now makes salient, and secondarily by leaving less room to withdraw.

Theoretical Predictions

Taken together, this body of research suggests that visibility reshapes feedback by altering the social conditions under which evaluative work is carried out. By heightening tacit or implied accountability, visibility channels behavior toward shared norms. The literature thus offers a framework which allows us to formulate three expectations about the effects of increased visibility in employee feedback systems.

1
2
3 *Convergence and divergence in feedback engagement.* Greater visibility of feedback to the
4
5 individuals evaluated should reduce gender disparities in engagement, though not evenly across
6
7 the dimensions of feedback. Because the standards an enlarged audience makes salient are shared
8
9 on some dimensions and gendered on others, visibility can be expected to produce convergence
10
11 where the operative norm is common to men and women but to produce divergence where the
12
13 norm is itself gendered. On both counts the larger movement falls to men, who under low
14
15 visibility have the most room to move.
16
17

18
19 Convergence is likely to narrow the gap chiefly by drawing men's engagement upward
20
21 toward the investment women already show. Because under low visibility men could disengage
22
23 from low-promotability work with little scrutiny (Simpson and Lewis 2005; Lewis and Simpson
24
25 2010), they enter with the most room to move. An observable, therefore creditable, task is then
26
27 one that men become more willing to take on (Babcock et al. 2017a; Babcock, Recalde, and
28
29 Vesterlund 2017b), such that the burden of low-promotability feedback work is no longer borne
30
31 so disproportionately by women.
32
33

34
35 Divergence follows from the gendered-norm logic set out above. Where conduct before
36
37 an audience is judged against standards that differ by gender, the same widening of the audience
38
39 moves men and women apart rather than together: the authoritative manner an audience sees in a
40
41 man as competence is read in a woman as overreaching, with the woman bearing social costs a
42
43 man does not bear (Brescoll 2011). A wider audience thus moves men and women differently
44
45 depending on the dimensions or aspect of the evaluative feedback. In regard to the effort
46
47 feedback receives or the warmth with which it is pitched to its recipient, convergence would be
48
49 likely to occur as men move toward the investment women already show (Hochschild 1979;
50
51 Tyler and Cohen 2010). Where it concerns the display of authority, however, the literature would
52
53
54
55
56
57
58
59
60

lead us to predict divergence, with men freer to project more of it under scrutiny and women constrained by the social costs that fall on authoritative women but not on authoritative men (Brescoll 2011; Cardador et al. 2022; Conzon 2023).

Rank and downward exposure. In directly testing the consequences of downward exposure, visibility may be expected to lead to the most pronounced change among higher-ranked evaluating observers. When supervisors' feedback practices are visible to the very subordinates over whom they hold authority, they face strong reputational and relational pressures to justify their behavior in ways that appear fair, consistent, and, at times, positive (Klimoski and Inks 1990; Mero, Guidice, and Bownlee 2007). Because accountability intensifies with rank, and because visibility to subordinates introduces new audience members whose judgments carry relational weight, visibility should have its strongest equalizing effects among managers giving downward feedback.

These accountability-based pressures can logically be expected to produce larger changes among men rather than women. Because under low visibility women tend to engage more intensively with low-promotability tasks, a more visible audience finds them already closer to the standard it makes salient, with little distance left to move. Because men have greater latitude to disengage when the work is less observable, given that invisibility tends to shield men's normative status from scrutiny (Simpson and Lewis 2005; Lewis and Simpson 2007, 2010), they stand farther from a new visibility standard; any movement would likely bring them closer to that standard. The equalizing effect of visibility should therefore operate primarily by raising men's engagement toward the standard women already meet.

Manner or content? As the manner of a relational act carries social work that its content may not (Malinowski 1923; Goffman 1967; Weeks 2004), a change in the conditions under which

1
2
3 feedback is given may induce change in how the message is delivered as much as, or more than,
4
5 in what is communicated. If so, the changes induced by visibility will be reflected in the tone,
6
7 confidence, and sensitivity reflected in the language, rather than in the evaluative content or
8
9 practical utility of the feedback itself. This reasoning follows the accountability literature's
10
11 emphasis on anticipatory audience management (Tetlock 1983; Lerner and Tetlock 1999). When
12
13 givers know their feedback will be read by its recipient, they are likely to adjust the presentation
14
15 of their assessments to appear thoughtful and fair, even if the underlying judgments remain
16
17 unchanged.
18
19
20

21 **METHODS**

22 *Research Setting*

23
24 Data for this study comes from a multinational pharmaceutical company ("Pharma Co.") whose
25
26 employment presence spans more than 20 countries and several U.S. states. Its workforce covers
27
28 a wide range of occupations, including scientists, engineers, manufacturing workers, managers,
29
30 and functional staff. Roughly 57% of the feedback in our sample came from feedback givers
31
32 based in the United States, where Pharma Co. is headquartered.
33
34
35
36

37 Pharma Co. administers its performance evaluation process through a "360 review"
38
39 platform built by Workday, a large enterprise HR software provider. As part of this process,
40
41 which runs throughout the year, employees provide 360-degree feedback to colleagues.
42
43 Feedback givers include anyone who worked with the recipient during the study period,
44
45 including peers, subordinates, and direct and indirect supervisors. There are no formal
46
47 requirements or quotas for providing feedback. Most feedback is requested by the recipient's
48
49 manager and once a request is made, the prospective feedback giver is reminded often (See
50
51 Appendix Table A1 for distribution of feedback types).
52
53
54

55 *The Quasi-Natural Field Experiment*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Following the intuition that increased visibility might reduce existing disparities in giving evaluative feedback in Pharma Co., the Executive Committee (EC) explored the possibility of removing a particular privacy feature from its 360-degree feedback platform. In July 2018, they asked an HR representative to determine whether removing the option “Hide the Feedback from the Employee” could be done internally or would require a formal request to the platform vendor. By the time the question was received by the tech personnel managing the platform, it was interpreted as a request for removal, and the tech team implemented the change. The HR staff member then shared news of the change with the EC. After some deliberation, the EC decided to retain the change (that is, an absence of the “Hide the Feedback” button). However, they also made sure that everyone in the company was aware of the change by adding the following sentence to the platform interface: “All feedback will always be visible to the employee and their manager.” In sum, this change came from an offhand executive comment, was unanticipated in its timing, and was a discrete shift from feedback that could be concealed from the recipient to feedback that was inherently visible to them. Therefore, it serves as a quasi-natural field experiment: observed shifts in gendered engagement can be credibly attributed to visibility rather than to other factors, with participants performing routine work in situ and unaware that their behavior was being studied (Harrison and List 2004). See Exhibit 1 for illustrations of this platform change.

Insert Exhibit 1 about here

Before this subtle change, whenever feedback givers had clicked the “Hide the Feedback from the Employee” button, their feedback and identity were only visible to the feedback recipient’s manager (i.e., not visible to the feedback recipients themselves). After the change in platform design, this button was no longer available, so both the content of the feedback and the givers’

identities were always visible to both the recipient's manager and the recipient of the feedback themselves. This design has a feature worth making explicit. The recipient's manager could read the feedback in both periods, so the presence of that observer was constant across the change. Comparing givers' compensation brackets with those of recipients' managers shows that this ever-present audience stood at or above the giver's own broad hierarchical tier in nearly all dyads: roughly half of givers sat below the recipient's manager, roughly half at the same tier, and fewer than 2% above (see Appendix Table A6). That relative positioning was itself stable, unchanged across the two periods for 96.5% of the dyads in both years, and as a fixed attribute of the pairing it is absorbed by the dyad fixed effects in our models. The platform change thus altered nothing about observation from above or alongside. It added a single new reader, the recipient, the audience our theory concerns, whose standing relative to the giver varies across dyads, a variation we exploit in the rank analyses below. Our estimates accordingly identify the effect of adding the recipient as an audience, net of the observation the recipient's manager already exercised in both periods.

One limitation of our data is that Pharma Co.'s records do not include an identifier for when feedback givers did or did not use the "Hide the Feedback" option during the low-visibility pre-period. As discussed below, we make several attempts to parameterize the likely frequency of use of this button in the low-visibility pre-period.

Analytic Strategy

To test whether increased visibility differentially affects men's and women's feedback-giving behavior, we compared engagement in the feedback provided after the removal of the "Hide the Feedback" button with engagement in the feedback between the same giver and recipient pair from the same calendar months in the previous year. The exact date of the button's removal in July 2018 is unknown, so we took a conservative approach by excluding July 2018 entirely and

focusing on August–December 2018, comparing this period to August–December 2017. This adjusts for seasonality effects, including Pharma Co.’s year-end performance evaluation process, especially pertinent as the five months from August to December capture over 75% of all feedback left on the platform in both years. Our results remain robust to including July 2018 in the analysis.

We analyzed 7,195 English-language evaluations across the pre-and post-visibility change window within 931 matched pairs of feedback givers and recipients (‘dyads’), from 704 unique feedback givers. This dyadic data represents approximately 11% of the total corpus of intra-company feedback from this period. The feedback distribution shows 54% of instances occurring during the low-visibility period and 46% during the higher-visibility period, a ratio that closely mirrors the overall distribution of feedback given companywide across the two periods. (See Appendix Table A2).

Since we restrict our analyses to feedback exchanged within the same giver–recipient dyads across the two periods, our results are not driven by shifts in who is giving or receiving feedback but rather by how individual givers adjust their behavior when faced with increased visibility. However, feedback givers of both genders left, on average, about 14% fewer instances of feedback in the higher-visibility post-period (See Table 1; $p < 0.001$), which suggests that during the earlier, low-visibility period, the “Hide the Feedback” option was being used by those in our dyad sample. This decline did not differ significantly by gender ($p = 0.41$) and therefore is unlikely to drive any of the gendered patterns we report.

This decline indicates that some feedback previously provided under lower visibility may have been discretionary, and once visibility increased, such feedback was withheld. While this pattern may indicate that a subset of low-visibility feedback reflected optional or sensitive

exchanges that givers opted not to share when visibility increased, it also means that our estimates of visibility's effects are likely conservative rather than inflated, since the instances most likely to be withheld would be the low-engagement ones where the visibility effect would otherwise have been largest. Because we analyze all available feedback in each period, rather than attempting to infer which specific instances were hidden in the low-visibility pre-period, our approach provides a cautious lower-bound estimate of visibility's impact.

Insert Table 1 about here

The volume of the qualitative evaluative feedback text makes manual, iterative coding infeasible. For this reason, and to enhance external validity, we employ the Linguistic Inquiry and Word Count program (LIWC-22) as our primary analytical tool. LIWC identifies systematic linguistic patterns by scanning text for predefined words and phrases associated with specific psychological and social constructs (Pennebaker, Mehl, and Niederhoffer 2003). LIWC has been widely validated for analyzing gendered language elements (Lin, Yu, and Dowell 2020; Newman et al. 2008; Kilis and Yildirim 2018; Yu 2014) and has proven valuable for studying organizational phenomena such as cultural fit, emotional communication patterns, and symbolic boundary maintenance between organizational groups (Goldberg et al. 2016; Mobasseri 2018; Bhatt, Goldberg, and Srivastava 2022). This dictionary-based approach is particularly appropriate for our study because our main theoretical prediction, that increased visibility reduces gender differences in feedback engagement, concerns linguistic dimensions that LIWC reliably captures.

Measures of Engagement

We use 'engagement' as our overarching construct for capturing how much a feedback giver invests in the task. We distinguish two dimensions of engagement: the first is effort, the

observable intensity of the feedback reflected in features such as length and structural elaboration; the second is recipient-orientation, the extent to which the giver pitches the feedback to how its recipient will receive it, reflected in linguistic properties associated with interpersonal attunement. This construct describes the properties of the feedback text rather than the internal states of the giver, which we cannot directly measure, nor do we have available data related to how recipients read and interpreted it; we therefore observe the text itself and speak of it as having a certain degree of orientation toward the recipient. We capture recipient-orientation primarily through three of LIWC’s summary variables, Tone, Clout, and Authenticity, and we treat the fourth, Analytical Thinking, as a related but weaker indicator for the reasons given below.¹ To measure effort, on the other hand, we use the two variables: Word Count and Linguistic Dimensionality. Excluding Word Count, these composite variables, widely applied in communication, marketing, management, and sociology research (e.g., Germann et al. 2024; Holm, Fong, and Anteby 2024; Jordan et al. 2019; Kyprianou et al. 2025; Newman et al. 2003; Sauder, Shi, and Lynn 2024), are algorithmically derived from multiple underlying LIWC categories to capture broader linguistic, psychological, and social dimensions of communication (Pennebaker et al. 2015).

Tone, conceptually the most straightforward variable, measures the relative frequency of positive versus negative emotion words in feedback text by comparing words associated with positive emotions (e.g., “love, nice, sweet”) against those associated with negative emotions (e.g., “hurt, ugly, nasty”) (Pennebaker et al. 2015: 3). Prior research has used this variable to capture affective valence in diverse evaluative contexts, such as predicting performance in quantitative college courses (Carrell, Page, and West 2010), identifying from reflection essays students’ likelihood of passing online STEM courses (Lin et al. 2020), and assessing from

1
2
3 supervisor feedback the readiness of healthcare residents for new responsibilities (Heath et al.
4
5 2024).

6
7
8 Clout measures the extent to which feedback gives project social status, confidence, or
9
10 leadership in their language (cf. Kacewicz et al. 2014). It captures the authority a giver pitches
11
12 into the feedback, which is one way of orienting to a recipient Clout (Germann et al. 2024):
13
14 feedback delivered with greater confidence and authority is, on average, perceived as more
15
16 credible and actionable by recipients (Germann et al. 2024; Toma and D'Angelo 2015). This
17
18 average, however, conceals variation in how authoritative language is received depending on
19
20 who delivers it. Work on status incongruity finds that the confident, directive style that reads as
21
22 competent from a man can draw a cooler reaction from a woman, especially one addressing a
23
24 subordinate (Eagly and Karau 2002; Rudman and Glick 2001). We therefore treat Clout as the
25
26 authority dimension of recipient-orientation, whose recipient-appropriate level depends on the
27
28 giver's status, rather than as a measure on which higher is uniformly better.
29
30
31

32
33 Higher Authenticity scores signal more genuine and personally-grounded observations,
34
35 rather than detached commentary, thereby strengthening recipient-orientation by establishing
36
37 trust and increasing the likelihood that recipients will accept and act upon the feedback.
38
39 Linguistic and psychological research demonstrates that truthful communication exhibits higher
40
41 cognitive complexity and more frequent self-references compared to deceptive communication
42
43 (Knapp, Hart, and Dennis 1974; Newman et al. 2003). LIWC's Authenticity score captures this
44
45 pattern by weighting linguistic markers such as first-person pronouns and insight words
46
47 positively, while weighting discrepancy words and third-person pronouns negatively. This
48
49 measure thus reflects the perceived sincerity of evaluative statements, and authentic language of
50
51 this kind has been linked to greater interpersonal connection and more favorable social and
52
53
54
55
56
57
58
59
60

business outcomes across settings ranging from online reviews to interpersonal and entrepreneurial encounters (Zhang 2019; Markowitz et al. 2023).

Analytical Thinking captures the formality and logical structure of text. In feedback research, it has generally operated inversely, with lower scores taken as the more recipient-oriented direction. Language that scores lower tends to read as warmer, more personable, and less rigid (Jordan et al. 2019; Kyprianou et al. 2025), a style associated with the warmth and accessibility of recipient-oriented feedback. Research on entrepreneurial feedback confirms this pattern, finding that feedback with lower Analytical Thinking scores was seen as more beneficial by receivers (Germann et al. 2024). Still, we treat this direction as suggestive rather than settled, as a related body of work on feedback specificity reaches a more mixed verdict: more specific, directive feedback can sharpen immediate performance while doing less for learning and exploration (Goodman, Wood, and Hendrickx 2004; Shute 2008). We therefore treat Analytical Thinking as one indicator among several rather than a clean marker of recipient-orientation on its own.

Finally, we measure *effort* using two constructs for which precise construction details are most straightforward: Word Count and Linguistic Dimensionality. Word Count reflects the total number of words in a feedback entry and serves as a proxy for the time and effort invested in providing feedback.² Linguistic Dimensionality is a composite indicator that aggregates 20 LIWC-22 measures of grammatical and interpersonal language features, including personal and impersonal pronouns, function words, verbs, conjunctions, and determiners (see Appendix Table A3). Together, these components capture the structural elaboration of written text, that is, the extent to which feedback is developed as complete, conversational prose rather than terse or telegraphic notes dominated by technical terminology. Higher scores reflect feedback that is

more linguistically developed and conversationally engaged. Together, these measures provide a comprehensive and transparent framework for assessing engagement, in both effort and recipient-orientation, across visibility conditions.

Difference-in-Differences Analysis of Gendered Feedback Across Visibility Conditions

We investigate how the baseline gap between men and women's feedback engagement narrows or widens after the increase in visibility on the feedback platform. Specifically, we first compare the gender gap in engagement prior to the visibility change across the two facets we measure, recipient-orientation (Tone, Clout, Authenticity, alongside Analytical Thinking) and effort (Word Count, Linguistic Dimensionality). We then assess how men and women adjust along these dimensions after visibility increases. Analytically, we estimate the differential impact of visibility on each linguistic construct, stratified by gender, using a difference-in-difference specification of the following form:

$$Y_{idt} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 (\text{Post}_t \times \text{Female_Giver}_d) + \alpha_d + \varepsilon_{idt}$$

Where, Y_{idt} measures each linguistic construct for dyad d at time t , Post_t is an indicator for the post-visibility period, Female_Giver_d indicates whether the feedback giver in dyad d is female, and α_d are dyad fixed effects that control for all time-invariant characteristics of each giver-recipient pair. Standard errors are clustered at the dyad level to account for within-dyad correlation. The coefficient of interest, β_2 , captures the differential effect of increased visibility on female versus male feedback givers, holding constant all stable characteristics of the dyad relationship. Because women are the reference category, a negative β_2 indicates that men's feedback rose more than women's over the period, and a positive β_2 the reverse. The dyad fixed effects absorb the main effect of giver gender, as gender is time-invariant within each dyad. In fact, because we compare feedback exchanged within the same giver-recipient pair across both periods, the dyad fixed effects absorb every characteristic of that pair that is stable over the study

window. This includes the giver and recipient’s geographic location, business unit, and tenure, along with any fixed feature of the giver and of the working relationship itself. The estimate of interest therefore captures how a given giver adjusted their feedback toward a given recipient as visibility changed, net of all such stable relationship characteristics, rather than any compositional difference in who gave feedback, to whom, or where they sat.

RESULTS

Two sets of results help contextualize the main results to follow. First, in support of framing this 360-feedback as low-promotability work, we find little correspondence between feedback givers’ engagement in providing feedback and their likelihood of receiving a promotion in the following year. In giver-level models predicting promotion from the six main engagement constructs, net of giver gender, rank, compensation grade, and number of recipients, the constructs are jointly unrelated to whether the giver was promoted. Nor does any individual construct survive a Benjamini-Hochberg correction for multiple comparisons. This lack of correspondence is robust to using either time period in our sample (2017 language predicting 2018 promotion: $\chi^2(5) = 3.57, p = .61$; 2018 language predicting 2019 promotion: $\chi^2(5) = 7.96, p = .16$), and to a linear probability specification and separate estimates for men and women (see Appendix B). Engagement in giving feedback, therefore, is not rewarded with promotion, and nothing about making it visible alters that.

Second, the shift from a low-visibility to high-visibility platform had little effect on the language differences *received* by men and women. Overall, feedback became more positive and detailed, but this change did not alter the gender gaps in receiving these attributes. Furthermore, as detailed in Appendix C, our analysis of gender pairings between feedback givers and recipients revealed that men’s increased engagement, relative to women, after the visibility

change occurred regardless of whether evaluators were giving feedback to men or women. In other words, the gender of the feedback recipient was not a significant factor in driving the convergence in engagement between men and women.

Baseline Differences in Low-Visibility Period

We first establish baseline gender differences in engagement during the low-visibility pre-period. In this period, women's feedback showed greater engagement than men's across multiple measures, in both effort and recipient-orientation. Apart from Word Count, all LIWC constructs used in this study are scored on a 0-100 scale where 100 represents the maximum possible score on any given construct. For three of the four summary measures, women feedback givers displayed values consistent with greater recipient-orientation in the low-visibility pre-period. Specifically, women's feedback scored directionally higher on Tone (difference = 2.00, $p = 0.07$), a marker of the warmth associated with recipient-oriented feedback. Women's feedback also scored lower on Analytical Thinking difference = -3.38, $p < 0.01$; See Table 2), a direction the feedback literature tends to link to warmer, more accessible language. Women's feedback also carried more authority, scoring higher on Clout (difference = 4.39, $p < 0.001$). We do not find baseline differences in Authenticity between men and women feedback givers. In addition, we also find that women feedback givers use on average 7.61 more words than men ($p < 0.001$) in the low-visibility pre-period. Further, women use more linguistically dimensional language in their feedback in the low-visibility pre-period as well (Linguistic Dimensionality difference = 2.18 [$p < 0.001$]).³

Insert Table 2 about here

Difference-in-Difference Between Male and Female Feedback across Visibility Conditions

After the visibility change, feedback shifted, but not uniformly across measures, and the gendered part of the shift was strongest in the recipient-oriented dimensions (see Table 3). On Tone, men’s feedback rose by 2.94 points ($p < 0.01$) while women’s was essentially unchanged (-0.83 , n.s.), a significant divergence ($\text{DiD} = -3.76$, $p < 0.05$). On Clout, the divergence was larger and took the form of a crossover: men’s Clout rose by 3.05 points ($p < 0.01$) while women’s fell by 3.32 ($p < 0.01$), so that men, who scored lower at baseline, now scored higher ($\text{DiD} = -6.36$, $p < 0.001$). Figure 1 traces each measure, in standardized z-scores for cross-measure comparison, from the low- to the high-visibility period, illustrating especially the recipient-oriented measures, Tone and Clout, where men close and cross the engagement measurement gap. Authenticity, the third recipient-oriented measure, showed no significant gender difference in change ($\text{DiD} = 2.01$, n.s.).

A parallel, if weaker, gendered pattern appears in the effort measures. Feedback grew longer in Word Count for both genders, by 11.02 words among men ($p < 0.001$) and 4.82 among women ($p < 0.05$); men’s larger increase is consistent with a gendered effort response, though it falls just short of conventional significance ($\text{DiD} = -6.20$, $p = 0.07$). The Linguistic Dimensionality composite showed no overall gender difference ($\text{DiD} = -0.80$, n.s.) (See Appendix A for the Linguistic Dimensionality subcomponent analysis and figures).

Not every dimension moved in a gendered way. Analytical Thinking rose for both men (4.17 , $p < 0.001$) and women (2.40 , $p < 0.05$) with no significant divergence ($\text{DiD} = -1.77$, n.s.), suggesting more of a general platform-wide shift rather than a gendered one.

Insert Table 3 about here
Insert Figure 1 about here

Differences by Organizational Rank

1
2
3 The redesign made feedback visible to its recipient regardless of the recipient's rank, so the
4
5 mechanical change was the same throughout the hierarchy. What differed across the hierarchy
6
7 was who this newly present audience was: a subordinate, a peer, or a superior. Our account
8
9 predicts that this should matter. If visibility works by making a feedback giver answerable to the
10
11 person they evaluate, its effects should vary with the giver's standing relative to that person. The
12
13 accountability literature identifies an audience with standing to judge as an activating condition
14
15 (Lerner and Tetlock 1999), and when the newly present audience is a subordinate, the authority
16
17 relation itself supplies that standing: there, authority becomes answerable to those subject to it
18
19 (Simmel 1950). Visibility to superiors or peers, by contrast, was present in both periods, so the
20
21 comparison across ranks isolates what adding the downward audience contributes.
22
23
24
25

26 We therefore examine whether the effects of increased visibility varied with the giver's
27
28 rank and the direction of feedback. Following Pharma Co. convention, we categorize the most
29
30 junior givers as "Individual Contributors" if their compensation falls among the lowest brackets,
31
32 and we classify employees whose brackets place them between Individual Contributors and
33
34 "Senior Managers" as "Managers", whether or not they directly supervise anyone. Across rank
35
36 increased visibility reduced the number of evaluations. Individual Contributors left an average of
37
38 0.31 fewer evaluations in the high-visibility period compared to the low-visibility period (n.s., p
39
40 < 0.20). Managers left an average of 1.04 fewer evaluations in the high-visibility period
41
42 compared to the low-visibility period ($p < 0.001$), and Senior Managers 0.73 fewer ($p < 0.001$);
43
44 see Table 1.
45
46
47
48

49 Just over half (51%) of our cross-period dyadic data come from peer dyads, where the
50
51 feedback giver and recipients share the same compensation grade bracket (See Table 4). Most of
52
53 the remaining, non-peer data come from Senior Managers giving downward feedback to
54
55
56
57
58
59
60

Managers or Individual Contributors (17% of the total). Upward feedback accounts for smaller shares: 6% from Individual Contributors and 9% from Managers, followed by downward feedback from Managers (7%).

Insert Table 4 about here

We begin with peers, the most common configuration. Here the gendered divergence is limited: only Clout shows a significant difference in men increasing their Clout more than women ($DiD = -4.79, p < 0.05$), while Tone, Word Count, Linguistic Dimensionality, Authenticity, and Analytical Thinking show no significant differences. The divergence is narrower and smaller than in downward feedback, which we turn to next.

In downward feedback, men’s feedback changed more positively than women’s, reflected in the negative interaction coefficients in Table 5 (negative values signify a greater increase among male than female givers). The largest increase was concentrated among male Managers giving feedback to Individual Contributors, consistent with our theorized mechanism. This is the configuration where accountability pressures were expected to be strongest: Managers delivering downward feedback now faced the subordinates they evaluate as part of their audience, creating reputational and relational pressures to appear fair, engaged, and authoritative. These pressures are weaker when feedback flows between peers or upward, where the gender differences are smaller. Among Manager-to-Individual-Contributor dyads, men increased their Clout ($DiD = -13.79, p < 0.05$) relative to women’s decrease, and their Tone ($DiD = -13.90, p < 0.05$), a shift toward confident, higher-power language and greater positivity.

Insert Table 5 about here

The pattern also holds at the top of the hierarchy. In Senior-Manager-downward feedback, men again showed greater increases than women in Tone ($b = -9.63, p = 0.014$) and Linguistic Dimensionality ($b = -3.55, p < 0.05$; the subcomponents underlying these Linguistic Dimensionality estimates appear in Appendix Table A5), with a marginal difference in Word Count ($b = -15.72, p < 0.10$). Because the recipients of downward feedback are not evenly divided by gender, a natural question is whether these patterns reflect the recipient's gender rather than the direction of feedback; we test this directly in Appendix C and find they do not.

Additional Analyses Investigating Potential Shifts in Substance or Content of Feedback

The focus of this paper, and the analyses presented so far, pertain to the level of engagement in the task of giving feedback. The evidence points to male feedback givers, especially men in management, increasing their feedback engagement to a greater extent than female givers over the change in visibility, thereby 'catching up' to the baseline engagement which women were giving prior to the increased visibility.

A natural question follows: does this shift in engagement also bring about a substantive change in the content of the feedback? To address this question, we supplemented our LIWC analysis by applying to our anonymized corpus of dyadic feedback a validated dictionary of 88 codes (see Appendix Exhibit D1 for a full list of coding scheme and a discussion of why we chose this analytical strategy³), developed by Correll et al. 2020 for performance evaluations of technical employees at a Silicon Valley technology company. These codes capture substantive evaluative dimensions, such as mentions of technical skills, communal behaviors, agentic language, and standout descriptors, that bear directly on career outcomes for the feedback recipient.

The main takeaway from these results is that the visibility change moved the substantive content of feedback far less than it moved its manner of engagement. In the low-visibility pre-period, 5 of the 88 Correll et al. (2020) codes (6%) showed statistically significant gender differences in their occurrence across the dyadic corpus, with women accounting for greater usage in 4 out of these 5 codes. After the change, 8 codes (9%) showed statistically significant gender differences, with women accounting for greater usage in 7 out of 8 of the codes. Three of these codes (technical expertise, visionary, and communal), were common to both periods and similar in magnitude and significance. Overall then, these substantive codes did show some movement, so these results are not a flat null. But against the sizable shifts in Tone, Clout, and effort dimensions documented in our main analyses, the content changes are modest. In other words, givers changed how they communicated their assessments much more than they changed what those assessments were.

DISCUSSION

This study shows how a small change to the design of a feedback platform, making feedback visible to its recipient rather than only to the recipient’s manager, significantly altered differences in how men and women engage with the low-promotability task of giving feedback. Under the low-visibility conditions of the earlier period, men provided feedback that was less recipient-oriented, less linguistically engaged, and shorter than women’s across multiple dimensions. Following the platform redesign, men increased their recipient orientation to a significantly greater extent than women, particularly in Tone, along with a rise in Clout that women did not share. Men’s effort rose as well, reflected in Word Count and several subcomponents of Linguistic Dimensionality. The gap narrowed primarily because men increased their effort across multiple dimensions, even as women’s Clout indicator declined.

1
2
3 Women's adjustments were otherwise comparatively modest, and together these patterns suggest
4
5 that increased visibility drew men toward women's pre-existing level of investment in this work.
6
7

8 The concentration of these effects among managers giving downward feedback is
9
10 consistent with our account of what the visibility change added: an audience with standing to
11
12 judge, the condition the accountability literature identifies as activating (Lerner and Tetlock
13
14 1999). Visibility to superiors and peers, the channels prior work has emphasized, operated in
15
16 both periods, since the recipient's manager could always read the feedback; the change newly
17
18 introduced the downward recipient. That audience is one over whom the feedback givers hold
19
20 authority, but authority is not a one-way imposition: it draws on recognition from those subject
21
22 to it, so that even within a hierarchy a superior remains in part answerable to the response of
23
24 those below (Simmel 1950). Visibility to subordinates activates that latent answerability, and the
25
26 act of evaluation becomes, in part, an exercise of authority conducted before the people subject
27
28 to it. Male managers, who had greater latitude to disengage under low-visibility conditions,
29
30 responded to this new audience with the largest increases in engagement.
31
32
33
34

35 Some of the linguistic changes are more suggestive than definitive. Men's Clout
36
37 increased while women's declined, indicating that visibility did not always move men and
38
39 women toward identical forms of engagement. We interpret this pattern cautiously, and in line
40
41 with the status logic set out earlier. Men could therefore project more authoritative language at
42
43 little relational cost, while women had reason to ease theirs (Eagly and Karau 2002; Cardador et
44
45 al. 2022). We do not read women's lower Clout as disengagement: their effort stayed near its
46
47 already higher baseline, and the gap closed mainly because men's increased. This pattern is
48
49 consistent with the interpretation that while both groups seek good outcomes in employee
50
51 behavior, men pursue this by slightly increasing Clout while women reduce Clout to achieve a
52
53
54
55
56
57
58
59
60

persuasive connection while protecting themselves from the costs an authoritative display carries for them. A similar selectivity appears in Authenticity, which did not rise with Tone and Clout and if anything declined among men. That the move under visibility favored warmth and standing over self-revelation suggests men calibrated on the dimensions that manage how the feedback would land, rather than on the one that exposes the giver, which is what an audience-management reading would predict. That the gender of the feedback recipient did not significantly moderate these effects further supports the interpretation that what changed was men's orientation to the task itself, not their behavior toward particular individuals.

One change ran the other way. Analytical Thinking rose for both men and women. As we note in describing the measure, the value of this direction is not settled: a higher score moves away from the warmth and accessibility usually tied to well-received feedback, yet it also reflects the more formal, specific language whose value the feedback literature treats as mixed. Given that the increase was parallel across genders, modest in size, and ambiguous in what it implies for how the feedback would be received, we read it as a shared response to an audience of recipients rather than part of the gendered pattern at issue here.

The broader pattern across our measures is one of manner of engagement shifting more than content. The wider audience raised the pressure on everyone to give feedback that would read as responsive and effortful, and both groups responded with more formal, more elaborated language (Bourdieu 1991). A change in the visibility conditions under which feedback is given surfaced less in what was assessed than in how the assessment was delivered (Malinowski 1923; Goffman 1967; Weeks 2004).

CONCLUSIONS

These findings contribute to three streams of scholarship. First, within the literature on gender and low-promotability work, this study demonstrates that gendered engagement patterns are not fixed properties of individuals or even of socialization paradigms but are responsive to how tasks are designed. Prior work has established that women do more low-promotability work than men (Babcock et al. 2017a; Kaplan 2022; Misra et al. 2012) and that they engage more intensively even when assignment is equalized (Conzon 2023; Feldberg 2022; Ranganathan and Shivaram 2021). The predominating explanations for these phenomena operate at the individual level, rooted in socialization or gendered expectations (Bellas 1999; El-Alayli et al. 2018). Our findings extend this literature by showing that a modest design change, a single change who can see the work, narrowed the gender gap in engagement, and did so by drawing men more fully into the relational work. Importantly, the gap narrowed primarily because men increased their effort, not because women decreased theirs. This suggests that the gendered engagement gap in low-promotability work is at least partly sustained by design features that typically allow men to disengage without consequence, rather than by deeply rooted dispositional differences alone. By addressing the underlying structure of such tasks, organizations can reduce gender disparities through systemic change rather than behavior-change interventions.

Second, this study extends theories of transparency and accountability in organizations by connecting them to classical sociology of audience and conduct. The insight that an act takes its character from who is positioned to observe it (Goffman 1959; Simmel 1950) gives purchase on a question the organizational literature has named but not fully examined. Prior work has traced how evaluations change when they are exposed upward, to the authorities who monitor decision-makers, or laterally, to peers, but has attended chiefly to the fairness of outcomes for those being evaluated rather than to the conduct of those doing the evaluating (Castilla 2008, 2011, 2015;

Bernstein 2012; Mas and Moretti 2009). What none of this work examines is the case of making an evaluative act visible to the person being evaluated. We find that adding the recipient to the audience changed how givers carried out the evaluation and that men’s response narrowed the gender gap in engagement seen in the period when feedback could be kept from its recipient.

Third, our findings speak to *why* this downward form of visibility carries the force it does. Authority is not a one-way imposition but a relation that remains in part answerable to those subject to it (Simmel 1950): a manager’s standing depends on being credited by subordinates, not only by superiors. When feedback could be hidden from the recipient, a manager’s conduct toward that subordinate went unseen by the one person whose recognition that authority rests on. Making the feedback visible closes that gap, so that the exercise of authority is now carried out before its subject and must answer to their response. This downward answerability matches the conditions the accountability literature identifies as an activating condition, an audience with standing to judge (Lerner and Tetlock 1999), but it locates the source of that standing in the authority relation itself rather than in the audience’s evaluative role alone (Simmel 1950; Weber 1978; Bourdieu 1991). That the pressure fell most heavily on men is what the account predicts: they had the most room to disengage when the work was less visible, and so the most room to change once it was more visible.

The Nature of the Observed Changes

An important feature of our findings is that the changes induced by visibility were concentrated in how feedback was delivered. The 88-code analysis using the Correll et al. (2020) validated dictionary revealed that the substantive content of feedback, the evaluative categories invoked and the types of performance discussed, remained largely unchanged across visibility conditions. What changed was the tone of the feedback, the confidence with which it was conveyed, and the effort invested in crafting it. This pattern is consistent with our theoretical expectation that

1
2
3 visibility triggers anticipatory audience management (Goffman 1959; Tetlock 1983) rather than a
4
5 change in the underlying evaluative judgment.
6

7
8 That the manner of feedback can change while its content holds constant does not make
9
10 the change merely cosmetic. The way a relational act is delivered does work that its propositional
11
12 content does not, sustaining the relationship between the parties even when the assessment itself
13
14 is unchanged (Goffman 1967; Weeks 2004). Read this way, men's move toward warmer, more
15
16 invested delivery is a move toward the relationship-fostering that women's feedback already
17
18 carried under conditions of lower visibility. Whether such a shift in presentation represents a
19
20 meaningful improvement depends on what we expect visibility to accomplish. From the
21
22 perspective of the feedback recipient, more carefully delivered feedback may be easier to absorb
23
24 and act upon even if the underlying evaluative content is the same.
25
26
27

28
29 A further implication concerns the removal of choice itself. Because the redesign
30
31 eliminated the option to hide feedback, it closed off a channel through which the gendered
32
33 pattern could operate. Where givers could previously decide whether their feedback would reach
34
35 its recipient, that decision was itself potentially a site of inequality. Our design cannot observe
36
37 who used the concealment option, but the pattern we document is consistent with concealability
38
39 having done gendered work: men's engagement sat below women's when feedback could be
40
41 hidden and rose once it could not, which is what one would expect if women were more inclined
42
43 to make their feedback visible under pressure to demonstrate their investment at each
44
45 opportunity (Feldberg 2022) while men retained the latitude to keep theirs hidden. Whether
46
47 givers used that latitude in this gendered way is a question our data cannot settle and one worth
48
49 settling: platforms of this kind are in wide organizational use, and visibility settings could be
50
51 varied deliberately to test it. What our results do show is that removing the option placed the
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

same condition on everyone. Whatever use givers had made of that latitude, the redesign foreclosed it, and under the imposed visibility men’s engagement rose toward the level women had maintained all along. Removing the optionality of visibility thus worked as a structural constraint on inequality.

At the same time, these findings reveal a limit of visibility as an intervention. Making feedback visible to recipients altered how men carried out the task of giving feedback, but it did not change the low-promotability nature of the work itself. Feedback engagement remained unrelated to the giver’s subsequent promotion in the high-visibility post-period. Thus, visibility equalized engagement without altering the structural devaluation of the task. This distinction matters for organizational design: visibility can narrow gender gaps in how low-promotability work is carried out, but closing those gaps does not, on its own, make the work more valued or more rewarded. As long as organizations treat giving evaluative feedback as incidental to career advancement, even equalized engagement will remain a form of unrewarded labor for both men and women.

Notes

1. Although the exact formulas for these four composite LIWC constructs are proprietary, given the level of detail provided in the manual (Boyd et al. 2022) and corpus of studies using the same variables (Germann et al. 2024; Jordan et al. 2019; Kyprianou et al. 2025; Newman et al. 2003), we are confident that this approach offers a transparent way of measuring our constructs of interest.

2. Without timestamp data, we cannot assess the time spent giving feedback, which may differ depending on the intensity of effort to generate the same feedback. Women wrote 7.61 (14%) more words on average than men in the low-visibility pre-period. Since feedback givers leave 5.5 evaluations during this period, if we use average typing speed to estimate the difference in time spent on the task, back-of-the-envelope calculations suggest that women spend roughly 1 additional minute more dedicated to this task than their male counterpart in this five-month period (using a roughly 40 words per minute (American Society of Administrative Professionals 2023)). However, this time-based calculation represents a crude proxy for effort, as word count alone is widely recognized as an inadequate measure that fails to capture the substance, coherence, and psychological dimensions of communication that distinguish effective from ineffective feedback (cf. He and Gao 2023).

3. We also considered word-embedding approaches, which are useful for measuring theoretically specified semantic dimensions in text, such as gendered associations in organizational language (Castilla and Rho 2023; Kozlowski, Taddy, and Evans 2019). Because our goal here was to assess changes across multiple forms of evaluative content using categories already developed and applied in prior research, Correll et al.'s (2020) coding schema offered a more direct fit.

REFERENCES

Abraham, Mabel, Tristan L. Botelho, and Gabrielle Lamont-Dobbin. 2024. “The (Re)Production of Inequality in Evaluations: A Unifying Framework Outlining How Gender and Race Shape Evaluative Outcomes.” *Research in Organizational Behavior*.

American Society of Administrative Professionals. 2023. “Average Words per Minute Typing: How Fast Is Fast Enough?” Retrieved (<https://www.asaporg.com/efficiency-skills/average-words-per-minute-typing-how-fast-is-fast-enough>).

Atwater, Leanne E., David A. Waldman, and Joan F. Brett. 2002. “Understanding and Optimizing Multisource Feedback.” *Human Resource Management* 41(2):193–208.

Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart. 2017a. “Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability.” *American Economic Review* 107(3):714–47.

Babcock, Linda, Maria P. Recalde, and Lise Vesterlund. 2017b. “Gender Differences in the Allocation of Low-Promotability Tasks: The Role of Backlash.” *American Economic Review* 107(5):131–35.

Babcock, Linda, Brenda Peyser, Lise Vesterlund, and Laurie Weingart. 2022. *The No Club: Putting a Stop to Women’s Dead-End Work*. New York: Simon and Schuster.

Bandelj, Nina. 2020. “Relational Work in the Economy.” *Annual Review of Sociology* 46(1):251–72.

Baron, James N., and David M. Kreps. 1999. *Strategic Human Resources: Frameworks for General Managers*. New York: John Wiley & Sons.

Bellas, Marcia L. 1999. “Emotional Labor in Academia: The Case of Professors.” *The Annals of the American Academy of Political and Social Science* 561(1):96–110.

Benschop, Yvonne, and Hans Doorewaard. 1998. “Covered by Equality: The Gender Subtext of Organizations.” *Organization Studies* 19(5):787–805.

Bernstein, Ethan S. 2012. “The Transparency Paradox: A Role for Privacy in Organizational Learning and Operational Control.” *Administrative Science Quarterly* 57(2):181–216.

Bernstein, Ethan, and Shelley Xin Li. 2026. "The performance effects of giving front-line employees direct access to performance data and thereby limiting the supervisor’s feedback-intermediation role: evidence from a field experiment." *Management Science* 72, no. 2: 805-835.

Bhatt, Anjali M., Amir Goldberg, and Sameer B. Srivastava. 2022. “A Language-Based Method for Assessing Symbolic Boundary Maintenance between Social Groups.” *Sociological Methods & Research* 51(4):1681–1720.

Blau, Peter M. 1964. *Exchange and Power in Social Life*. New York: Wiley.

Botelho, Tristan L. 2024. “From Audience to Evaluator: When Visibility into Prior Evaluations Leads to Convergence or Divergence in Subsequent Evaluations among Professionals.” *Organization Science*.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Translated by R. Nice. Cambridge, MA: Harvard University Press.

- Bourdieu, Pierre. 1991. *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.
- Boyd, Ryan L., Ashwini Ashokkumar, Sarah Seraj, and James W. Pennebaker. 2022. "The Development and Psychometric Properties of LIWC-22." Austin, TX: University of Texas at Austin.
- Brewer, Alexandra, Melissa Osborne, Anna S. Mueller, Daniel M. O'Connor, Arjun Dayal, and Vineet M. Arora. 2020. "Who Gets the Benefit of the Doubt? Performance Evaluations, Medical Errors, and the Production of Gender Inequality in Emergency Medical Education." *American Sociological Review* 85(2):247–70.
- Brescoll, Victoria L. 2011. "Who Takes the Floor and Why: Gender, Power, and Volubility in Organizations." *Administrative Science Quarterly* 56(4):622–41.
- Brody, Leslie R. 1993. "On Understanding Gender Differences in the Expression of Emotion: Gender Roles, Socialization, and Language." Pp. 87–121 in *Human Feelings: Explorations in Affect Development and Meaning*, edited by S. L. Ablon, D. Brown, E. J. Khantzian, and J. E. Mack. Hillsdale, NJ: Analytic Press.
- Buckingham, Marcus, and Ashley Goodall. 2015. "Reinventing Performance Management." *Harvard Business Review* 93(4):40–50.
- Cappelli, Peter, and Anna Tavis. 2016. "The Performance Management Revolution." *Harvard Business Review* 94(10):58–67.
- Cardador, M. Teresa, Patrick L. Hill, and Arghavan Salles. 2022. "Unpacking the Status-Leveling Burden for Women in Male-Dominated Occupations." *Administrative Science Quarterly* 67(1):237–84.
- Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125(3):1101–44.
- Castilla, Emilio J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113(6):1479–1526.
- Castilla, Emilio J. 2011. "Bringing Managers Back In: Managerial Influences on Workplace Inequality." *American Sociological Review* 76(5):667–94.
- Castilla, Emilio J. 2015. "Accounting for the Gap: A Firm Study Manipulating Organizational Accountability and Transparency in Pay Decisions." *Organization Science* 26(2):311–33.
- Castilla, Emilio J., and Hye Jin Rho. 2023. "The Gendering of Job Postings in the Online Recruitment Process." *Management Science* 69(11):6912–39.
- Cech, Erin A. 2013. "Ideological Wage Inequalities? The Technical/Social Dualism and the Gender Wage Gap in Engineering." *Social Forces* 91(4):1147–82.
- Chan, Curtis K., and Michel Anteby. 2016. "Task Segregation as a Mechanism for Within-Job Inequality: Women and Men of the Transportation Security Administration." *Administrative Science Quarterly* 61(2):184–216.
- Chaplin, Tara M., Pamela M. Cole, and Carolyn Zahn-Waxler. 2005. "Parental Socialization of Emotion Expression: Gender Differences and Relations to Child Adjustment." *Emotion* 5(1):80–88.

- Cleveland, Jeanette N., Audrey S. Lim, and Kevin R. Murphy. 2007. "Feedback Phobia? Why Employees Do Not Want to Give or Receive Performance Feedback." Pp. 168–86 in *Research Companion to the Dysfunctional Workplace: Management Challenges and Symptoms*, edited by J. Langan-Fox, C. L. Cooper, and R. J. Klimoski. Cheltenham, UK: Edward Elgar.
- Connell, R. W., and Julian Wood. 2005. "Globalization and Business Masculinities." *Men and Masculinities* 7(4):347–64.
- Conzon, Vanessa M. 2023. "The Equality Policy Paradox: Gender Differences in How Managers Implement Gender Equality–Related Policies." *Administrative Science Quarterly* 68(3):648–90.
- Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106(6):1691–1730.
- Correll, Shelley J., Katherine R. Weisshaar, Alison T. Wynn, and JoAnne Delfino Wehner. 2020. "Inside the Black Box of Organizational Life: The Gendered Language of Performance Assessment." *American Sociological Review* 85(6):1022–50.
- De Pater, Irene E., Annelies E. M. Van Vianen, and Myriam N. Bechtoldt. 2010. "Gender Differences in Job Challenge: A Matter of Task Allocation." *Gender, Work & Organization* 17(4):433–53.
- Di Fiore, Alessandro, and Marcio Souza. 2021. "Are Peer Reviews the Future of Performance Evaluations?" *Harvard Business Review*, January 12.
- Eagan, Thomas, and Patricia Garvey. 2015. "Stressing Out: Connecting Race, Gender, and Stress with Faculty Productivity." *The Journal of Higher Education* 86(6):923–54.
- Eagly, Alice Hendrickson, and Steven J. Karau. 2002. "Role Congruity Theory of Prejudice toward Female Leaders." *Psychological Review* 109(3):573–98.
- El-Alayli, Amani, Ashley A. Hansen-Brown, and Michelle Ceynar. 2018. "Dancing Backwards in High Heels: Female Professors Experience More Work Demands and Special Favor Requests, Particularly from Academically Entitled Students." *Sex Roles* 79(3–4):136–50.
- Elvira, Marta M., and Mary E. Graham. 2002. "Not Just a Formality: Pay System Formalization and Sex-Related Earnings Effects." *Organization Science* 13(6):601–17.
- Emirbayer, Mustafa. 1997. "Manifesto for a Relational Sociology." *American Journal of Sociology* 103(2):281–317.
- Feldberg, Alexandra C. 2022. "The Task Bind: Explaining Gender Differences in Managerial Tasks and Performance." *Administrative Science Quarterly* 67(4):1049–92.
- Fletcher, Joyce K. 1999. *Disappearing Acts: Gender, Power, and Relational Practice at Work*. Cambridge, MA: MIT Press.
- Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Germann, Frank, Stephen J. Anderson, Pradeep K. Chintagunta, and Naufel Vilcassim. 2024. "Frontiers: Breaking the Glass Ceiling: Empowering Female Entrepreneurs through Female Mentors." *Marketing Science* 43(2):244–53.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday Anchor.

- Goffman, Erving. 1967. *Interaction Ritual: Essays on Face-to-Face Behavior*. New York: Pantheon Books.
- Goldberg, Amir, Sameer B. Srivastava, V. Govind Manian, William Monroe, and Christopher Potts. 2016. "Fitting In or Standing Out? The Tradeoffs of Structural and Cultural Embeddedness." *American Sociological Review* 81(6):1190–1222.
- Goodman, Jodi S., Robert E. Wood, and Margaret Hendrickx. 2004. "Feedback Specificity, Exploration, and Learning." *Journal of Applied Psychology* 89(2):248–62.
- Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42(4):1009–55.
- Hart, Peter, and William Cress. 2008. "Are Women Faculty Just 'Worrywarts'? Accounting for Gender Differences in Self-Reported Stress." *Journal of Human Behavior in the Social Environment* 17(1):175–93.
- He, Wenjing, and Ying Gao. 2023. "Explicating Peer Feedback Quality and Its Impact on Feedback Implementation in EFL Writing." *Frontiers in Psychology* 14:1177094.
- Heath, Janae K., Jennifer R. Kogan, Eric S. Holmboe, Lisa Conforti, Yoon Soo Park, and C. Jessica Dine. 2024. "Gender Differences in Work-Based Assessment Scores and Narrative Comments after Direct Observation." *Journal of General Internal Medicine* 39(10):1795–1802.
- Heilman, Madeline E. 1983. "Sex Bias in Work Settings: The Lack of Fit Model." Pp. 269–98 in *Research in Organizational Behavior*, Vol. 5, edited by B. Staw and L. Cummings. Greenwich, CT: JAI.
- Heilman, Madeline E. 2001. "Description and Prescription: How Gender Stereotypes Prevent Women's Ascent Up the Organizational Ladder." *Journal of Social Issues* 57(4):657–74.
- Hibbard, David R., and Duane Buhrmester. 1998. "The Role of Peers in the Socialization of Gender-Related Social Interaction Styles." *Sex Roles* 39:185–202.
- Hochschild, Arlie Russell. 1979. "Emotion Work, Feeling Rules, and Social Structure." *American Journal of Sociology* 85(3):551–75.
- Hochschild, Arlie Russell. 1983. *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press.
- Holm, Audrey L., Bella T. Fong, and Michel Anteby. 2024. "The Perils of Voice Veneer: The Case of Disneyland Puppeteers' Unionization Efforts." *Academy of Management Discoveries* 10(4):527–42.
- Huising, Ruthanne, and Susan S. Silbey. 2021. "Accountability Infrastructures: Pragmatic Compliance inside Organizations." *Regulation & Governance* 15:S40–S62.
- Jordan, Kayla N., Joanna Sterling, James W. Pennebaker, and Ryan L. Boyd. 2019. "Examining Long-Term Trends in Politics and Culture through Language of Political Leaders and Cultural Institutions." *Proceedings of the National Academy of Sciences* 116(9):3476–81.
- Kacewicz, Ewa, James W. Pennebaker, Matthew Davis, Moongee Jeon, and Arthur C. Graesser. 2014. "Pronoun Use Reflects Standings in Social Hierarchies." *Journal of Language and Social Psychology* 33(2):125–43.

- Kaplan, Adi. 2022. “‘Just Let It Pass By and It Will Fall on Some Woman’: Invisible Work in the Labor Market.” *Gender & Society* 36(6):838–68.
- Kelly, Erin L., Samantha K. Ammons, Kelly Chermack, and Phyllis Moen. 2010. “Gendered Challenge, Gendered Response: Confronting the Ideal Worker Norm in a White-Collar Organization.” *Gender & Society* 24(3):281–303.
- Kilis, Selcan, and Zahide Yıldırım. 2018. “Investigation of Community of Inquiry Framework in Regard to Self-Regulation, Metacognition and Motivation.” *Computers & Education* 126:53–64.
- Klimoski, Richard, and Lawrence Inks. 1990. “Accountability Forces in Performance Appraisal.” *Organizational Behavior and Human Decision Processes* 45(2):194–208.
- Knapp, Mark L., Roderick P. Hart, and Harry S. Dennis. 1974. “An Exploration of Deception as a Communication Construct.” *Human Communication Research* 1(1):15–29.
- Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. “The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings.” *American Sociological Review* 84(5):905–49.
- Kyprianou, Christina, Siddharth Vedula, Wenxi Pu, and Markus Fitza. 2025. “Shifts in National Entrepreneurial Culture: The Promise of Linguistic Cultural Artifacts and Machine Learning Analysis.” *Strategic Entrepreneurship Journal* 19(1):29–82.
- Lerner, Jennifer S., and Philip E. Tetlock. 1999. “Accounting for the Effects of Accountability.” *Psychological Bulletin* 125(2):255–75.
- Lewis, Patricia, and Ruth Simpson. 2007. “Gender and Emotions: Introduction.” Pp. 1–15 in *Gendering Emotions in Organizations*, edited by P. Lewis and R. Simpson. Basingstoke, UK: Palgrave Macmillan.
- Lewis, Patricia, and Ruth Simpson, eds. 2010. *Revealing and Concealing Gender: Issues of Visibility in Organizations*. New York: Springer.
- Lin, Yiwen, Renzhe Yu, and Nia Dowell. 2020. “LIWCs the Same, Not the Same: Gendered Linguistic Signals of Performance and Experience in Online STEM Courses.” Pp. 333–45 in *Artificial Intelligence in Education*, edited by I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, and E. Millán. Cham, Switzerland: Springer.
- Malinowski, Bronislaw. 1923. “The Problem of Meaning in Primitive Languages.” Pp. 296–336 in *The Meaning of Meaning*, edited by C. K. Ogden and I. A. Richards. London: Kegan Paul.
- Markowitz, David M., Maryam Kouchaki, Francesca Gino, Jeffrey T. Hancock, and Ryan L. Boyd. 2023. “Authentic First Impressions Relate to Interpersonal, Social, and Entrepreneurial Success.” *Social Psychological and Personality Science* 14(2):107–16.
- Mas, Alexandre, and Enrico Moretti. 2009. “Peers at Work.” *American Economic Review* 99(1):112–45.
- Mero, Neal P., Rebecca M. Guidice, and Amy L. Brownlee. 2007. “Accountability in a Performance Appraisal Context: The Effect of Audience and Form of Accounting on Rater Response and Behavior.” *Journal of Management* 33(2):223–52.

- Misra, Joya, Jennifer Hickes Lundquist, and Abby Templer. 2012. "Gender, Work Time, and Care Responsibilities among Faculty." *Sociological Forum* 27(2):300–323.
- Mobasser, Sanaz. 2018. "Emotions at Work: Norms of Emotional Expression and Gender Dynamics in Workplace Communication." PhD dissertation, Haas School of Business, University of California, Berkeley, Berkeley, CA.
- Murphy, Kevin R. 2020. "Performance Evaluation Will Not Die, but It Should." *Human Resource Management Journal* 30(1):13–31.
- Nelson, Laura K., Alexandra Brewer, Anna S. Mueller, Daniel M. O'Connor, Arjun Dayal, and Vineet M. Arora. 2023. "Taking the Time: The Implications of Workplace Assessment for Organizational Gender Inequality." *American Sociological Review* 88(4):627–55.
- Newman, Matthew L., James W. Pennebaker, Diane S. Berry, and Jane M. Richards. 2003. "Lying Words: Predicting Deception from Linguistic Styles." *Personality and Social Psychology Bulletin* 29(5):665–75.
- Newman, Matthew L., Carla J. Groom, Lori D. Handelman, and James W. Pennebaker. 2008. "Gender Differences in Language Use: An Analysis of 14,000 Text Samples." *Discourse Processes* 45(3):211–36.
- Pennebaker, James W., Matthias R. Mehl, and Kate G. Niederhoffer. 2003. "Psychological Aspects of Natural Language Use: Our Words, Our Selves." *Annual Review of Psychology* 54(1):547–77.
- Pennebaker, James W., Ryan L. Boyd, Kayla N. Jordan, and Kate Blackburn. 2015. *The Development and Psychometric Properties of LIWC2015*. Austin: University of Texas at Austin.
- Pfeffer, Jeffrey. 1994. *Competitive Advantage through People: Unleashing the Power of the Work Force*. Boston: Harvard Business School Press.
- Ranganathan, Aruna, and Ranjitha Shivaram. 2021. "Getting Their Hands Dirty: How Female Managers Motivate Female Worker Productivity through Subordinate Scut Work." *Management Science* 67(5):3299–3320.
- Reskin, Barbara, and Patricia A. Roos. 1990. *Job Queues, Gender Queues: Explaining Women's Inroads into Male Occupations*. Philadelphia: Temple University Press.
- Ridgeway, Cecilia L. 2001. "Gender, Status, and Leadership." *Journal of Social Issues* 57(4):637–55.
- Ridgeway, Cecilia L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. New York: Oxford University Press.
- Ridgeway, Cecilia L., and Shelley J. Correll. 2004. "Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations." *Gender & Society* 18(4):510–31.
- Rivera, Lauren A., and András Tilcsik. 2019. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84(2):248–74.

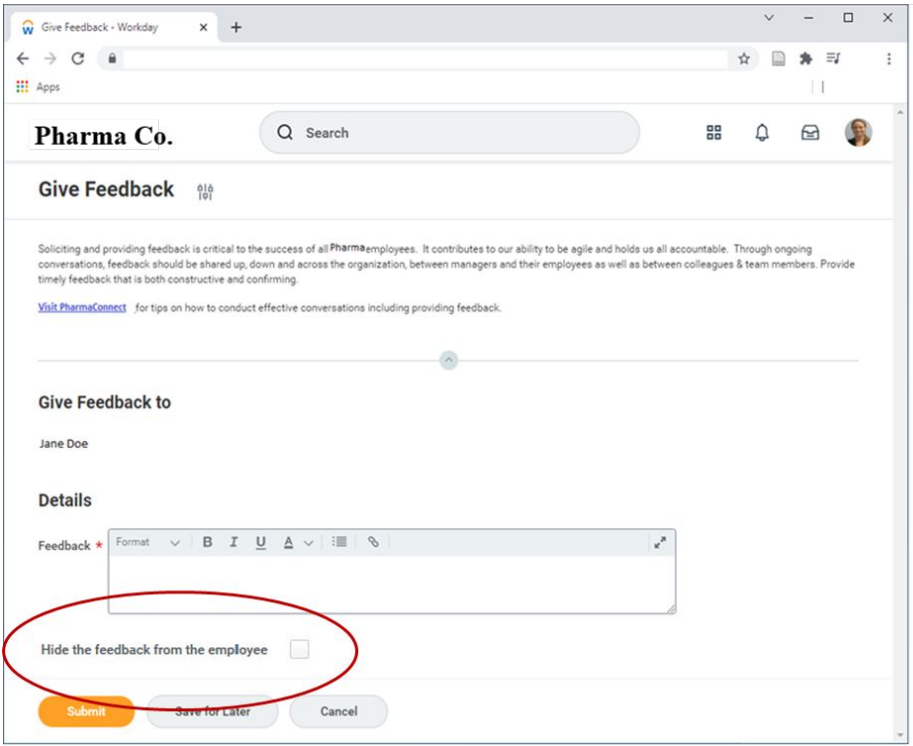
- Rivera, Michael, Liangfei Qiu, Subodha Kumar, and Tony Petrucci. 2021. "Are Traditional Performance Reviews Outdated? An Empirical Analysis on Continuous, Real-Time Feedback in the Workplace." *Information Systems Research* 32(2):517–40.
- Rudman, Laurie A., and Peter Glick. 2001. "Prescriptive Gender Stereotypes and Backlash toward Agentic Women." *Journal of Social Issues* 57(4):743–62.
- Sauder, Michael, Yongren Shi, and Freda Lynn. 2024. "Multiple Meritocracies: A Text-Based Analysis of Personal Narratives Revealing Distinct Frames of Success." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 10(5):86–117.
- Scott, Marvin B., and Stanford M. Lyman. 1968. "Accounts." *American Sociological Review* 33(1):46–62.
- Shute, Valerie J. 2008. "Focus on Formative Feedback." *Review of Educational Research* 78(1):153–89.
- Simmel, Georg. 1950. *The Sociology of Georg Simmel*. Translated and edited by K. H. Wolff. Glencoe, IL: Free Press.
- Simpson, Ruth, and Patricia Lewis. 2005. "An Investigation of Silence and a Scrutiny of Transparency: Re-examining Gender in Organization Literature through the Concepts of Voice and Visibility." *Human Relations* 58(10):1253–75.
- Tetlock, Philip E. 1983. "Accountability and Complexity of Thought." *Journal of Personality and Social Psychology* 45(1):74–83.
- Tetlock, Philip E. 1985. "Accountability: The Neglected Social Context of Judgment and Choice." *Research in Organizational Behavior* 7:297–332.
- Tetlock, Philip E. 1992. "The Impact of Accountability on Judgment and Choice: Toward a Social Contingency Model." *Advances in Experimental Social Psychology* 25:331–76.
- Toma, Catalina L., and Jonathan D. D'Angelo. 2015. "Tell-tale words: Linguistic cues used to infer the expertise of online medical advice." *Journal of Language and Social Psychology*, 34(1), 25–45.
- Tomaskovic-Devey, Donald, and Dustin Avent-Holt. 2019. *Relational Inequalities: An Organizational Approach*. New York: Oxford University Press.
- Tyler, Melissa, and Laurie Cohen. 2010. "Spaces That Matter: Gender Performativity and Organizational Space." *Organization Studies* 31(2):175–98.
- Vial, Andrea C., and Colleen M. Cowgill. 2022. "Heavier Lies Her Crown: Gendered Patterns of Leader Emotional Labor and Their Downstream Effects." *Frontiers in Psychology* 13:849566.
- Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*, Vol. 1. Berkeley: University of California Press.
- Weeden, Kim A. 2002. "Why Do Some Occupations Pay More Than Others? Social Closure and Earnings Inequality in the United States." *American Journal of Sociology* 108(1):55–101.
- Weeks, John. 2004. *Unpopular Culture: The Ritual of Complaint in a British Bank*. Chicago: University of Chicago Press.

- West, Candace, and Don H. Zimmerman. 1987. "Doing Gender." *Gender & Society* 1(2):125–51.
- Wharton, Amy S. 2009. "The Sociology of Emotional Labor." *Annual Review of Sociology* 35(1):147–65.
- Williams, Christine L. 1992. "The Glass Escalator: Hidden Advantages for Men in the 'Female' Professions." *Social Problems* 39(3):253–67.
- Wrong, Dennis H. 1979. *Power: Its Forms, Bases, and Uses*. New York: Harper & Row.
- Yu, Bei. 2014. "Language and Gender in Congressional Speech." *Literary and Linguistic Computing* 29(1):118–32.
- Zenger, Jack. 2016. "How Effective Are Your 360-Degree Feedback Assessments?" *Forbes*, March 10. Retrieved (<https://www.forbes.com/sites/jackzenger/2016/03/10/how-effective-are-your-360-degree-feedback-assessments/>).
- Zhang, Jurui. 2019. "What's Yours Is Mine: Exploring Customer Voice on Airbnb Using Text-Mining Approaches." *Journal of Consumer Marketing* 36(5):655–65.

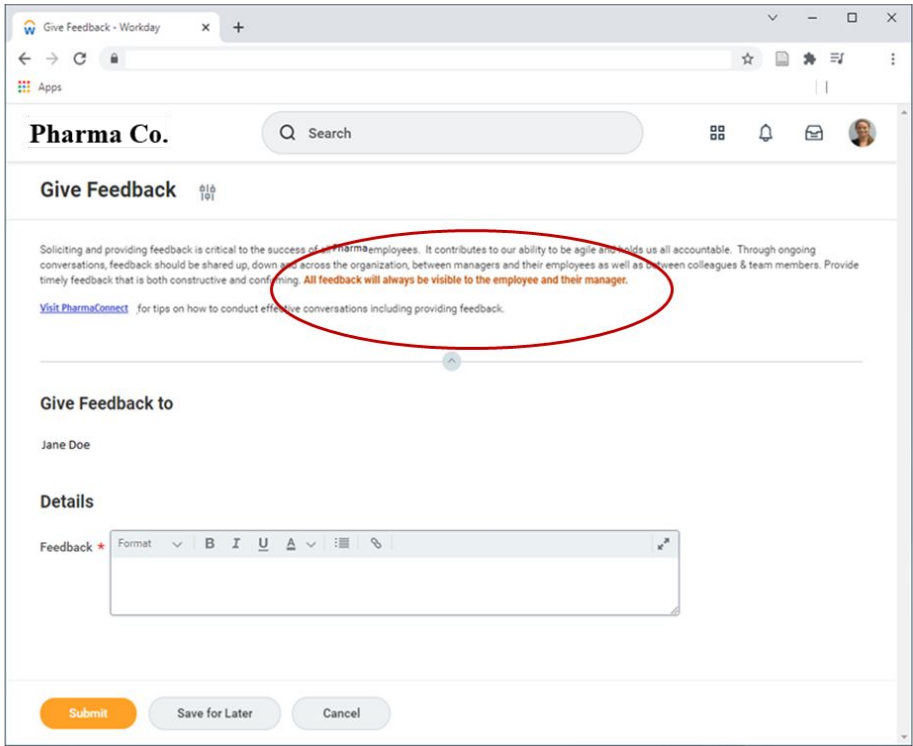
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Exhibit 1. Example screen of Pharma Co.’s 360-degree colleague feedback portal in Pre-Change, Privacy-Enabled, Low-Visibility Condition & Post-Change, High-Visible Condition (note: all colors and emphasis are present in the actual portal itself; only alterations this research called for was using a pseudonym for the name of the company depicted.)

“Pre-Period”: **Low Visibility** Condition



“Post-Period”: **High Visibility** Condition



TABLES

Table 1. Statistics for Feedback Givers and Evaluative Feedback Effort in 2017 and 2018

Giver group	n	Evals 2017	Evals 2018	Change	WC 2017	WC 2018	Change
Men	370	5.55	4.70	−0.85***	66.3	75.3	+8.98*
Individual contributors	53	4.91	4.66	−0.25	61.5	57.2	−4.32
Managers	162	5.89	4.68	−1.21***	62.3	75.2	+12.91*
Senior Managers	138	5.44	4.72	−0.72**	75.3	83.5	+8.22
Women	334	5.44	4.75	−0.69***	70.9	82.2	+11.36**
Individual contributors	63	4.57	4.21	−0.37	65.9	73.0	+7.11
Managers	151	5.47	4.60	−0.87***	69.9	84.9	+15.08*
Senior Managers	107	6.12	5.38	−0.74**	75.2	83.9	+8.69
Total	704	5.50	4.72	−0.77***	68.5	78.6	+10.11***

Note: Entries summarize feedback at the giver-level for the 704 givers (370 men, 334 women) who provided evaluative feedback within the 931 matched dyads in both periods. The evaluation columns report the average number of feedback evaluations a giver provided each year, and the word-count columns the average number of words per feedback evaluation. Change is the within-giver difference between periods, and significance comes from paired t-tests, since the same givers appear in both periods. Two-tailed: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 2. Baseline Gender Differences in Feedback Engagement, Low-Visibility Pre-Period

Engagement Measure	Women	Men	Difference (SE)	Larger Level
	Mean (SE)	Mean (SE)		
Tone	70.48 (0.73)	68.48 (0.71)	2.00 (1.12)	Women
Clout	59.10 (0.71)	54.71 (0.68)	4.39*** (1.23)	Women
Authenticity	33.19 (0.78)	33.64 (0.74)	-0.45 (1.24)	Men
Analytical Thinking	60.45 (0.71)	63.83 (0.66)	-3.38** (1.27)	Men
Word Count	62.53 (1.38)	54.93 (1.19)	7.61*** (2.27)	Women
Linguistic Dimensionality	65.53 (0.29)	63.36 (0.30)	2.18*** (0.56)	Women

Note: Entries in the Women and Men columns are mean LIWC scores for feedback given by women and men during the lower-visibility pre-period, with the standard error of each mean in parentheses, computed across the 3,870 evaluations (1,816 from women, 2,054 from men) exchanged within the 931 matched giver–recipient dyads in the low-visibility, pre-period. The Difference column reports the female-minus-male difference in means with its standard error in parentheses; positive values indicate higher scores among women. The difference and its significance test come from a regression of each measure on a female-giver indicator, with standard errors clustered at the dyad level to account for the multiple evaluations a giver provides within a dyad, the same treatment of within-dyad dependence used in the difference-in-differences models reported below. For every measure except Analytical Thinking, a higher score indicates more engaged, more recipient-oriented feedback; for Analytical Thinking, a higher score indicates cooler, more impersonal language, so men’s higher value reflects less personal engagement rather than more. * $p<.05$, ** $p<.01$, *** $p<.001$ (two-tailed).

Table 3. Change in Feedback Engagement from the Low- to High-Visibility Period, by Giver Gender

Engagement Measure	Men's Change	SE	Women's Change	SE	DiD (W – M)	SE
Tone	2.94**	1.06	–0.83	1.21	–3.76*	1.61
Clout	3.05**	0.98	–3.32**	1.12	–6.36***	1.49
Authenticity	–2.68*	1.18	–0.67	1.25	2.01	1.73
Analytical Thinking	4.17***	1.08	2.40*	1.10	–1.77	1.54
Word Count	11.02***	2.40	4.82*	2.41	–6.20	3.40
Linguistic Dimensionality	–1.36**	0.52	–2.16***	0.56	–0.80	0.77

Note: The first four columns report the within-dyad change in mean LIWC score from the low-visibility pre-period to the high-visibility post-period for male and female givers separately, estimated at the evaluation level with dyad fixed effects and standard errors clustered at the dyad level ($N = 7,195$ evaluations within 931 dyads). The difference-in-differences column reports the Post $\times$ Female Giver coefficient; women are the reference category, so the estimate equals women's change minus men's, and a negative value indicates that men's feedback changed more positively than women's. For every measure except Analytical Thinking, a higher score indicates more engaged, more recipient-oriented feedback; for Analytical Thinking, a higher score indicates cooler, more impersonal language. Linguistic Dimensionality decomposes into twenty subcomponents, reported in Appendix Table A4. * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

Table 4. Dyad Configuration by Hierarchical Relationship and Gender Pairing

Dyad Configuration	F → F	F → M	M → F	M → M	Total
Peers (same level)	109 (22.8%)	116 (24.2%)	90 (18.8%)	175 (35.1%)	479 (51.5%)
IC → upward	15 (27.3%)	18 (32.7%)	5 (9.1%)	17 (30.9%)	55 (5.9%)
Manager → upward	30 (37.0%)	19 (23.5%)	9 (11.1%)	23 (28.4%)	81 (8.7%)
Manager → downward	19 (29.2%)	11 (16.9%)	15 (23.1%)	20 (30.8%)	65 (7.0%)
Sr Manager → downward	43 (27.2%)	25 (15.8%)	45 (28.5%)	45 (28.5%)	158 (17.0%)
Unclassified (missing rank)	16 (17.2%)	22 (23.7%)	22 (23.7%)	33 (35.5%)	93 (10.0%)
Total	232 (24.9%)	211 (22.7%)	186 (20.0%)	302 (32.4%)	931 (100%)

Note: Cell entries are dyad counts with row percentages in parentheses, showing the gender composition of giver-to-recipient pairs within each hierarchical configuration. Columns run from giver gender to recipient gender; for example, M → F denotes a male giver and a female recipient. The Total column and Total row report counts with each category’s share of all 931 dyads. Upward configurations are those in which the giver holds a lower rank than the recipient and downward configurations the reverse; peers share the same rank. Fifty-eight dyads (6.2%) lacked a clean rank assignment for the giver or recipient and are shown as unclassified. Same-gender dyads (F → F and M → M) account for 534 of the 931 (57.4%).

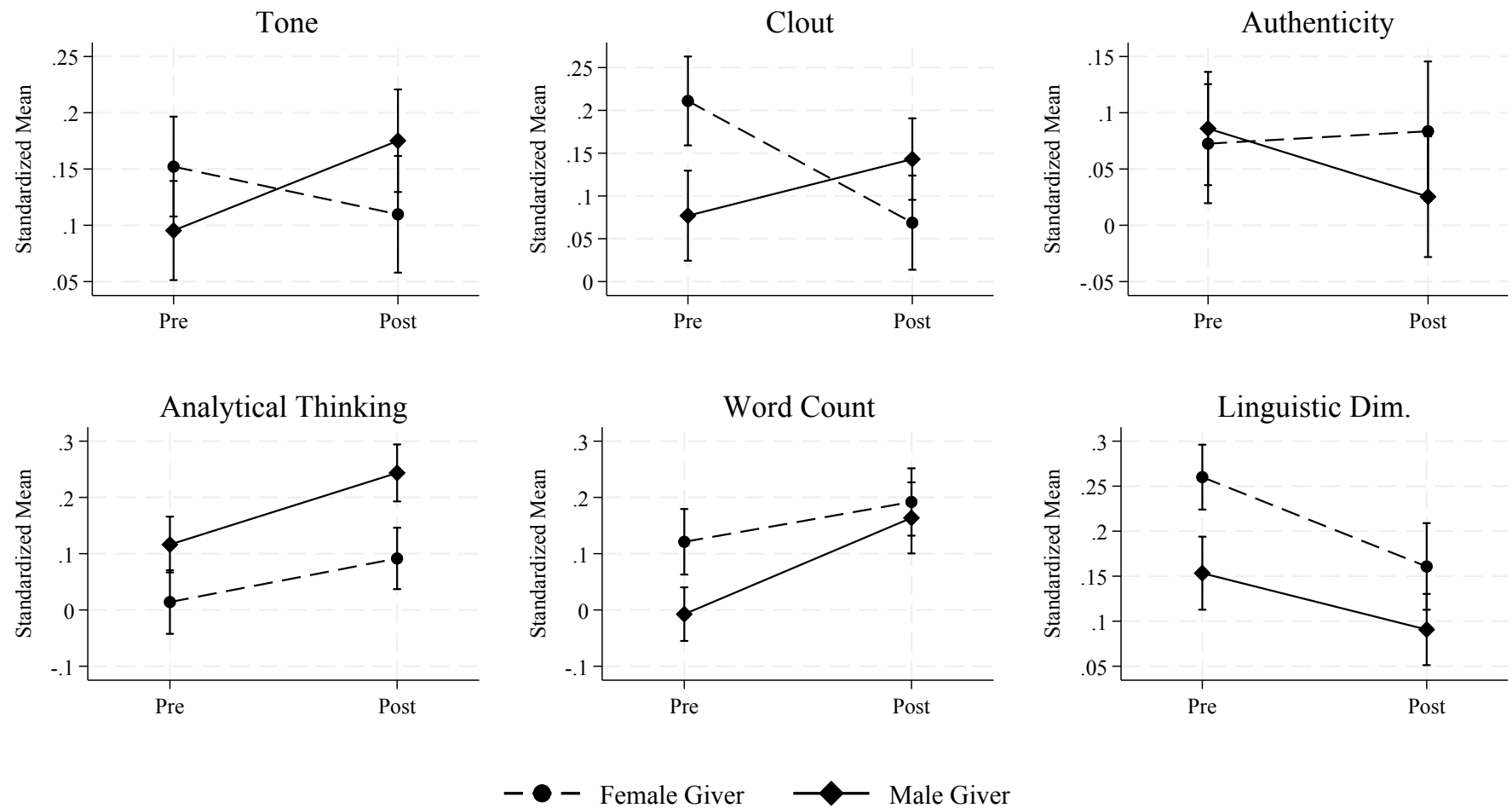
Table 5. Difference-in-Differences Estimates of Feedback Engagement by Hierarchical Direction

Measure	All Peers	IC → Upward	Manager → Downward	Manager → Upward	Sr Manager → Downward
Tone	−2.23 (2.24)	0.96 (6.80)	−13.90* (5.97)	−2.47 (5.22)	−9.63* (3.88)
Clout	−4.79* (2.01)	−6.76 (6.62)	−13.79* (5.61)	−5.89 (5.19)	−3.32 (3.88)
Authenticity	3.21 (2.40)	−7.30 (8.18)	6.88 (6.60)	3.22 (5.45)	−2.40 (4.16)
Analytical Thinking	−2.25 (2.14)	1.88 (5.94)	−4.48 (6.32)	−5.08 (4.64)	−1.02 (3.97)
Word Count	−0.61 (4.47)	−0.93 (8.16)	−11.18 (8.88)	8.84 (16.19)	−15.72 (8.31)
Linguistic Dimensionality	0.23 (1.09)	−0.51 (2.57)	−2.28 (3.36)	0.89 (2.33)	−3.55* (1.63)
<i>Dyad n</i>	479	55	65	81	158

Note: Each cell reports the difference-in-differences estimate (the post-period change among female givers minus male givers) with its standard error below, from a dyad-fixed-effects regression fit within the indicated configuration and clustered at the dyad level. A negative coefficient indicates that men's feedback changed more positively than women's. The All Peers column covers same-level dyads, which have no hierarchical direction; the remaining columns run from a lower to a higher rank (upward) or the reverse (downward). Givers are grouped as Individual Contributor, Manager, and Senior Manager and above, consistent with Tables 1 and 4. The twenty subcomponents of Linguistic Dimensionality are reported in Appendix Table A5. The Dyad n row gives the number of dyads in each model. * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

FIGURES

Figure 1. Pre- to Post-Period Change in Feedback Engagement Indicators, by Giver Gender



Note: Points are standardized group means; bars are 95% confidence intervals.

Appendixes

Appendix A. Additional Descriptives and Results for the Linguistic Dimensionality Subcomponents

Tables A1 and A2 describe the composition of feedback across the two visibility periods, first for the full company corpus and then for the analytic dyad sample used in our models. Two features are stable across the change. In the analytic sample, the gender and rank composition of recipients barely moves between periods, which follows from how the sample is built: because we hold the giver–recipient dyad fixed across periods, the people receiving feedback, and their rank relative to the giver, are largely the same before and after. The linguistic shifts we document therefore cannot be traced to changes in who was receiving feedback.

One feature does move. In both the full corpus and the analytic sample, the share of feedback requested by others fell sharply while the share requested by the giver rose (see Table A2): among analytic-sample givers it dropped from 89% to 54% for men and from 88% to 55% for women. The shift was close to identical for the two groups, a drop of 35 points for men and 33 for women. It is unlikely to account for the gender difference we report, for two reasons. Controlling for solicitation type, and allowing it to shift with the visibility change, leaves the gender interaction essentially unchanged (for Tone and Clout, -3.9 and -6.3 , against -3.8 and -6.4 in our main models). And when we allow the gender effect itself to differ between self-requested and other feedback, the three-way interaction is not significant for any focal measure, so men's larger increase is not concentrated in the feedback givers chose to solicit. The solicitation shift may contribute to the parallel, non-gendered changes we report, such as the rise in feedback length for both groups, but not to the divergence between men and women that is our focus.

Table A3 lists the components of the Linguistic Dimensionality measure together with Word Count, with a brief description and examples of each. Linguistic Dimensionality aggregates 20 LIWC-22 categories that capture the grammatical and interpersonal texture of written feedback, the function words, pronouns, determiners, conjunctions, and verbs that distinguish elaborated, conversational prose from terse, content-only notes. We report Word Count alongside them because it is the other effort measure in our framework and shares this concern with the form of feedback rather than its evaluative substance.

Table A4 reports the difference-in-differences estimates for Linguistic Dimensionality and each of its twenty subcomponents, decomposing the composite result from our main analysis. The composite shows no significant gender difference in change ($DiD = -0.80$, n.s.), and we report the subcomponents here for completeness rather than as separate findings. A few components reach significance on their own, but the composite is the measure our framework relies on, and it does not move by gender.

Table A5 breaks the subcomponent estimates out by hierarchical direction, matching the configurations in Table 5. As with the composite, no clear gendered pattern emerges at the component level within any single configuration and the composite itself is significant only in

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Senior-Manager-downward feedback ($DiD = -3.55, p < .05$). We include the full breakdown so that the configuration-level composite results in Table 5 can be traced to their parts.

Figure A1 plots the pre-to-post trajectories for the Linguistic Dimensionality subcomponents by giver gender, in standardized units for comparability across components. The figure is included alongside Table A4 to show the direction of movement across components at a glance. Consistent with the composite null, the component trajectories do not separate cleanly by gender.

Table A1. Composition of Feedback by Giver Gender and Visibility Period, Whole-Company Sample

	Male Givers		Female Givers	
	Low-Vis.	High-Vis.	Low-Vis.	High-Vis.
<i>Evaluations (N)</i>	17,687	15,347	15,670	15,123
<i>Recipient gender</i>				
% to men	62	55	52	42
% to women	38	45	48	58
<i>Recipient rank</i>				
% to peers	64	55	62	56
% to superiors	12	13	16	17
% to subordinates	25	31	22	26
<i>Feedback request type</i>				
% requested by others	78	53	77	48
% requested by self	10	45	10	49
% anytime	0	2	0	3
% missing	13	0	13	0

Note: Cells are column percentages. Each panel uses its own base: recipient gender among evaluations with a known recipient gender, recipient rank among evaluations with known giver and recipient rank, and request type across all evaluations, with missing shown as a category. Columns may not sum to 100 due to rounding. Low-visibility is August–December 2017 and higher-visibility is August–December 2018. The sample is the full company corpus, before the restriction to dyads present in both periods.

Table A2. Composition of Feedback by Giver Gender and Visibility Period, Analytic Dyad Sample

	Male Givers		Female Givers	
	Low-Vis.	High-Vis.	Low-Vis.	High-Vis.
<i>Evaluations (N)</i>	2,054	1,738	1,816	1,587
<i>Recipient gender</i>				
% to men	64	64	49	49
% to women	36	36	51	51
<i>Recipient rank</i>				
% to peers	59	60	56	57
% to superiors	13	13	19	20
% to subordinates	28	27	25	23
<i>Feedback request type</i>				
% requested by others	89	54	88	55
% requested by self	11	44	12	43
% anytime	0	2	0	2
% missing	0	0	0	0

Note: Cells are column percentages, defined as in Table A1. The sample is the analytic dyad sample used in the paper, restricted to evaluations exchanged within giver–recipient dyads present in both periods. Because the dyads are fixed across periods, recipient gender and recipient rank are nearly constant between the two columns within each giver gender; the request-type panel, by contrast, shifts across the visibility change for both men and women givers. Columns may not sum to 100 due to rounding.

Table A3. LIWC's Linguistic Dimensionality Measure Sub-Construct Composition

LIWC variables	Examples
WC	Word count
function	Total function words (the, to, and, I)
pronoun	Total pronouns (I, you, that, it)
ppron	Personal pronouns (I, you, me, my)
i	1st person singular (I, me, my, myself)
we	1st person plural (we, our, us, lets)
you	2nd person (you, your, u, yourself)
shehe	3rd person singular (he, she, her, his)
they	3rd person plural (they, their, them, themsel*)
ipron	Impersonal pronouns (that, it, this, what)
det	Determiners (the, at, that, my)
article	Articles (a, an, the, a lot)
number	Numbers (one, two, three)
prep	Prepositions (to, of, in, for)
auxverb	Auxiliary verbs (is, was, be, have)
adverb	Adverbs (so, just, about, there)
conj	Conjunctions (and, but, so, as)
negate	Negations (not, no, never, nothing)
verb	Common verbs (is, was, be, have)
adj	Common adjectives (more, very, other, new)
quantity	Quantities (all, one, more, some)

Table A4. Change in Linguistic Dimensionality and Its Subcomponents, by Giver Gender

Engagement Measure	Men's Change	SE	Women's Change	SE	DiD (W – M)	SE
<i>Linguistic Dimensionality (composite)</i>						
Linguistic Dimensionality	–1.36**	0.52	–2.16***	0.56	–0.80	0.77
<i>Subcomponents</i>						
function	–0.83	0.47	–2.46***	0.50	–1.63*	0.69
pronoun	–0.59**	0.22	–1.24***	0.25	–0.64	0.33
ppron	–0.33	0.18	–0.83***	0.20	–0.49	0.27
i	–0.09	0.11	0.03	0.13	0.12	0.17
we	0.09	0.06	–0.21**	0.07	–0.30**	0.09
you	–0.03	0.06	0.02	0.08	0.05	0.10
shehe	–0.27*	0.13	–0.60***	0.13	–0.34	0.18
they	–0.07*	0.03	–0.08**	0.03	–0.02	0.04
ipron	–0.26*	0.12	–0.41**	0.13	–0.15	0.17
det	0.27	0.21	–0.55*	0.23	–0.83**	0.32
article	0.10	0.15	–0.37*	0.16	–0.47*	0.22
number	0.25*	0.10	–0.08	0.12	–0.33*	0.16
prep	0.15	0.26	0.29	0.33	0.14	0.42
auxverb	–0.15	0.17	–0.57**	0.19	–0.42	0.26
adverb	–0.46**	0.17	–0.66***	0.17	–0.20	0.24
conj	–0.16	0.14	–0.27	0.16	–0.11	0.22
negate	–0.08	0.12	0.18	0.11	0.25	0.16
verb	–0.87**	0.27	–0.34	0.32	0.53	0.41
adj	0.28	0.18	0.13	0.19	–0.15	0.27
quantity	0.25	0.16	0.14	0.16	–0.10	0.22

Note: Each row reports the within-dyad change in mean LIWC score from the low-visibility pre-period to the high-visibility post-period for male and female givers, with the difference-in-differences (Post × Female Giver, women the reference category) in the final columns; a negative value indicates that men’s feedback changed more positively than women’s. Estimates are at the evaluation level with dyad fixed effects and dyad-clustered standard errors (N = 7,195 evaluations within 931 dyads). The Linguistic Dimensionality composite is the measure of interest and shows no significant gender difference in change; its subcomponents are reported here for completeness. * p<.05, ** p<.01, *** p<.001 (two-tailed).

Table A5. Difference-in-Differences Estimates for Linguistic Dimensionality Subcomponents, by Hierarchical Direction

Measure	All Peers	IC → Upward	Manager → Downward	Manager → Upward	Sr Manager → Downward
<i>Linguistic Dimensionality (composite)</i>					
Linguistic Dimensionality	0.23 (1.09)	-0.51 (2.57)	-2.28 (3.36)	0.89 (2.33)	-3.55* (1.63)
<i>Subcomponents</i>					
function	-1.16 (1.02)	0.38 (2.34)	-4.74 (2.58)	-0.45 (2.02)	-2.56 (1.39)
pronoun	-0.33 (0.48)	0.08 (1.18)	-1.41 (1.20)	-0.10 (1.11)	-1.25 (0.84)
ppron	-0.43 (0.37)	0.88 (1.11)	-1.72 (0.90)	-0.06 (1.02)	-0.54 (0.74)
i	0.10 (0.23)	0.84 (0.81)	-0.26 (0.55)	0.29 (0.60)	0.00 (0.41)
we	-0.21 (0.12)	-0.17 (0.49)	-0.61 (0.43)	-0.51 (0.34)	-0.22 (0.22)
you	0.12 (0.13)	0.07 (0.46)	-0.26 (0.56)	0.27 (0.33)	-0.01 (0.23)
shehe	-0.43 (0.26)	0.23 (0.57)	-0.44 (0.67)	-0.49 (0.68)	-0.14 (0.50)
they	0.00 (0.05)	-0.04 (0.14)	-0.04 (0.12)	0.36* (0.15)	-0.18 (0.12)
ipron	0.10 (0.25)	-0.80 (0.62)	0.31 (0.58)	-0.04 (0.49)	-0.71 (0.46)
det	-1.07* (0.42)	2.42* (1.12)	-1.15 (1.23)	-1.29 (1.33)	-0.96 (0.82)
article	-0.76* (0.30)	1.69* (0.81)	-0.81 (1.00)	-0.84 (0.89)	-0.43 (0.54)
number	-0.54* (0.24)	0.28 (0.27)	-0.05 (0.26)	-0.10 (0.81)	-0.46* (0.23)
prep	0.14 (0.54)	-0.46 (2.11)	-0.86 (2.11)	0.45 (1.21)	0.43 (1.14)
auxverb	-0.31 (0.38)	0.17 (1.04)	-1.09 (0.92)	-0.18 (0.66)	-0.23 (0.63)
adverb	-0.18 (0.32)	-1.29 (1.02)	0.09 (1.46)	0.22 (0.86)	-0.45 (0.51)
conj	0.39 (0.31)	-0.94 (0.88)	-1.64* (0.80)	-0.05 (0.76)	-0.33 (0.47)
negate	0.28 (0.21)	0.21 (0.53)	0.45 (0.69)	0.45* (0.20)	0.28 (0.34)
verb	1.16* (0.55)	0.44 (2.27)	1.40 (2.14)	0.78 (1.09)	-0.47 (0.96)
adj	-0.11 (0.37)	-0.57 (1.26)	-0.39 (1.00)	0.49 (0.61)	-1.06 (0.63)
quantity	-0.04 (0.29)	-0.77 (1.62)	0.31 (0.81)	-0.15 (0.57)	-0.94* (0.40)
<i>Dyad n</i>	<i>479</i>	<i>55</i>	<i>65</i>	<i>81</i>	<i>158</i>

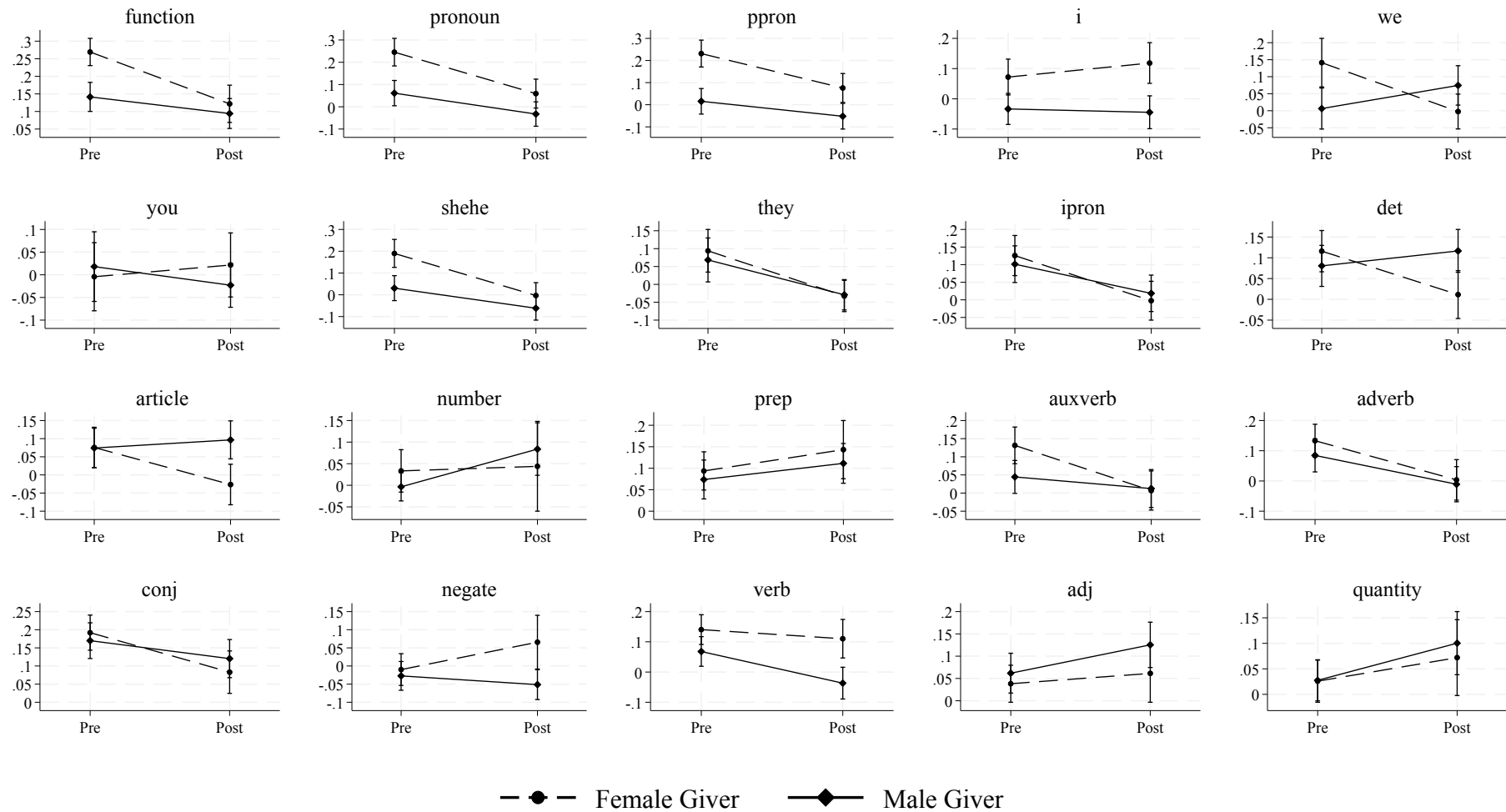
Note: Each cell reports the difference-in-differences estimate (the post-period change among female givers minus male givers) with its standard error below, from a dyad-fixed-effects regression fit within the indicated configuration and clustered at the dyad level. A negative coefficient indicates that men's feedback changed more positively than women's. These are the twenty subcomponents of the Linguistic Dimensionality composite, whose configuration-level estimates appear in the top row and in Table 5. The Dyad n row gives the number of dyads in each model. * p<.05, ** p<.01, *** p<.001 (two-tailed).

Table A6. Feedback Giver’s Position Relative to the Recipient’s Manager

	2017, n (%)	2018, n (%)
<i>Panel A. Position of feedback giver, by year</i>		
Below recipient’s manager	406 (48.4)	401 (49.9)
Same broad tier as recipient’s manager	414 (49.4)	392 (48.8)
Above recipient’s manager	18 (2.1)	11 (1.4)
Dyads with usable comparison	838	804
Missing comparison	93	127
<i>Panel B. Stability across years (dyads with usable comparison in both years, n = 789)</i>		
Below recipient’s manager in both years	380 (48.2)	
Same broad tier in both years	370 (46.9)	
Above recipient’s manager in both years	11 (1.4)	
Changed between below and same tier	22 (2.8)	
Changed involving an above position	6 (0.8)	

Note: The giver’s position is measured by comparing the giver’s compensation bracket with that of the recipient’s manager, using the same three-tier rank classification employed in the main analyses (Individual Contributor, Manager, Senior Manager and above); “same broad tier” therefore indicates membership in the same tier, not an identical compensation grade. Each dyad contributes once per year, and classifications were internally consistent within every dyad-year. Percentages are computed over dyads with a usable comparison in the given year. Hierarchy information is available in at least one year for 853 of the 931 dyads (91.6%). In Panel B, 761 of 789 dyads (96.5%) retain the same relative position across the two years. Because this positioning is effectively time-invariant, it is absorbed by the dyad fixed effects in all models.

Figure A1. Sub-Component Variables to the Linguistic Dimensionality Construct Pre (Low-Visibility) vs. Post (High-Visibility) Differences between Female and Male Feedback Givers



Appendix B. Feedback Language and the Giver’s Subsequent Promotion

In this appendix, we ask whether the language a person used when giving feedback predicted that person’s promotion in the following year. We build a giver-level file with one record per giver per period, averaging each LIWC construct across the comments a giver wrote, and define promotion as an increase in compensation grade from the feedback year to the next. We estimate the relationship in both windows the data allow: 2017 feedback predicting promotion by 2018 (N = 662) and 2018 feedback predicting promotion by 2019 (N = 556). In each, we regress promotion on the five engagement constructs, standardized so that each coefficient reflects a one standard deviation change, adjusting for giver gender, rank, compensation grade in the feedback year, and the logged number of recipients, with a third model adding Linguistic Dimensionality.

Table B1 reports the first window (2017 feedback, promotion by 2018) as odds ratios. The engagement constructs carry no joint association with promotion (likelihood-ratio $\chi^2(5) = 3.57$, $p = .61$), and adding them worsens fit on both information criteria (AIC rises from 419.9 to 426.3; BIC from 446.9 to 475.8). No construct is significant individually, and none survives a Benjamini–Hochberg adjustment for the five tests. The magnitudes are small: a one standard deviation change in a construct moves the predicted probability of promotion by less than two percentage points in either direction (Tone -0.7 , Clout -0.9 , Authenticity -0.8 , Analytical Thinking $+1.7$, Word Count $+1.0$ points). Because a non-significant coefficient is not on its own evidence of no effect, we test each construct against a band of negligibility; treating a shift of more than five percentage points per standard deviation as the threshold for a meaningful effect, all five fall inside the band, with every 90% confidence interval interior to ± 5 points. The two controls that move are the expected ones: women are somewhat more likely to be promoted in this window net of the other terms, and givers already at higher grades are less likely to advance further.

Table B2 reports the second window (2018 feedback, promotion by 2019) and repeats the pattern. The engagement constructs are again jointly unrelated to promotion ($\chi^2(5) = 7.96$, $p = .16$), no construct is significant individually or survives the Benjamini–Hochberg adjustment, and the estimated effects remain small, the largest a 3.2 percentage point change for Analytical Thinking.

1
2
3 The verdict is stable across specifications in both windows. Linear probability models return the
4 same joint null (2017–2018: $F(5, 651) = 0.74$, $p = .59$; 2018–2019: $F(5, 545) = 1.75$, $p = .12$), as
5 do separate estimates for men and women (2017–2018 men: $\chi^2(5) = 3.47$, $p = .63$; women: $\chi^2(5)$
6 $= 4.23$, $p = .52$; 2018–2019 men: $\chi^2(5) = 2.86$, $p = .72$; women: $\chi^2(5) = 7.97$, $p = .16$). Pooling
7 both periods and aligning each to promotion in the following year, which gives 1,218 giver-years
8 with standard errors clustered on the giver, likewise leaves the engagement block jointly non-
9 significant ($\chi^2(5) = 8.45$, $p = .13$).
10
11

12
13
14 Taken together, across both time windows the language of feedback bears no meaningful
15 relationship to the giver's own advancement. Feedback that is more engaged, warmer in tone, or
16 more carefully reasoned does not raise the giver's promotion odds, which is what it means,
17 concretely, for this to be low-promotability work.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table B1. 2017 feedback language and the giver’s promotion by 2018 (odds ratios)

	(1)	(2)	(3)
	Controls	+ Engagement	+ Linguistic
Female giver	1.706* [1.006, 2.893]	1.820* [1.061, 3.120]	1.856* [1.081, 3.188]
Manager	1.620 [0.520, 5.045]	1.549 [0.491, 4.888]	1.452 [0.456, 4.625]
Senior manager and above	3.339 [0.560, 19.927]	3.076 [0.505, 18.726]	2.941 [0.479, 18.061]
Compensation grade	0.738* [0.553, 0.985]	0.730* [0.544, 0.981]	0.731* [0.543, 0.983]
Recipients (logged)	0.967 [0.592, 1.580]	1.052 [0.613, 1.805]	1.106 [0.643, 1.901]
Tone		0.924 [0.694, 1.230]	1.016 [0.750, 1.376]
Clout		0.902 [0.663, 1.228]	0.958 [0.698, 1.314]
Authenticity		0.912 [0.672, 1.238]	1.005 [0.731, 1.381]
Analytical Thinking		1.217 [0.928, 1.595]	1.118 [0.841, 1.487]
Word Count		1.121 [0.827, 1.519]	1.168 [0.864, 1.580]
Linguistic Dimensionality			0.749 [0.558, 1.005]
N	662	662	662
Pseudo-R ²	0.030	0.039	0.047
Log-likelihood	-203.9	-202.2	-200.3
AIC	419.9	426.3	424.7
BIC	446.9	475.8	478.6
LR χ^2 (engagement block)		3.57 (p = .61)	

Note. Logit, giver level; 2017 feedback predicting promotion by 2018. Odds ratios with 95% confidence intervals in brackets. Language predictors standardized (per 1 SD). Individual contributor is the omitted rank category. The likelihood-ratio test compares the engagement block against the controls-only model. * p < .05, ** p < .01, *** p < .001 (two-tailed).

Table B2. 2018 feedback language and the giver's promotion by 2019 (odds ratios)

	(1)	(2)	(3)
	Controls	+ Engagement	+ Linguistic
Female giver	1.116 [0.784, 1.590]	1.164 [0.809, 1.674]	1.148 [0.798, 1.654]
Manager	2.127 [0.975, 4.642]	2.018 [0.921, 4.417]	2.105 [0.958, 4.627]
Senior manager and above	3.289 [0.995, 10.876]	2.882 [0.867, 9.578]	2.962 [0.889, 9.869]
Compensation grade	0.676*** [0.552, 0.828]	0.689*** [0.562, 0.845]	0.689*** [0.562, 0.845]
Recipients (logged)	0.773 [0.544, 1.098]	0.631* [0.422, 0.944]	0.632* [0.422, 0.946]
Tone		0.875 [0.726, 1.055]	0.865 [0.717, 1.045]
Clout		0.969 [0.783, 1.199]	0.968 [0.782, 1.198]
Authenticity		1.074 [0.881, 1.308]	1.038 [0.846, 1.273]
Analytical Thinking		1.157 [0.957, 1.398]	1.209 [0.989, 1.478]
Word Count		0.888 [0.736, 1.072]	0.875 [0.724, 1.058]
Linguistic Dimensionality			1.150 [0.937, 1.411]
N	556	556	556
Pseudo-R ²	0.058	0.069	0.071
Log-likelihood	-352.3	-348.3	-347.4
AIC	716.5	718.5	718.7
BIC	742.4	766.1	770.6
LR χ^2 (engagement block)		7.96 (p = .16)	

Note. Logit, giver level; 2018 feedback predicting promotion by 2019. Odds ratios with 95% confidence intervals in brackets. Language predictors standardized (per 1 SD). Individual contributor is the omitted rank category. The likelihood-ratio test compares the engagement block against the controls-only model. * p < .05, ** p < .01, *** p < .001 (two-tailed).

Appendix C. Sensitivity of Visibility Effects to Recipient Gender

Our main analyses document that male managers delivering downward feedback exhibited the largest visibility-induced improvements in feedback quality. Because recipients of downward feedback are not guaranteed to be evenly divided by gender, a plausible alternative is that the observed effects reflect recipient composition rather than rank-based accountability. We test this directly.

We augment our dyad fixed effects models with an interaction between the post-visibility indicator and a female recipient indicator ($\text{Post} \times \text{Female Recipient}$). This term captures whether the visibility effect on feedback quality differs depending on whether the recipient is a woman or a man. If recipient gender were driving our results, we would expect this interaction to be statistically significant, indicating that givers changed their feedback differentially based on who was receiving it.

Table C1 reports these results for our five focal outcomes across all five rank configurations. The $\text{Post} \times \text{Female Recipient}$ coefficient is substantively small and statistically nonsignificant for Tone, Word Count, and Linguistic Dimensionality across every rank configuration (all $p > .17$). The same pattern holds for Clout and Authenticity. Critically, within both downward feedback contexts where we observe the strongest visibility effects (Manager-to-IC and Sr. Manager-downward), the interaction terms are uniformly nonsignificant, confirming that male givers improved their feedback quality comparably regardless of whether their recipient was a man or a woman.

These results support our interpretation that visibility-induced changes in feedback quality among male managers reflect the accountability pressures associated with hierarchical rank rather than differential treatment by recipient gender. The consistency of this null finding across rank configurations reinforces the conclusion that the gendered patterns we document are driven by giver-side behavioral change in response to increased observability, not by the gender composition of recipients within each feedback direction.

Table C1. Sensitivity of Visibility Effects to Recipient Gender, by Hierarchical Direction

Measure	All Peers	IC → Upward	Manager → Downward	Manager → Upward	Sr Manager → Downward
<i>Panel A. Visibility effect (Post)</i>					
Tone	1.69 (1.40)	0.30 (4.63)	0.99 (4.66)	0.56 (3.29)	2.95 (3.01)
Clout	-0.47 (1.28)	-3.53 (4.02)	4.34 (3.87)	-1.37 (3.80)	-2.42 (2.80)
Authenticity	-1.56 (1.59)	-0.74 (5.45)	-6.31 (4.08)	-0.61 (3.57)	4.25 (3.16)
Analytical Thinking	4.50*** (1.33)	2.06 (3.89)	6.36 (4.88)	2.74 (3.32)	3.49 (2.89)
Word Count	6.83* (2.91)	0.63 (5.21)	1.95 (6.24)	12.05 (7.06)	13.18* (6.07)
Linguistic Dimensionality	-1.24 (0.69)	-1.40 (1.88)	-3.45 (2.40)	-3.47* (1.57)	-2.06 (1.09)
<i>Panel B. Post × Female Recipient</i>					
Tone	-1.63 (2.30)	2.22 (6.50)	-4.14 (6.25)	0.62 (5.11)	0.16 (3.93)
Clout	1.91 (2.03)	7.11 (7.08)	-4.71 (5.77)	-3.86 (5.08)	3.73 (3.91)
Authenticity	-1.65 (2.40)	-2.76 (7.94)	-0.05 (6.34)	-1.41 (5.15)	-4.32 (4.18)
Analytical Thinking	-3.63 (2.22)	4.59 (5.83)	-2.57 (6.34)	0.60 (4.58)	1.47 (3.95)
Word Count	-2.31 (4.43)	-7.46 (8.61)	12.15 (8.88)	1.53 (17.39)	-1.63 (8.57)
Linguistic Dimensionality	-1.70 (1.08)	1.38 (2.21)	2.19 (3.26)	3.59 (2.29)	1.05 (1.65)
<i>Dyad n / evaluations</i>	<i>479 / 3,753</i>	<i>55 / 449</i>	<i>65 / 501</i>	<i>81 / 608</i>	<i>158 / 1,169</i>

Note: Each column reports a separate dyad-fixed-effects regression fit within the indicated rank configuration, with standard errors clustered at the dyad level in parentheses. Panel A gives the average within-dyad change from the low-visibility to the high-visibility period (Post); Panel B gives the Post × Female Recipient interaction, which tests whether that change differs when the recipient is a woman. Outcomes are in raw LIWC units. None of the recipient-gender interactions in Panel B reaches significance in any configuration. The bottom row gives the number of dyads and the number of evaluations in each model. All models include dyad fixed effects. * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix D. Additional Analyses using Correll et al. 2020’s Viewing and Valuing Codes for Performance Evaluations

LIWC-22 reliably captures differences in linguistic engagement and gendered patterns of language use, but it is not designed to summarize changes in the substantive nature of feedback. To assess whether the visibility change altered what givers wrote about, as distinct from how much they engaged with the task, we needed a content-oriented method. Modern text analysis offers several: topic modeling, supervised learning, and generative large language models (LLMs), each with distinct advantages and limitations.

Each alternative posed obstacles for structured, theory-driven comparison. Topic modeling can surface broad themes in feedback, such as “communication skills” or “technical competence,” but as an unsupervised method it offers no benchmark for whether the discovered topics are “correct,” requires post hoc interpretation that introduces subjectivity, and yields results that are difficult to validate or compare across samples and periods. Supervised machine learning is better suited to classifying substantive differences in text (see Nelson et al. 2023), but it demands substantial resources, including qualitative coding by a team of domain experts, which the terms of our data use agreement did not permit. Free-form LLM theme generation carries a related set of tradeoffs: fluent and context-aware output, but opaque inner workings, results that are not always reproducible, and interpretations that can drift from the conceptual boundaries our research design requires and that are difficult to audit against an established standard.

We therefore drew on the coding schema developed by Correll et al. (2020), whose 88 codes were built for year-end performance evaluations at a technology company. Adopting an established codebook rather than constructing one from scratch gives us descriptive variables suited to testing for substantive change in feedback and, at the same time, direct comparability to

the setting Correll and colleagues studied. We applied the codebook to the full corpus of dyad feedback ($n = 931$ dyads) entirely within our university's Microsoft Copilot environment, so that all protected data remained inside our institution's secure infrastructure (Exhibit D1 reproduces the prompt). This approach scales the application of codes across a large corpus while preserving the consistency, transparency, and theoretical alignment of the coding scheme, and it offers a more interpretable and replicable basis for comparing language across groups and periods than unsupervised or free-form LLM methods. Table D1 reports, for each period, mean code usage by the gender of the feedback giver for the subset of codes showing a statistically significant gender difference in at least one period; Figure D1 plots the corresponding standardized within-dyad means. Codes are labeled throughout with the terminology of the Correll et al. (2020) codebook.

Two features of the design shape how these results should be read. First, Correll et al. (2020) analyzed how evaluative language varied with the gender of the employee being evaluated, that is, the recipient of the assessment; usage by the gender of the evaluation's author was not measured in their study. The analysis here inverts that lens and compares code usage by the gender of the feedback giver, so the patterns in Table D1 are not a test of their findings and should not be expected to reproduce them. Second, recipient gender is a fixed attribute of each dyad, and the dyad pairs themselves are stable across the two periods, mirroring the dyad fixed effects in the main analysis. Sample composition therefore cannot generate change between the periods; any movement in the table reflects change in what givers write, not change in who is writing or being written about.

Read this way, the table serves as a test of whether the visibility change altered the substance of feedback, and it shows considerable stability. Gendered patterns of code usage present in the lower-visibility pre-period largely persist in the higher-visibility period, and only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

three codes, TECHNICAL EXPERTISE, VISIONARY, and COMMUNAL, show a statistically significant gender difference in both periods. That the content of feedback stays largely fixed while the engagement measures in the main analysis move is the pattern the paper’s argument predicts: the shift we document operates through how givers engage with the feedback task, not through a change in what the feedback is about. The analysis thus helps rule out an alternative reading of the main result, in which the visibility change altered the topical composition of feedback rather than the level of engagement with it.

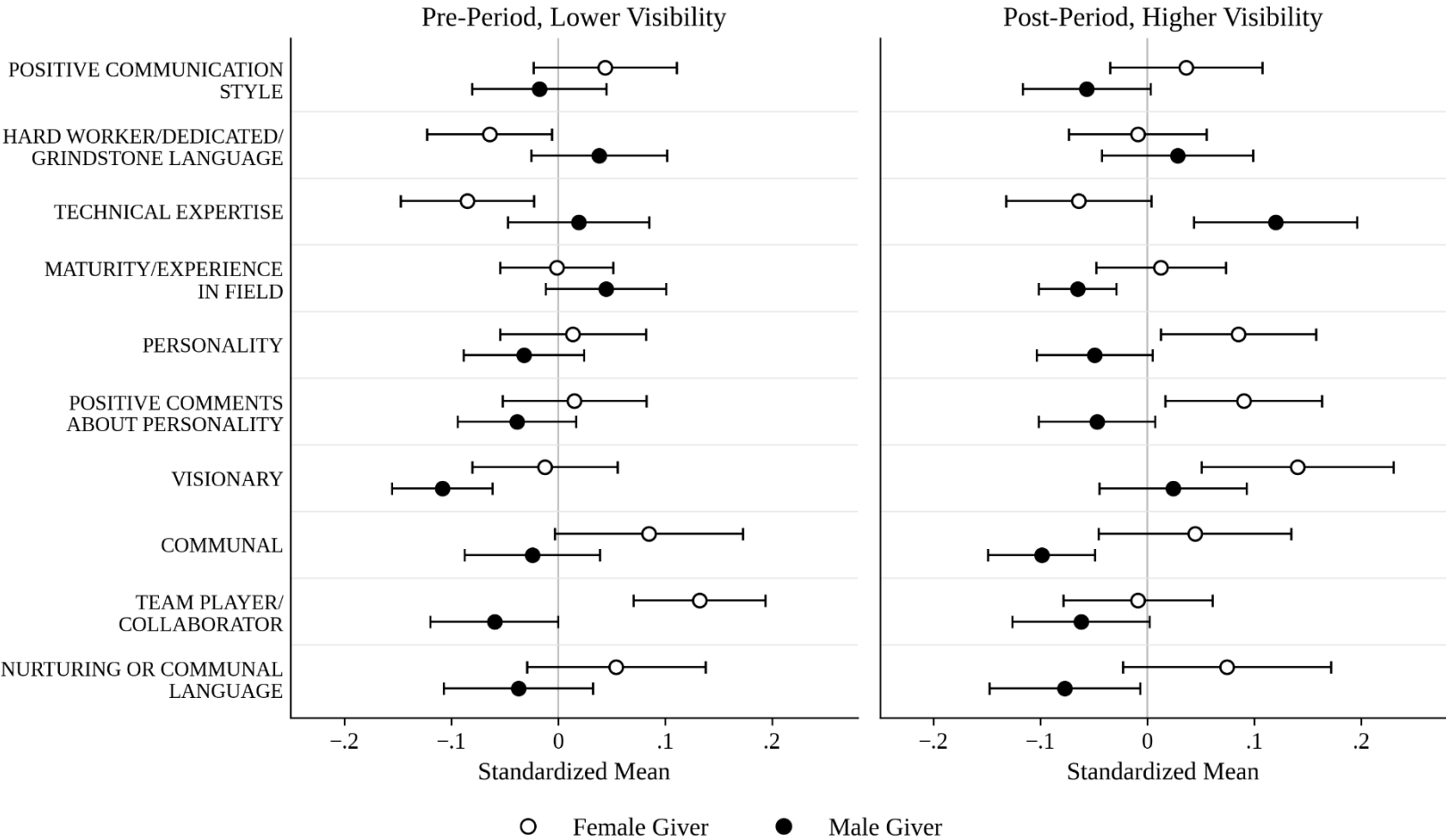
Appendix Table D1. Gender Differences in Usage of Significant Correll et al. (2020) Codes within Dyad Feedback (n = 931 Dyads) across the Visibility Change.

	Pre-Period, Lower Visibility							Post-Period, Higher Visibility						
Correll et al. (2020) Code	Mean Female	Mean Male	Difference	t	Std. Err	P> t	Sig	Mean Female	Mean Male	Difference	t	Std. Err	P> t	Sig
POSITIVE COMMUNICATION STYLE	0.126	0.107	0.019	1.33	0.015	0.182		0.124	0.094	0.029	1.99	0.015	0.047	*
HARD WORKER/DEDICAT ED/GRINDSTONE	0.134	0.171	−0.037	−2.36	0.016	0.018	*	0.154	0.168	−0.014	−0.78	0.018	0.437	
LANGUAGE TECHNICAL EXPERTISE	0.141	0.181	−0.040	−2.29	0.017	0.022	*	0.149	0.219	−0.070	−3.58	0.019	0.000	***
MATURITY/EXPERI ENCE IN FIELD	0.019	0.026	−0.006	−1.20	0.005	0.232		0.021	0.011	0.011	2.20	0.005	0.028	*
PERSONALITY	0.122	0.107	0.015	1.04	0.014	0.301		0.144	0.101	0.043	2.95	0.015	0.003	**
POSITIVE COMMENTS ABOUT PERSONALITY	0.119	0.102	0.017	1.23	0.014	0.219		0.143	0.099	0.044	2.99	0.015	0.003	**
VISIONARY	0.071	0.046	0.025	2.31	0.011	0.021	*	0.111	0.081	0.031	2.04	0.015	0.041	*
COMMUNAL TEAM PLAYER/COLLABO RATOR	0.088	0.061	0.027	2.00	0.014	0.046	*	0.078	0.042	0.036	2.77	0.013	0.006	**
NURTURING OR COMMUNAL LANGUAGE	0.548	0.453	0.096	4.44	0.022	0.000	***	0.478	0.452	0.027	1.12	0.024	0.264	
	0.078	0.055	0.022	1.66	0.013	0.097		0.083	0.045	0.037	2.52	0.015	0.012	*

Note: Codes are labeled with Correll et al. (2020) codebook terminology. Difference is the female-minus-male mean. *p<.05, **p<.01, ***p<.001 (two-tailed tests).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Figure D1. Frequency of Correll et al. (2020) Code Usage by Female and Male Feedback Givers, for the Subset of Codes with a Statistically Significant Gender Difference ($p < .05$) in Either the Lower-Visibility Pre-Period or the Higher-Visibility Post-Period.



Note: Standardized within-dyad means with 95% confidence intervals; standard errors clustered at the dyad level.

Exhibit D1. Prompt given to Private Microsoft Enterprise CoPilot for coding completely anonymized data set of feedback given among matched giver-recipient dyads across the shift in visibility

Correll et al. 2020 Coding Instructions

For all conversations in this chat, you will act as an expert qualitative coder familiar with the Viewing and Valuing Social Cognitive Processing Model developed by Correll et al. (2020).

I will share performance evaluation text for coding. Your task is to identify ALL applicable codes from the codebook of 88 codes developed for analyzing gendered language in workplace evaluations. These codes were carefully developed to capture subtle patterns in how managers view and value employee behaviors.

Coding Process:

1. Analyze each comment thoroughly for ALL applicable codes
2. Consider the full context and nuance of the statement
3. Report codes separated by semicolons using EXACT terminology from the codebook
4. Format your response as: "Code1 ; Code2 ; Code3"
5. Do not add any commentary, explanations, or text beyond the applicable codes
6. Return each coded comment on a single line with no extra spaces between lines for direct Excel import

Important Coding Guidelines:

- Be thorough - apply multiple codes when appropriate (evaluations often contain several coded elements)
- Be precise - use only the exact codes from the 88-code system
- Focus on both explicit statements and implicit messages
- Consider both VIEWING aspects (how behaviors are observed) and VALUING aspects (how behaviors are evaluated)
- Pay special attention to language around:
 - Technical/business skills and contributions
 - Communal/group-oriented behaviors
 - Agentic/leadership behaviors
 - Communication style and personality traits
 - Standout language vs. improvement language

- Future-oriented evaluations vs. past performance

Coding Tips from Research Findings:

- Look for subtle patterns in how performance is described (e.g., “taking charge” vs. “helpful” language)
- Note when communication styles are described as “too aggressive” or “too soft”
- Identify instances of “standout language” (visionary, genius, ideal worker)
- Be attentive to future-oriented feedback about needed improvements
- Consider if the evaluation focuses on personality vs. technical ability

Codebook:

General performance: positive valence ; Praise ; Vague praise ; Hard worker/dedicated/grindstone language ; Promotion ; Mentorship ; Sponsorship ; Experience ; Maturity/experience in field General performance: neutral valence ; Feedback from someone other than reviewer ; Evaluator’s tone ; Reframe/spin ; Shout-out/thanking someone ; Service work ; Networking ; Description of the work/project status ; State of the project ; Status/visibility of project ; General performance: negative valence ; Criticism ; Vague criticism ; Criticism not linked to business outcome ; Doubt-raising language ; Work/family conflict ; Needs to get on tech. project in the future ; Technical and business-related codes ; Technical skills ; Has a technical skill ; Technical projects ; Technical expertise ; Criticism linked to specific business outcome ; Criticism of general business outcome ; Communal and group-oriented language Communal ; Nurturing or communal language ; Team player/collaborator ; Team accomplishments ; Leader in organizational skill ; Agentic and managerial language Agentic language ; Managerial skills ; Leader of/in teams ; Hiring or firing ; Attributes others’ accomplishments to their leadership ; Directly mentions promotion opportunities ; Standout language vs. questioning language Visionary ; Genius/game-changer language ; Ideal worker ; Increased scrutiny ; Improvement ; Future changes ; Non-technical improvement ; Vague improvement; Vague call for future improvement; Specific suggestion for future improvement ; Needs to develop a technical skill ; Technical improvement ; Communication style, personality Communication style ; Has worked on changing communication style ; Needs to change communication style ; Negative communication style ; Positive communication style ; Too aggressive communication style ; Too distant/detached communication style ; Too soft communication style ; Assertive communication style ; Personality ; Positive comments about personality ; Negative comments about personality ; Likability penalty ;

Begin coding when I share evaluation text. I only need the codes, formatted precisely as requested, with one evaluation per line and no extra spaces between lines for seamless Excel import.