



## EMPIRICAL RESEARCH

# What Students Think and Do in Classroom Teams When Peer Evaluations Are Highly Consequential: A Two-Stage Study

Avan R. Jassawalla<sup>†</sup>

*Professor of Management, School of Business, State University of New York at Geneseo, Geneseo, NY, 14454, e-mail: jassawal@geneseo.edu*

Hemant C. Sashittal

*Professor of Marketing, School of Business, St. John Fisher College, Rochester, NY, 14618, e-mail: sashi@sjfc.edu*

## ABSTRACT

The article presents findings from a two-stage study that examined student perceptions of peer evaluations (PEs) conducted in undergraduate business classroom teams. In stage 1, we used qualitative research to identify constructs focal in students' PE-related cognitive schemas and developed grounded measurement scales and hypotheses about their relationships. Then, we implemented PEs in 17 sections of undergraduate business courses taught over seven semesters. The PEs were highly consequential; i.e., they entirely determined the grade each student received on her/his team project. At the end of each semester, we surveyed student perceptions and behaviors using measurement scales we developed after stage 1 of the study. We find that the knowledge of impending PEs leads students to exercise a great deal of care in terms of what they say and do while working with others. The higher levels of care trigger both impression management behaviors and perceptions that others are contributing more. The perceptions that others are contributing more seem instrumental in shaping students' decision to contribute more themselves. Implications for instructors and future research are discussed.

***Subject Areas: Classroom Teams, Peer Evaluations.***

## INTRODUCTION

Many business school instructors assign students to classroom teams and require them to collaborate on class-related projects (Brutus & Donia, 2010). Many also require students to complete peer evaluations (PEs) because they help reduce social loafing (Falchikov & Goldfinch, 2000) and serve as reliable and valid ways of assessing individual contributions to teamwork (Murphy & Cleveland, 1995). While legitimate concerns are raised about their efficacy (e.g., Fellenz, 2006; see Ohland et al., 2012 for detailed discussion), scholars are convinced that PEs

---

<sup>†</sup>Corresponding author.

promote teamwork in classroom teams (e.g., Brooks & Ammons, 2003; Brutus & Donia, 2010; Brutus, Donia, & Ronen, 2013).

Despite the advocacy, an opportunity for new thinking and research into PEs in classroom teams emerges as a result of two important gaps in the literature. First, the bulk of current insights reflect instructors' and scholars' perspectives; the students' perspective is poorly represented. The literature is silent when it comes to explaining what students think about and do as a result of PEs in classroom teams. Theory development is severely inhibited if the voices of the key constituency for whom PEs are designed are unrepresented in the literature. Second, the logic of PEs in classroom teams rests in the premises that participants will not contribute equally to teamwork; hence, a common grade unfairly rewards some and unfairly punishes others. Current thinking suggests there is merit to: (a) mid-semester PEs that provide formative feedback to team members and enable midcourse correction, and (b) end-of-semester PEs that provide summative feedback to instructors to ensure fairness in grading. However, the literature remains vague on the nature of consequences they must optimally produce for students who receive positive or negative PEs. In particular, the effects of highly consequential PEs that produce a heavy stake in teamwork-related outcomes remain unknown.

The article reports findings from a two-stage study we conducted that aimed to address the gap created by the unreported students' perspectives and the absence of research into highly consequential PEs. The first stage was devoted to exploring students' perspectives into PEs in classroom teams and deriving a grounded conceptual model, measurement scales, and hypotheses. The second stage was devoted to surveys of students who had participated in highly consequential PEs and to testing the hypotheses that emerged from stage 1 of the study.

The article begins with the case that students' perspective into PEs and their thinking and actions as a result of highly consequential PEs are worthy of study (i.e., when they shape a 100% of their team project grade). Then, we describe our two-stage study and discuss findings and implications. In so doing, we aim to make four key contributions to current thinking about PEs in classroom teams. First, we discuss evidence that favors implementation of highly consequential PEs; we learn that students overwhelmingly prefer them over other types of PEs that produce a low or no grade consequence. Second, we present evidence to suggest that PEs focus students' attention toward other members' behaviors. When required to complete PEs, team members adopt a *wait and see* approach; they calibrate their own behaviors based on their perceptions of others' behaviors. We argue in favor of making explicit and forewarning students about the tendency to *wait and see*, and the perils of triggering the suboptimal cycle of inaction as a result of *let's see what others are doing first*. Third, when PEs are used, social loafing does not appear to go away; it merely changes form. That is, instead of slacking off, students resort to manipulative behaviors and impression management to garner favorable PEs without expending effort toward teamwork. Fourth, when highly consequential PEs are implemented, we learn that students are unjustifiably optimistic or pessimistic about their team's performance. We argue in favor of new research that can help students better predict teamwork-related outcomes, and help them temper their tendency toward unjustifiable optimism or pessimism. Our overriding intent is to channel underrepresented student perspectives into PEs,

identify new constructs focal in their experiences, present new measurement scales, and discuss confirmatory evidence to trigger new thinking and research. Causal analysis and experimental designs that assess the impact on students' perceptions and behaviors as a result of varying the consequences of PEs are left to future research.

### **A Theory-Derived Case for the Study**

Our purpose here is to present a theory-derived case for the study of: (a) students' perspectives into PEs in classroom teams and (b) highly consequential PEs—versus the attempt to replicate discussions about PEs that have occurred elsewhere. We argue that both are meritorious, without precedent, deserving of empirical attention, and call for a two-stage study; i.e., the first stage for grounded hypotheses generation via qualitative research followed by the second stage for hypotheses testing based on survey data.

First, the gap in current thinking that strongly implicates an initial exploration of students' perspectives into PEs relates to the following. The literature is rich with insights into what PEs can do for instructors based on theory-derived conceptualizations and scales (see valuable contributions of Baker, 2008; Bowes-Sperry, Kidder, Foley, & Chelte, 2005; Brutus & Donia, 2010; Chen & Lou, 2004; Fellenz, 2006; Murphy & Cleveland, 1995; Paswan & Gollakota, 2004; Pfaff & Huddleston, 2003). For instance, the merits of PEs in classroom teams are widely discussed (e.g., Baker, 2008; Brutus & Donia, 2010; Fellenz, 2006; Murphy & Cleveland, 1995). Their role in assessing individual performance (e.g., Chen & Lou, 2004; Fellenz, 2006) and usefulness to instructors in assessment and grading are well documented (e.g., Paswan & Gollakota 2004; Pfaff & Huddleston 2003; Verzat, Byrne, & Fayolle, 2009). Similarly, how PEs prevent problems of free riding and aid student learning (Brutus & Donia, 2010), allow allocation of fair grades (Fellenz, 2006), and provide developmental feedback are issues that have invited inquiry (Dominick, Reilly, & McGourty, 1997; Mayo, Kakarika, Pastor, & Brutus, 2012). These writings have shaped the design of PE instruments suitable for classroom teams (see CATME developed by Loughry, Ohland & Moore, 2007; and CATME-B by Ohland et al., 2012).

Nearly all empirical evidence related to PEs in classroom teams: (a) currently speaks to scholars' and instructors' concerns for reducing social loafing and improving assessment (Falchikov & Goldfinch, 2000), and (b) is linked to survey or experimental research with literature-derived scales (see Brooks & Ammons, 2003 and Bowes-Sperry et al., 2005, for survey research; see Brutus & Donia, 2010 and Chen & Lou, 2004), for evidence from experimental research). Even the widely cited PE instrument (CATME) was initially designed using impressionistic data gathered from scholars and teachers and not from students; only later was it tested using sample of students (see Ohland et al., 2012). Hence, current conceptions, hypotheses, and measurement scales are not reflective of students' perspectives; little is known about how and why PEs shape their thinking and actions.

Second, the case for a study of highly consequential PEs rests in the following arguments. Scholars agree that PEs shape thinking and behavior. That is, they agree that *a priori* knowledge of criteria for PEs: (a) shapes student behaviors

(Brooks & Ammons, 2003; Falchikov & Goldfinch, 2000); (b) motivates collaborative behaviors (Bowes-Sperry et al., 2005); (c) promotes hard work (Brutus & Donia, 2010); and (d) allows participants to notice specific contributions made by others and to provide high-quality information to instructors (Chen & Lou, 2004). Scholars also agree that PEs do not deter social loafing—one of the primary reasons for their use (Chen & Lou, 2004; Fellenz, 2006)—unless they produce significant consequences for slackers (Clinebell & Stecher, 2002). Hence, scholars agree that PEs must produce heavy consequences, particularly grade-related consequences, to shape behaviors (e.g., Bowes-Sperry et al., 2005; Clinebell & Stecher, 2002; Fellenz, 2006). For instance, based on the knowledge that PEs will contribute to just 10% of their final course grade, one study shows that students expend more effort and cooperate more with others (see Erez, LePine, & Elms, 2002). The study demonstrates that grade-consequential PEs matter (see Psenicka, Vendemia, & Kos, 2013). Chen and Lou (2004, p. 281) write: “students perceive determining peers’ grades as the most attractive outcome for the use of peer evaluation.”

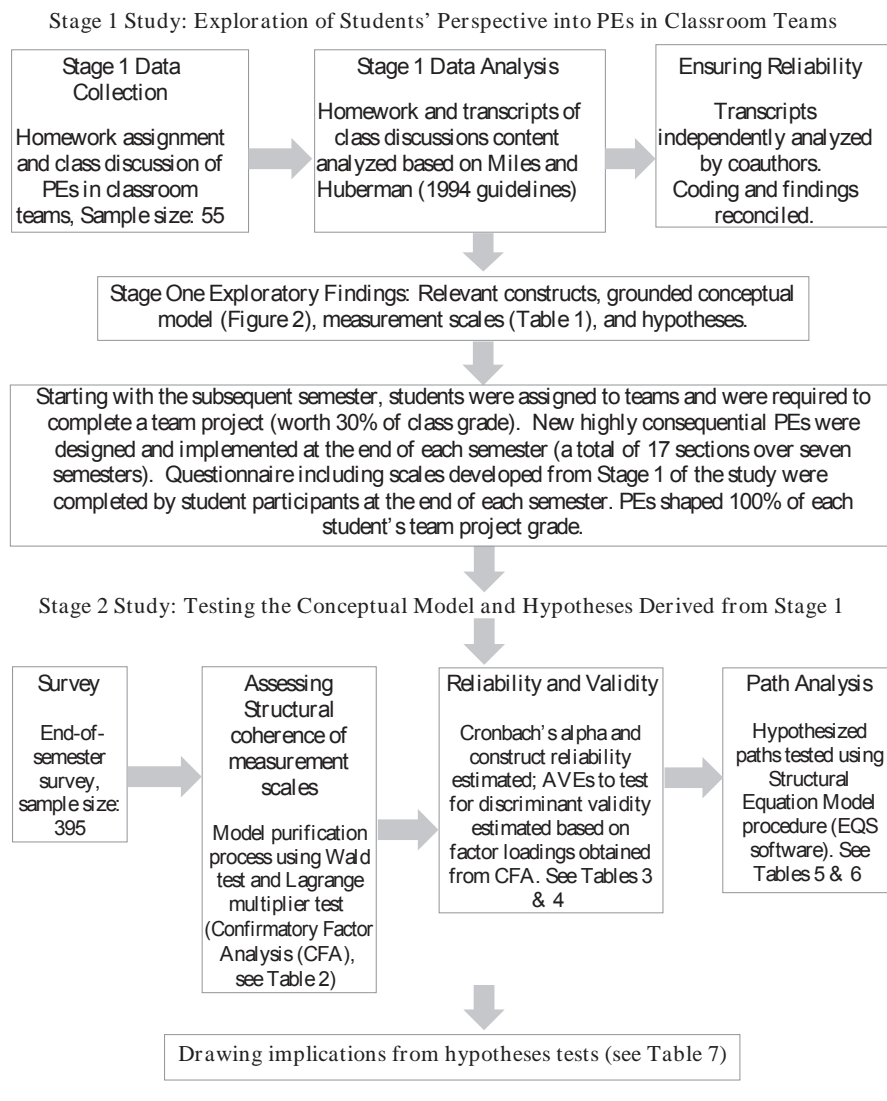
The following literature illustrates the gap in current thinking that suggests the need for a study of highly consequential PEs. Scholars agree that PEs *should* produce a consequence in general (Erez et al., 2002), and a grade consequence in particular (Chen & Lou, 2004). Yet, there is no evidence in the literature about how students think and respond when the PEs are highly consequential; i.e., when they have the potential to shape 100% of their individual grade on the team assignment. A potentially useful understanding of how and why PEs work the way they do in classroom teams can emerge when a lot is at stake and the students’ team project grade is determined entirely by the feedback-generating, team-building, behavior-modifying mechanism.

## STAGE 1 EXPLORATORY STUDY

Figure 1 illustrates the flow chart of the two-stage study we conducted after a review of the literature. Briefly, the gap in research precluded a one-shot, theory-derived hypotheses testing study. Hence, we began with an exploration of the underreported students’ perspectives into PEs with the intent of developing a grounded conceptual model, measurement scales, and hypotheses. After the first stage was completed, we designed and implemented highly consequential PEs in undergraduate business classroom teams. At the end of each semester, we surveyed students using the measurement scales we had developed at the end of stage 1. The second stage focused on analyzing survey data collected via end-of-semester questionnaires administered to students in 17 sections over seven semesters.

### Stage 1 Data Collection

As an initiating step, one of the coauthors polled instructors teaching in an undergraduate business program offered by a B-School located in Northeastern United States to learn that 90% of sections require students to work in teams and conduct PEs (program accredited by AACSB-International). Given the high likelihood that students had conducted PEs at least once, the coauthor assigned the following questions for homework in two sections of undergraduate *Organizational*

**Figure 1:** A flow chart of the research methodology.

*Behavior* classes: (a) how did the knowledge of impending PEs and your experience with using PEs shape perceptions and behaviors in your team, and (b) what are, and should be, some of the important outcomes that PEs should produce for participants? Out of the 63 students enrolled in the two classes, 55 turned in typed homework papers on the assigned questions because they had participated in at least one PE process in a previous classroom team (33 males, 22 females; mostly accounting and business administration majors in their junior year). All participants indicated that they had *a priori* knowledge that PEs would occur in their teams.

On the day the assignment was due, students were instructed to discuss their responses in small groups, report the results to the class, and engage in class discussion. The instructor sought clarifications, asked for specific examples, and recorded the key points from the class discussion on the white board.

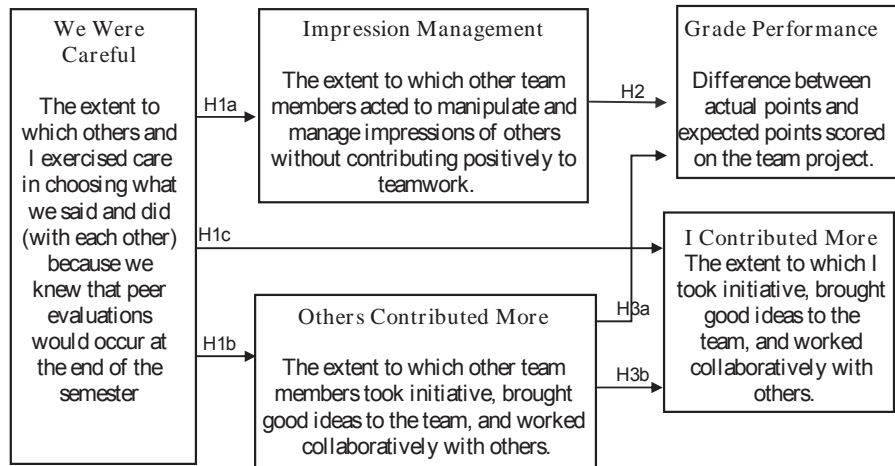
### **Stage 1 Data Analysis**

The students' homework responses, the transcript of whiteboard notes, and instructor's notes were content-analyzed. To ensure intercoder reliability, each coauthor conducted independent content analysis adhering closely to the guidelines of Miles and Huberman (1994). Briefly, we began by constructing a data matrix, each line devoted to a student and each column devoted to specific issues which students had identified in their homework and later during class discussions. Rooted in the data matrix and the transcribed class notes, we began by structural theorizing; i.e., we identified key constructs in student voices and hypothesized about their interrelationships.

Then, the two coauthors met to reconcile independently derived findings; after accounting for differences in terminologies, we found over 95% concurrence in the constructs we had identified. The independently derived findings concurred that a wide variance existed in: (a) the specificity of criteria for PEs; i.e., some instructors required evaluation of others' contribution on specific, predetermined criteria, while others required a global evaluation, and (b) the grade consequences of PE; i.e., in some classes, they produced a negligible grade impact (less than 5%), in others they produced mid-level grade impact (up to 50%). Both evaluations highlighted that students overwhelmingly preferred highly consequential PEs. There was a strong concordance around the notion that students' responses were clustered into concepts associated with: (a) the high level of carefulness triggered by the knowledge of impending PEs, (b) the immediate focus on what others were doing, (c) the calibration of one's contribution based on the contribution of others, and (d) the concern with grades, i.e., whether they would get the grade they individually deserved. To further insure reliability (i.e., data-inference concordance), we supported each of these constructs with excerpts of student responses. Then, we jointly derived a conceptual model and hypotheses.

### **Stage 1 Exploratory Findings and Conceptual Model**

Figure 2 illustrates the conceptual model derived from student voices. It identifies key constructs and makes explicit our learning about convergence, i.e., the directionality of linkages we hypothesize as significant (Cavusgil, Deligonul, & Griffith, 2008). The measurement scales included in Table 1 reflect student voices and were derived from the processes advocated by Anderson and Gerbing (1988), Churchill (1979), and Hinkin (1995). To address concerns about content validity, we ensured that the observed indicators for our measurement scales reflected the breadth of student voices, and reflected the terminologies used by students (based on guidelines of Cavusgil, Deligonul, & Yaprak, 2005, and Hinkin, 1995). The next discussion focuses on the key constructs of our conceptual model.

**Figure 2:** Grounded conceptual model (result of stage 1 study).

### We were careful

When students are informed about impending PEs, and their impact on grades, the most immediate and focal cognitive and emotional response relates to carefulness. Students are immediately on guard; they begin to exercise care in terms of what they say and do, and they find that others exercise similar care. Their past experience with PEs has made them wary; it leads them to strategize about their own participation in teamwork. Consider the following voice:

The members were very careful of what to say or do. Our meetings were quiet and no one said anything shocking or disrespectful to another member. Typically, our conversation only was about the task at hand or future assignments. I feel each member conducted himself or herself in this way to receive a good peer evaluation.

Students *decide* to exhibit their best behaviors during teamwork, well before such decisions translate into actions:

During this (forming) stage, members were trying to determine behavior that is acceptable. Due to the peer evaluation system we also wanted to make sure that we were on each other's good side so as to achieve a more favorable score. This was achieved by filtering what we said until we learned new members' perspective on things as well as being of use to the team . . .

Students draw inferences about the care others are exercising upon awareness of PEs as well:

Some of the group members (seemed) pressured to put in work due to the peer evaluation. They knew that if they did not contribute before the assignment ended it would negatively affect their grade. If they were not to put in their share they would receive a poor score on the peer evaluation.

**Table 1:** Measurement scales (five-point Likert)

Factor 1 We were careful	<p><i>Once we became aware that we would evaluate each other at the end of the semester, and that our evaluation would affect our individual grades on the team project, I can say that:</i></p> <p>C1. I was very careful about what I said during our team meetings.</p> <p>C2. I was very careful about what I did during our team meetings.</p> <p>C3. My team members seemed very careful about what they said during the team meeting.</p> <p><i>C4. My team members seemed very careful about what they did during the team meetings. [Discarded after step 1 of the scale purification process]</i></p>
Factor 2 Impression Management  (Negative contribution made by others)	<p><i>Based on what other members of my team said and did, I can honestly say that:</i></p> <p><i>IM1. There was a lot of fake-niceness among team members. [Discarded after step 3 of the scale purification process]</i></p> <p>IM2. Team members seemed more interested in managing other people's impression instead of authentically participating in teamwork.</p> <p>IM3. Team members increased the quantity of participation without increasing quality of participation.</p> <p>IM4. Team members reduced the spontaneity of their contributions.</p>
Factor 3 Others Contributed More  (Positive contribution made by others)	<p><i>Because we knew that we would be evaluated by others, I can say that:</i></p> <p>OCM1. Other members of the team worked more collaboratively as a team.</p> <p>OCM2. Other members of the team took more initiative in the team and did not wait to be asked.</p> <p><i>OCM3. Other members of the team took leadership positions more often. [Discarded after step 2 of the scale purification process]</i></p> <p>OCM4. Other team members brought new, well researched ideas more often to the team meeting.</p>
Factor 4 I Contributed More	<p><i>Because I knew that I would be evaluated by others, I can say that:</i></p> <p>ICM1. I attended team meetings held outside class more regularly.</p> <p>ICM2. I contributed a lot more during the team meetings than I expected.</p> <p>ICM3. I took more initiative in the team and did not wait to be asked.</p> <p><i>ICM4. I took a leadership position in the team more often. [Discarded after step 3 of the scale purification process]</i></p> <p>ICM5. I brought new, well researched ideas more often to the team meetings.</p>
Variable 1 Grade I expected for the team project	<p>What grade are you expecting to earn in this team project? (Choose 1 of A, A-, B+, B, B-, C+, C, C-, D, F.)</p>



“We were careful” translates to strategically calibrated behaviors:

Everyone knew that we were going to be grading each other on their work and effort at the end of the semester. So, I think that subtly everyone tried to be very nice, flexible, and easy to deal with throughout the project. For example, everyone was very polite and easy going when scheduling meetings and splitting up work. Nobody wanted to come off as slightly mean so as to possibly lower the grade they’d eventually receive from each other at the end of the semester.

Figure 2 illustrates the key learning we derive from students’ perspectives; i.e., the *we were careful* construct serves as the key antecedent for the thinking and behavior that occurs in teams when students learn that they are required to conduct PEs. The five-point Likert scale we developed that reflects the voices and sentiments of students as they describe carefulness is as follows (see Table 1):

*Once we became aware that we would evaluate each other at the end of the semester, and that our evaluation would affect our individual grades on the team project, I can say that:*

... I was very careful about what I said during our team meetings.

... I was very careful about what I did during our team meetings.

... My team members seemed very careful about what they said during the team meeting.

... My team members seemed very careful about what they did during the team meetings.

Participants note that varying directly with the care exercised, others engage in both negative, noncontributive, social-loafing-related behaviors (*impression management*) and positive, contributing, collaborative behaviors (*contributing more*). We discuss these findings next.

### **Impression management**

Carefulness appears to trigger, albeit to varying extents, impression management behaviors among participants. The connotation of *impression management* is negative; the reference is to behaviors designed to manipulate one’s team members into gaining positive PEs without the burden of contributing positively. The negative, noncontributory types of behaviors relate to fake-niceness, not challenging others or engaging in constructive conflict, and resorting to practiced (versus spontaneous) ways of engaging others to garner positive PEs. Social loafing does not go away as a result of PEs; rather we find evidence to suggest that people adapt their behaviors in response to PEs and attempt to manipulate others without contributing to teamwork. Consider student voices about acting artificially rather than in authentic ways:

I found myself frequently complementing my fellow group members, inquiring about their social and sporting lives, as well as cracking jokes in an effort to keep the mood light and keep their perceptions of me positive.

Their premise seemed to be, “We, i.e., others and I, actively managed impressions, not by working harder, but by being duplicitous, because we knew others were judging us:”

The ever-present knowledge in each of our minds that we were being silently and constantly judged for every nuance of our behavior stymied our productivity in some respects. I got the distinct impression that certain team members were behaving with a stilted, forced niceness that made me wonder if they were being genuine.

Spontaneity and constructive conflict is curbed as a result of impression management. Participants tactically refuse to confront others even when the team is not making progress and even when some confrontation is needed for improved performance. Consider the following voices:

People did not bring up as much conflict as they normally would. I personally had instances where I thought that the conflict was not worth getting a bad grade from my peers so ultimately I did not express my conflict.

In a sense, we were all pandering to each other, lest we grade one another poorly. This palpable charade precluded some of the conflict that is so necessary to team effectiveness.

The five-point Likert scale we derived from participant voices to assess *impression management* (noncontributory behaviors designed to influence others’ PEs *but not* help teamwork) was:

*Based on what other members of my team said and did, I can honestly say that:*

... There was a lot of fake-niceness among team members.

... Team members seemed more interested in managing other people’s impression

instead of authentically participating in teamwork.

... Team members increased the quantity of participation without increasing quality of participation.

... Team members reduced the spontaneity of their contributions.

### **Others contributed more**

Not all behaviors triggered by carefulness are negative; many participants speak about the positive contributions by all participants as a direct result of exercising care. Carefulness motivates team members, albeit to varying extents, to take more initiative, take a leadership position more often, do their homework, and attend team meetings prepared with thoughtful, well-researched ideas. These contributory actions help the team’s collective pursuits and are attributed as motivators for their own contributions. Consider the following voices:

Often, team members would voluntarily take on roles that the group as a whole may not have seen as desirable. An example of this is volunteering to organize

all the files and Google documents the team was currently working with and condense them into one organized folder to help benefit the entire team.

Everyone (in the team) was on completely different schedules and the only times we could really meet were really obscure times. (Yet) Everybody made it appear as though they were more than willing to go out of their way to help the group.

The five-point Likert scale we derived to assess *others contributed more* (i.e., contribution that aided teamwork) was:

*Because we knew that we would be evaluated by others, I can say that:*

... Other members of the team worked more collaboratively as a team.

... Other members of the team took more initiative in the team and did not wait to be asked.

... Other members of the team took leadership positions more often.

... Other team members brought new, well researched ideas more often to the team meeting.

### **I contributed more**

The heightened carefulness triggered by the knowledge of PEs, coupled with the evidence that others are contributing more positively, leads participants to contribute more to teamwork. One participant describes the improved contribution she/he makes because she/he knows that PEs will occur:

I knew that we were submitting a peer evaluation survey so I was acting on my best behavior to maximize my grade for it. I would make sure that I had all of my work done before every meeting and that I was on time to make the best impression on my team.

Perceptions of other participants' positive contributions also matter; when others are contributing more, team members choose to scale up their contribution. Improved quality of contribution is defined in terms such as: *I attended meetings held outside class more regularly, I contributed more than I expected, I took initiative, I led more frequently, and I did my research before coming to the team meetings*. Consider the following voices:

I would be the first person to volunteer or would offer and take on more responsibility for the team . . . . By me always being open to take on tasks that needed to be completed, I was hoping for a good peer evaluation. This helped the team be effective as it allowed us to not have to spend a long time figuring out who was going to write particular parts and waste time dividing up.

The evidence that others were contributing more, not just faking it, was a strong motivator for stepping up their own contributions. Consider the following response:

(Name of person) showed up to a meeting with 102-degree fever. He was very dedicated and I admired him for keeping everyone motivated. The way he handled himself in and out of meetings was impressive. Furthermore, he was also easy to reach and took his time to explain his part of the project. I was motivated by watching him as I wanted to come to every meeting prepared with quality work.

The five-point Likert scale we derived from participants' voices for measuring *I contributed more*, was:

*Because I knew that I would be evaluated by others, I can say that:*

- . . . I attended team meetings held outside class more regularly.
- . . . I contributed a lot more during the team meetings than I expected.
- . . . I took more initiative in the team and did not wait to be asked.
- . . . I took a leadership position in the team more often.
- . . . I brought new, well-researched ideas more often to the team meetings.

### Grade performance

Grades students expect to receive on the team project are focal in the consciousness. Students speak about grades in two interrelated ways. First, students are unequivocal that PEs should matter; their individual contribution—when recognized by their peers—should ensure that the grade they receive for the team project is adjusted upward. Moreover, they strongly feel that PEs are useless if the input they provide to others and to the instructor about others' behavior is without real consequence or does not substantially affect anyone's grade. Even though scholars have yet to advocate for highly consequential PEs, students seem unequivocal in their preference for PEs that fully impact grades—both their own and those of others. Participating in PEs that hold few grade consequences triggers cynicism and consternation. Describing the frustration with inconsequential PEs that were used in previous classes, a student notes:

The peer evaluation was a mere 20% of the project grade which was 20% of the final (course) grade. The peer evaluation system hardly increased team effectiveness because it was such a minimal part of the grade. Social loafing and a general low quality of work were evident because peer evaluation only counted towards 4% of the final grade, it did not mean much to us.

Second, the PE grade threat connection explains *much* of what is perceived and done in the team environment. In other words, if PEs are insignificant, they are unlikely to shape thinking or behavior in significant ways, defeating their formative and summative purposes:

After receiving our peer evaluation grades back (mid-semester), team members began acting differently in hopes of getting higher scores on the next one . . . . Team member H showed little initiative in the first half of the semester and it was clear that he was trying to impress us now so he can achieve a higher grade on the peer evaluation.

I knew that we were submitting a peer evaluation survey so I was acting on my best behavior to maximize my grade for it.

After we received the peer feedback from other team members (mid-semester), we all became much more aware of what each of us was doing individually and how that was going to impact our grade.

The construct of grade performance was therefore operationalized as the difference between the grade they said they expected to receive on the team project and the grade they actually received. Hence, the scale shown in (A) below was

included in the survey we administered during stage 2 of our study, and the scale shown in (B) below was added to record the actual grade received by the student on the team project *after* the adjustments resulting from PEs completed by others. Grade performance was calculated as “actual grade less expected grade.”

A. What grade are you expecting to earn in this team project?

(Choose 1 of A, A–, B+, B, B–, C+, C, C–, D, F.)

B Actual Grade for the Team Project: \_\_\_\_\_

### Hypotheses from Stage 1 Exploration

The hypotheses that guided the second-stage survey in the context of classroom teams, team projects, undergraduate business programs, and highly consequential PEs are as follows:

H1: The higher the reported *we were careful* score:

H1a: The stronger the perception of impression management in the team.

H1b: The stronger the perceived contribution of others in the team.

H1c: The higher the reported level of contribution made by the participant (self-reported contribution).

H2: The stronger the perception of impression management on the team, the more likely that the actual grade received on the team project is *lower* than the grade expected by the participant (expect positive slope, a positive  $\beta$  value).

H3: The stronger the perceived contribution of others in the team:

H3a: The more likely that the actual grade received on the team project is *higher* than the grade expected by the participant (expect negative slope, negative  $\beta$  value).

H3b: The higher the reported level of contribution made by the participant (self-reported contribution).

## STAGE 2 STUDY

### Purpose of Stage 2 Study

The purpose of the second-stage study was to collect data from students who had participated in highly consequential PE processes and test the hypotheses using the measurement scales yielded by the first-stage study. Hence, we began the following semester by assigning students enrolled in sections of undergraduate *Organizational Behavior* and *Human Resource Management* to teams, and requiring them to work collaboratively on a team project worth 30% of the final grade. On the first day of class, students were informed about the highly consequential PEs; students were informed that the grade awarded by the instructor to the *team* could be changed based on the PEs they received (i.e., a student could receive zero or a better grade than other team members; see Appendix A for key information included in every syllabus).

## Stage 2 Data Collection

At the end of the semester in which highly consequential PEs were assigned, students made their final presentations of the team project during the last week of classes and turned in the team paper (one per team) as well as the PEs they had conducted (each student turned in as many PEs as the number of other members on her/his team). Students completed the survey that contained questions included in Table 1 after they had completed the highly consequential PE.

After finals week, the instructor graded the team project and adjusted each individual's grade based on the PEs. This adjusted grade was recorded as the *actual grade* on each questionnaire (which was confidential but not anonymous). This enabled comparison between actual and expected grade on the team project. This process was repeated in 17 sections taught over seven semesters. It is important to note the time gap between stage 1 and stage 2. Stage 2 began in the semester following the one during which the stage 1 study was conducted. Stage 1 ended in semester one, and stage 2 began in semester two and continued up to semester eight. No student who participated in stage 1 also participated in stage 2 of the study; hence, this should preclude concerns about bias introduced by polling the same sample of students.

## Stage 2 Sample

Of the 417 students enrolled over seven semesters, 395 completed responses were used for data analysis; 23 questionnaires were discarded because they were incomplete (response rate: 94.7%). No extra credit was awarded for completing surveys; no student was sanctioned because of incomplete surveys. About half of the sample (48.7%) was 20 years old; 99% of the sample was under 25 years of age. Most (64%) of the participants were juniors, 24% were seniors, and 5.3% were sophomores. The average grade differential (i.e., actual grade less expected grade, on 100 points) was a negative number ( $-2.983$ ), suggesting that on an average, students overestimated the grade they would receive for the team project (A = 100–96, A– = 95–90, B+ = 89–87, etc.). Nearly all (98%) of the students were full-time students; males were slightly more represented (52.9%) than female students (46.8%). About half the sample was pursuing a Business Administration major (45.7%), followed by Accounting (30.6%), Communication (16.7%), Economics (1%), and Psychology (1%).

## Stage 2 Data Analysis

We used EQS 6.2 to conduct the following analysis based on Anderson and Gerbing (1988). In stage 2, we fitted a confirmatory factor analysis on the data to assess whether any structural model existed based on goodness-of-fit indices, and purified the measurement model. The confirmatory factor analysis (CFA) model included covariances among all latent factors (in our case, the four factors (1) we were careful, (2) impression management, (3) they contributed more, and (4) I contributed more). Because multivariate estimation procedures are often biased due to nonnormality, we used robust estimation for the goodness-of-fit statistics for the CFAs (based on Anderson & Gerbing, 1988; Bentler & Wu, 2002). The overall goodness of fit for the first iteration of the CFA (all 17 variables

**Table 2:** Purification of measurement model (results of CFA).

Steps	Normed Fit Index (NFI)	Non-normed Fit Index (NNFI)	CFI	Incremental Fit Index (IFI)	Root Mean Square Error of Approximation (RMSEA)	Action
Step 1	.843	.853	.88	.881	.082	Remove M4 from factor 1
Step 2	.884	.904	.923	.924	.066	Remove OCM1 from factor 3
Step 3	.895	.912	.931	.932	.064	Remove IM1 from factor 2, and remove ICM4 from factor 4
Step 4	.931	.955	.967	.967	.046	None. 90% confidence interval of RMSEA = .033–.059.

**Table 3:** Key statistics and correlations among factors/variable.

	Alpha	CR	AVE	Correlations				
				Factor 1 WWC	Factor 2 IM	Factor 3 OCM	Factor 4 ICM	Grade Diff
Factor 1 We were careful	.84	.83	.62	1				
Factor 2 IM	.735	.74	.49	.205	1			
Factor 3 OCM	.831	.84	.63	.38	.084	1		
Factor 4 ICM	.831	.84	.56	.504	.127	.450	1	
Grade Diff		1	1	-.162	.103	-.275	-.046	1

CR: Construct Reliability; AVE: Average Variance Explained, IM: Impression Management, OCM: Other team members contributed more; ICM: I Contributed More; Grade Diff: Actual grade on Team Project *less* Expected Grade on Team Project.

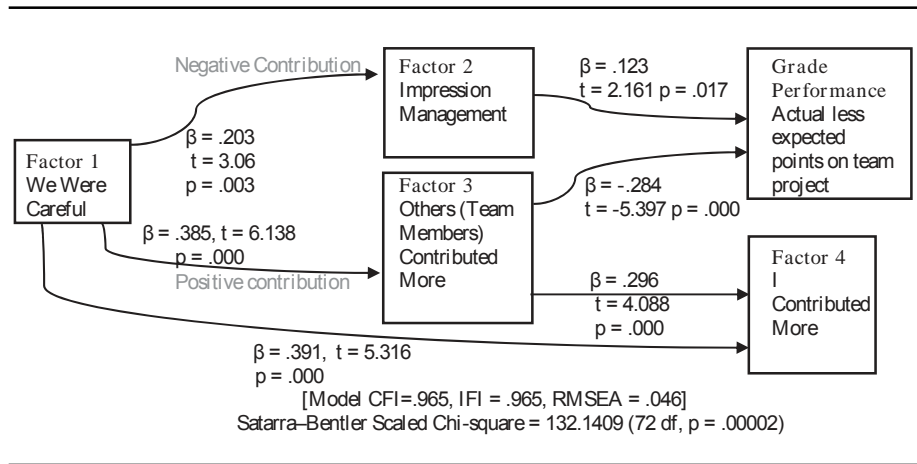
included) was reasonable, however, not acceptable (e.g., Bentler, 1995’s comparative fit index [CFI] was .88, i.e., below the acceptable .9). To modify the model, we used Lagrange Multiplier (LM) tests to identify variables that cross-loaded on multiple latent factors. Table 2 highlights the improvements in fit indices (particularly RMSEA) as a result of removing one or two variables at a time. After the fourth iteration of the CFA, the fit indices were as follows: NFI = .931, NNFI = .955, CFI = .967, IFI = .967, and RMSEA = .046.

Table 3 serves as a guide for the following discussion of reliability and validity of scales we eventually used while estimating our theoretical model. Cronbach’s alphas range from .735 to .84 (e.g., Churchill, 1979), and construct reliability ranges from .74 to .84 (e.g., Hair, Black, Babin, & Anderson, 1998)—attesting to reliability of the scales. Discriminant validity was established in two ways. First, the correlations between factors are less than .504, i.e., less than the .8 that is often associated with problems of construct discriminance (e.g., Yanamandram & White, 2010). Second, all estimates of the average variance explained for each of the four factors (based on factor loadings) are greater than the squared correlations between each pair of factors and attest to discriminant validity of scales (see Table 4). Hence,

**Table 4:** Comparison of average variance explained (AVE) and squared correlations among factors (and variable) for establishing discriminant validity.

	F1 We were careful	F2 Impression Management	F3 Others Contributed More	F4 I Contributed More	Grade Diff Difference between Actual and Expected Grade
F1: We were careful	AVE = .62				
F2: Impression Management	$R^2 = .042$	AVE = .49			
F3: Others Contributed More	$R^2 = .144$	$R^2 = .007$	AVE = .63		
F4: I Contributed More	$R^2 = .254$	$R^2 = .016$	$R^2 = .203$	AVE = .56	
Grade Diff:	$R^2 = .026$	$R^2 = .011$	$R^2 = .076$	$R^2 = .002$	AVE = 1

**Figure 3:** Student perceptions of highly consequential performance evaluations in undergraduate business classroom teams (results of stage 2 study).



the purification process yielded a structurally coherent, valid measurement model, ready for further refinement, and assessment of path coefficients.

In the second stage, we specified the hypothesized paths and used Wald tests to examine whether any could be dropped and LM tests to examine whether additional parameters could be added to the theoretical model (based on Bentler, 1995). The model converged, however, with highly acceptable parameters in the first iteration with significant path coefficients (NFI = .926; NNFI = .955, CFI = .965, IFI = .965, RMSEA = .46). Figure 3 illustrates the path coefficients ( $\beta$  values and  $t$ -statistics), Table 5 reports the standardized solution for the hypothesized model, and Table 6 reports the measurement and structural parameters from the revised theoretical model.



**Table 5:** Standardized solution for the hypothesized model.

Latent Variable	Variable	$\beta$	Error	$R^2$
Factor 1 We were careful	C1. I was very careful about what I said during our team meetings.	.908	.418	.825
	C2. I was very careful about what I did during our team meetings.	.816	.578	.666
	C3. My team members seemed very careful about what they said during the team meeting.	.612	.791	.374
Factor 2 Impression Management	IM2. Team members seemed more interested in managing other people's impression instead of authentically participating in teamwork.	.745	.667	.556
	IM3. Team members increased the quantity of participation without increasing quality of participation.	.779	.627	.606
	IM4. Team members reduced the spontaneity of their contributions.	.568	.823	.323
Factor 3 Others Contributed More	OCM1. Other members of the team worked more collaboratively as a team.	.807	.59	.651
	OCM2. Other members of the team took more initiative in the team and did not wait to be asked.	.87	.494	.756
	OCM4. Other members brought new, well researched ideas more often to the meeting.	.693	.721	.481
Factor 4 I Contributed More	ICM1. I attended team meetings held outside class more regularly.	.703	.711	.495
	ICM2. I contributed a lot more during the team meetings than I expected.	.836	.549	.699
	ICM3. I took more initiative in the team and did not wait to be asked.	.717	.697	.515
	ICM5. I brought new, well researched ideas more often to the team meetings.	.732	.681	.536

## Stage 2 Findings

As Figure 3 and Tables 5 and 6 report, in the context of PEs, we find the following. First, higher reported levels of *we were careful* produced: (a) higher levels of *we managed impressions* ( $\beta = .203$ ,  $t = 3.06$ , supporting H1a), (b) higher levels of *others contributed more* ( $\beta = .385$ ,  $t = 6.138$ , supporting H1b), and (c) higher levels of *I contributed more* ( $\beta = .391$ ,  $t = 5.316$ , supporting H1c). Second, higher *impression management* was linked to lower than expected grade ( $\beta = .123$ ,  $t = 2.161$ , supporting H2). Third, the higher the level of *others contributed more*: (a) the lower the expected grade compared to actual ( $\beta = -.284$ ;  $t = -5.397$ ,

**Table 6:** Measurement and structural parameters from the revised theoretical model.

Structural Model	$\beta$ s ( <i>t</i> -Values) [ <i>p</i> Value] <sup>a</sup>
Impression Management → We were careful	.203 (3.06) [.003]
Others Contributed More → We were careful	.385 (6.138) [ <i>p</i> = .000]
I Contributed More → We were careful	.391 (5.316) [ <i>p</i> = .000]
Grade Difference → Impression Management	.123 (2.161) [ <i>p</i> = .017]
Grade Difference → Others Contributed More	-.284 (-5.397) [ <i>p</i> = .000]
I Contributed More → Others Contributed More	.296 (4.08) [ <i>p</i> = .000]
We were careful → I was very careful about what I said during our team meetings.	1 <sup>b</sup>
We were careful → I was very careful about what I did during our team meetings.	.862 (15.793) [ <i>p</i> = .000]
We were careful → My team members seemed very careful about what they said during the team meeting.	.683 (11.751) [ <i>p</i> = .000]
Impression Management → Team members seemed more interested in managing other people's impression instead of authentically participating in teamwork.	1 <sup>b</sup>
Impression Management → Team members increased the quantity of participation without increasing quality of participation	1.06 (9.029) [ <i>p</i> = .000]
Impression Management → Team members reduced the spontaneity of their contributions.	.751 (8.516) [ <i>p</i> = .000]
Others Contributed More → Other members of the team worked more collaboratively as a team.	1 <sup>b</sup>
Others Contributed More → Other members of the team took more initiative in the team and did not wait to be asked.	1.101 (15.88) [ <i>p</i> = .000]
Others Contributed More → Other members brought new, well researched ideas more often to the meeting.	.825 (13.601) [ <i>p</i> = .000]
I Contributed More → I attended team meetings held outside class more regularly.	1 <sup>b</sup>
I Contributed More → I contributed a lot more during the team meetings than I expected.	1.09 (13.727) [ <i>p</i> = .000]
I Contributed More → I took more initiative in the team and did not wait to be asked.	.911 (11.719) [ <i>p</i> = .000]
I Contributed More → I brought new, well researched ideas more often to the team meetings.	.896 (13.535) [ <i>p</i> = .000]

<sup>a</sup>*p* Values are estimated based on Satorra-Bentler scaled chi-square test with 72 degrees of freedom.

<sup>b</sup>Indicant loading fixed at 1 to set the scale (*t*-values, all are significant at *p* = .05).

supporting H3a), and (b) the higher the *I contributed more* ( $\beta = .296$ ,  $t = 4.088$ , supporting H3b). The grounded model and all hypothesized relationships were supported by the survey data.

In other words, when undergraduate business students think about PEs in their classroom team context, they think about: I should care and others seem to care. The carefulness triggers both duplicitous, manipulative, impression management behaviors, and sensitivity to evidence that others are contributing more. When noncontributory, manipulative impression management behaviors increase, the gap between actual grade received and the grade expected on the team project worsens; i.e., the actual grade is lower than what students say they expect. Conversely, when other participants are perceived as engaging less in impression management, and contributing more, students not only contribute more themselves; the difference between the actual and expected grade changes in the students' favor. Their actual grades are higher than what they say they expected.

## IMPLICATIONS

The purpose of this discussion is to highlight the key contributions made by our study of students' perspectives into highly consequential PEs and the contributions to future theory and practice (see Table 7 for an encapsulation). To frame this discussion in its appropriate perspective, we begin with a brief discussion of limitations. First, we examined the impact of highly consequential PEs on student perceptions. The evidence we generate can serve as a foundation for future experimental and causal research with control groups; the current study is not a substitute for such efforts. Studying the impact produced by varying the consequences of PEs is left to future research. Second, we acknowledge the following: the merits of PEs can be further discussed, alternative PE instruments can be designed, less consequential PEs can be used, alternative ways of collecting data exist (e.g., longitudinal surveys), and random samples can be taken from all students across all schools. The pursuit of these alternatives is left to future research. Third, the directionality of the arrows in Figure 2 result from qualitative research and grounded data and should preclude concerns about investigating other relationships. The Wald test parameters did not indicate the presence of relationships other than the ones we had hypothesized based on qualitative data. Fourth, the Structural Equation Modeling (SEM) procedure was used to simultaneously assess unidirectional relationships; the evidence of discriminant validity provides sufficient grounds for future attempts to estimate causal relationships using controlled experiments (consistent with Anderson & Gerbing, 1988 and Fornell & Larcker, 1981). Finally, the grade performance construct is independently assessed and should address concerns about common methods variance (e.g., Podsakoff, Mackenzie, Lee, & Podsakoff, 2003). To further preempt concerns, we followed Conway and Lance's (2010) guidelines; we assessed composite reliability and the discriminant validity of data-derived constructs and scales.

### **Salience of *We Were Careful***

The *we were careful* construct is salient; it serves as the significant antecedent of every other construct focal in students' PE-related cognitive schemas. This finding adds value to what is known in the following way. Current thinking, reflecting instructors' and scholars' points of view, advocates for PEs in classroom teams.

**Table 7:** Overview of the contributions of our study.

The literature devoted to Peer Evaluations in classroom teams suggests:	In the context of highly consequential Peer Evaluations implemented in classroom teams, we learn:
PEs improve student learning and promote hard work (Brutus & Donia, 2010)	PEs raise students' sensitivity about the scrutiny from others; they stand on guard knowing that others are evaluating their behaviors. The inferences students draw from the behavior of others' lead them to adopt a <i>wait and see</i> approach. If others are perceived to contribute more, they contribute more; if others are perceived to contribute less, they contribute less. Making the perils of <i>wait and see</i> explicit to team members and designing PEs to encourage, assess, and reward initiative and risk taking in teams are implications of this study.
Effective team members regularly attend team meetings, are cooperative and prepared, and demonstrate leadership (Baker, 2008; Paswan & Gollakota, 2004).	PEs focus students' attention on what other team members are doing (i.e., are they attending regularly, are they prepared, do they demonstrate leadership, etc.). They evaluate whether others are contributing more or engaging in manipulative behaviors to elicit positive PEs.
PEs make students expend more effort (Erez, LePine, & Elms, 2002).	Their inferences about others' behaviors produce unjustified optimism (if they infer others are contributing positively to teamwork) or unjustified pessimism (if they infer others are engaged in manipulative impression management). Cautioning students about their inability to predict the team's performance and the unjustifiable optimism and pessimism as a result of focusing on what others are doing is an implication of this study.
PEs reduce social loafing (Falchikov & Goldfinch, 2000) and reduce free riding (Brutus & Donia, 2010).	PEs change the nature of, but do not reduce or eliminate, social loafing, and free riding. Students reduce social loafing as it relates to slacking off; they devote attention to manipulating the impressions of others without contributing more to teamwork. Expanding the conceptual domain of social loafing in undergraduate business classroom teams in which PEs are used is an implication of this study.

PEs are known to help instructors; they measure individual performance when students are engaged in teamwork (e.g., Chen & Lou, 2004; Fellenz, 2006) and allow for fairer grading (e.g., Paswan & Gollakota 2004; Pfaff & Huddleston 2003; Verzat, Byrne, & Fayolle, 2009). PEs are known to help team members by providing them with developmental feedback and improving overall effectiveness of teams (Dominick, Reilly, & McGourty, 1997; Mayo et al., 2012).

Our findings add value to this line of thinking by pointing to the differing students' perspective. To them, the knowledge of impending PEs triggers high levels of carefulness; i.e., they exercise care in terms of what they say and do on the

team and are on guard. Their voices mirror current concerns raised in contexts other than PEs in classroom teams. When one can regret making wrong choices, decision making is careful (see Jochen, 2008). As an antecedent carefulness sharpens focus on what other team members are doing and contributing to the team and triggers inference drawing that shapes their thinking about the team's likely performance outcomes and helps them calibrate their own contributions. This finding is new; the attention-focusing effect of PEs in classroom teams has invited no discussion in the literature.

Future research that alters the consequences of PEs and assesses its impact, likely through experimental designs, is essential before notions of optimum consequence—PEs emerge.

### **Implications of the *Grade Performance Construct***

Our findings contribute to current thinking by drawing attention to the inferences students draw from the behaviors of other team members when highly consequential PEs are used, and how their inferences shape their optimism or pessimism about their grades. In the highly consequential PE context, we learn that students are unable to fairly estimate the performance outcomes of the team based on the behaviors of other team members. That is, we learn that their predictions or their expected grade performance are either unjustifiably pessimistic or optimistic. New thinking and research that enable students to accurately gauge and predict their team's performance in terms of grades is implicated by our findings. The explanation for this inference is as follows.

We included the construct of *grade performance* in the study because the concern with grades was focal in students' responses to our questions during the exploratory stage and because scholars highlight students' concerns with grades as inseparable from PEs (e.g., Chen & Lou, 2004; Psenicka, Vendemia, & Kos, 2013). Students want to ensure that: (a) the instructor is aware of the true extent of their contribution as recognized by their team members, so that their personal grade accurately reflects their contribution, and (b) the instructor is aware of the true extent of their team members' contributions, so that others' personal grades accurately reflect their contributions. Hence, we defined the construct of *grade performance* as a comparison between actual grade (that we calculated after the end of the semester for each student) and expected grade (that the student indicated on each questionnaire they completed).

The *grade performance* construct essentially assesses participants' optimism and pessimism. The *grade performance* score for a student is positive (or a greater positive) number if the grade a student receives is higher than what the student thinks she/he will earn for the team project. Moreover, a positive score suggests that the student is pessimistic, but unjustifiably so. She/he expects a low score when she/he actually earns a higher one. A negative score suggests that the student is optimistic, but unjustifiably so. She/he expects a higher score when she/he actually earns a lower one.

As Figure 3 illustrates, when students see others exhibiting manipulative impression management behaviors, they are unjustifiably pessimistic; i.e., the  $\beta$  is positive, indicating that they expect a lower grade than the one they earn.

A similar inference is drawn from the inverse link (negative  $\beta$ ) between *others contributed more* and *grade performance*. When team members infer that others are contributing more, they are unjustifiably optimistic; i.e., the  $\beta$  is negative, indicating that they expect a higher grade than the one they earn. Even though the knowledge of impending highly consequential PEs focuses their attention on what others are doing, our findings suggest that the inferences they draw about their own grades are unjustifiably pessimistic or optimistic. In other words, their perceptions of what others are doing serves to mislead them about the team's performance and their own score.

The question arising from this finding is: why are students unjustifiably optimistic or pessimistic about their grade performance based on the inferences they drew from other team members' behaviors? We identify two plausible explanations that deserve testing before generalizations are drawn. First, it is likely that the focus on other team members' behaviors triggered by highly consequential PEs renders students sensitive. They read too much into others' behaviors and draw an exaggerated inference about their consequences on grades. This inference deserves additional testing. Second, it seems likely that currently, popular indicators of positive contribution to teamwork, such as regular attendance and participation in teamwork, preparedness, taking initiative, and willingness to lead (e.g., Baker, 2008; Paswan & Gollakota, 2004), are insufficient predictors of how well the team will perform in terms of grades. We infer this because students are misled by the inferences they draw from these cues. Our findings point to the merit of identifying via further exploration the cues participants can draw from the behaviors of others that allow for realistic inferences about grade performance (versus their currently unjustified optimism or pessimism). In practical terms, there is merit in instructing students that their inferences about others' positive and negative contributions to the team may mislead their predictions of their team's performance and their own grade.

### Managing Impressions

In the context of highly consequential PEs, we learn that students can identify manipulative, duplicitous behaviors of others designed to garner positive PEs without the burden of contributing positively to the team (i.e., impression management). This finding produces two implications that deserve additional consideration and research.

First, while current thinking indicates that PEs can inhibit social loafing (Brutus & Donia, 2010), our findings suggest that in the context of highly consequential PEs, social loafing transforms from *slacking off* (e.g., Brutus & Donia, 2010) and *disruptive behaviors* (Jassawalla, Sashittal, & Malshe, 2008, 2009) to include manipulative impression management behaviors (see Figure 4). Each type of social loafing such as slacking off, disruption, and manipulative impression management shares a common link with behaviors designed to get away without honest contribution to teamwork and leave others to pick up the slack. This finding deserves additional exploration and empirical confirmation.

Second, our learning about impression management as a negative contribution that produces pessimism among participants contrasts with current views.

**Figure 4:** New thinking about social loafing in classroom teams.

CONCEPTUAL DOMAIN OF THE SOCIAL LOAFING CONSTRUCT		
Traditional definition (Latane et al., 1979, Williams, & Harkins, 1979)	Expanded definition after Jassawalla, Sashittal and Malshe (2008, 2009)	Expanded definition in the context classroom teams in which PEs are used
Social loafing is synonymous with slacking off.	Social loafing includes disruptive behaviors that hurt teamwork.	Social loafing includes manipulative impression management behaviors.

Scholars suggest that impression management—even when it is duplicitous and inauthentic—is conducive to organizational life; the issue is widely studied in the supervisor-subordinate interaction contexts. For instance, current theory holds that if subordinates are managing impressions to seem more like their superiors, the latter’s rating of the former improves (see Wayne & Liden, 1995). When socially astute people manage other people’s impressions, others are known to rate them better (see Brouer, Badaway, Gallagher, & Haber, 2015). The Bolino and Turnley (1999) scale of impression management assesses the extent to which team members talk proudly about their experience or education, talent or qualifications, value to the organization, and accomplishments. Steps taken by a team member to inflate others’ perceptions of their contributions help elicit better ratings from others. Scholars also suggest that impression management behaviors aimed at shaping other people’s perception of one’s contributions produce positive outcomes for teams (see Drory & Zaidman, 2007; Rozell & Gundersen, 2003). The literature, if implicitly, suggests that impression management behaviors designed to shape others’ perceptions of one’s contribution—versus actual improvements in contributions without regard to perceptions—helps the impression manager in particular and the firm in general. For example, talking up one’s contribution or pretending to share interests with supervisors are not inherently contributive behaviors; they are manipulative even when benign. Contrasting with these notions in the literature, our findings caution that team members can discern noncontributory, negative impression management and exacerbate the difference between what team members expect to achieve as a team, and render them unjustifiably pessimistic. Instructors implementing highly consequential PEs may find it useful to make explicit to students that team members are known to resort to impression management as a form of social loafing and that such impression management is transparent to others and is unlikely to help the impression manager. Instructors may similarly benefit if the PE instrument they design explicitly requires students to assess each other based on their authentic contribution and impression management.

**Managing *Wait and See* in Classroom Teams**

When highly consequential PEs are used in classroom teams, the level of individual contribution seems likely to increase if instructors inform students about the

suboptimal consequences of *wait and see* that can lead to a cycle of *waiting for others to contribute*, and hence design PE instruments and processes that encourage, assess, and reward students for taking initiative and setting precedence in teams. Our findings imply that: (a) the knowledge of impending highly consequential PEs makes team members highly careful about what they say and do in teams, and (b) the carefulness focuses attention on what others are doing and triggers a *wait and see*. If others are seen to contribute less, team members contribute less; if others are seen to contribute more, team members contribute more; and if others are viewed to engage in manipulative impression management, team members lower their expectations of the team's performance.

These findings about *wait and see* resonate with current thinking in the literature in two important ways (e.g., Cowart, Gilley, Avery, Afton, & Gilley, 2014). First, it is well documented that people define their own behaviors in positive ways when they see evidence of others behaving positively (Weaver, Ellen, & Mathiassen, 2015). The way that people learn from social engagement with others is also well-researched territory (see Bandura, 1986). While the literature has focused more on workplace mentors and senior managers (see Brown & Trevino, 2014 and Chin, 2015's meta-analysis), our study is the first to identify the complex roots of the decision to contribute more in the classroom team context.

Second, *wait and see* responsiveness is reported in multiple contexts such as investments (Miller & Folta, 2002), innovation (Day & Shoemaker, 2000), and options trading (Chatterjee, Lubatkin, & Schulze, 1999; Tiwana, Wang, Keil, & Ahluwalia, 2007). In all instances, they relate to human behavior in high-uncertainty situations. For instance, firms are known to delay investments in times of macroeconomic uncertainty (e.g., Campa, 1993; Dixit, 1989; Goldberg & Kolstad, 1995) and delay innovation when confronted with market uncertainty (Day & Shoemaker, 2000). To overcome or preempt *wait and see* responsiveness, scholars advocate for actions designed to reduce uncertainty in the contexts of options trading (e.g., Folta & O'Brien, 2004) and entrepreneurship (e.g., Metzger & King, 2015).

Our findings coupled with the literature's view of *wait and see* responsiveness produce the following implications for scholars and instructors interested in highly consequential PEs in classroom teams. Implementation of highly consequential PEs likely introduces high levels of uncertainty in teams. It seems to focus participants' attention on what others are doing and the suboptimal cycle of *waiting for others to take initiative*. Instructors may find it useful to: (a) make explicit the "uncertainty-*wait and see*" linkage, the perils of waiting for others to take initiative, and (b) design and implement highly consequential PEs that assess and reward initiative and risk taking by individual team members (i.e., include items related to initiative taking and risk taking as explicit criteria on the PE instrument).

## CONCLUSION

While scholars seem to favor the notion of consequential PEs, our study is the first to provide evidence of student perceptions and behaviors associated with ones that entirely determine a student's team project grade. To the extent that students respond to the knowledge of impending PEs by raising their guard and strategizing



about their behaviors, their voices resonate with current literature. However, highly consequential PEs seem to focus attention on what others are doing, trigger a wait-and-see approach to contributing positively to the team, and make students unjustifiably optimistic or pessimistic about the team's performance.

## REFERENCES

- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411–423.
- Baker, D. (2008). Peer assessment in small groups: A comparison of methods. *Journal of Management Education, 32*(2), 183–209.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software Inc.
- Bentler, P. M., & Wu, E. J. C. (2002). *EQS for Windows user guide*. Encino, CA: Multivariate Software Inc.
- Bolino, M. C., & Turnley, W. H. (1999). Measuring impression management in organizations: A scale development based on the Jones & Pittman taxonomy. *Organizational Research Methods, 2*(2), 187–206.
- Bowes-Sperry, L., Kidder, D., Foley, S., & Chelte, A. (2005). The effect of peer evaluations on student reports of learning in a team environment: A procedural justice perspective. *Journal of Behavioral and Applied Management, 7*(1), 4–24.
- Brooks, C., & Ammons, J. (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*(5), 268–272.
- Brouer, R. L., Badaway, R. L., Gallagher, V. C., & Haber, J. A. (2015). Political skill dimensionality and impression management choice and effective use. *Journal of Business and Psychology, 30*(2), 217–233.
- Brown, M. E., & Trevino, L. K. (2014). Do role models matter? An investigation of role modeling as an antecedent of perceived ethical leadership. *Journal of Business Ethics, 122*(4), 587–598.
- Brutus, S., & Donia, M. B. L. (2010). Improving the effectiveness of students in groups with a centralized peer evaluation system. *Academy of Management Learning and Education, 9*(4), 652–662.
- Brutus, S., Donia, M., & Ronen, S. (2013). Can business students learn to evaluate better? Evidence from repeated exposure to a peer evaluation process. *Academy of Management Learning and Education, 12*(1), 18–31.
- Campa, J. M. (1993). Entry by foreign firms in the United States under exchange rate uncertainty. *Review of Economics and Statistics, 75*(4), 614–622.

- Cavusgil, T., Deligonul, S., & Griffith, D. A. (2008). Rigor in international business research: A review and methodological recommendations. In J. J. Boddewyn (Ed.), *International business scholarship: AIB fellows on the first 50 years and beyond (research in global strategic management, volume 14)*. Emerald Group Publishing Limited, 229–246.
- Cavusgil, T., Deligonul, S., & Yaprak, A. (2005). International marketing as a field of study: A critical assessment of earlier development and a look forward. *Journal of International Marketing, 13*(4), 1–27.
- Chatterjee, S., Lubatkin, M. H., & Schulze, W. S. (1999). Toward a strategic theory of risk premium: Moving beyond CAPM. *Academy of Management Review, 24*, 556–567.
- Chen, Y., & Lou, H. (2004). Student perceptions of peer evaluations: An expectancy perspective. *Journal of Education for Business, 79*(5), 275–279.
- Chin, R. J. (2015). Examining teamwork and leadership in the fields of public administration, leadership, and management. *Team Performance Management: An International Journal, 21*(3/4), 199–216.
- Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*(1), 64–73.
- Clinebell, S., & Stecher, M. (2002). Teaching teams to be teams: An exercise using the Myers-Briggs type indicator and the five-factor personality traits. *Journal of Management Education, 27*, 362–383.
- Conway, J. L., & Lance, C. E. (2010). What reviewers should expect from authors regarding common methods bias in organizational research. *Journal of Business Psychology, 23*, 325–334.
- Cowart, T., Gilley, A., Avery, S., Afton, B., & Gilley, J. (2014). Ethical leaders: Trust, work-life balance, and treating individuals as unique. *Journal of Leadership, Accountability and Ethics, 11*(3), 70–81.
- Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The effects of peer feedback on team member behavior. *Group & Organization Management, 22*(4), 508–520.
- Day, G. S., & Shoemaker, P. J. H. (2000). Avoiding the pitfalls of emerging technologies. *California Management Review, 42*(2), 8–33.
- Dixit, A. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy, 97*(3), 620–638.
- Drory, A., & Zaidman, N. (2007). Impression management behavior: Effects of organizational systems. *Journal of Managerial Psychology, 22*(3), 290–308.
- 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*(4), 929–948.
- Falchikov, N., & Goldfinch, N. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.

- Fellenz, M. (2006). Toward fairness in assessing student groupwork: A protocol for peer evaluation of individual performance. *Journal of Management Education, 30*(4), 570–590.
- Folta, T. B., & O'Brien, J. P. (2004). Entry in the presence of dueling options. *Strategic Management Journal, 25*(2), 121–138.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39–50.
- Goldberg, L. S., & Kolstad, C. D. (1995). Foreign direct investment, exchange rate variability, and demand uncertainty. *International Economic Review, 36*(4), 855–873.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (1998). *Multivariate data analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 969–988.
- Jassawalla, A. R., Sashittal, H. C., & Malshe, A. (2009). Students' perceptions of social loafing, its antecedents and consequences in undergraduate business classroom teams. *Academy of Management Learning & Education, 8*(1), 42–54.
- Jassawalla, A. R., Sashittal, H. C., & Malshe, A. (2008). Student perceptions of social loafing in undergraduate business classroom teams. *Decision Sciences Journal of Innovative Education, 6*(2), 392–415.
- Jochen, R. (2008). Regret aversion and decision process quality: Effects of regret salience on decision process carefulness. *Organizational Behavior and Human Decision Processes, 105*(2), 169–182.
- Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 823–832.
- Loughry, M. L., Ohland, M. W., & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement, 67*(3), 505–524.
- Mayo, M., Kakarika, M., Pastor, J. C., & Brutus, S. (2012). Aligning or inflating your leadership self-image? A longitudinal study of responses to peer feedback in MBA teams. *Academy of Management Learning & Education, 11*(4), 631–652.
- Metzger, M. L., & King, J. S. (2015). Extending constructivist perspectives on opportunity production through an incorporation of effectual logics. *Entrepreneurship Research Journal, 5*(4), 323–345.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage Publications Inc.
- Miller, K. & Folta, T. B. (2002). Options value and entry timing. *Strategic Management Journal, 23*(7), 655–665.

- Murphy, K. R., & Cleveland, J. N. (1995). *Performance appraisal*. Needham Heights, MA: Allyn & Bacon.
- Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Finelli, C. J., Layton, R. A., Pomeranz, H. R., & Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning & Education, 11*(4), 609–630.
- Paswan, A., & Gollakota, K. (2004). Dimensions of peer evaluation, overall satisfaction and overall evaluation: An investigation in a group task environment. *Journal of Education for Business, 79*(4), 225–231.
- Pfaff, E., & Huddleston, P. (2003). Does it matter if I hate teamwork? What impacts student attitudes toward teamwork. *Journal of Marketing Education, 25*(1), 37–45.
- Podsakoff, P. M., Mackenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common methods bias in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.
- Psenicka, C., Vendemia, W., & Kos, A. (2013). The impact of grade threat on effectiveness of peer evaluations of business presentations: An empirical study. *International Journal of Management, 30*(1), 168–174.
- Rozell, E. J., & Gundersen, D. E. (2003). The effect of leader impression management on group perceptions of cohesion, consensus, and communication. *Small Group Research, 34*(4), 197–222.
- Tiwana, A., Wang, J., Keil, M., & Ahluwalia, P. (2007). The bounded rationality bias in managerial valuation of real options: Theory and evidence from it projects. *Decision Sciences, 38*(1), 157–181.
- Verzat, C., Byrne, J., & Fayolle, A. (2009). Tangling with spaghetti: Pedagogical lessons from games. *Academy of Management Learning and Education, 8*(3), 356–369.
- Wayne, S., & Liden, R. C. (1995). Effects of impression management on performance ratings: A longitudinal study. *Academy of Management Journal, 38*(1), 232–260.
- Weaver, S. T., Ellen, P. S., & Mathiassen, L. (2015). Contextualist inquiry into organizational citizenship: Promoting recycling across heterogeneous organizational actors. *Journal of Business Ethics, 129*(2), 413–428.
- Yanamandram, V., & White, L. (2010). Are inertia and calculative commitment distinct constructs: An empirical study in the financial sector. *International Journal of Bank Marketing, 28*(7), 569–584.

**APPENDIX A: PEER EVALUATION FORM**

Students were informed that they would provide mid-semester formative feedback to team members, conduct a summative end-of-semester PE, and submit that to the instructor (the table below was included in the syllabus). The summative PE would determine 100% of each student’s team project grade. A student could score 0 or a full 100, regardless of the collective grade assigned by the instructor for the team project depending on the PEs. The instrument for providing the formative and summative PE was as follows:

Assign points for each team member (you do not rate yourself) on 100 points (one sheet per team member you evaluate):

Name of team member to whom your feedback is addressed: \_\_\_\_\_

Criteria	SCORE (0-100)
Attendance at team meetings	
Participation in team meetings	
Effort shown outside team meetings	
Demonstrated interest and willingness to work in group situations	
Initiative and leadership demonstrated	
Quantity of contributions to the team project	
Quality of contributions to the team project	
<i>Total on 700 points</i>	Total: _____
<i>Total divided by 7</i>	<i>Percent:</i> _____

Comments for Team member: \_\_\_\_\_

**Avan R. Jassawalla** (PhD, Syracuse University) is a Professor of Management at the School of Business, State University of New York at Geneseo, NY. Her research interests are pedagogy, teamwork and collaboration, product innovation, strategy and implementation, HRM, and social media.

**Hemant C. Sashittal** (PhD Syracuse University) is a Professor of Marketing at the School of Business, St. John Fisher College, Rochester, NY. His research interests are pedagogy, social media marketing, strategy and implementation, and product innovation.