

Leveraging Crowdsourced Peer-To-Peer Assessments to Enhance the Case Method of Learning

Jill Avery

ABSTRACT

Purpose of the Study: Many marketing educators use the case method to help their students strengthen their decision making skills. Rigorous class participation is essential to achieving the learning objectives in case method learning. One challenge for case method instructors is the assessment of students' class participation, particularly in large classes. This article offers a solution that mines the practices of peer-to-peer feedback and crowdsourcing to enhance the assessment of learning in face-to-face class sessions.

Method/Design and Sample: The article outlines a technique used in an MBA marketing course for crowdsourced peer-to-peer assessment of class participation during case discussions, and empirically validates how crowdsourced peer-to-peer assessment compares to students' self-assessment and to the instructor's assessment of class participation performance, based on five years of data ($N=7,025$) across ten sections.

Results: The article demonstrates that crowdsourced peer-to-peer assessment (unlike self-assessment) offers ratings that are highly correlated with instructor assessment and demonstrate strong inter-rater reliability. Results show that crowdsourced peer-to-peer assessments are perceived by students as fair and accurate.

Value to Marketing Educators: Educators can leverage crowdsourcing and peer-to-peer feedback to enhance the assessment of class participation during face-to-face case discussions.

Keywords: Case Method, Peer-to-Peer Assessment, Crowdsourcing, Class Participation, Student-to-Student Interaction

Jill Avery (email javery@hbs.edu) is a Senior Lecturer, Harvard Business School

Many marketing educators use the case method to help their students strengthen their critical thinking and decision making skills and to deliver marketing content in a real world context (Bonomo & Koznik, 1989; Clow & Wachter, 1996; Corey, 1996; Crespy, Rosenthal, & Stearns, 1999). During a typical face-to-face case discussion in the MBA classroom, thirty to ninety students are vying to analyze, generate alternatives, decide, recommend, persuade, constructively criticize, and debate with their classmates about how they would solve the problems facing the case protagonist. Rigorous class participation from all students is essential to achieving the learning objectives in case method learning. The instructor gently guides the discussion by 1.) posing probing questions to draw out students' critical thinking, 2) encouraging debate, and 3.) playing the devil's advocate, while, at the same time, 4.) assessing students on the quality of their class participation.

One of the biggest challenges for instructors using the case method is the fair and thorough assessment of students' class participation, particularly for classes with large numbers of students (Peterson, 2001). This article offers a solution to enhance the assessment of

learning in face-to-face MBA case method class sessions and to better deliver against the key principles of student-to-student interaction outlined in Standard 9 in the AACSB's Accreditation Standards (2012). In this article, I discuss how educators can leverage crowdsourcing and peer-to-peer feedback to enhance the delivery and assessment of class participation during face-to-face case discussions.

The article begins with conceptual foundations based in the pedagogical and psychological literatures. Then, empirical validation from the MBA classroom is provided to show how crowdsourced peer-to-peer assessment compares to students' self-assessments and to instructor assessment of class participation performance. The results demonstrate that crowdsourced peer-to-peer assessment (unlike self-assessment) produces ratings that are highly correlated with instructor assessment, and offers support that these assessments are perceived by students as fair and accurate. Finally, a logistical outline provides detailed instructions and facilitation materials for instructors interested in using the technique.

CONCEPTUAL FOUNDATIONS

Assessing Class Participation in a Case Discussion

The traditional case method, which originated at Harvard Business School, is used by many marketing instructors worldwide (Christensen, 1989). In the traditional case method, instructors “choreograph” case discussions (Rangan, 1995) utilizing a Socratic approach, using questions to encourage students to put themselves in the shoes of the case protagonist and to make the complex real world decisions facing them. Within this pedagogical approach, the instructor does not lecture to achieve the learning objectives, but rather relies on students to inductively derive frameworks and learnings from their own and their classmates’ participation in the discussion. Thus, broad and diverse class participation is critical to achieving the learning goals in a case discussion (Corey, 1996). It is only through rigorous and intense participation that each student experiences the simulated real world experience the case method offers, and thus, learns by doing, acting in the decision making role of a manager, rather than by listening. To learn, students must not only be present, but also be mentally engaging with the course material, struggling with difficult decisions, recognizing the risks in their recommendations, and debating the merits of their arguments. Particularly in disciplines such as marketing where there is rarely a right or wrong answer to the case issues, but rather, where students must struggle to make and justify decisions, the process that students go through to come to a decision in a class discussion is often more important than the final decision itself (Bonoma & Kosnik, 1989). It is this process of active learning, where students are responsible for managing their own learning, rather than relying on an instructor to impart information to them via lecture, that forms the heart of the case method (Corey, 1996).

Active learning pedagogies like the case method are only successful if certain conditions are met. According to Barnes et al. (1994), a case discussion course is a shared partnership in which both students and teacher take responsibility for the learning process. To achieve a successful partnership, the class must move from a loose collection of individual voices at the beginning of the semester to a communal collaborative with shared goals and values by the end. As Ellett (2007, p. 12) reminds, students collectively own the learning that emerges from a case discussion, and those who do not participate or who come unprepared risk the learning for everyone else, “Students provide most of the content of a case discussion. They are indispensable to the creation of knowledge. In fact, if they don’t come to class well prepared, the case method will fail because the people responsible for making meaning from the case are not equipped to do it.” Recognizing the importance of participation to the successful achievement of case method learning objectives, instructors often allocate significant portions of students’ final grades to class

participation. At the world’s leading business schools, MBA students’ grades in marketing courses are often comprised of 25-50% participation. Providing accurate feedback to students about their performance in these settings is essential to delivering a quality learning experience (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Graeff, 1998; Hounsell, 2003; Ramsden, 2003).

Instructors often struggle to assess students’ class participation during a case discussion. Facilitating a case discussion puts instructors under heavy cognitive load, diminishing their resources available for reliable assessment (Chylinski, 2010). Because case discussions do not follow a linear path, but rather primarily take their direction from student comments, case instructors must be flexible and nimble to ensure that learning objectives are met, and must concomitantly orchestrate the discussion and assess students’ class participation.

Instructors often suffer from reconstructive memory effects that often cause them to use faulty heuristics (e.g., attendance) to assess participation when they cannot remember more significant information (Chylinski, 2010). Frequency and recency effects distort instructors’ assessments, and global assessments may dominate over the course of the semester, rather than specific evidence demonstrated each week in the classroom. Subjective biases plague all instructors and efforts to hold them at bay diminish, but do not eliminate them. Particularly for new instructors, grading participation during a case discussion is a challenging endeavor.

Students often misunderstand how class participation is graded in the MBA classroom (Desiraju & Gopinath, 2001; Peterson, 2001). What they see as strong class participation (showing up for class, reading the case, and demonstrating a knowledge of case facts during the case discussion) is often viewed as “table stakes” by case instructors (Peterson, 2001), who value critical thinking, qualitative and quantitative analysis, contrast and comparison, and synthesis of the case’s critical issues, all of which demonstrate metacognition, when grading class participation. Students are thus often resistant to class participation grades provided by their instructors, as students assess their own participation based on frequency rather than quality (Desiraju & Gopinath, 2001). As a result, students often dismiss or discount instructors’ assessments and fail to incorporate the feedback contained within them and, hence, fail to improve their performance.

This lack of understanding also contributes to many of the bad behaviors students exhibit during case discussions, such as hogging airtime, engaging in one-upsmanship with their classmates, and speaking even when they have nothing original to contribute or when they are unprepared (Barnes, et al., 1994). Often, instead of engaging in the active listening that is so critical to successful learning in the case method, students focus inwardly on what they are going to say, rather than processing their classmates’ contributions as part of their learning. Instead of building a collaborative communal learning

environment, students tend to focus on themselves, rather than the group's learning trajectory.

The Problem with Self-Assessment

One way instructors have dealt with the challenge of assessing students' class participation during case discussions is to have students assess their own performance. Peterson (2001), for example, espouses a pedagogical technique by which students take responsibility for documenting their own course participation. The use of student portfolios in which students document their work to provide evidentiary support for learning outcomes is gaining traction (Arter & Spandel, 1992). Studies show students are more invested in their participation grade when they can assess their performance rather than relying on the instructor's grade (Peterson, 2001).

However, it may be difficult for students to objectively assess their own class participation, due to self-serving biases that hamper their perspective. People suffer from positive illusions (Taylor & Brown, 1988) that make it difficult to offer an objective view of their own behaviors. These illusions enable people to overemphasize their positive qualities while underemphasizing their negative qualities. People believe they are more capable than they actually are, and they attribute their successes in life to their innate abilities, while attributing their failures to external factors (Zuckerman, 1979). As Greenwald's (1980) classic study shows, "[the self] distorts the past and manipulates the present. Whatever makes it look good is exaggerated. Whatever makes it look bad is downplayed, discredited, and denied." (Baumeister, 1998, p. 690). Thus, asking students to self-assess their level of participation is likely to yield ratings that are artificially high.

Leveraging Peer-to-Peer Assessments to Assess Class Participation

Peer assessment is another alternative to instructor assessment. Marketing instructors have embraced peer assessment in online learning environments, leveraging the common practice of peer assessment on the digital platforms students use outside of class, such as Facebook and YouTube. Massive open online courses (MOOCs) commonly use peer feedback to manage the unwieldy task of grading the thousands of students enrolled in the course. Many instructors in blended learning environments ask students to assess each other's online posts in discussion forums by "liking" them when the post contributes to their learning of the course material. Others ask students to rate their peers' contributions on more rigorous quality rating scales. Others quantitatively track the social response to discussion posts, measuring the number of replies a post generates as a behavioral response metric that measures the impact of the post.

Marketing instructors also use peer assessments in face-to-face courses. One area that is popular with instructors is to use peer assessments to help grade small group work, such as project teams working on marketing plan projects. In this arena, peer

assessments have been demonstrated to decrease social loafing, to increase student satisfaction with the experience of working in groups, and to raise the fairness perceptions associated with a group grade (Aggarwal & O'Brien, 2008; Poddar, 2010). Duverger and Steffes (2012) found that peer ratings of a student's creative project work were no less rigorous nor stringent than instructors' ratings.

Some early pedagogical studies in diverse fields, including management, deemed that peer assessment was not a reliable substitute for instructor grading (c.f. Gopinath, 1999) and that students did not like it. However, others showed strong correlations (in the range of 0.72-0.9) between peer and professor ratings (Burchfield & Sappington, 1999; Melvin, 1988); while still others showed that students' performance improved when they knew that they were being assessed by their peers (Rust, Price, & O'Donovan, 2003). A review of the relevant literature demonstrates these contrasting views; 18 out of 31 studies across a diverse set of disciplines (including management, psychology, engineering, math, optometry, social science, art and design) and across diverse types of assignments (including writing, presentations, group work, creative output, and other skilled professional behaviors), deemed peer assessment a reliable alternative to instructor assessment, while seven studies deemed it a poor substitute with low reliability and validity (Topping, 1998).

However, in the time since many of these empirical investigations have been conducted, the world has drastically changed with the rise of today's digital culture, which may make peer assessment a more effective and palatable pedagogical technique than ever before. Peer feedback forms the backbone of the new digital economy, where social impact, as measured by the response of one's peers, is often the measure of the quality of content and guides people's choices of what to read, watch, and buy (Li & Bernoff, 2007). On the web, people constantly evaluate content by voting with a click of their mouse, using the "Like" button on Facebook to indicate their approval of a friend's post, re-tweeting information that they find useful on Twitter, rating books on Amazon, and giving videos on YouTube a thumbs up or thumbs down. People have emerged as ardent arbiters, raters, rankers, and commentators, providing authoritative judgment and critique of each other's content. The content that makes up today's culture is judged and authenticated by the masses; cultural populism determines the value of information (Qualman, 2009).

Thus, today's students, as digital natives, may be more willing and better equipped than those that have come before them to assess their peers, given their immersion in this new digital culture with its focus on peer evaluation and feedback. As Davidson cautions, "Education is way behind just about everything else in dealing with these [media and technology] changes. It's important to teach students how to be responsible contributors to evaluations and assessment. Students are contributing and assessing each other on the Internet anyway, so why not make that a part of

learning?"(Hendry, 2009, p. 2). Educators today have the opportunity to rethink the findings raised by prior empirical studies of peer assessment that showed mixed results. Today's students may be more open to peer assessment than students of the past because they have been conditioned to assess each other through their immersion in social media.

Some marketing instructors remain skeptical that students can accurately and objectively assess each other's performance during a case discussion in the rich social environment that a face-to-face class offers, without the bias that arises from interacting live with others in shared physical space. Students, too, are uneasy about peer-to-peer assessment, fearing that their peers are not capable of providing fair and informed feedback, which lowers their confidence in the result (Bloxham & West, 2004). Instructors fear that students may unfairly bring grudges and other biases from outside the classroom into their ratings. Social psychological research has documented a strong in-group favoritism (and resulting out-group derogation), where people evaluate the actions of those that are similar to them more favorably than the actions of those perceived as being different (Tajfel & Turner, 1979). Gender, ethnic, and age stereotypes and their associated prejudices may color the ratings of students, leading to over or under-inflation of ratings as students discriminate against those who are unlike them (Fiske, 1998). Students, like instructors, may also suffer from frequency, recency, and other reconstructive memory effects that bring bias into their assessments of others.

Using Crowdsourcing to Mitigate Bias in Peer Assessments

Crowdsourcing may provide a solution to the issue of bias in peer assessments. Crowdsourcing, soliciting the opinions and feedback of individuals and aggregating the information gleaned from each of them into a mean or median estimate, has become a popular way to harness the wisdom of crowds to enhance prediction and assessment tasks. As Surowiecki (2004, p. xiii) notes, "groups are remarkably intelligent, and are often smarter than the smartest people in them. Groups do not need to be dominated by exceptionally intelligent people in order to be smart. Even if most people within a group are not especially well-informed or rational, it can still reach a collectively wise decision." Groups outperform the vast majority of individuals, even experts, on decision making tasks, by harnessing the collective intelligence of their members. Mauboussin (2007) explains that "a diverse crowd will always predict more accurately than the average individual. So, the crowd predicts better than the people in it...the collective is often better than even the best of the individuals."

Crowdsourcing helps eliminate the bias that exists in any individual estimate or assessment by balancing it out against the biased estimates of others. When the ratings of a diverse group of assessors are aggregated together and averaged, mathematical theory, specifically the diversity prediction theorem (Page,

2007; Treynor, 1987), predicts that each person's errors will be canceled out by the errors of others, assuming that the assessors are thinking and acting independently and that their biases are personal and not systematic. For example, in making an assessment of class participation, each individual assessor relies upon both concrete data and his/her subjective perceptions, e.g., how much someone participated, the quality of their comments, and/or how much their comments drove the discussion forward. Their rating, thus, is informed by both objective information and subjective, potentially biased, information. The subjective part of the rating may be thought of as the error portion of the rating, i.e., the deviance from an objective rating. Thus, if we crowdsource peer-to-peer assessments by allowing a number of students to assess each student's performance, we can mitigate some of the bias that will creep into any one individual's assessment.

As the size of the rating group increases, the idiosyncratic biases of individuals get canceled out and extreme values become less influential. Therefore, we do not need each individual assessor to be a perfect rater. To crowdsource class participation assessments to yield the wisdom of crowds, one must ensure diversity of thought in the rater pool and rater independence to ensure that errors are uncorrelated (Surowiecki, 2004). As long as we establish conditions to ensure that students complete their rating tasks independently of each other, our "imperfect raters" can collectively come up with a more accurate picture of class participation than any one individual can alone.

Below, a pedagogical technique for crowdsourced peer-to-peer assessment of class participation during case discussions is outlined. Empirical validation from the MBA classroom is provided to support how crowdsourced peer-to-peer assessment compares to students' self-assessment and to instructor assessment of class participation performance. Following that, a more thorough discussion of the technique is provided that outlines specific procedures for instructors interested in using it in an MBA course.

EMPIRICAL VALIDATION

In this section, the outcome of crowdsourced peer-to-peer assessment of class participation is assessed and its reliability is compared to self-assessment and instructor assessment. The source of data is instructor-, peer-, and self-assessments from students in ten sections of a required MBA marketing management course at a small, AACSB-accredited business school in the United States. The course was taught face-to-face once per week for three hours for 14 weeks and sections included full-time (5 sections) and part-time (5 sections) MBA students. The instructor evaluated student participation in case discussions after each class session. The mean of these weekly evaluations was used as the point of comparison for the instructor assessment. At mid-semester, students were required to complete a self-

assessment and a peer-to-peer assessment, using a rubric. The grading rubric covered four content areas that students rated on five point scales: analytical quality of contributions, tone of contributions, level of engagement in case discussions, redundancy of contributions, and frequency of contributions (see Table 1). These rubric items were selected and pretested in the classroom by three multidisciplinary case method instructors prior to their use in the current study and were designed to capture the learning outcomes instructors desire from class participation during a case discussion: mental engagement and

active listening, analytical and critical thinking to make meaning from data, the active use of comparison and contrast during debate to empirically derive general principles from diverse situations, decision making under uncertainty and within complex conditions, and effective and authoritative persuasiveness (Barnes, et al., 1994; Corey, 1996). They were also designed to mitigate detrimental student behaviors such as hogging airtime, making redundant or tangential comments that fail to move the discussion forward, and creating an unsafe, unprofessional classroom atmosphere by disrespecting others.

TABLE 1: GRADING RUBRIC FOR ASSESSMENT OF CONTRIBUTION TO GROUP LEARNING

Grading Criteria	1-2 = Deficient	3 =Acceptable	4-5 = Exemplary
Analytical Quality of Contributions	Talks loosely, sometimes pulling facts from the case, theoretical materials, and relevant experience, and sometimes not. Contributes facts, but no analysis to the discussion. Does not take a point of view or make recommendations. Never takes and defends positions that are different from the positions of others in the class.	Draws comments from facts of case and/or talks about articles or relevant experience, but does not integrate across. Contributes only qualitative analysis to the case discussion. Makes recommendations, but analysis is sometimes lacking. Sometimes takes and defends positions that are different from the positions of others in the class.	Analyzes facts of the case and integrates marketing theory into discussion; provides relevant insights from experience. Contributes both qualitative and quantitative analysis to the discussion. Always takes a point of view or makes a recommendation backed up by analysis. Frequently takes and defends positions that are different from the positions of others in the class.
Tone of Comments	Talks down to people or speaks disrespectfully. Is not able to respectfully disagree with others or does so in a non-constructive manner.	Talks with hesitation and uncertainty. Sometimes is able to respectfully disagree with the recommendations of others in a constructive manner.	Speaks in a collegial tone but with authority. Frequently is able to respectfully disagree with the recommendations of others in a constructive manner.
Level of Engagement	Not prepared for the case discussions. Does not analyze the case and other readings. Appears disconnected from the conversation, engages in class-unrelated activities, or leaves the room during case discussion.	Somewhat prepared for the case discussions. Adequately analyzes the case and other readings. Engagement ranges from active to inactive but present.	Always well prepared for discussions. Creatively and thoroughly analyzes the case and other readings closely. Actively involved throughout the class.
Redundancy	Repeats perspectives already discussed and/or is unnecessarily long-winded. Contributions never build upon, synthesize, or otherwise recognize the comments of others. Comments never encourage other people to jump in by introducing a new topic or area of analysis.	Sometimes contributes new perspectives but is at other times repetitive; is sometimes succinct and at other times long-winded. Contributions sometimes build upon, synthesize, or otherwise recognize the comments of others. Comments sometimes encourage other people to jump by introducing a new topic or area of analysis.	Contributes new perspectives and/or insights and keeps comments to the point and succinct. Contributions always build upon, synthesize, or otherwise recognize the comments of others. Comments frequently encourage other people to jump by introducing a new topic or area of analysis.
Frequency of Contributions	Rarely speaks OR dominates the discussion.	Talks sometimes.	Speaks frequently but does not dominate the conversation.

Two hundred and eighty six students participated in the study and generated a total of 7,025 peer ratings (an average of 25 peer ratings per student). The mean of the numerical ratings given to a particular student

across all peer raters was used as the point of comparison for peer-to-peer assessment. All assessments were completed on 5-point scales, with 1=deficient and 5=exemplary. For the peer

assessment, students were permitted to assign any rating to any student; no forced distributions were enforced. All students and the instructor were blind to other ratings until all assessments were completed. Following the completion of the course, a random sample of 150 students across sections was asked to voluntarily complete an online survey to express their opinions and perceptions of the peer-to-peer assessment. The questions were derived from previous research (Orsmond & Merry, 1996; Ryan, Marshall, Porter, & Jia, 2007). Of those asked, 47% completed the survey (N=71).

Table 2 summarizes the descriptive statistics of the data used to conduct the analysis. A look at the histograms (see Figure 1) and skewness/kurtosis statistics shows that the self-assessments demonstrate significant right skewness and kurtosis, while the instructor and peer assessment ratings do

not deviate significantly from the normal distribution. The analysis began with t-tests to analyze the differences between the means of pairs of raters (i.e., peer vs. instructor, peer vs. self, and self vs. instructor). The means for the peer and the instructor groups only exhibited significant differences in two of the ten sections, with eight sections showing no significant differences between the average peer rating and the instructor's rating. In contrast, the means for the self and instructor groups exhibited significant differences in seven of the ten sections, and the means for the self and peer groups exhibited significant differences in seven of the ten sections. This suggests that students' ratings match the instructor's ratings more closely when the students are rating their peers than when they are rating themselves.

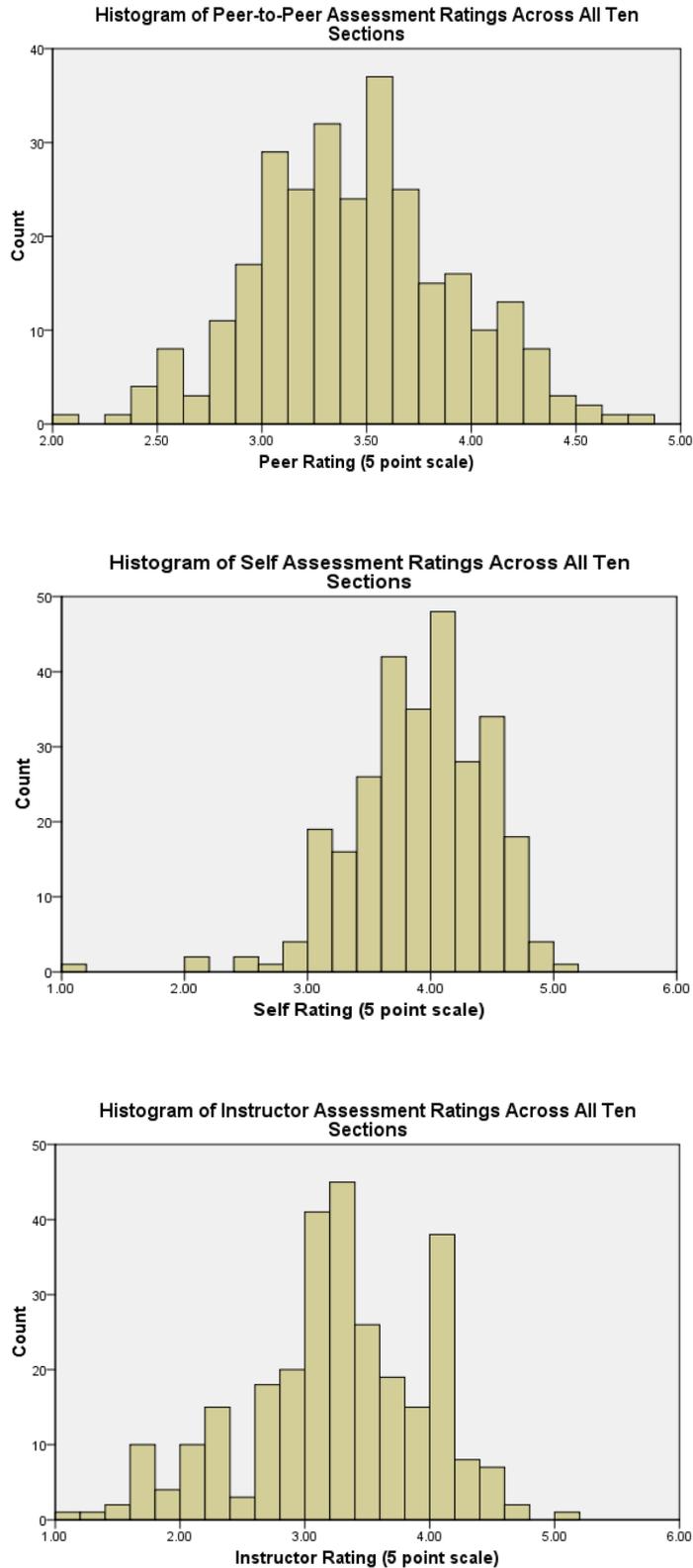
TABLE 2: DATA DESCRIPTIVES AND COMPARISON OF MEANS

Section Number	n	Mean Rating and Standard Deviation						Mean Difference ^a		
		Self		Peer		Instructor		Self vs. Instructor	Peer vs. Instructor	Self vs. Peer
		Mean	SD	Mean	SD	Mean	SD			
1	10	3.80	(.42)	3.79	(.54)	3.86	(.44)	-0.06	-0.07	0.00
2	23	3.70	(.39)	3.54	(.37)	3.62	(.51)	0.07	-0.08	0.15
3	21	4.35	(.43)	3.71	(.56)	3.43	(.75)	0.92***	0.28	0.64***
4	20	4.18	(.51)	3.52	(.38)	3.34	(.58)	0.84***	0.18	0.65***
5	26	4.00	(.59)	3.36	(.39)	3.33	(.76)	0.67**	0.03	0.64***
6	47	3.71	(.41)	3.23	(.32)	3.16	(.63)	0.55***	0.06	0.48***
7	48	3.88	(.46)	3.39	(.41)	3.06	(.69)	0.82***	0.34**	0.48***
8	33	4.07	(.46)	3.28	(.42)	3.04	(.72)	1.03***	0.24	0.79***
9	30	3.85	(.70)	3.43	(.54)	2.94	(.80)	0.91***	0.49**	0.42*
10	28	3.62	(.71)	3.69	(.61)	3.39	(.79)	0.23	0.30	-0.07

^a Statistical results based upon t-tests on the mean differences between groups.

^b *** t-statistic is significant at the p < .001 level, ** t-statistic is significant at the p < .01 level, * t-statistic is significant at the p < .05 level.

FIGURE 1: HISTOGRAMS OF SELF, PEER, AND INSTRUCTOR RATINGS



Then, the interrater reliability of the students as raters was analyzed. Reliability analysis was used to calculate the Cronbach's alpha of the peer raters in each section to see if they were consistent in the way

they rated each peer, following the method outlined in Duverger and Steffes (2012). Across all sections, Cronbach's alpha exceeded .85, above the standard for judging interrater reliability (see Table 3). Hence,

MBA students demonstrate internal reliability when rating their peers on class participation.

TABLE 3: INTERRATER RELIABILITY ANALYSIS

Section Number	n	Cronbach's Alpha
1	10	0.914
2	23	0.905
3	21	0.952
4	20	0.858
5	26	0.911
6	47	0.954
7	48	0.963
8	33	0.931
9	30	0.948
10	28	0.943

Correlation coefficient analysis was used to determine the accuracy and/or reliability of these methods versus an instructor assessment, using the instructor's assessment as a proxy for an objective assessment. Table 4 summarizes the results of the Pearson's correlation coefficient statistical tests. First, the data show that there is a significant and positive correlation between crowdsourced peer-to-peer assessments and instructor assessment of class participation performance in all ten sections with Pearson's correlation coefficient values ranging from .59 to .84. Crowdsourced peer assessments yield ratings that are highly correlated with the ratings of the instructor. Second, the data show that there is a significant and positive correlation between self-assessments and instructor assessment of class participation performance in only five of the ten sections. Pearson's correlation coefficient values

range from .01 to .73, with five sections failing to demonstrate statistically significant correlation. In an earlier study, Burchfield and Sappington (1999) found a similar result, showing no correlation between a professor's grade and a student's self-assessment. Consistent with psychological theory, students' self-ratings are overinflated versus the instructor's assessments, sometimes by as much as 34%, indicating the effect of positive illusions. There is also less correlation between self-assessments and crowdsourced peer-to-peer assessments of class participation performance. Significant correlations between self- and peer assessments were only found in five of the ten sections with Pearson's correlation coefficient values ranging from .02 to .60. Collectively, peers are more capable of objectively assessing a student's performance than that student is himself or herself.

TABLE 4: CORRELATION ANALYSIS

Section Number	N	Pearson's Correlation Coefficient (<i>r</i>)		
		Self vs. Instructor	Peer vs. Instructor	Self vs. Peer
1	10	0.17	0.77**	0.25
2	23	-0.06	0.70**	0.02
3	21	0.73**	0.84**	0.60**
4	20	0.44	0.68**	0.47*
5	26	0.41*	0.76**	0.38
6	47	0.38**	0.59**	0.43**
7	48	0.44**	0.80**	0.46**
8	33	0.33	0.70**	0.28
9	30	0.68**	0.81**	0.56**
10	28	0.01	0.76**	0.07

** Correlation is significant at $p < .01$. * Correlation is significant at $p < .05$.

Interesting patterns emerge from the data. An analysis of the residuals shows that both peer and self-ratings are higher, on average, than instructor ratings. Peer ratings are, on average, .21 points higher ($SD = .50$) than instructor ratings, while self-ratings are, on average, .64 points higher ($SD = .74$) than instructor ratings. A new variable was computed and labeled "residual valence," which indicated whether the residual value of a peer or self-rating versus the instructor's rating was positive or negative, with positive values indicating that the peer assessor or self-assessor rated higher than the instructor, and negative values indicating that the peer assessor or self-assessor rated lower than the instructor. Ratings that exhibited no residual, i.e., the peer or self-rating exactly matched the instructor rating, were excluded from the analysis. Descriptive statistics showed that peer assessors rated students higher than the instructor 66% of the time, while self-assessors rated themselves higher than the instructor 84% of the time.

To understand whether peer assessors rated highly performing students differently than poorly performing students, the degree of correlation between the residual valence and the instructor's rating was tested, using the instructor's rating as a proxy for actual performance of the student. A Pearson's correlation coefficient statistical test showed that there is a significant and negative correlation ($r = -.547$, $p < .01$) between peer-to-peer assessments and instructor assessment and actual class participation performance. This indicates that peer raters tend to rate higher than the instructor for poorly performing students, and lower than the instructor for well performing students. Peers rating poorly performing students may be inflating their ratings to protect the grades of their classmates. Peers rating well performing students may be deflating their ratings, perhaps because they view class participation as a zero-sum game and, thus, bring a competitive mindset to their assessment (Aylesworth, 2008; Gonzalez, Ingram, LaForge, & Leigh, 2004). This would cause them to assess each other too harshly in an effort to save their own grades. This result is consistent with an earlier empirical study by Sadler and Good (2006) in which they found that student raters gave significantly lower grades to top performing students than the instructor did.

Similarly, to understand whether self-assessors rated highly performing students differently than poorly performing students, the degree of correlation between the residual valence and the instructor's rating was tested. A Pearson's correlation coefficient statistical test showed that there is a significant and negative correlation ($r = -.460$, $p < .01$) between self-assessments and instructor assessment and actual class participation performance. This indicates that self-assessors tend to rate higher than the instructor when they themselves are poorly performing students, and lower than the instructor when they themselves are well performing students. This result is consistent with that found by Falchikov (1986) in his study of psychology students, where stronger students tended

to under mark themselves, while average students tended to over mark themselves. Poorly performing students may be inflating their own ratings to protect their grades or they may be suffering from self-serving biases. The more interesting result is that well performing students appear to be undervaluing their own performance. A similar result was found by Saavedra and Kwum (1993) who demonstrated that the best performers on work teams were excellent peer assessors, but less reliable at rating themselves. This is contrary to what psychological literature on positive illusions would predict and, perhaps, suggests that good students are using another self-presentation tactic, sandbagging (i.e., pretending to be less than you are) (Shepperd & Socherman, 1997), or modesty, to curry sympathy or favor with the instructor. Gibson and Sachau (2000) demonstrate that sandbagging is not limited to competitive settings, but rather can appear as a tactic used to influence an evaluator. Modesty is a trait that allows people to maintain perceptions of competence, while increasing others' liking of them (Stires & Jones, 1969). There is an opportunity for future research to further explore the motivations behind this undervaluation by top performers.

Second, the correlation between peer-to-peer assessments and instructor assessments were higher for the full-time day sections ($r = .776$) than the part-time night sections of the course ($r = .706$). Night students seemed to have a harder time accurately assessing students who were absent; their ratings of these students tended to be higher than the instructor's. While absenteeism was negligible in the day sections, it was more prevalent in the night sections, as part-time students were juggling school and job commitments and so needed to miss class sessions more often. Peers did not appear to judge absences as harshly as the instructor, perhaps because they were not actively tracking them as rigorously as the instructor was.

Students' perceptions of fairness are an important determinant of whether they are receptive to peer-to-peer assessment and find value in the results. The student survey data (see Table 5) indicates that the majority of students found that the peer-to-peer assessment was fair in helping to determine their grade and that their peers evaluated them fairly. However, only 20% agreed that the peer assessment was more accurate than the professor's assessment, while 47% disagreed, indicating that students still have more faith in the instructor's ability to objectively grade their class participation than in their peers' abilities. This may stem from the perceived difficulty of the peer rating task; 38% of students disagreed that it was easy to evaluate their peers and many of the open ended comments discussed how difficult it was to assess peers, especially in large classes and especially when rating those students who did not speak frequently. It may also stem from the uncomfortable emotions peer assessment generates, as several students noted in the open ended responses to the survey. It was also sometimes difficult for students to accept peer

feedback; as one student noted in the survey, “The truth hurts.” This difficulty was also reported by psychology students in a peer assessment study conducted by Falchikov (1995). In that study, students reported experiencing cognitive challenges, strain, and social embarrassment from peer assessment. Falchikov suggested that social embarrassment might be heightened for students in smaller groups or in groups that have a long tenure together. In this article’s study, students in half of the sections were day students who were grouped in a cohort with whom

they took all of their required courses. The course was taken in their second semester together, so the “groupiness” of the section was well established. This may have heightened their social anxiety. Students in the remaining five sections were night students who were randomly assigned to the course and had not been together as a unified group prior to the course and were less likely to be together in the future. This may have made the peer assessment task socially easier for them.

TABLE 5: STUDENT OPINION SURVEY RESULTS (n=71)

Question	Mean	SD	Disagree (%)	Agree (%)
I feel that the peer-to-peer assessment was fair in helping to determine my class participation grade.	3.54	(1.2)	21.1%	63.4%
I participated more because I knew that my peers were evaluating me.	3.10	(1.3)	35.2%	45.1%
I found that my peers assessed me fairly and accurately.	3.55	(1.0)	19.7%	63.4%
I found that the professor assessed me fairly and accurately.	4.34	(1.0)	8.5%	85.9%
I found it easy to evaluate my peers on their contribution to group learning.	3.20	(1.3)	38.0%	50.7%
I would recommend using peer-to-peer assessment in this class in the future.	3.69	(1.3)	22.5%	63.4%
I found the peer assessment to be more accurate than the professor's assessment.	2.59	(1.0)	46.5%	19.7%
Having the peer-to-peer assessment improved my class participation in the second half of the course.	3.23	(1.2)	26.8%	50.7%

Questionnaire Scale: 1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Neutral, 4 = Somewhat Agree, 5 = Strongly Agree

The reported anxiety is an indication that providing and receiving feedback is a skill set with which MBA students need further practice; it is not something that comes naturally to them and must be developed and practiced before they can demonstrate proficiency with it in the workplace as a manager. Instructors using peer assessment should be prepared to spend time during class discussing the emotional aspects of providing and receiving feedback; ideally, this would happen prior to the rating task so that students can be better prepared for the perceived difficulties. Topping recommends that a “collaborative and trusting ethos [be] fostered,” to reduce anxiety (Topping, 1998, p. 265). Instructors may also want to explain the concept of crowdsourcing and how it can be used to reduce biased judgments. This may make students more comfortable with the rating task and also more accepting of its results.

Despite their misgivings about the accuracy of their peers’ ratings of them, half of the students were motivated to improve their class participation performance following peer assessment, with 51%

agreeing that having their peers assess them improved their participation in the second half of the course (with 27% disagreeing). Indicating a greater level of support for peer-to-peer assessment than students involved in previous studies in other fields, 63% of the students recommended using peer-to-peer assessments in the marketing management class in the future, while 23% did not. This compares quite favorably to students in the Ryan et al. study (2007), only 10% of whom recommended using peer-to-peer assessments in a future course, and may reflect changing attitudes towards peer assessment and crowdsourcing due to the cultural changes brought by digital culture.

CROWDSOURCED PEER ASSESSMENT IN MARKETING CASE DISCUSSIONS

This section outlines specific procedures¹ for how instructors can use crowdsourced peer-to-peer

¹ These procedures were followed across all ten sections in the empirical validation.

assessment in their marketing courses. Successfully executing crowdsourced peer-to-peer assessment requires three steps.

Step 1: Establishing a Collective Classroom Culture Where Crowdsourcing is Valued. During the first class session of the semester, it is essential for the instructor to establish collective ownership of the learning experience for the course. A brief primer on the case method of learning is useful to communicate the importance of active learning and student-led exploration. Linguistically, the instructor can set a collective tone by re-labeling “class participation” as “contribution to group learning,” to demonstrate the desire for students to contribute content during case discussions that is of value to other students in their learning process, rather than just participating to get their voices heard by the instructor. This shifts the emphasis from a process measure (speaking during a case discussion) to an outcome measure (helping other students and yourself learn). This distinction is essential for students to understand upfront, as it conveys the transference of responsibility for teaching, learning and assessment from the professor to the collective body of students. When students embrace a collective culture, their attention during case discussion quickly shifts from pleasing the professor to enriching the learning experience of their classmates. The affirmation of a job well done comes not from the instructor, but rather from their peers. Rather than speaking to earn a check in the professor’s grade book, students craft their comments in ways that help bring the case discussion to a higher level of quality. The peer-to-peer assessment method outlined in this paper is designed to switch students’ focus from “class participation” which is about measuring how much I speak to “contribution to group learning” which is about measuring how much my peers hear me and learn from me. Using peer assessment reminds students that their audience is their fellow students, not the case instructor, during a traditional case discussion.

Step 2: Conducting the Crowdsourced Peer-to-Peer Assessment. At certain points during the semester, each student can be required to provide a numerical rating for each of his/her classmates based on an assessment of each classmate’s contribution to group learning. Instructors can customize the ratings task to take into account the size of the class. In the empirical validation above, the number of raters per student varied from nine to forty-seven and results held up across these varying group sizes. Within this range of values, peer-assessments were significantly correlated with instructor assessments. Instructors may want to use caution if using less than nine raters, as the benefits of crowdsourcing may diminish at lower levels. However, in a classic study of crowdsourcing, Gordon (1924) found that the benefits of crowdsourcing exist even at very small group sizes, although the efficacy of the crowd diminishes as the number of people in the group falls. Gordon’s students were individually asked to do a task that involved

arranging a series of ten weights in weight order. The researcher then grouped the students’ estimates into groups of varying size and calculated the correlation between the group’s average estimate and the actual true order. She found that aggregating the estimates of a group of 50 students yielded a correlation of .94, aggregating the estimates of a group of 20 students yielded a correlation of .86, aggregating the estimates of a group of 10 students yielded a correlation of .79, and aggregating the estimates of a group of 5 students yielded a correlation of .68 with the actual true order.

Instructors may also want to use caution if using or more than forty-seven peer raters, as, at higher levels, students may be unable to track the performance of too many others. As class size increases, it becomes increasingly difficult for students to rate all of the other students in the class. For smaller classes, instructors may ask each student to rate everyone. However, for larger classes, instructors may want to randomly assign students to assess a small number of their peers. To achieve the benefits of crowdsourcing, it is critical to maintain a number of multiple raters for each student to minimize the effect of outlier opinions and to assure that errors in judgment cancel each other out. Instructors must strike a comfortable balance between having enough raters to take advantage of the mathematical benefits of crowdsourcing as a way to allow the uncorrelated errors of biased estimates to cancel each other out, and not having raters rate too many students so that they tire of the task. Randomly assigning students as raters also diminishes the chance that their errors will be correlated.

A comprehensive rubric should be provided to students at the beginning of the semester and during each peer rating task. The rubric is used to ground peer ratings in the measurement of explicit behaviors desired by the instructor and to avoid global assessments that may be biased by stereotypes or personal relationships outside the classroom. The use of a detailed rubric is essential to obtaining reliable and consistent peer assessments (Orsmond & Merry, 1996).

To accompany the numerical rating, students can be asked to provide qualitative feedback to each student that helps illuminate the quantitative assessment. The qualitative feedback can be structured by two prompts: *Areas of Strength* that the student should continue to capitalize on, and *Areas of Weakness* that the student should work to address. This step serves the function of forcing students to provide rationale for their numerical rating. As Peterson (2001) found, forcing students to document evidence supporting their assessment is key to generating high quality assessments, as students often rate themselves highly, but then fail to produce concrete documentation for their rating.

One challenge faced by instructors using peer assessment is that students may fail to discriminate among different levels of performance, thereby rating all students with a similar value, leading to ratings that show little variance across students (Pond, Uihag, &

Wade, 1995). This pattern may be a reflection that the rater does not deem himself/herself sufficiently qualified or motivated to make distinctions among students. This may arise from rating fatigue when students are asked to rate too many peers and tire of the task, resorting to a common rating across all students. It may also arise when students feel unprepared to assess their fellow classmates, either due to the lack of a clear rubric by which to assess their performance, or due to a lack of their own engagement in active listening during case discussions. Students who are frequently absent or distracted during case discussions, or those who are solely focused on their own contributions, will often be unable to generate quality ratings. To solve this problem, instructors may wish to force students to assign grades in a normally distributed pattern or to limit students to a certain number or percentage of high and low grades, so that the assessments are dispersed across the grading scale. This eliminates the lack of variance problem. This is particularly important if the instructor has similar limitations on assigning grades and is using a forced distribution. In this dataset, a small minority of raters exhibited these patterns. Oftentimes, these raters were people who were less engaged in the course (frequently absent or physically present, but mentally disengaged). In a very few instances, these raters were actively fighting against the idea of peer-to-peer assessments by refusing to differentiate among their classmates as a matter of principle because they thought it was unfair to judge other people.

The second challenge that may arise is when all raters rate a particular student with the same rating (i.e., where student X's ratings from all of the raters demonstrates no variance). Instructors who are experiencing this problem may want to consider a more expansive rating scale, e.g., a 5 or 7-point scale versus a 3-point scale. In this empirical validation, a 5 point scale was sufficient to generate variance among ratings for MBA students. Across sections, the students interpreted the 5-point scale as centered around a grade of B (i.e., 3 = B) with higher numbers documenting above average performance and lower numbers documenting below average performance. The computation of an inter-rater reliability measure (i.e., Cronbach's alpha) can give instructors visibility as to the level of consistency of relative ratings across raters, while a visual inspection of the intra-student standard deviation can identify cases in which inter-rater variance is low.

Students should then submit their ratings to the instructor, who tabulates the numerical ratings into a crowdsourced average for each student. The instructor then can communicate each student's average and provide a compilation of the qualitative comments received. Importantly, this feedback should be delivered anonymously to the students, so that individual ratings and comments cannot be traced back to a particular classmate. This addresses another problem faced by instructors who use peer assessments. Students may be hesitant to accurately

rate the performance of a fellow student, due to social pressures that arise from being classmates and facing each other in the classroom each week. Today's students feel pressured to maintain high grade point averages, so peers may feel uncomfortable providing a low rating for a student that might hurt his or her course grade. Peer ratings, therefore, may be inflated, as students try to play nice and preserve the grades of their classmates. Grade inflation has been documented in other studies of student peer assessment (Brindley & Scofield, 1998; Falchikov, 1995; Orsmond & Merry, 1996; Topping, 1998) and studies suggest that peer assessments will supply higher ratings than professor evaluations. Keeping their ratings confidential frees students from many of the social fears that can bias their results, such as the fear of social retribution for their assessment of a classmate. However, making each student identify themselves as the rater to the instructor allows the instructor to have visibility of how each student has rated the others, providing an incentive for students to put time and effort into the rating task. This serves several important functions. It eliminates the tendency of students to provide sloppy or unconstructive feedback, as the instructor will review their rating work, and it allows the instructor to assess intra-rater reliability.

Step 3: Tabulating the Data and Identifying and Dealing with Outliers.

The final step in the process involves the instructor checking the input data, tabulating and disseminating the results. An important sub-process in this step involves identifying and dealing with data problems and outliers. First, the instructor should review the intra-rater reliability of each student to judge their performance as a rater. A key warning sign of low intra-rater reliability is no to little variance among ratings (e.g., the rater just chooses a rating of 4 for every student). The class attendance of these raters should be assessed; students who are frequently absent may be ill equipped to rate their fellow students' performance. If intra-rater reliability is deemed to be low, the instructor can decide to remove the rater's data from the pool to be analyzed. Second, the instructor should assess the normality of the ratings and identify any extreme outliers (>3+ standard deviations from the mean) that may distort the averages. Charting a histogram of the data is helpful to determine the shape of the distribution. If the size of the class is large enough, most outliers can be safely ignored, as their effect will be drowned out by the ratings of other students; however, in smaller classes, it may be best to eliminate extreme outliers before calculating the average to limit their undue impact on the results.

DISCUSSION

Marketing educators can leverage crowdsourcing and peer-to-peer feedback to enhance the assessment of class participation during face-to-face case discussions and to support professors trying to deliver

an active learning environment via the case method. Crowdsourced peer-to-peer assessments solve many of the problems that plague case method instructors in the marketing classroom: 1.) how to reliably assess students' class participation while under the heavy cognitive load of leading a case discussion, 2.) how to mitigate personal bias in the assessment of student performance, 3.) how to encourage students to be open to and accept their class participation grades as fair and accurate constructive feedback, 4.) how to engage students in active listening during case discussions, 5.) how to shift students' perspective from the "I" to the "we" during case discussions, and 6.) how to help build a collaborative active learning environment.

First, crowdsourced peer-to-peer assessment, unlike self-assessment, offers ratings that are highly correlated with instructor assessment and demonstrate strong inter-rater reliability. Leveraging peer assessments can take some of the pressure off of the instructor to grade class participation in the moment while conducting a case discussion. They can also serve as an additional data point if the instructor is unsure about his/her own rating of a particular student. The qualitative comments that students provide can help instructors provide detail behind a numerical rating to further flesh out the feedback the instructor provides to an individual student on their performance, providing a more robust description of performance and suggestions for improvement.

Second, crowdsourced peer-to-peer assessments are perceived by students as fair and accurate. One benefit of using both peer and instructor assessments is the ability to boost students' receptivity to the class participation grade. If peer and instructor ratings are similar, students may perceive greater reliability in the outcome (Gopinath, 1999) and therefore, more fully incorporate the feedback into their learning process. As one student noted in a survey response, "I enjoyed having the feedback from the professor and from my fellow students. It was a nice check and balance on something that can often feel a bit arbitrary. It also took the adversarial nature of grading away from the professor/student relationship." From the instructor's perspective, the crowdsourced nature of peer assessments helps mitigate the bias inherent in any individual's rating of class participation, including the instructor's own biases, making them more accurate than an assessment received from any one particular individual.

Third, the process of having students complete peer assessments helps instructors generate a collaborative active learning culture among students that is critical to achieving the learning objectives in a case method course. As outlined in the above sections, an important first step in the process is establishing a collective classroom culture, where students better understand how their contributions to a case discussion help others and themselves learn. Shifting students' attention from an individual level of analysis ("Did I speak enough?"), to a collective level of analysis ("How well did we achieve group learning

today?"), is a critical part of shifting responsibility for learning from the professor to the students and encourages active listening during case discussions. As one student noted in the survey, "It [crowdsourced peer-to-peer assessment] made the class into an active learning environment, based on dynamic discussion. It emphasized that peers were as much a part of the learning process as the professor— the professor was not the sole source of knowledge or insights. It forced students to engage with each other." Assessing their peers and knowing that a crowd of peers will be assessing them leads students to have a greater sense of accountability, responsibility, and ownership for their own and others' learning (Topping, 1998).

Fourth, through objectively rating their peers, students learn the markers of strong participation. Having students actively engaged in rating class participation raises their awareness of what good and bad class participation is, which helps them craft their own participation in the future. As Topping reminds us, "When the criteria for assessment have been discussed, negotiated, used in practice, and clarified by all participants, greater clarity concerning what constitutes high quality work is likely, which focuses assessee (and assessor) attention on crucial elements." (Topping, 1998, p. 255) By forcing students to use assessment criteria in practice, peer assessment drives students to engage in meta-level processing of their own class participation, making them more thoughtful about the active learning process and their individual and collective roles in it. Topping describes peer assessment as an active learning activity that requires the assessor to engage in "cognitively demanding activities [such as active listening, processing, evaluating, comparing, contrasting, assessing deviations from the ideal, and providing feedback] that could help to consolidate, reinforce, and deepen understanding in the assessor" (Topping, 1998, p. 254). Conducting the peer assessment task at mid-semester allows it to be a formative assessment that can improve student performance while the learning process is unfolding over the course of the semester, rather than judging performance after the behavior is already complete at the end of the semester. This allows students to plan and manage their own learning by identifying and understanding those areas in which they are strong, and taking remedial actions to rectify their weaknesses.

The findings outlined above must be considered in light of the empirical validation study's limitations. The results were generated from one course, at a single university, with MBA student raters, so the generalization of the findings to other teaching contexts and other student populations may vary. It would be helpful for future research to duplicate these results in an undergraduate classroom to understand if younger students have the maturity to rate each other with the same accuracy and reliability as MBA students can. It seems likely that some of the social anxiety raised by the MBA students might be

heightened in an undergraduate classroom. The results are also highly sensitive to the development and communication of the class participation rubric items and students' buy-in to the rubric when forming their assessments. A lack of clarity about assessment criteria is likely to lead to lower correlations between student and instructor assessments. The results are also sensitive to the degree of independence among raters. For crowdsourcing to cancel out the errors contained in individual assessments, there must be enough raters, the raters must be thinking and acting independently, and their biases must be personal and not systemic, so that their errors are uncorrelated. As raters become more homogeneous, the degree of independence among them is likely to diminish. As the

number of raters goes down, the mathematical benefits of crowdsourcing may be diminished. Instructors should use caution when using this method with small, homogenous groups. Finally, an objective measure of a student's contribution to group learning is not used in the empirical validation; as discussed above, even the instructor's assessment is likely to reflect subjective biases and cognitive distortions.

In conclusion, leveraging peer-to-peer feedback and crowdsourcing in today's face-to-face classrooms enables instructors to solve some of the pedagogical challenges associated with creating an active, collaborative learning environment by teaching via the case method.

REFERENCES

- AACSB. (2012). Eligibility Procedures and Accreditation Standards for Business Accreditation. In.
- Aggarwal, P., & O'Brien, C. L. (2008). Social Loafing on Group Projects. *Journal of Marketing Education, 30*, 255-264.
- Arter, J., & Spandel, V. (1992). NCME Instructional Module: Using Portfolios of a Student's Work in Instruction and Assessment. *Educational Measurement: Issues and Practice, 11*, 36-44.
- Aylesworth, A. (2008). Improving Case Discussion with an Improv Mindset. *Journal of Marketing Education, 30*, 106-115.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The Instructional Effects of Feedback in Test-Like Events. *Review of Educational Research, 61*, 213-238.
- Barnes, L. B., Christensen, C. R., & Hansen, A. J. (1994). *Teaching and the Case Method*. Boston, MA: Harvard Business School Press.
- Baumeister, R. F. (1998). The Self. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *Handbook of Social Psychology* (pp. 680-740). New York: McGraw Hill, Oxford University Press.
- Bloxham, S., & West, A. (2004). Understanding the Rules of the Game: Making Peer Assessment as a Medium for Developing Students' Conceptions of Assessment. *Assessment and Evaluation in Higher Education, 29*, 721-733.
- Bonoma, T. V., & Kosnik, T. J. (1989). *Learning by the Case Method in Marketing*. Harvard Business School Publishing, No. 9-590-008.
- Bonomo, T., & Kosnik, T. (1989). Learning by the Case Method in Marketing. In *Harvard Business School Publishing*. Boston, MA.
- Brindley, C., & Scoffield, S. (1998). Peer Assessment in Undergraduate Programmes. *Teaching in Higher Education, 3*, 79-89.
- Burchfield, C. M., & Sappington, J. (1999). Participation in Classroom Discussion. *Teaching of Psychology, 26*, 290-291.
- Christensen, C. R. (1989). *Teaching and the Case Method*. Cambridge, MA: Harvard Business School Press.
- Chylinski, M. (2010). Participation Money as a Mechanism for Measurement, Reward, and Formative Feedback in Active Class Participation. *Journal of Marketing Education, 32*, 25-38.
- Clow, K., & Wachter, M. (1996). Teaching Methodologies Used in Basic Marketing: An Empirical Investigation. *Journal of Marketing Education, 18*, 48-59.
- Corey, R. E. (1996). The Uses of Cases in Management Education. *Harvard Business School Publishing, 9-376-240*.
- Crespy, C., Rosenthal, D., & Stearns, J. (1999). Change and Growth: Advances in the Use of Cases in Marketing. *Marketing Education Review, 9*, 1-3.
- Desiraju, R., & Gopinath, C. (2001). Encouraging Participation in Case Discussions: A Comparison of the MICA and the Harvard Case Methods. *Journal of Management Education, 25*, 394-408.
- Duverger, P., & Steffes, E. M. (2012). The Reliability of Crowdsourcing Grading in a Creative Marketing Project Context. *Journal for Advancement of Marketing Education, 20*, 39-49.
- Ellett, W. (2007). *The Case Study Handbook*. Boston, MA: Harvard Business School Press.
- Falchikov, N. (1986). Product Comparisons and Process Benefits of Collaborative Peer Group and Self Assessments. *Assessment and Evaluation in Higher Education, 11*, 146-166.
- Falchikov, N. (1995). Peer Feedback Marking: Developing Peer Assessment. *Innovation in Education and Training, 32*, 175-187.
- Fiske, S. T. (1998). Stereotyping, Prejudice, and Discrimination. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology* (Fourth Edition ed., Vol. II, pp. 357-411). Boston, MA: McGraw-Hill.
- Gibson, B., & Sachau, D. (2000). Sandbagging as a Self-Presentational Strategy: Claiming to Be Less

- Than You Are. *Personality and Social Psychology Bulletin*, 26, 56-70.
- Gonzalez, G. R., Ingram, T. N., LaForge, R. W., & Leigh, T. W. (2004). Social Capital: Building an Effective Learning Environment in Marketing Classes. *Marketing Education Review*, 14, 1-8.
- Gopinath, C. (1999). Alternatives to Instructor Assessment of Class Participation. *Journal of Education for Business*, 75, 10-14.
- Gordon, K. H. (1924). Group Judgments in the Field of Lifted Weights. *Journal of Experimental Psychology*, 4, 398-400.
- Graeff, T. R. (1998). Writing Behavior Learning Objectives for Marketing Courses: Meeting the Challenges of AACSB Outcomes Assessment. *Marketing Education Review*, 8, 13-25.
- Greenwald, A. G. (1980). The Totalitarian Ego: Fabrication and Revision of Personal History. *American Psychologist*, 35.
- Hendry, E. (2009). Duke Professor Uses 'Crowdsourcing' to Grade. In *The Chronicle of Higher Education*.
- Hounsell, D. (2003). Student Feedback, Learning and Development. In M. Slowey & D. Watson (Eds.), *Higher Education and the Lifecourse* (pp. 67-78). Berkshire, UK: SRHE & Open University Press.
- Li, C., & Bernoff, J. (2007). *Groundswell: Winning in a World Transformed by Social Technologies*. Harvard Business Review Press.
- Mauboussin, M. J. (2007). Explaining the Wisdom of Crowds. In *Mauboussin on Strategy*. Legg Mason Capital Management.
- Melvin, K. (1988). Rating Class Participation: The Prof/Peer Method. *Teaching in Psychology*, 15, 137-139.
- Orsmond, P., & Merry, S. (1996). The Importance of Marking Criteria in the Use of Peer Assessment. *Assessment and Evaluation in Higher Education*, 21, 239-249.
- Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ: Princeton University Press.
- Peterson, R. M. (2001). Course Participation: An Active Learning Approach Employing Student Documentation. *Journal of Marketing Education*, 23, 187-194.
- Poddar, A. (2010). Continuous Additive Peer Review: A New System to Control Social Loafing in Group Projects. *Journal for Advancement of Marketing Education*, 17, 1-12.
- Pond, K., Uihag, R., & Wade, W. (1995). Peer Review: A Precursor to Peer Assessment. *Innovations in Education and Teaching International*, 32, 314-323.
- Qualman, E. (2009). *Socialnomics: How Social Media Transforms the Way We Live and Do Business*. New York: Wiley.
- Ramsden, P. (2003). *Learning to Teach in Higher Education* (2nd Edition ed.). London: Routledge
- Rangan, V. K. (1995). Choreographing a Case Class. In *Harvard Business School Note*. Boston, MA: Harvard Business School Publishing.
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes. *Assessment and Evaluation in Higher Education*, 28, 147-164.
- Ryan, G. J., Marshall, L. L., Porter, K., & Jia, H. (2007). Peer, Professor and Self-Evaluation of Class Participation. *Active Learning in Higher Education*, 8, 49-61.
- Saavedra, R., & Kwun, S. K. (1993). Peer Evaluation in Self-Managing Work Groups. *Journal of Applied Psychology*, 78, 450-462.
- Sadler, P. M., & Good, E. (2006). The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment*, 11, 1-31.
- Shepperd, J. A., & Socherman, R. E. (1997). On the Manipulative Behavior of Low Machiavellians. *Journal of Personality and Social Psychology*, 72, 1448-1459.
- Stires, L. K., & Jones, E. E. (1969). Modesty Versus Self-Enhancement as Alternative Forms of Ingratiation. *Journal of Experimental Social Psychology*, 5, 172-188.
- Surowiecki, J. (2004). *The Wisdom of Crowds*. New York, NY: Doubleday.
- Tajfel, H., & Turner, J. (1979). An Integrative Theory of Intergroup Conflicts. In W. G. A. a. S. Worchel (Ed.), *The Social Psychology of Intergroup Relations* (pp. 33-47). Monterey, CA: Brooks-Cole.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and Well-Being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin*, 103, 193-210.
- Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, 68, 249-276.
- Treynor, J. L. (1987). Market Efficiency and the Bean Jar Experiment. *Financial Analysts Journal*, 50-53.
- Zuckerman, M. (1979). Attribution of Success and Failure Revisited, or: The Motivational Bias is Alive and Well in Attribution Theory. *Journal of Personality*, 47, 245-287.