© Academy of Management Learning & Education, 2010, Vol. 9, No. 4, 652–662.

Improving the Effectiveness of Students in Groups With a Centralized Peer Evaluation System

STÉPHANE BRUTUS
MAGDA B. L. DONIA
Concordia University

We describe the impact of a centralized electronic peer evaluation system on the group effectiveness of undergraduate business students over a pair of semesters. Using a quasi-experimental design, 389 undergraduate students evaluated, and were evaluated by, their peers using a web-based system that captures peer evaluations in quantitative and qualitative formats and allows for the reception of anonymous feedback. Results show that the effectiveness of students, as perceived by their peers, increased over semesters. This effect could be directly linked to the use of the system. The results of this study underscore the benefit of centralizing peer evaluations for the assessment of important skills and their development in higher education. The implication of these results and possible avenues of research are detailed.

A common challenge for business school professors is the assessment and management of individual contribution to group work (Brown, Rust, & Gibbs, 1994; Mello, 1993). While the outcomes of group work are typically under the purview of educators, much value resides in monitoring the individual activities that precede the final group report, project, or presentation. Peer evaluations are meant to capture information related to internal group processes—information that is inaccessible to anyone other than group members. Peer evaluations not only offer a useful window into "what happened" (evaluation tool) but also they represent a valuable learning tool in providing students with the experience of evaluating others and receiving feedback from their peers on their behaviors. These processes also expose students to an important and difficult organizational duty, that of evaluating, and being evaluated by, others. Although the reliance on students as a source of evaluation has become quite common in higher education (Chen & Lou, 2006; Fellenz, 2006; Li, 2001; Topping, 1998), little is known of student evaluations' impact on student learning. We investigate the benefits, for students, of using a standardized, web-based system to capture peer evaluations and provide feedback. More specifically, we seek to investigate the effects of using peer evaluations in one semester on students' group effectiveness in the subsequent semester. In the following paragraphs, we provide some context for the use of peer evaluations in higher education.

PEER EVALUATION IN HIGHER EDUCATION: SOME BACKGROUND

In organizations, a vast majority of workflow is structured around teams (e.g., Guzzo & Shea, 1992) and, as a result, significant interpersonal and social demands are placed on organizational members (Cappelli & Rogovsky, 1994; Zedeck & Goldstein, 2000). Although many critics have argued that business schools have been slow to respond to this need (Navarro, 2008; Porter & McKibbin, 1988), there has been increased interest in teaching group-

The authors would like to express their gratitude to Devasheesh Bhave and Gary Johns for their insightful critiques of previous drafts of this paper. We also thank John Longo, Mary Genova, and Mary Waterhouse of the Center for Instructional Technology of the John Molson School of Business for the technical support that they have provided for the peer evaluation system over the years.

Copyright of the Academy of Management, all rights reserved. Contents may not be copied, emailed, posted to a listserv, or otherwise transmitted without the copyright holder's express written permission. Users may print, download or email articles for individual use only.

related skills in higher education (e.g., Chen, Donahue, & Klimoski, 2004). At the very least, most university courses now provide students with the experience of team work (Johnson & Johnson, 1987; Topping, 1998), especially those offered in business schools (e.g., Baldwin, Bedell, & Johnson, 1997). Today, the typical undergraduate or MBA student works within a group in a majority of his or her courses (Peiperl & Trevelyan, 1997). A common supplement to group-based projects is to have students evaluate each other at the end of the exercise (Falchikov, 1986). Peer evaluation processes are relatively simple to develop, and their use follows an important trend in higher education of relying on peer relationships to support educational objectives. Examples of peer-supported systems include student-led workshops (Hess, 2007), student-to-student learning partnerships (Chio & Fandt, 2007), and peer mentoring (Parker, Hall, Kram, in press). Interestingly, the use of student peers as evaluators also builds upon contemporary multisource assessment practices found in industry (e.g., Conway, Lombardo, & Sanders, 2001; Drexler, Beehr, & Stetz, 2001).

Peers as Evaluators

Peer evaluations are well-accepted as accurate sources of performance assessment in organizations and higher education (Bernardin, Dahmus, & Redmon, 1993; Fox, Ben-Nahum, & Yinon, 1989; Huber, Neale, & Northcraft, 1987; Schumacher, Scogin, Howland, & McGee, 1992). In fact, the bulk of the research on peer evaluations has focused on their psychometric characteristics and, as a whole, paints a rather positive picture of peer evaluations as assessment devices. First, in contrast to evaluations that stem from a single evaluator (e.g., supervisor, professor, etc.), peers, by virtue of their numbers, allow for the aggregation of evaluations. The pooling of peer evaluations is related to an increase in reliability and a partial removal of idiosyncratic biases associated with singular raters (Conway & Huffcutt, 1997; Greguras & Robie, 1998; Kenny & Berman, 1980). Much evidence also supports the validity of peer evaluations in organizations and higher education (Gardner, Scogin, Vipperman, & Varela, 1998; Schumacher, Scogin, Howland, & McGee, 1992; Harris & Schaubroeck, 1988). By virtue of their relationship with the target and the unique opportunities to observe colleagues and fellow students, peers possess a privileged viewpoint to evaluate performance. This is especially true for the measurement of those performance dimensions related to interpersonal skills and team-work effectiveness (Falchikov & Goldfinch, 2000; London & Smither, 1995). The unique perspective of peers has led Murphy and Cleveland (1995: 144) to claim that "all sources may have insights regarding an individual's strengths and weaknesses but peers may represent the single best informed source."

In addition to being recognized as reliable and valid evaluation tools, peer evaluations have also been found to have a significant impact on individual and group processes. In this respect, peer evaluations are not only viewed as a data collection exercise, but also as an intervention that can have a positive influence on evaluators and evaluatees. In other words, peer evaluations can take the roles of both dependent and independent variables. Much research has described the many ways in which involvement in a peer evaluation process can influence the attitudes and behaviors of students. These studies, however, typically address the issue by isolating the effect of peer evaluations on the same group in which they are used, by using either correlational (e.g., Saavedra & Kwun, 1993; Villanova, Bernardin, Dahmus, & Sims, 1993); one-shot experimental (e.g., Dominick, Reilly, & McGourty, 1997); or longitudinal designs (e.g., Brooks & Ammons, 2003; Dominick, Reilly, & McGourty, 1997; Drexler, Beehr, & Stetz, 2001; Druskat & Wolff, 1999; Erez, LePine, & Elms, 2002). The fact that these investigations focus on the impact of peer evaluations within their groups underscores the immediate value of these processes but also neglects the long-term consequence for students using peer evaluations when they switch from one group to another. In this study, we seek to observe the effects of using a common peer evaluation process over multiple semesters or, in other words, the effect of participating in peer evaluations in Group 1 (at Time 1) on a subsequent group (Group 2 at Time 2).

Our interest in the distal influence of peer evaluations stems from a facultywide implementation of a centralized peer evaluation system. Before outlining a theoretical framework that delineates the effects of peer evaluations over multiple semesters, we use the following section to describe the basic characteristics of the peer evaluation system on which this investigation is based.

Standardized Peer Evaluation System

The peer evaluation system (PES) was first developed in 2004 by the first author with support from the educational technology unit of the business school. The system, based on a web interface, is now in its 6th version. The PES is made available, on a voluntary basis, to all professors in the busi-

ness school and is currently being introduced to other faculties at the university. In the 2007–2008 academic year, the system was used by 35 professors in 94 classes. Although the voluntary use of the system leads to some variation in how the PES is used by professors, a specific, step-by-step set of guidelines is provided to those professors in order to obtain some consistency in its administration.

At the beginning of the semester, a document that details how to use the system is provided to all future student users. In addition to describing in detail the many steps of the process, how the data is handled, and the competency model used, this document also outlines general guidelines on how to provide feedback to others (i.e., focus on behavior, constructive feedback, etc). The students' experience of the system consists of two distinct phases. The first one pertains to the assessment of their peers. For this phase, students use a password-protected access to evaluate each other in both quantitative and qualitative formats. The quantitative assessment relies on a point allocation system (i.e., forced-distribution) to evaluate peer performance across four dimensions (i.e., cooperation, practical contribution, conceptual contribution, and work ethic; see Appendix A). Students also evaluate their peers on two general performance items on an absolute scale. Following the quantitative assessment, students are asked to comment on their peer's performance in narrative format. At various points of the evaluation process, students are made aware that their anonymity is protected. Two days before the evaluation deadline, those students who have not yet rated their peers are sent reminders by way of electronic mail.

The second phase of the PES allows students to obtain their feedback by way of the same password-protected portal. Once the performance evaluation has been collected, and the evaluation period has concluded, professors have the opportunity to review evaluations before releasing them to their students. Once the professor opens the evaluations for review, students are notified by electronic mail that the feedback from their peers is available. This feedback is anonymous in that the quantitative data is aggregated and the comments are listed without identifying information.

Several safeguards are built into the system to protect students. First and as previously mentioned, professors are encouraged to screen all comments generated by their students before they are allowed to view them. This safeguard ensures that students do not receive any inappropriate or damaging feedback from their peers. Second, because anonymity may be compromised in small groups, students in groups composed of three students and less are made aware of this fact, on screen, before they submit their evaluations. Finally, the data is stored on a protected server purchased specifically for this purpose.

In order to give readers an overview of how the system is used and of the perceptions of the professors that rely on it, we surveyed the 35 professors who used the system last year. Overall, a vast majority of them held positive opinions of the PES (92%). In their comments, they reported that it offered a great introduction to evaluation, one that will be useful for students when they enter organizations. Professors also recognized the value of the system as a developmental tool aimed at improving group-related competencies. Many commented on the logistical support that the system provides in terms of facilitating the collection and use of peer evaluations.

The survey also asked professors to comment on barriers that they may have encountered in adopting the PES. These barriers were of three types. First, some commented on the difficulty to master a new technology that also needs to be introduced effectively to students. A fair amount of effort is needed to fully comprehend how the system works. Second, a few would have liked more flexibility in how the system operates. For example, the use of a fixed competency model was frustrating to some. Finally, the use of the evaluation results for grade allocation created issues for some professors. They felt that experience with using the system over multiple semesters is needed to devise just the right translation scheme. Related to this point, the survey also queried professors on the methods that they used for this translation, if any. A majority of them (74%) did use the PES to adjust grades. More specifically, 59% used the information to weight the final project (between 5% and 50% of the project grade); 10% used it to adjust the overall course grade (between 3% and 10%); and 5% integrated the information into students' participation marks. More detailed results from this survey are provided in Appendix B.

Of interest in here is how the repeated exposure to such a system influences students. More specifically, our research focuses on one purported outcome of experience with PES: subsequent effectiveness in group work. In the following sections, we describe the effects of engaging with a PES over multiple semesters.

BEING EVALUATED BY OTHERS: ENHANCING EFFECTIVENESS IN GROUPS

Probably the most noteworthy implication of using peer evaluations is their impact on the motivation of individual group members. *Loafing behavior*, or the reduction in individual motivation, is often associated with collective work and is less likely to occur when individual tasks are visible or distinguishable from the effort put forth by others (George, 1992; Liden, Wayne, Jaworski, & Bennett, 2004). Inherent to peer evaluations is the identification of individual contributions, and meta-analytical results indicate that loafing behavior does fade when the contributions of individual members are evaluated (Karau & Williams, 1993). Interestingly, evidence suggests that the expectation that peer evaluations are to be conducted is as impactful on group members' behaviors as the evaluations themselves. Erez, LePine, and Elms (2002), for example, found that group members told that they would be involved in peer evaluations subsequently reported higher group sharing, cooperation, and team performance.

In the present study we seek to investigate the distal effects of participating in a centralized peer evaluation process. Under this perspective, it is expected that not only will the effects of expectations outlined in the previous paragraph be present every time the PES is used but that additional effects, specific to the repeated use of a common system, are also likely to take place. For one, repeated experience with a same evaluation process allows one to be familiar with, and pay attention to, valued behaviors. An exposure to the evaluation framework alone is often sufficient to improve effectiveness (Dominick, Reilly, & McGourty, 1997; Reilly, Smither, & Vasilopoulis, 1996; Smither, London, Vasilopoulos, Reilly, Millsap, & Salvemini, 1995). Experience with a structured evaluation process provides individuals with a useful framework for understanding their behaviors and focusing their efforts on those that are valued (Dominick et al., 1997). This knowledge provides an incentive to make contributions that are not only valuable to peers but that are also visible (Erez, LePine, Helms, 2002).

An additional effect of the peer evaluation system studied here is the reception of feedback about one's behavior. Feedback information can be used to diagnose one's weaknesses and determine directions for behavior change (e.g., Kanfer, 1990). A principal idea underlying the developmental feedback processes is that recipients use the feedback information to alter their behavior and, ultimately, improve their performance (Carver & Scheier, 1981; Kluger & DeNisi, 1996). Such an effect is especially likely if the recipient cares about the source of the information, the source is credible, and if the information cannot readily be denied, ignored, or rationalized away (Beach, 1990; London, 1995). Much research has shown that the provision of feedback does influence behavior within the same groups in which it was collected (e.g., Druskat & Wolff, 1999). In these situations, evaluators recognize the feedback that they provided has had an influence on the performance of the recipient. In the current study we go a step further and investigate if the impact of this feedback can also be detected in a new group setting, with new evaluators. Thus, we are interested in determining how experience with a PES in Group 1 influences one's behavior in Group 2. Based on the above rationale, it is expected that students will become more effective team members in their second utilization of the peer evaluation system.

Hypothesis 1: The effectiveness of students in groups will increase with repeated exposure to the peer evaluation system.

To further refine the predictive value of our tests, it is necessary to consider and account for possible alternative explanations to the expected improvements in effectiveness at Time 2. For example, it is possible that maturation accounts for this change. In other words, students may perform better at Time 2 not as a result of previous use of the system, but as a reflection of their greater experience and understanding of what is required to succeed as they progress through their academic careers. It is also possible that there are actual differences in the difficulty of the tasks comprising the group projects at Time 2, whereby all students involved in group projects at this time perform better, regardless of previous experience with the system. For example, although the course content was the same across Time 2, discretion afforded professors in the design and implementation of group projects may have translated into less difficult tasks, especially considering the more advanced nature of the students. In order to rule out these alternative explanations, it is posited that students who used the system at Time 2 will also perform better than students who have not used the peer evaluation system in a previous class but who are in the same class and at the same stage of their academic careers as the Time 2 students.

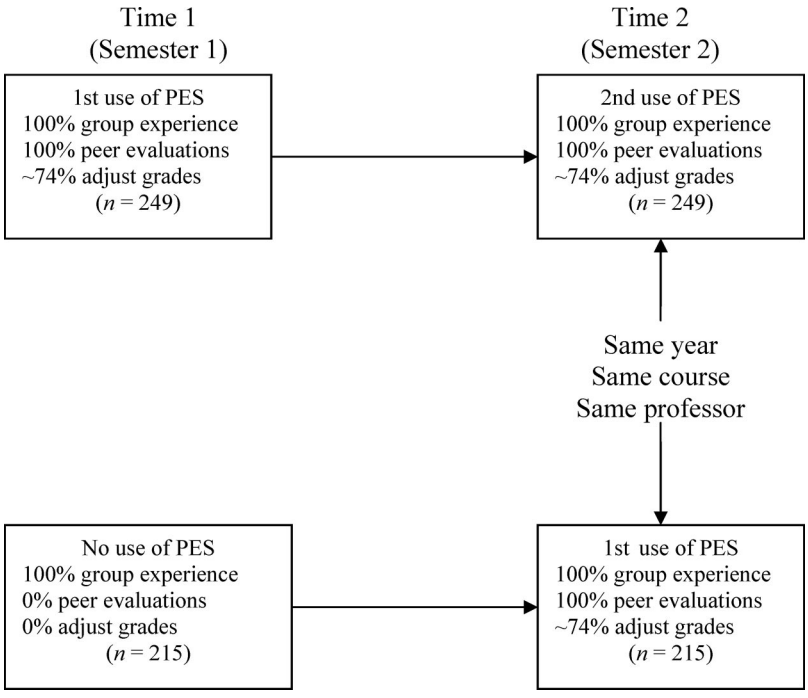

FIGURE 1
PES Study Design: Within-Subject With a Matched Comparison Group

Hypothesis 2: Students who have had prior exposure to the peer evaluation system will exhibit greater effectiveness in groups than students who have not had prior exposure to the system.

METHODOLOGY

The hypotheses were tested using a within-subject design with a matched comparison group (see Figure 1). The matched comparison group was relied upon to tease out maturation effects or the possibility that students improve their group-related skills naturally, as they progress through the curriculum.

Sample

The sample was composed of undergraduate business students drawn from a large public university in Canada. In 2007–2008, the system was used a total of 6,570 times (i.e., each representing a single student being evaluated by each one of his or her peers). For the purposes of our analyses, a very tight frame was used to select the final sample: Only first-year students who used the system in two consecutive semesters (only once per semester) were included. In their first semester, these students used the PES in a general introduction to business course. In their second semester, they used the system as part of an introduction to organizational behavior course. It is important to note that, in Canada, most undergraduate business programs are completed in 3 years and, in contrast with many U.S. programs, a strict sequence of business courses is followed in the first year and a half of studies. In our sample this sequence has the introduction to business course as a prerequisite to the organizational behavior course. While both of these courses are taught in multiple sections, their coordinated nature ensures consistency of content and evaluation across sections. Both courses included a group project that is similar in terms of content and scope. Furthermore, the possibility of self-selection into a particular section is quite low, given the tight structure for core courses whereby each cohort takes roughly the same courses, and sections fill up quickly given their compulsory and sequential nature. In this regard, we consider course assignment to be random or, at least, not systematically related to variables of interest in the study. The final data set consisted of 498 evaluations or 249 students in total.[1] In order to rule out maturation effects,

[1] Note that only 4.2% of students followed one of their peers from Time 1 to Time 2, reducing the probability of instrumentation bias.

we also selected a matched comparative sample composed of first time users ($N = 215$). Students in this sample participated in the evaluation system in exactly the same context as their experimental match. More specifically, these students used the PES in their 2nd semester as part of an organizational behavior course that was taught by the same professor who taught their experimental counterparts. When two or more matches were available, one was randomly selected.

Measures

Effectiveness in Groups

Students' effectiveness in groups was obtained from their group members' ratings. Specifically, students' performance ratings consisted of an average of their group members' evaluations of them on two general performance items using an absolute response scale. This was necessary because the data derived from forced-distribution scales violates the assumption of independence and does not allow for use of inferential statistics. These two items ("Overall, I am satisfied with the performance of this teammate in the group project" and "The performance of this teammate has been, in general, excellent"), were measured with a 7-point Likert scale (i.e., 1 = *Strongly Agree* and 7 = *Strongly Disagree*). The intraclass correlations ($ICC_{T1} = .54$, $p < .001$; $ICC_{T2} = .34$, $p < .01$; $ICC_{M} = .44$, $p < .001$) and reliability of the performance measure ($r_{xx\ T1} = .82$; $r_{xx\ T2} = .66$; $r_{xx\ M} = .75$) are similar to those found in other multisource performance assessments studies (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; Erez, LePine, & Elms, 2002), and justify aggregation across peers (Bliesse, 2000; James, 1982; Scherbaum & Ferreter, 2009). We therefore computed the average evaluations that each respondent received as an overall index of performance.

RESULTS

Paired sample *t* tests supported the first hypothesis that students received higher evaluations from their peers as a result of a second utilization of the system. The scores assessing peers' evaluations of students' performance increased from Time 1 ($M_{T1} = 5.81$, $SD = 1.13$) to Time 2 ($M_{T2} = 6.01$, $SD = .98$; $t(248) = -2.96$, $p < .01$; $d = .23$). A comparison between second time users and the matched sample also shows a significant difference, supporting the second hypothesis. The evaluation of students' performance was higher for students who were experienced users ($M_{T2} = 6.07$, $SD = .98$) than for those in the matched group ($M_{M} = 5.81$, $SD = 1.19$; $t(214) = 2.62$, $p < .01$; $d = .24$).

DISCUSSION

Although the development of group-related abilities is a key objective for the use of group work in higher education (Falchicov, 1986; Falchicov & Goldfinch, 2000), most business school curricula do not *explicitly* address the development of these skills (e.g., Navarro, 2008; Porter & McKibbin, 1988). Chen, Donahue, and Klimoski (2004: 29) stated that "the prevailing pedagogical priorities in higher education settings are inconsistent with the need for developing school-to-work programs that create a 'teamwork-capable' workforce." A recent AACSB report commented on the fact that, while interpersonal, leadership, and communication skills are highly important for success in business environments, skill development in these areas also represents the least effective components of business school curricula (AACSB, 2002). Assessment is at the core of any development effort, and the present study offers evidence that a centralized peer evaluation system can be an effective instrument to track the development of students and a powerful experience for the development of group-related behaviors.

Our results indicate that from one semester to the next the effectiveness of students in groups increases as a result of the experience with the PES. While research on peer evaluations has clearly demonstrated how they enable *concurrent* processes (i.e., benefits those same groups in which the evaluations are used), the results of the present study offer evidence of a yet uncovered trend in student development and more distal, and lasting, effects. Not only were students performing better at Time 2, but their observed performance at that time was also significantly higher than that of the matched sample. Students who participated in the PES for a second time knew, firsthand, on what dimensions they would be evaluated and, as a result of having previously received specific feedback on these same dimensions, adjusted their behavior accordingly. By virtue of its unique design, our study provides robust evidence of distal effects of peer evaluations. In contrast to previous studies on the effects of peer evaluations, our design relied on a different set of evaluators/feedback providers at Times 1 and 2. A comparison of group membership between these two periods confirmed that only a small percentage of the students in our samples experienced some overlap in group membership across the two time periods (4.2%). The independence between these two sets of raters

reduces the likelihood of response shift biases (i.e., beta and gamma changes; Atwater, Brett, & Waldman, 2003) and, thus, provides strong evidence of the impact of the PES on behavior (i.e., alpha change).

FUTURE RESEARCH AND LIMITATIONS

Our findings raise many interesting questions as to the effect of peer evaluations on students. First, this study focuses on trends over two consecutive semesters. As stated earlier, it would be interesting to investigate the impact of this exercise over a longer period of time and establish how the improvements detected in this study evolve over multiple semesters or over repeated uses of the system. Some research has demonstrated that the effects of feedback are not linear and, rather, fluctuate over time (e.g., Druskat & Wolff, 1999; Reilly, Smither, & Vasilopoulos, 1996). In addition, it would be worthwhile to further probe the boundaries of the detected trends and investigate whether they remain after graduation. Gardner, Scogin, Vipperman, and Varela (1998), for example, found that peer evaluations collected during police officer training significantly correlated with turnover 6 years after graduation from the academy. It would be interesting to investigate the extent to which transfer occurs and if experienced PES users also perform better in the workplace.

Second, our study remains silent as to the specific source of the effects detected. The choice to use the PES was left to the instructors and the lack of random assignment prevents us from teasing out the role of the professor in the PES effect. Questions also remain as to which component(s) of students' educational experience actually contributed to the detected increase in performance. The experience of peer evaluations is multifaceted and complex; students assume the roles of both evaluatees and evaluators. Moreover, our study does not tease out the discrete contributions of each element of the process to behavior change (e.g., expectation of being evaluated, evaluating others, receiving feedback, increased familiarity with the evaluation criteria, being graded based on group effectiveness, etc.). Given the variation in how the PES was administered (see Appendix B), an important follow-up to this study would be to investigate the consequences of these variations for users. For example, performance appraisal research has demonstrated that the purpose of appraisal greatly influences reactions to these systems (Jawahar, & Williams, 1997). It would be interesting to investigate whether students' improvement in performance is affected by whether the PES is used for evaluative or simply for developmental purposes.

Third, by focusing only on a single dependent variable, we took a rather narrow perspective of the impact of participating in the PES. It is very probable that experience with the PES has additional consequences for participants, above and beyond those detected in this study. For example, a recent study by Jassawalla, Sashittal, and Malshe (2009) delineated the construct of loafing behavior in student teams in two separate facets: performing poor quality work and engaging in distractive and disruptive behaviors. To the extent that peer evaluations are linked to a reduction in loafing behaviors, it would be worthwhile to investigate whether a PES operates differently on these distinct facets of loafing.

Also, future research on this topic should be particularly attentive to the impact of peer evaluations on poor performers or within dysfunctional group situations. While this represents a rather small subset of users (in our sample, only 7.8% of students received an evaluation below the midpoint of the scale), it is a critical one. Peer evaluations may actually hinder group dynamics and individual effectiveness in these situations (Bacon, Stewart, & Silver, 1999). A better understanding of the influence of peer evaluations in difficult circumstances may enable the development of appropriate interventions and a better learning experience for those students that need it the most. Finally, future research exploring the impact of a PES on performance improvements should consider the use of a more complex operationalization of performance in order to tease out, with more precision, the effects of such a complex and broad intervention.

PRACTICAL IMPLICATIONS

A centralized PES represents a valid tool to track a set of competencies that, by most accounts, is critical in business education. A recent report on the future of higher education stressed the need for the measurement and reporting of important learning outcomes (U.S. Department of Education, 2006). With increasing pressure from governments, external stakeholders, and accreditation bodies to establish learning goals and, more important, to assess whether learning actually occurs, interest in PES is likely to grow.

As a learning tool, such a system represents the cornerstone for a more elaborate initiative aimed at developing group-related skills. We should

point out that the effects observed in this study probably underestimate the potential value of peer evaluation systems because the educational support currently provided to the PES has been minimal. The integration of the PES with various educational components would undoubtedly increase its impact and allow for the development of positive norms in regard to group work (Saavedra & Kwun, 1993). Examples of such components include:

- *Provision of courses or workshops on group-related skills.* Recently, Chen, Donahue, and Klimoski (2004) presented the successful development, implementation, and evaluation of an undergraduate-level course aimed at developing group-work competencies. Training programs have also been found to help individuals with evaluation duties (Bernardin, Buckley, Tyler, & Wiese, 2000; Woehr & Huffcutt, 1994). Formal training on these soft skills is a natural complement to the PES.
- *Emphasis on individual accountability.* The literature on feedback is quite clear as to the role of individual accountability to improve its impact on individuals (London, Smither, & Adsit, 1997). While making peer evaluations public represents the most direct method to increase students' accountability for their behaviors in groups and their responses to the feedback, such a display of evaluations would likely be met with resistance due to privacy issues. A more appropriate alternative is to make peer evaluation transcripts available to those students who desire it. These transcripts would be very similar to a typical 360-feedback report in that they would provide individual performance information aggregated across peers for the different courses in which the system was used. Such a transcript would be valuable for job applicants, and its availability could motivate those that would otherwise neglect group-related behaviors. Also, some formal or informal acknowledgment of student improvement over time or a recognition program that rewards excellence in groups could also be provided by the administration as a way to raise accountability.

CONCLUSION

We offer supporting evidence for the benefits of using a centralized PES in an undergraduate business curriculum. Given the ubiquity and importance of group work in organizations, increased effectiveness in groups is key for success at work and, as a result, represents relevant educational objectives for business schools. Nielsen and Halfhill (2007) argued the need for quantifying the "softer side" of business education in order to better examine how effective we are in teaching these competencies and obtain the data necessary to demonstrate their value. Group-related competencies represent educational objectives for business schools because they are "transportable" across settings (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Cassidy, 2006). Requiring minimal cost for implementation and administration, a centralized PES can offer a real competitive advantage to business schools and their students.

APPENDIX A
Performance Dimensions

COOPERATION

1
- Actively participating in meetings
- Communicating within the group
- Co-operating within the group
- Assisting team-mates when needed
- Volunteering for tasks

CONCEPTUAL CONTRIBUTION

2
- Researching and gathering information
- Quality of individual contribution
- Suggesting ideas
- Tying ideas together
- Identifying difficulties
- Identifying effective approaches

PRACTICAL CONTRIBUTION

3
- Writing of the report(s)
- Reviewing others' report(s) or section(s)
- Providing constructive feedback on the report(s) or the presentation
- Contributing to the organization of the work
- Contributing to the preparation of presentation(s) (if appropriate)

WORK ETHIC

4
- Displaying a positive attitude
- Respecting team-mates
- Respecting commitments
- Respecting deadlines
- Respecting team-mates' idea

APPENDIX B
Results From a Survey of Professors Who Used the PES

Overall reactions to the PES	92% of professors reported favorable attitudes toward using the system. They found that it: • Provides a great introduction to the real-world peer/colleague evaluation system in place in corporations today. • Is easier to handle and more information rich than the paper and pencil format. • Provides an effective means in large classes for students to report on each other's performance. • Is simple to use, has a user friendly interface, and asks the right questions. • Gives students clear feedback on their performance. • Ensures anonymity to evaluators that helps in the delivery of feedback.
Perceived barriers to using the PES	Some were initially resistant to use the system because they see it as more work for them. Some felt uncomfortable with using on-line tools. A few felt that changes in group membership lead to more work in terms of updating the system. Some felt constrained with the use of prespecified criteria/ratings that cannot be modified. A few felt that the optimal use the system for grade adjustment involved some trial and error to strike the right balance.
Use of the PES for grade adjustment	74% of professors reported using the evaluations to adjust group grades. More specifically: • 59% reported using the evaluations to yield a weight for the final project (between 5 and 50% of group project). • 10% reported using the evaluations to adjust overall marks (between 3 and 10% of the overall grade). • 5% reported using the evaluations to adjust participation marks.
Method of group assignment	Almost 96% of professors allowed students to create their own groups.

REFERENCES

AACSB. 2002. ***Management education at risk: Report of the Management Education Task Force to the AACSB.***

Antonioni, D. 1994. The effects of feedback accountability on upward appraisal ratings. ***Personnel Psychology,*** 47: 349–356.

Atwater, L. E., Brett, J. F., & Waldman, D. A. 2003. Understanding the benefits and risks of multi-source feedback. In S. Murphy & R. Riggio (Eds.), ***The future of leadership development:*** 89-106. Mahwah, NJ: Erlbaum.

Bacon, D. R., Stewart, K. A., & Silver, W. S. 1999. Lesson from the best and worst student team experience: How a teacher can make the difference. ***Journal of Management Education,*** 23: 467–488.

Baldwin, T. T., Bedell, M. D., & Johnson, J. L. 1997. The social fabric of a team-based M.B.A. program: Network effects on student satisfaction and performance. ***Academy of Management Journal,*** 40: 1369–1397.

Beach, L. R. 1990. ***Image theory: Decision making in personal and organizational contexts.*** Chichester, UK: Wiley.

Bernardin, H. J., Buckley, M. R., Tyler, C. L., & Wiese, D. S. 2000. A consideration of strategies in rater training. In J. Ferris (Ed.), ***Research in personnel and human resource management,*** 18: 221–274. San Francisco: Jossey-Bass.

Bernardin, J. H., Dahmus, S. A., & Redmon, G. 1993. Attitudes of first line supervisors toward subordinate appraisals. ***Human Resource Management,*** 32: 315–324.

Bliese, P. D. 2000. Within-group agreement, non-independence, and reliability: Implications or data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), ***Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions:*** 349–381. San Francisco: Jossey-Bass.

Brooks, C. M., & Ammons, J. L. 2003. Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessments. ***Journal of Education for Business,*** 78: 268–272.

Brown, S., Rust, C., & Gibbs, G. 1994. ***Strategies for diversifying assessment in higher education.*** Oxford, England: Oxford Center for Staff Development.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. 1995. Defining competencies and establishing team training requirements. In R. A. Guzzo & E. Salas (Eds.), ***Team effectiveness and decision making in organizations:*** 333–381. San Francisco: Jossey-Bass.

Cappelli, P., & Rogovsky, N. 1994. New work systems and skill requirements. ***International Labour Review,*** 133: 205–220.

Carver, C. S., & Scheier, M. F. 1981. ***Attention and self regulation: A control theory to human behavior.*** New York: Springer-Verlag.

Cassidy, S. 2006. Developing employability skills: Peer assessment in higher education. ***Education + Training,*** 48: 508–517.

Chen, G., Donahue, L. M., & Klimoski, R. J. 2004. Training undergraduates to work in organizational teams. ***Academy of Management Learning and Education,*** 3: 27–40.

Chen, Y., & Lou, H. 2006. Students' perceptions of peer evaluation: An expectancy perspective. ***The Journal of Education for Business,*** 79: 275–282.

Chio, V. C. M., & Fandt, P. M. 2007. Photovoice in the diversity classroom: Engagement, voice, and the "Eye/I" of the Camera. ***Journal of Management Education,*** 31: 484–504.

Conway, J. M., & Huffcutt, A. I. 1997. Psychometric properties of multi-source performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. ***Human Performance,*** 10: 331–360.

Conway, J. M., Lombardo, K., & Sanders, K. C. 2001. A meta-analysis of incremental validity and nomological networks

for subordinate and peer ratings. ***Human Performance***, 14(4): 267–303.

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. 2004. A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. ***Journal of Applied Psychology,*** 89: 1035–1056.

Dominick, P.G., Reilly, R. R., & McGourty, J. 1997. The effects of peer feedback on team member behavior. ***Group and Organization Management,*** 22: 508–520.

Drexler, J. A., Beehr, T. A., & Stetz, T. A. 2001. Peer appraisals: Differentiation of individual performance on group tasks. ***Human Resource Management,*** 40(4): 333.

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.

Erez, A., LePine, J. A., & Elms, H. 2002. Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. ***Personnel Psychology,*** 55: 929–948.

Falchicov, N. 1986. Product comparisons and process benefits of collaborative self and peer group assessments. ***Assessment and Evaluation in Higher Education,*** 11: 146–166.

Falchicov, N., & Goldfinch, J. 2000. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. ***Review of Educational Research,*** 70: 287–322.

Fellenz, M. R. 2006. Toward fairness in assessing student groupwork: A protocol for peer evaluation of individual contributions. ***Journal of Management Education,*** 30: 570–591.

Fox, S., Ben-Nahum, Z., & Yinon, Y. 1989. Perceived similarity and accuracy of peer ratings. ***Journal of Applied Psychology,*** 74: 781–786.

Gardner, J., Scogin, F., Vipperman, R., & Varela, J. G. 1998. The predictive validity of peer assessment in law enforcement: A 6 year follow-up. ***Behavioral Sciences and the Law,*** 16: 473–478.

George, J. M. 1992. Extrinsic and intrinsic origins of perceived social loafing in organizations. ***Academy of Management Journal,*** 35(1): 191–202.

Greguras, G. J., & Robie, C. 1998. A new look at within-source interrater reliability of 360-degree feedback ratings. ***Journal of Applied Psychology,*** 83: 960–968.

Guzzo, R. A., & Shea, G. P. 1992. Group performance and intergroup relations in organizations. In M. D. Dunnette & L. M. Hough (Eds.), ***Handbook of industrial and organizational psychology,*** 3: 269–313.

Harkins, S. G., & Jackson, J. M. 1985. The role of evaluation in eliminating social loafing. ***Personality and Social Psychology Bulletin,*** 11(4): 457–465.

Harris, M., & Schaubroeck, J. 1988. A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. ***Personnel Psychology,*** 41: 43–62.

Hess, P. W. 2007. Enhancing leadership skill development by creating practice/feedback opportunities in the classroom. ***Journal of Management Education,*** 31(2): 195–213.

Huber, V. A., Neale, M. A., & Northcraft, G. B. 1987. Judgment by heuristics: The effects of performance standards and rater and ratee characteristics on performance-related judgments. ***Organizational Behavior and Human Decision Processes,*** 40: 149–169.

James, L. R. 1982. Aggregation bias in estimates of perceptual agreement. ***Journal of Applied Psychology,*** 67: 219–229.

Jassawalla, A., Sashittal, H., & Malshe, A. 2009. Students' perceptions of social loafing, its antecedents and consequences in undergraduate business classroom teams. ***Academy of Management Learning and Education,*** 8: 42–54.

Jawahar, I. M., & Williams, C. R. 1997. Where all the children are above average: The performance appraisal purpose effect. ***Personnel Psychology,*** 50: 905–925.

Jelley, R. B., & Goffin, R. D. 2001. Can performance-feedback accuracy be improved? Effects of rater priming and rating-scale format on rating accuracy. ***Journal of Applied Psychology,*** 86(1): 134–144.

Johnson, D. W., & Johnson, R. T. 1987. ***Learning together and alone: Cooperative, competitive, and individualistic.*** Englewood Cliffs, NJ: Prentice Hall.

Kanfer, R. 1990. Motivation theory and industrial/organizational psychology. In M. D. Dunnette & L. Hough (Eds.), ***Handbook of industrial and organizational psychology:*** 75–170. Palo Alto, CA: Consulting Psychologists Press.

Karau, S. J., & Williams, K. D. 1993. Social loafing: A meta-analytic review and theoretical integration. ***Journal of Personality and Social Psychology,*** 45(4): 681–706.

Kenny, D. A., & Berman, J. S. 1980. Statistical approaches to the correction of correlational bias. ***Psychological Bulletin,*** 88: 288–295.

Kluger, A. N., & DeNisi, A. 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis and a preliminary feedback intervention theory. ***Psychological Bulletin,*** 119: 254–284.

Li, L. K. 2001. Some refinements in peer assessment of group projects. ***Assessment and Evaluation in Higher Education,*** 26: 6–18.

Liden, R. C., Wayne, S. J., Jaworski, R. A., & Bennett, N. 2004. Social loafing: A field investigation. ***Journal of Management,*** 30: 285–304.

London, M. 1995. Giving feedback: Source-centered antecedents and consequences of constructive and destructive feedback. ***Human Resource Management Review,*** 5: 159–188.

London, M., & Smither, J. W. 1995. Can multi-source feedback change perception of goal accomplishment, self-evaluation and performance-related outcomes? Theory-based applications and directions for research. ***Personnel Psychology,*** 48: 803.

London, M., Smither, J. W., & Adsit, D. J. 1997. Accountability: The Achilles Heel of multi-source feedback. ***Group and Organization Management,*** 22(2): 162–184.

Mello, J. 1993. Improving individual member accountability in small group work settings. ***Journal of Management Education,*** 17(2): 253–259.

Murphy, K. R., & Cleveland, J. N. 1995. ***Performance appraisal: An organizational perspective***. Boston: Allyn & Bacon.

Navarro, P. 2008. The MBA core curricula of top-ranked US business schools: A study in failure? ***Academy of Management Learning and Education,*** 7: 108–123.

Nielsen, T. M., & Halfhill, T. R. 2007. Quantifying the "softer side" of management education: An example using teamwork competencies. ***Journal of Management Education,*** 31: 64–80.

Parker, P., Hall, D. T., & Kram, K. E. In press. Peer coaching: A

relational process for accelerating career learning. ***Academy of Management Learning and Education.***

Peiperl, M. A., & Trevelyan, R. 1997. Predictors of performance at business school and beyond–Demographic factors and the contrast between individual and group outcomes. ***Journal of Management Development,*** 16: 354–367.

Porter, L. W., & McKibbin, L. E. 1988. ***Management education and development: Drift or thrust into the 21st century?*** New York: McGraw Hill.

Reilly, R., Smither, J., & Vasilopoulos, N. 1996. A longitudinal study of upward feedback. ***Personnel Psychology,*** 49: 599–612.

Saavedra, R., & Kwun, S. K. 1993. Peer evaluation in self-managing workgroups. ***Journal of Applied Psychology,*** 78(3): 450–462.

Scherbaum, C. A., & Ferreter, J. M. 2009. Estimating statistical power and required sample sizes for organizational research using multilevel modeling. ***Organizational Research Methods,*** 12(2): 347–367.

Schumacher, J. E., Scogin, F., Howland, K., & McGee, J. 1992. The relation of peer assessment to future law enforcement performance. ***Criminal Justice and Behavior,*** 19: 286–293.

Smither, J. W., London, M., Vasilopoulis, N. L., Reilly, R. R., Millsap, R. E., & Salvemini, N. 1995. The examination of the effects of an upward feedback program over time. ***Personnel Psychology,*** 48(1): 1–32.

Topping, K. 1998. Peer assessment between students in colleges and universities. ***Review of Educational Research,*** 68: 249–276.

U.S. Department of Education. 2006. ***A test of leadership: Charting the future of U.S. higher education.*** Washington, DC: Author.

Villanova, P., Bernardin, H. J., Dahmus, S. A., & Sims, R. L. 1993. Rater leniency and performance appraisal discomfort. ***Educational and Psychological Measurement,*** 53: 789–799.

Woehr, D. J., & Huffcutt, A. I. 1994. Rater training for performance appraisal: A quantitative review. ***Journal of Occupational and Organizational Psychology,*** 67: 189–197.

Zedeck, S., & Goldstein, I. L. 2000. The relationship between I/O psychology and public policy: A commentary. In J. F. Kehoe (Ed.), ***Managing selection in changing organizations:*** 371–396. San Francisco: Jossey-Bass.

Stéphane Brutus is the chair of the Management Department at the John Molson School of Business in Montreal, Quebec. He obtained his PhD in industrial-organizational psychology from Bowling Green State University in 1995 and completed a postdoctoral fellowship at the Center of Creative Leadership. Stéphane's research focuses mainly on feedback processes in organizations.

Magda B. L. Donia is a doctoral candidate in organizational behavior at the John Molson School of Business. She also holds an MSc in administration from the John Molson School of Business. Magda's research interests include the transferability of skills developed during university to work, organizational citizenship behaviors, citizenship motives, and role definitions.