

Electronic performance appraisals: The effects of e-mail communication on peer ratings in actual and simulated environments

Terri R. Kurtzberg ^{a,*}, Charles E. Naquin ^b, Liuba Y. Belkin ^c

^a *Rutgers Business School, Rutgers University, 94 Rockafeller Road, Piscataway, NJ 08854, USA*

^b *Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556, USA*

^c *Rutgers Business School, Rutgers University, 111 Washington Street, Newark, NJ 07102, USA*

Received 2 November 2004

Available online 18 August 2005

Abstract

Across three empirical studies, this paper explores the effect of using e-mail as a communication medium (versus pen-and-paper) when conducting performance appraisals of peers. The notion put forth by Media Richness Theory that paper-form and e-mail media should be considered identical for conveying this information was theoretically challenged and differences were empirically supported. Using two different settings, results demonstrate that evaluators offered more negative appraisals of their peers when using e-mail than when using traditional paper-form methods. Reduced feelings of social obligation in the e-mail condition were found to mediate this relationship, indicating that social psychological processes can influence the effects of different media on peer ratings.

© 2005 Elsevier Inc. All rights reserved.

Keywords: E-mail; Technology; Performance appraisal; Social obligation

Effectively evaluating employees and communicating feedback to them is a constant challenge in organizational life. Generally speaking, a performance appraisal is an evaluation of an employee's performance along pertinent dimensions (e.g., results, participation, etc.), and feedback is the communication of the appraisal results to the person being appraised. Performance appraisals are essential because they often influence vital issues such as the focus of individual improvement, pay raises, promotions, and next work assignment. Though clearly necessary for a high performing work system (see Butler, Ferris, & Napier, 1991; Druskat & Wolff, 1999; Ouchi, 1981; Staw, 1980), two recent changes in the workplace have presented even greater obstacles for the performance appraisal

process: (1) the rise of teams and laterally based organizational structures, and (2) communication technology, or more specifically, changes in how employees work and communicate due to the proliferation of electronic communication opportunities. These challenges have been recognized in a recent review of the performance appraisal literature as being important but lacking research attention (see Levy & Williams, 2004).

Along with the widespread shift from hierarchical organizational structures (sometimes referred to as control-based organizations) to more lateral- or team-based structures (sometimes referred to as participatory organizations) (Katzenbach & Smith, 1993; Nadler, Gerstein, & Shaw, 1992; Thompson, 2004) comes a shift in both the content and the structure of the employee evaluation process. Whereas in traditional hierarchical (i.e., control-focused) organizations, performance was often rewarded by such factors as

* Corresponding author. Fax: +1 732 445 6987.

E-mail address: tk@business.rutgers.edu (T.R. Kurtzberg).

seniority or level of education, team- or laterally-based organizations tend to reward according to more dynamic measures such as demonstrated competencies or performance delivered (Druskat & Wolff, 1999; Saavedra & Kwun, 1993). This shift in organizational design has also resulted in a change in *who* does the appraising (Levy & Williams, 2004) in many organizations. Traditionally, only one person might have been involved in performance feedback—the supervisor. In contemporary team- or laterally-based organizations, peer reviews are becoming commonplace to provide the necessary feedback most accurately (Druskat & Wolff, 1999; Shore, Shore, & Thornton, 1992; Yammarino & Waldman, 1993). Some organizations have employees evaluated by as many as 10 of their peers to determine how much they contribute to their team (Thompson, 2004). Following this trend, many graduate business schools now also have mandatory peer-based performance appraisals among their graduate- and executive-level students.

Yet another consequence of this shift concerns not just *who* does the evaluation but *how* it is conducted. Specifically, the proliferation of the Internet has changed the way individuals communicate and exchange information both within and between organizations, especially considering recent increases in globalization and the emergence of the virtual workplace. Rapidly changing technological developments have made e-mail an indispensable tool for managers (Markus, 1994), and one which employees are now expected to use as commonplace communication (Nadler & Shestowsky, 2004). The research community has also followed suit with explorations of the effects of using electronic communication media (such as e-mail) for business communications and the resulting effects on trust- and relationship-building (e.g., Landry, 2000; Moore, Kurtzberg, Thompson, & Morris, 1999; Morris, Nadler, Kurtzberg, & Thompson, 2002; Naquin & Paulson, 2003). However, as noted in a review of the performance appraisal literature (Levy & Williams, 2004), the use of electronic communication in performance appraisal processes has yet to be explored. This leaves open the question of whether there are any potential detriments to using this type of communication medium in performance appraisals that must be balanced against the potential benefits (e.g., convenience, ease of use, etc.).

Addressing these two challenges to performance appraisal processes, the presented research seeks to extend the literature by empirically examining how e-mail communication may influence peer-based performance appraisals. Specifically, with three empirical experiments, we explore how individuals may evaluate one's colleagues differently depending upon whether the communication channel is electronic or paper-based.

Theoretical perspectives comparing communication media

Much previous work, dominated by Media Richness Theory (see Daft & Lengel, 1986) has focused on communication media tools and has compared their features and capabilities. Generally, these studies conclude that face-to-face communication should be more comprehensive than voice-only communication (such as communicating via telephone), which should in turn be more comprehensive than text-only communication such as either paper-text letters or e-mail (considered equivalent by these theories), especially in ambiguous situations. This is generally attributed to the increased number of channels available for cues (verbal, aural, and visual) to be sent and received in a face-to-face context.

A second school of thought, however, argues that there are serious conceptual and empirical limitations to the categorization of media based primarily on the capabilities of the tool itself (Rice, 1992; Yates & Orlikowski, 1992). For example, the Social Influence Perspective is based on the idea long held by sociologists and cultural anthropologists that symbols have tremendous power for creating meaning for people in organizational life (e.g., Swidler, 1986). Comprehensive understanding, these theorists would argue, comes from the need to appreciate the subjective nature of interpretation of even the most seemingly objective elements of the empirical world (Martin, 2003). For example, elements such as peer pressure, supervisor attitudes and behaviors, and organizational and group norms can each affect one's attitudes towards a given communication medium (Fulk, Schmitz, & Steinfield, 1990). Thus, media choice and its subsequent impact may be even more a function of norms and attitudes rather than more objective characteristics of the media channel itself (see, e.g., Lee, 1994—Social Construction of Reality; Ngwenyama & Lee, 1997—Critical Social Theory; DeSanctis & Poole, 1994—Adaptive Structuration Theory; Carlson & Zmud, 1999—Channel Expansion Theory). Communication media thus serve “not only technical, efficiency-focused functions but also instrumental symbolic purposes” (Barry & Fulmer, 2004, p. 276). More specifically, we argue that current norms surrounding e-mail suggest that it is a more informal channel, appropriate for many forms of communication but not for all (King & Xia, 1997). Examples of inappropriate use of e-mail may include hiring, firing, or the transmission of serious personal information (Poe, 2001; Siddle, 2003). The effects of normative influence on behavior in the online environment can also be seen through differences in feelings of anonymity, social obligation, and deindividuation that can result (e.g. Dubrovsky, Kiesler, & Sethna, 1991; McKenna & Bargh, 2000; Wilkerson, Nagao, & Martin, 2002; Yates, Wagner, & Suprenant, 1997).

Taken together, it is likely that the objective medium characteristics, coupled with the social influences present

in the environment, shape both attitudes and behaviors. Here, we explore the application of these ideas in the performance appraisal context, and theoretically argue for a difference between e-mail communication and paper-form communication based on these types of social influences. We now turn to the limited empirical evidence that has contrasted various communication channels in the performance appraisal context, and then present our own comparison of these two text-only media.

E-Mail versus paper-based performance appraisals

Research on how e-mail appraisals compare to *face-to-face* interaction has revealed distinct differences; unfortunately, the research comparing e-mail appraisals to other *written* forms (i.e., the focus of the presented research) is limited, and conclusions are mixed at best. A wealth of previous research has demonstrated that people tend to give more negative appraisals when communicating in a computer mediated mode than when communicating face-to-face (Herbert & Vorauer, 2003; Siegel, Dubrovsky, Kiesler, & McGuire, 1986; Sussman & Sproull, 1999). For example, Herbert and Vorauer (2003) observed that when participants assigned to be judges evaluated other participants' written essays and communicated their feedback via e-mail or in a face-to-face context, interpersonal judgments were more positive in face-to-face exchanges compared to computer mediated exchanges. Other research has shown similar effects in the group context: groups that used computer-mediated communication exhibited more uninhibited (negative) behavior (had stronger and more inflammatory expressions in interpersonal interactions) when compared to face-to-face condition study participants (Siegel et al., 1986). In all, there seems to be a consistent trend in the empirical literature suggesting that people are more negative in an online environment than they are face-to-face.

To help explain these results, previous research has suggested that social desirability distortion may be lessened in the online context (Cannavale, Scarr, & Pepitone, 1970; Zerbe & Paulhus, 1987). In other words, while speaking face-to-face, one may feel a certain degree of social obligation to present more positive comments, while in the more detached online context, this sense of obligation is lessened (if not removed entirely). McKenna and Bargh (2000) present two possible results for this effect, based on two different mechanisms—first, they argue that the more negative comments may actually be more accurate because people feel more anonymous in the online context and are thus freed from the constraints of social obligation; second, they argue that the negative results may also be more harsh because of the deindividuation process by which people feel less

personally responsible for their comments. Other research has resulted in mixed conclusions on the topic of anonymity: some has indicated that people often feel more private or anonymous when communicating through a computer (Dubrovsky et al., 1991; Kiesler & Sproull, 1986; King & Miles, 1995; Shellie & Gilbert, 1995), while other recent work has suggested that people have become so accustomed to working with computers that they may actually feel *less* anonymous using online communication at this point (Griffith & Northcraft, 1994; Lautenschlager & Flaherty, 1990; Rosenfeld, Booth-Kewley, Edwards, & Thomas, 1996; Yates et al., 1997).

In contrast to the above-mentioned findings, the research specifically comparing e-mail to *paper-form* communication is notably smaller and less consistent than that which compares computer-mediated communication to face-to-face communication. This imbalance is particularly unfortunate for the understanding of performance appraisals, as they have been traditionally done via paper-form. Current trends in organizations might tend to pair the appraisals written on paper with a (face-to-face or verbal) conversation, but most would recognize the paper form to be an integral part of appraisal processes.

A central controversy in this small stream of research is whether or not e-mail and paper-form conditions produce identical responses from participants when used for any purpose, not just for appraisals. Some research has shown equivalence in the answers generated by computer-administered questionnaires and paper-form versions, further hypothesizing that both of these conditions yielded “more honest” answers from participants than would face-to-face interviews (Martin & Nagao, 1989; Wilkerson et al., 2002). These results are consistent with the predictions of Media Richness Theory, demonstrating that e-mail and paper media can be considered one and the same thing in terms of the behaviors they elicit (Booth-Kewley, Edwards, & Rosenfeld, 1992; Finegan & Allen, 1994; Potosky & Bobko, 1997). On the other hand, several empirical studies have shown that impression management is higher for respondents in a paper-form condition than in a computerized communication (Kiesler & Sproull, 1986) when both conditions guaranteed anonymity, and thus, people seemed willing to evaluate *themselves* less favorably via computer (Joinson, 1999; King & Miles, 1995; Richman, Kiesler, Weisband, & Drasgow, 1999).

Thus, this small stream of empirical results suggests potential differences in computer-generated versus paper-form media for *self*-reports based on a lessened need for impression management, and allows us to theorize a similar mechanism for *peer* evaluations, based on a lessened sense of social obligation. We argue that similar to the more negative evaluations people may give themselves online, they may also be more willing to rate

others more negatively online (as compared to ratings on a paper-form), based on the decrease in social obligation that they feel when using this medium, as has been documented in the comparison between online and face-to-face media (McKenna & Bargh, 2000). In other words, we are stating here that the electronic media context may not only influence how people feel about themselves, but also how they feel about others. In this way we aim to extend the knowledge in this area by moving from the realm of the self to the realm of the other in the online-versus-paper ratings context.

To summarize, the conflicting evidence noted above leaves open the possibility that individuals will provide different appraisal data in the e-mail condition than by paper, even though both paper-form and e-mail media may technically offer the same communication opportunity (as stated by Media Richness Theory) for one-way, text-only communication. Thus, in the present research we predict that more negative evaluations will occur on e-mail as opposed to on a written evaluation. This argument is based on (1) the findings that e-mail seems to elicit more negativity in general (Kiesler & Sproull, 1992) and more negative appraisal evaluations in particular, when compared to face-to-face alternatives (Herbert & Vorauer, 2003; Siegel et al., 1986), and (2) based on the idea that feelings of social obligation will be lessened in the online environment as compared to a paper-based medium, and so people may tend to self-censor their comments/evaluations less in this context than they would on a paper form.

Hypothesis 1. Individuals will provide more negative evaluations by e-mail than by paper-form.

Hypothesis 2. Feelings of social obligation will mediate the relationship between communication medium and evaluations.

We now turn our attention towards three experiments that test these hypotheses. The first experiment, outlined below, was designed to directly test the relationship between e-mail and paper-based appraisals (H1), the second experiment will also present a test of the proposed mediation effect (H2), and the third experiment will reinforce the main hypothesis (H1) in a more controlled setting, allowing for more confidence in the results.

Study 1

Study 1 method

Participants were 73 fulltime graduate level business students who completed the study as part of a class assignment. All participants were informed that they were to provide a performance review evaluating each of their teammates' performance during a simulated nego-

tiation. The experimental design had one manipulation—they either completed their performance review on-line via e-mail ($n=36$) or by traditional paper-form ($n=37$). All participants were randomly assigned to one of the two conditions.

All participants completed a team task comprised of working on one side of a long-term labor-management team negotiation. Participants were randomly assigned to teams composed of six members with the exception of one team having seven members. The teams were then assigned to one role of the negotiation (management or labor), and then negotiated against another team. The negotiation exercise, ABC-Local 190 (Valley & Medvec, 1996), involved three distinct rounds of negotiation and was conducted over a span of 10 weeks. The participants were asked to assume the role of either (1) management as a Senior Personnel Officer on a negotiating team representing the Adam Baxter Company (ABC), or (2) labor as a union official on a negotiating team representing the employees of ABC, and all members of the same team received the exact same materials.

After the completion of the entire negotiation (all three rounds), participants were asked to complete a performance review of their teammates (again, only those on the exact same side of the negotiation as themselves). Participants were explicitly informed that their ratings would be held in the strictest of confidence, and would be worth 10% of the ratee's final grade.¹ Participants in the paper-form condition received their material via their school mailbox. Participants in the e-mail condition received their material in electronic format via an e-mail attachment. In both conditions, participants had one week in which to complete their peer evaluations. Those in the paper-form condition returned their completed evaluations to the instructor's campus mailbox (a secure location which was managed and monitored by fulltime university employees) while those in the online condition returned their completed evaluations to the instructor via e-mail.

Dependent measure

Participants evaluated their peers along three dimensions: participation, contribution, and an overall category. Participants rated their peers on an A, B, C, D, and F scale in which "A" is excellent and "F" is failure, with no pluses or minuses permitted. All participants were familiar with this scale, as that is how all their courses were officially graded. These grades were translated into a number scale where A=4, B=3... F=0. An overall peer review index measure was compiled for each

¹ In no case did the inclusion of this variable change any participants' final grade; i.e., every participant received an identical final grade with or without the inclusion of the peer ratings variable.

Table 1
Intraclass correlations (ICC) for the evaluated dimensions by experimental condition for Study 1

Dimension	Experimental condition							
	Pen-and-paper				Online			
	ICC (2, 1)	ICC (2, 5)	<i>M</i>	<i>SD</i>	ICC (2, 1)	ICC (2, 5)	<i>M</i>	<i>SD</i>
Participation	.72	.96	3.94	.23	.84	.97	3.51	.56
Contribution	.66	.80	3.92	.27	.73	.86	3.51	.61
Overall	.68	.82	3.96	.23	.85	.92	3.56	.60

participant by aggregating the three categories across raters (Cronbach's $\alpha = .92$).

We also analyzed the inter-rater reliability of the reported evaluations by calculating intraclass correlations (ICC) to ensure that most of the variance in measurements was due to the target variable, the ratee, as opposed to the raters (see ShROUT & Fleiss, 1979).² As displayed in Table 1, the ICCs are sufficient to suggest that most of the variance in measurements was due to the target variable.

Results

The role of the participants in the negotiation (labor or management) did not influence the dependent variable, so the data were collapsed across role.

Significant differences were found in how individuals rated their peers as a function of media. Confirming Hypothesis 1, participants in the e-mail condition ($M = 3.51$, $SD = .51$) evaluated their peers significantly lower than those in the paper-based condition ($M = 3.94$, $SD = .23$), $F(1, 71) = 21.61$, $p < .001$, $\eta^2 = .23$. As a post hoc analysis we examined evaluations as a function of communication media with role (management or labor) as a covariate, and the basic finding remained the same: $F(1, 70) = 21.50$, $p < .001$, $\eta^2 = .23$.

Study 2

Results of Study 1 revealed that media type did significantly influence peer evaluations such that individuals were more likely to evaluate their peers negatively in the e-mail condition than in the paper-form condition. In this follow-up study we attempt to better understand

why this may happen through an exploration of our hypothesized mediator: feelings of social obligation. In this context, these feelings may well tap into one of the recognized problems with peer-based appraisals—the use of evaluators who may feel uncomfortable influencing their peers' rewards and punishments, something traditionally reserved for those hierarchically superior (Druskat & Wolff, 1999; Saavedra & Kwun, 1993). We here explore whether the perception of social obligation may differ as a function of media type. More specifically, we rely on the theory, described earlier, that social obligation is generally lowered in the online environment. That is, individuals may feel more freedom to assess one another as they see fit, and not artificially inflate their ratings based on a sense of courtesy, and thus may feel more justified in giving negative evaluations when communicating online than with the paper-form method. Thus, our second study was designed to both confirm Hypothesis 1 and also test Hypothesis 2 by examining the degree of social obligation that participants felt as a possible mediating variable that may help to explain this relationship.

Study 2 method

One hundred ten fulltime graduate level business students completed the study as part of a class assignment. The experimental design had one manipulation—they either completed their performance review on-line via e-mail ($n = 56$) or by traditional paper-form ($n = 54$). All participants were randomly assigned to one of the two conditions.

Using the same activity as Study 1, all participants completed a team task comprised of a long-term labor-management team negotiation. Participants were randomly assigned to teams composed of 4 or 5 members in this case (10 teams had 4 members and 14 teams had 5). The teams were then randomly assigned a role (management or labor) and then negotiated against another team. After the completion of the entire negotiation (all three rounds), participants were asked to complete a performance review of their teammates. The structure of the evaluation procedure was essentially the same as that used in Study 1: participants were explicitly informed that their ratings would be held in the strictest of confidence; those in the paper-form condition received the materials via their school mailbox, while those in the e-mail condition

² Specifically, following the guidance set forth by McGraw and Wong (1996) in selecting an appropriate ICC given data and conceptual purposes, we used what is referred to as Case 2, which is a two-way random effects model. In this model, not only are the targets (i.e., ratees) considered random, but the raters are deemed a random effect also (as they are randomly selected from a larger pool of judges). Furthermore, the Case 2 ICCs we present are for both the single measure and that of the average measure of raters. This is referred to in Table 1 by the nomenclature ICC(2, 1), the "2" being Case 2 and "1" indicating the reliability measure is for a single rating (versus the expected average rating) and ICC(2, k) where "k" denotes the average number of raters. Table 1 also presents the ICCs when using the absolute agreement index (versus a consistency index).

received their material in electronic format via an e-mail attachment. In both conditions, participants had 48 h in which to complete their peer evaluations, with those in the paper-form condition returning their completed evaluations to the instructor’s campus mailbox in the monitored room, and those in the online condition returning their completed evaluations to the instructor via e-mail. Differing from Study 1, in class on the following day, all participants in both conditions received and filled out an identical paper-form questionnaire (containing the social obligation items and two masking questions) during class time and submitted it to the instructor.

Dependent measures

Performance

As in Study 1, participants evaluated their peers along three dimensions: participation, contribution, and an overall category on the A, B, C, D, and F scale used in their own courses in which “A” is excellent and “F” is failure, with no pluses or minuses permitted. These grades were translated into a number scale where A = 4, B = 3 ... F = 0. An overall peer review index measure was again compiled for each participant by aggregating the three categories (Cronbach’s $\alpha = .93$). As in the first study, we explored the inter-rater reliability of the reported evaluation measures. Table 2 displays the intra-class correlations (ICC) used to measure inter-rater reliability through the same procedures as in Study 1, which (as in Study 1) were sufficient to suggest that most of the variance in measurements was due to the target variable, the ratee (as opposed to the raters).

Social obligation

Two items from the paper-based questionnaire were used to create the scale on feelings of social obligation. First, participants were asked to rate their thoughts on the following item: “To what degree do you feel that team members owe each other the courtesy of reasonably good ratings?” on a 7-point Likert-type scale, with 1 = Not at All and 7 = Extremely. In addition, they were asked, “How justified would it be if you rated your teammates in the peer review critically (i.e., less than perfect if it was appropriate)?” on a 7-point Likert-type scale, with 1 = Not at All and 7 = Extremely. This latter question was reverse scored, so a higher overall rating indicates feeling a greater degree of social obligation. Responses

to these two questions were aggregated into a composite measure for social obligation (Cronbach’s $\alpha = .89$).

Masking questions

In addition to the two questions measuring feelings of social obligation, participants were also asked two masking questions to hide the true intention of the questionnaire: “How do you feel you performed relative to other teams?” and “To what degree are you willing to work with this team in the future?” These are referred to as masking questions “A” and “B” respectively in Table 3. It is important to note that these two questions, though somewhat evaluative in nature, were asked on the *paper form* for all participants in all conditions; therefore, we would not expect any systematic differences in their evaluations on these items based on experimental condition.

Study 2 results

Participant responses did not vary as a function of negotiation role and thus, results were collapsed across roles. In addition, participants’ responses to the masking questions did not vary as a function of experimental manipulation. Please see Table 3 for all descriptive statistics from Study 2.

Performance

Replicating the findings of Study 1, participants who completed the appraisal electronically tended to provide lower evaluations ($M = 3.50, SD = .61$) than those who completed the appraisal via paper form ($M = 3.78, SD = .42$), $F(1, 108) = 7.81, p < .01, \eta^2 = .07$. As with Study 1, we examined evaluations as a function of com-

Table 3
Means, standard deviations, correlations, and reliabilities of variables in Study 2

Variable	<i>M</i>	<i>SD</i>	1	2	3	4
1 Evaluation	3.64	.54	(.93)			
2 Social Obligation	4.92	.88	.33**	(.89)		
Masking Questions						
3 A	5.01	.94	.04	.14		
4 B	4.83	1.67	.09	.07	.01	

Note. Coefficients alpha are displayed on the diagonal in parentheses. $N = 110$. Evaluation Scores are on a scale of 1–4, all others are on a scale of 1–7.

** $p < .01$.

Table 2
Intraclass correlations (ICC) for the evaluated dimensions by experimental condition for Study 2

Dimension	Experimental condition							
	Pen-and-paper				Online			
	ICC (2, 1)	ICC (2, 5)	<i>M</i>	<i>SD</i>	ICC (2, 1)	ICC (2, 5)	<i>M</i>	<i>SD</i>
Participation	.83	.94	3.77	.42	.78	.89	3.50	.69
Contribution	.79	.91	3.80	.41	.84	.94	3.57	.53
Overall	.76	.89	3.76	.47	.77	.87	3.48	.63

munication media with role (management or labor) as a covariate, and again the basic finding remained the same: $F(1,107) = 8.40, p < .01, \eta^2 = .07$.

Social obligation

E-mail respondents were significantly lower ($M = 4.63, SD = .93$) than paper respondents ($M = 5.22, SD = .72$) in their judgments of social obligation $F(1,108) = 13.87, p < .001, \eta^2 = .11$ (in other words, they felt a reduced sense that they owed others the courtesy of good ratings and felt more justified in giving their ratings). Post hoc analysis using role as a covariate revealed no change in the effect, $F(1,107) = 14.42, p < .001, \eta^2 = .11$.

Mediation analyses were conducted as per the steps set forth by *Kenny, Kashy, and Bolger (1998)*. First, communication medium was shown to be a significant predictor of ratings, $\beta = -.26, t(107) = -2.80, p < .01$, as described above. Next, medium was used to predict participant's reported social obligation, $\beta = -.34, t(107) = -3.86, p < .001$. Third, the relationship between reported social obligation and performance appraisal was identified, while controlling for the medium used. Using final ratings as the criterion variable in a regression equation with (a) the media type and (b) reported social obligation, it was found that perceptions of social obligation, the hypothesized mediator, was correlated with the dependent variable of performance rating, $\beta = .34, t(107) = 3.61, p < .001$, and remained significant even when controlling for the communication medium, $\beta = .28, t(106) = 2.81, p < .01$. This suggests that the reported social obligation mediates this relationship, confirming Hypothesis 2. Finally, when controlling for reported social obligation in this regression equation, the significant relationship between media type and performance ratings, $\beta = -.26, t(107) = -2.80, p < .01$, was reduced to nonsignificance, $\beta = -.17, t(106) = -1.72, ns$, indicating that reported social obligation does indeed mediate this relationship.

We next tested whether the indirect effect was significant, the indirect effect being the relationship between the communication medium and performance appraisal via the social obligation measure. Using the more conservative version of the Sobel test (*Goodman I*) recommended by *Baron and Kenny (1986)* which is distributed as *Z*, it was demonstrated that this indirect effect was significant, $Z = 2.22, p < .05$. Thus, not only does the multi-item measure for social obligation medium the reported effect but the indirect effect is also significant (i.e., the mediator carries the influence of an independent variable).

Study 3

Two items of concern remained based on the methods employed by the first two studies, so an entirely new design was implemented as Study 3. First, in Studies 1 and 2, we did not actually know where the participants

completed the appraisal task, and so the online-versus-paper context may have spuriously overlapped with anonymity in the study as it was executed, as a function of location. In other words, if it was the case that those in the e-mail condition systematically filled out their forms at home or away from other people somewhere, whereas those in the paper-form condition tended to sit in the classroom or in a school lounge to fill out the forms, then those in the e-mail condition may have actually had more anonymity—a condition that might change how free respondents felt to offer negative evaluations (and potentially even influence their feelings of social obligation). In addition, though the random assignment used in both Studies 1 and 2 gives us some confidence in the overall effect, participants rated a variety of targets instead of a single one. Thus, we cannot be absolutely sure about the relative ratings of the various stimuli. In other words, it is at least possible that those rated in the online condition actually performed worse as a group than those in the paper-form condition, and therefore the appraisal differences that we observed could possibly have accurately reflected reality instead of a systematic tendency based on the medium used. To address both of these concerns, Study 3 used a common stimulus (all participants rated the exact same performance—that of a person whom nobody knew personally) and a common setting for ratings (school computer classrooms during a fixed and monitored timeframe during class), thereby removing possible concerns about these sources of variation. While the stimuli in both Studies 1 and 2 were real classmates whom each participant knew and with whom they had worked personally, Study 3 deviates from this pattern and allows us to test Hypothesis 1 under conditions in which there is no apparent “social obligation” to provide a lenient assessment, to begin to map out the generalizability of our findings.

Study 3 method

Participants were 122 graduate level students enrolled in an executive MBA program, who had all recently completed a course in Negotiations the semester prior to participating in this experiment.

Participants were informed that as follow-up to their Negotiations course, they were to watch a documentary video of a “real life” negotiation and complete a decision making exercise upon its conclusion. Participants watched the video in one of two classrooms, identical in layout and facilities. The two (identical) rooms, it is important to note, were computerized classrooms (used to teach programming languages) in that each seat had a desk with a glass window in which a computer screen is displayed (the computer screen, keyboard, and CPU are located underneath the desk). The desk, along with the glass window, acts as a writing surface. In short, the separate classrooms in which participants watched the video were identical

and, furthermore, each student had access to both a writing desk and individual access to a computer.

The stimulus material was a video entitled “Waco: the inside story”³ that documents the actual 51-day FBI negotiation with the Branch Davidians in Waco, Texas during the spring of 1993. Although many documentaries have been made of the interaction between the FBI and the Branch Davidians and the tragic outcome that resulted, this is the only documentary to focus exclusively upon the *negotiation* process. Specifically, the documentary focused upon the FBI negotiation team that was led by Agent Byron Sage. While some of the negotiation tactics portrayed in the video had positive results, other negotiation tactics portrayed had negative consequences, mimicking actual performance in most organizational tasks which most often contain some positive and some negative elements. Further, as described above, the mixed-element performance seen by participants is specifically the most appropriate for examining variation in peer evaluations, as even more of the judgment is left to the discretion of the individual than would be in the case of an entirely positive or an entirely negative sequence of behaviors.

After watching the video, participants were asked to complete the evaluation exercise. Participants were reminded that they had just recently completed their Negotiations course. They were asked to imagine themselves being employed by the FBI as a negotiator, and asked to evaluate the performance of one of their peers, Agent Byron Sage, as a negotiator.

The experimental design had one manipulation: participants either completed their performance review online via e-mail ($n=60$) or by traditional paper-form ($n=62$), based on which room they saw the video—in one classroom the participants all completed paper-based evaluations, while in the other participants completed online evaluations at the computers in the desks. All participants were randomly assigned to one of the two conditions. Participants were explicitly informed that their ratings would be held in the strictest of confidence. Participants in the paper-form condition had their paper-form passed out to them, while those in the e-mail condition were asked to log in to their accounts via the computer in front of them, and received their material in electronic format via an e-mail attachment. In both conditions, participants were asked to complete their assignment before leaving the classroom. Those in the paper-form condition returned their completed evaluations to the instructor by hand. Those in the online condition returned their completed evaluations to the instructor via e-mail.

Dependent measure

Performance

Participants evaluated Byron Sage’s performance along two dimensions that were aggregated into a single measure of overall performance. One question asked: “Overall, based on this documentary, how would you rate Byron Sage’s performance as a negotiator?” Participants’ responses were measured on a scale of 1 (poor) to 7 (excellent) with a midpoint at 4 (average). A second question asked participants the following:

“As you may know pay raises are based on performance. Given Byron Sage’s performance for this particular negotiation, what do you recommend to Sage’s superiors? Specifically, given that pay raises can vary between 0 to 6% with an average of 3%, what do you recommend Sage’s performance deserves?”

Participants’ responses were measured on a scale from 0 (no raise) to 6 (maximum) with a midpoint of 3 (average). This scale was then transformed into a 1–7 scale to assure that the two scales received equal weight in the composite measure.

An overall performance index measure was compiled for each participant by aggregating the two categories ($\alpha = .94$).

Involvement

As a control, participants were asked whether they had any involvement in the events portrayed in the video. This was done by asking participants, “Were you involved (either directly or indirectly) with the incident portrayed in the video?” with a spot to check either “yes” or “no.” Participants were then asked to elaborate in an open-ended response if they answered yes. Not a single participant (0%) reported having any involvement in the events portrayed in the documentary.

Results

Performance

Significant differences were found in how individuals rated Agent Byron Sage’s negotiation ability as a function of media. Confirming our original hypothesis and replicating the pattern of results from Studies 1 and 2, participants in the e-mail condition ($M=4.46$, $SD=.98$) evaluated Byron Sage significantly lower than those in the paper-based condition ($M=4.95$, $SD=1.13$), $F(1, 120) = 6.56$, $p < .05$, $\eta^2 = .05$.

General discussion

The three studies presented a clear pattern of results suggesting that the communication media used for performance appraisals of peers does indeed matter.

³ This video was produced by the news program Frontline and is available through PBS.

Specifically, e-mail seemed to result in more negative ratings overall, as well as a lessened sense of social obligation, than did paper-form ratings. In demonstrating this, we contribute empirical evidence to a small but growing stream of literature that describes the potential differences resulting from the use of various communication media.

All three studies demonstrate that people give more negative ratings over e-mail as compared to those given on a paper form. It is important to realize, however, that the degree of negativity measured here is relevant only in a relative sense as compared to the less-negative evaluations offered by those using paper forms, and we cannot draw specific conclusions about these ratings being either needlessly harsh or more honest. However, in many modern organizational jobs, performance is often indeed something of a subjective experience (hence the need for peer-based impressions and appraisals at all—truly objective performance could likely be measured by some definitive and quantitative standard such as units produced or sold, for instance). In the realm of more subjective performance, ratings are likely to fall within a range, instead of converging at a single “true” value, even without the added complication of the use of different tools for the assessment process. Given this, it is crucial to understand the impact of differing media on the appraisal process—the systematic tendency for appraisals offered via e-mail to be more negative than those offered on a written form is a type of bias that can and should be recognized by organizations.

In addition, employees may well feel a sense of conflicting loyalties about having to evaluate their peers, so it is particularly worthy of additional research attention in this context to discover more definitively whether e-mail may allow for a window of greater honesty, a chance to be more negative than is fair, or, as we suggest, simply swing one’s ratings systematically towards the negative end of a reasonable range of responses. Our results do at least suggest that one of the purported challenges of the peer review process—the resistance to critically evaluate a peer as rewards and punishments are traditionally reserved for the hierarchy—may be attenuated in the electronic environment. However, it is important to recognize that in addition to this subjective bias within a reasonable range of ratings, it is also possible that some degree of inaccuracy is more likely to emerge in one medium than in another. To test this, a future study would have to explore stimuli with an outcome that could be objectively defined, to see if people consistently under- or over-rate it based on a particular medium used.

Study 2 results demonstrated that those in the e-mail condition admitted feeling a lesser sense of social obligation in giving more negative ratings, and these feelings fully mediated the relationship between medium and ratings. While it is possible that participants later reported

greater feelings of justification or lesser feelings of courtesy in giving negative ratings merely to support their previous decisions to give negative ratings, this likely does not explain the entire phenomenon. Our findings in this regard are consistent with previous research that has observed a reduced sense of impression management for online communicators in terms of self-report, which lends some confidence to the link between these constructs beyond the need to present one’s decisions in a consistent manner. That said, we also wish to note that our investigation of social obligation does not rule out other potential mediators. Indeed, it is likely that there exist other mediators influencing this process and our exclusive focus on social obligation is a limitation of our presented research.

Future research can aim to further tease apart other possible causal mechanisms in this process and more closely pinpoint the contribution of influences such as perceived anonymity, deindividuation, and self-awareness in online responses (McKenna & Bargh, 2000). For example, it is possible that despite the experimental controls utilized, participants continued to feel less anonymous in one or the other conditions, based perhaps on the need to write on the form in their own handwriting, or based on the fact that their e-mail comments get mailed with their e-mail address necessarily associated with the review. It may be a worthwhile avenue for future research to attempt to tease these effects apart.

In fact, our Study 3 methodology lends itself to the idea that the effect in question may also be encouraged by other psychological processes in addition to (or instead of) the feelings of social obligation and justification identified in Study 2. One of the key differences in the third study as presented was the fact that the “peer” relationship in question was simulated as opposed to actual; that is, unlike in our first two studies, participants in Study 3 did not actually know the focal actor whom they were to evaluate, and so likely would not have felt the same degree of social obligation towards that individual. Yet, ratings of that individual’s performance in the e-mail condition were still systematically lower than those reported in the paper condition. As described above, this opens an avenue for future research to help establish mechanisms other than social constraint that may also generate this effect, relating more to people’s reactions to electronic communication in general (such as anonymity, deindividuation, or even generalized affective responses) and less to their feelings directed towards one specific other person.

Thus, theoretically, our findings further the line of thinking which states that to truly understand a tool, you must understand not only its physical properties, but also the way people think about it, treat it, and its symbolic representation in an organization (Fiol, 2002; Weick, 1979). In this case, though we cannot know precisely why people think of written text on a computer

screen as a systematically different opportunity than writing on a piece of paper, our results suggest that these differences do indeed exist. And, as organizations increasingly transfer tasks that were formerly paper-based to paper-free or online versions, the reasons behind these differences become even more critical to understand. Indeed, the first tasks to become electronic are likely to be those that were formerly written tasks, as opposed to those typically accomplished on the phone or in a face-to-face conversation. Perhaps the difference we saw here relates to the ease of typing instead of handwriting, but since our analyses here were done only on the letter-grades that participants gave each other, this seems less likely to be influenced by ease of creation.

From the results of our study several practical implications for managers can be suggested. First, as our research indicates, people seem to behave more negatively online (a finding consistent with previous research on this topic—see Siegel et al., 1986; Sussman & Sproull, 1999). A concrete prescription based on our results is that in this performance appraisal context, either e-mail should be used entirely or not at all—if some people use this method for providing performance appraisals and others do not, the possibility for systematic bias emerges. In a more general sense, if some people in an organization feel that e-mail is a primarily informal tool, effective for “venting” and thinking through issues but not for formal decisions, while others assume it is exactly equivalent to paper-forms or other communication tools, miscommunications could arise. Thus, it seems necessary to create a consistent understanding of what e-mail is for and how it is to be used in an organization’s culture, to avoid possible systematic bias of the type seen here.

On a more societal level, a popular belief exists that the technological revolution of the last decades changed the way we perceive the world—in particular, how we work and communicate with each other. Indeed, technology did simplify many routine tasks. Further, communication technology has tremendously sped up information processes, as well as dramatically increased information quantity. However, the human mind was and still is the main tool we use for information handling and it has its natural boundaries with respect to information processing and absorption. In addition, human rationality is limited by many psychological and cognitive mechanisms that significantly influence our reactions to events we encounter. Nevertheless, as stated by Simon (1997), “new problems created (or made visible) by our new scientific knowledge are symptoms of progress, not omens of doom” (p. 249). Therefore, an understanding of the ways in which technology changes our assumptions and behaviors is clearly necessary for successful organizational practices. These studies provide an additional step towards mapping the new organizational landscape that electronic communication technology has created, at least in the realm of performance appraisals.

References

- Barry, B., & Fulmer, I. S. (2004). The medium and the message: The adaptive use of communication media in dyadic influence. *Academy of Management Review*, 29, 272–292.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychology research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.
- Booth-Kewley, S., Edwards, J. E., & Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference?. *Journal of Applied Psychology*, 77, 562–566.
- Butler, J., Ferris, G., & Napier, N. (1991). *Strategy and human resource management*. Cincinnati: South-Western Publishing.
- Cannavale, F. J., Scarr, H. A., & Pepitone, A. (1970). Deindividuation in the small group: Further evidence. In W. L. Richman, S. Kiesler, S. Weisband, & F. Drasgow (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology* 84, 754–775.
- Carlson, J. R., & Zmud, R. W. (1999). Channel expansion theory and the experiential nature of media richness perceptions. *Academy of Management Journal*, 42, 153–170.
- Daft, R. L., & Lengel, R. J. (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32, 554–571.
- DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive Structuration Theory. *Organization Science*, 5, 121–147.
- Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, 84, 58–74.
- Dubrovsky, V., Kiesler, S., & Sethna, B. (1991). The equalization phenomenon: Status effects in computer-mediated and face-to-face decision groups. *Human Computer Interaction*, 6, 119–146.
- Finegan, J. E., & Allen, N. J. (1994). Computerized and written questionnaires: Are they equivalent?. *Computers in Human Behavior*, 10, 483–496.
- Fiol, M. C. (2002). Intraorganizational cognition and interpretation. In J. A. C. Baum (Ed.), *Companion to Organizations*. Malden, MA: Blackwell.
- Fulk, J., Schmitz, J., & Steinfield, C. W. (1990). *A social influence model of technology use*. Newbury Park, CA: Sage.
- Griffith, T. L., & Northcraft, G. B. (1994). Distinguishing between the forest and the trees: Media, features, and methodology in electronic communication research. *Organization Science*, 5, 272–285.
- Herbert, B. G., & Vorauer, J. D. (2003). Seeing through the screen: Is evaluative feedback communicated more effectively in face-to-face or computer-mediated exchanges. *Computers in Human Behavior*, 19, 25–38.
- Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, 31, 433–438.
- Katzenbach, J. R., & Smith, D. K. (1993). *The wisdom of teams: Creating the high performance organization*. Boston, MA: Harvard Business School Press.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 233–265). Boston, MA: McGraw-Hill.
- Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*, 50, 402–413.
- Kiesler, S., & Sproull, L. (1992). Group decision making and communication technology. *Organizational Behavior and Human Decision Processes*, 52, 96–123.

- King, W. C., & Miles, E. W. (1995). A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pen measurements: A test of measurement equivalence. *Journal of Applied Psychology*, *80*, 643–651.
- King, R. C., & Xia, W. (1997). Media appropriateness: Effects of experience on communication media choice. *Decision Sciences*, *28*, 877–910.
- Landry, E. M. (2000). Scrolling around the new organization: The potential for conflict in the on-line environment. *Negotiation Journal*, *133*–142.
- Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology*, *75*, 310–314.
- Lee, A. S. (1994). Electronic mail as a medium for rich communication: An empirical investigation using hermeneutic interpretation. *MIS Quarterly*, *18*, 143–157.
- Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*, 881–905.
- Markus, M. (1994). Electronic mail as the medium of managerial choice. *Organization Science*, *5*, 502–527.
- Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology*, *74*, 72–80.
- Martin, J. (2003). *Meta-theoretical controversies in studying organizational culture*. Oxford, UK: Oxford University Press.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46.
- McKenna, K. Y. A., & Bargh, J. A. (2000). Plan 9 from cyberspace: The implications of the Internet for personality and social psychology. *Personality and Social Psychology Review*, *4*, 57–75.
- Moore, D. A., Kurtzberg, T. R., Thompson, L. L., & Morris, M. W. (1999). Long and short routes to success in electronically mediated negotiation: Group affiliations and good vibrations. *Organizational Behavior and Human Decision Processes*, *77*, 22–43.
- Morris, M. W., Nadler, J., Kurtzberg, T. R., & Thompson, L. (2002). Schmooze or lose: Social friction and lubrication in e-mail negotiations. *Group Dynamics: Theory, Research, and Practice*, *6*, 89–100.
- Nadler, D. A., Gerstein, M. S., & Shaw, R. B. (1992). *Organizational architecture: Designs for changing organizations*. San Francisco: Jossey-Bass Publishers.
- Nadler, J., & Shestowshy, D. (2004). Negotiation, information technology, and the problem of the faceless other: Dispute Resolution Research Center, Working paper #304. Kellogg School of Management.
- Naquin, C. E., & Paulson, G. D. (2003). Online bargaining and interpersonal trust. *Journal of Applied Psychology*, *88*, 113–120.
- Ngwenyama, O. K., & Lee, A. S. (1997). Communication richness in electronic mail: Critical Social Theory and the contextually of meaning. *MIS Quarterly*, *145*–167.
- Ouchi, W. (1981). *Theory Z*. New York: Addison-Wesley.
- Poe, A. C. (2001). Don't touch that 'send' button. *HR Magazine*, 2001.
- Potosky, D., & Bobko, P. (1997). Computer versus paper-and-pen administration mode and response distortion in noncognitive selection tests. *Journal of Applied Psychology*, *87*, 293–299.
- Rice, R. E. (1992). Task analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. *Organization Science*, *3*, 475–500.
- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, *84*, 754–775.
- Rosenfield, P., Booth-Kewley, S., Edwards, J. E., & Thomas, M. D. (1996). Responses on computer survey: Impression management, social desirability, and the Big Brother syndrome. *Computers in Human Behavior*, *12*, 263–274.
- Saavedra, R., & Kwun, S. K. (1993). Peer evaluations in self-managing work groups. *Journal of Applied Psychology*, *78*, 450–462.
- Shellie, L. D., & Gilbert, B. O. (1995). Method of psychological assessment, self-disclosure, and experiential differences: A study of computer, questionnaire, and interview assessment formats. *Journal of Social Behavior and Personality*, *10*, 255–263.
- Shore, T. H., Shore, L. M., & Thornton, G. C. (1992). Construct validity of self and peer evaluations of performance dimensions in an assessment center. *Journal of Applied Psychology*, *77*, 42–54.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.
- Siegel, J., Dubrovsky, V., Kiesler, S., & McGuire, T. W. (1986). Group processes in computer mediated communication. *Organizational Behavior and Human Decision Processes*, *37*, 241–262.
- Siddle, P. F. (2003). Mind your e-mail manners. *Journal of Accountancy*.
- Simon, H. A. (1997). *Administrative behavior* (4th ed.). New York: Free Press.
- Staw, B. M. (1980). *Rationality and justification in organizational life*. Greenwich, CT: JAI Press.
- Sussman, S. W., & Sproull, L. (1999). Straight talk: Delivering bad news through electronic communication. *Information Systems Research*, *10*, 150–166.
- Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review*, *51*, 273–286.
- Thompson, L. (2004). *Making the team: A guide for managers* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Valley, K. L., & Medvec, V. H. (1996). *ABC-local negotiation case*. Boston, MA: Harvard Business School.
- Weick, K. E. (1979). *The social psychology of organizing* (2nd ed.). Addison-Wesley.
- Wilkerson, J. M., Nagao, D. H., & Martin, C. L. (2002). Socially desirable responding in computerized questionnaires: When questionnaire purpose matters more than the mode. *Journal of Applied Social Psychology*, *32*, 544–559.
- Yammarino, F. J., & Waldman, D. A. (1993). Performance in relation to job skill importance: A consideration of rater source. *Journal of Applied Psychology*, *78*, 242–249.
- Yates, J., & Orlikowski, W. J. (1992). Genres of organizational communication: A structural approach to studying communication and media. *Academy of Management Review*, *17*, 299–326.
- Yates, B. T., Wagner, J. L., & Suprenant, L. M. (1997). Recall of health-risky behaviors for the prior 2 or 4 weeks via computerized versus printed questionnaire. *Computers in Human Behavior*, *13*, 83–110.
- Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, *12*, 250–264.