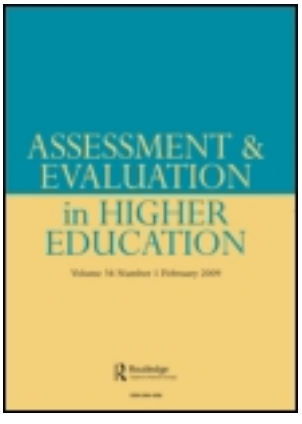
2010

Peer Assessment of Class Participation: Applying Peer Nomination to Overcome Rating Inflation

James Heyman
University of St. Thomas, Minnesota, jeheyman@stthomas.edu

John Sailors
University of St. Thomas, Minnesota, jjsailors@stthomas.edu



Assessment & Evaluation in Higher Education

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/caeh20

Peer assessment of class participation: applying peer nomination to overcome rating inflation

James E. Heyman [a] & John J. Sailors [a]

[a] Department of Marketing, Opus College of Business, University of St. Thomas, 1000 La Salle Avenue (TMH 443), Minneapolis, MN, 55403, USA

Available online: 27 Jul 2011

To cite this article: James E. Heyman & John J. Sailors (2011): Peer assessment of class participation: applying peer nomination to overcome rating inflation, Assessment & Evaluation in Higher Education, 36:5, 605-618

To link to this article: http://dx.doi.org/10.1080/02602931003632365

Assessment & Evaluation in Higher Education
Vol. 36, No. 5, August 2011, 605–618

Routledge
Taylor & Francis Group

Peer assessment of class participation: applying peer nomination to overcome rating inflation

James E. Heyman* and John J. Sailors

Department of Marketing, Opus College of Business, University of St. Thomas, 1000 La Salle Avenue (TMH 443), Minneapolis, MN 55403, USA

Having students actively engaged with each other in discussions has become an increasingly important and common aspect of the classroom environment. This increased emphasis has also meant that instructors need to find ways to effectively and efficiently evaluate class participation. In this paper, we describe the most common method used for these assessments and highlight some of its inherent challenges. We then propose an alternative method based on peer nominations. Two case studies illustrate the advantages of this method; we find that it is both easy for students to complete and provide instructors with valuable diagnostic information with which to provide feedback and assign grades.

Keywords: peer nomination; peer assessment; class participation; ratings; rankings

Introduction

Class participation is a form of active learning in which students publicly discuss the course material. The purpose of these conversations is to introduce multiple perspectives on the topic at hand. It is an opportunity for students to explore their own understanding and to advance other students' learning by introducing them to alternative perspectives (Jones 2008). To a lesser extent, it also includes closed-ended answers to specific questions posed by the instructor (Moguel 2004), although the crux remains the avoidance of expected responses (Mehan 1979). Class participation has been studied and encouraged in diverse academic disciplines such as law (Hayes and Hayes 1973; Kamvounias 1996), marketing (Wooldridge 2008), psychology (Elliot 1993; Hodge and Nelson 1991), mathematics (Jansen 2006) and physics (Mamola 2005). Across disciplines, class participation has been shown to improve critical reading and understanding of material (Dancer and Kamvounias 2005; French 1992), increase classroom engagement (Elliot 1993), advance active learning (Bean and Peterson 1998; Peterson 2001; Petress 2006; Wooldridge 2008), improve work habits (Turner and Patrick 2004), enhance a variety of professional skills (Hayes and Hayes 1973; Kamvounias 1996; Lord and Melvin 1994; Wooldridge 2008) and liven up dry material (Magel 1996). The list of participation methods and their benefits is extensive and there is broad agreement that students benefit from this pedagogic approach (Jones 2008).

However, many instructors struggle with assessing class participation.[1] Many students likewise struggle when it comes to accepting those assessments. Instructors and students alike recognise there is a large subjective component to class participation assessment not only due to the nature of the judgements being made but also due to

*Corresponding author. Email: jeheyman@stthomas.edu

ISSN 0260-2938 print/ISSN 1469-297X online

DOI: 10.1080/02602931003632365
http://www.informaworld.com

the assessment procedures themselves. For example, many instructors find regular record keeping of participation difficult and are reduced to a holistic assessment at the end of the term (Bean and Peterson 1998). This can make it hard to justify the participation score if a student challenges it (Jacobs and Chase 1992). One way to mitigate many of these problems is to employ peer assessment of participation to supplement or even replace instructor evaluations.

Peer assessment in the classroom has been investigated and found to be both meritorious and rigorous as a tool. Meritorious because it emphasises the goal of communal learning (Gopinath 1999), encourages process transparency (Brindley and Scoffield 1998) and leverages the perspective of others operating in the same environment (Kane and Lawler 1978). In addition to students benefiting from being graded by their peers, the graders also benefit because it gives them experience in evaluating others, which is an increasingly important skill (Mainkar 2008) particularly as more organisations adopt various forms of personnel (Antonioni 1996; Tyler 2007), work group (Urch Druskat and Wolff 1999) and work product (Russell 2006) rating systems. Evaluating peers can be uncomfortable (Mowl and Pain 1995; Orsmond and Merry 1996; Sherrard and Raafat 1994) and it is only with practice that students will gain more confidence in their ability to do so.

Peer assessment is also rigorous in that it has been shown to be an accurate indicator of performance (Falchikov 1995; Falchikov and Goldfinch 2000; Gagne 1998; Hollander 1965; Kane and Lawler 1978; Liu and Carless 2006; Love 1981; Magin and Helmore 2001; Topping 1998). Importantly, the validity of peer assessment has been demonstrated with respect to concurrent validity with instructor evaluations (e.g. Melvin 1988) and predictive validity based on subsequent objective performance (e.g. Henry 2006; Schwarzwald, Koslowsky, and Mager-Bibi 1999).

In exploring different types of peer assessment, ranking has been found to have advantages over rating. For example, when rating one's peers there is a tendency to inflate marks and to end pile on the positive end of the scale (McCarty and Shrum 1997, 2000).[2] This effect is exacerbated by students' reluctance to assign low grades to their classmates (Ballantyne, Hughes, and Mylonas 2002; Brindley and Scoffield 1998). The result is that many receive the same mark even when differences in performance actually exist (Ovadia 2004).

However, ranking is not devoid of issues. One is that it forces differentiation even where none may exist (Maio et al. 1996). Another is that it is taxing on the respondents (Alwin and Krosnick 1985); comprehensively ranking more than 15, let alone 30, classmates will require a very high level of motivation and effort, making it ill-suited to peer assessment (or for that matter, instructor assessment) of participation (Beatty et al. 1996; Durbin 1951; Ovadia 2004; Russell and Gray 1994). A third is that, because it is so much more difficult for respondents than rating, it is also much more time consuming (Munson and McIntyre 1979). In this paper, we demonstrate the application of an alternative ranking method that largely overcomes the problems with both ratings and rankings while maintaining the merits and rigour associated with traditional peer assessment techniques.

Peer nominations

Peer nominations are a type of partial ranking that differ from both ratings and full rankings in important ways. In a partial ranking system, n objects are assigned to k

different bins, where $k < n$. Within each bin, objects are not ordered, rather, they are considered equivalent on the dimension(s) of interest. For example, if 30 items were to be evenly ordered across three bins then each bin would have 10 of the items and there would be no ranking within each bin. In other words, in a partial ranking system, ties among the objects are allowed, differentiating it from full ranking, which prohibits ties. The greater the difference between k and n, the more ties are allowed. Evenly sorting 30 items into three bins results in three sets of 10 items that are tied while evenly sorting the same items into five bins generates five sets of six tied items. However, as long as $k \neq 1$ partial ranking, as opposed to rating, requires at least some discrimination among the evaluated objects. The two approaches overlap in that rating is a form of partial ranking, where the number of bins into which people are sorted is limited to the number of points in the rating scale; in practice, given typical scales, this would mean an upper limit of between 5 and 10 ranked categories into which rating places people.

In practice, peer nominations most often require respondents to nominate the top m of their peers and also the bottom m of their peers, which we will hereafter refer to as HLm, where H simply refers to 'high', L to 'low' and m being the number of nominations made for each. The HLm method is equivalent to a partial ranking system with $k = 3$ bins. Such peer nominations have been fruitfully applied in diverse areas such as research into bullying (Henry 2006) and evaluation of uniformed personnel (Schwarzwald, Koslowsky, and Mager-Bibi 1999) but have not been extensively used to assess class participation, although Melvin's (1988) method, while not labelled peer nomination, can be viewed as such.

A criticism of peer nominations has been that it only provides information regarding students who were considered to be at the top or bottom end of the group on the dimension of interest. This is certainly true with regard to the responses supplied by the individual respondents. At first this seems problematic with regard to using peer nominations to evaluate class participation given the finer differences in student performance. What this criticism misses, and what we shall demonstrate, is that algebraic manipulation of the positive and negative nominations across the graders yields just such demarcations.

The field of sociometrics has developed into a domain with a large emphasis on practical tools to determine the position of individuals within their social group (Cillessen and Bukowski 2000). Much of this work involves peer nominations and because the emphasis is on people's standing within a group, these researchers frequently use various methods to combine top and bottom peer nominations (Coie, Dodge, and Coppotelli 1982; Newcomb and Bukowski 1983; Peery 1979). The variance across the individual peer nominations typically allow, in aggregate, classifying the group into more than just two or three categories. For example, Coie, Dodge, and Coppotelli (1982) use such data to place all the children studied into five sociometric categories. While it is possible to generate this type of discrete classification scheme, the algebraic manipulations themselves result in a measure that is continuous across a range of values.

In the sociometric literature, positive and negative peer nominations are typically combined in two ways. First, the total number of positive nominations and the total number of negative nominations that each individual receives are counted. Social performance is calculated by subtracting an individual's total number of negative nominations from their total number of positive nominations. Social salience or impact is calculated by adding the two instead of subtracting (Coie, Dodge, and

Coppotelli 1982; Newcomb and Bukowski 1983; Peery 1979). For the present paper, we focus on the calculation underlying the social performance measure.

Tables 1 and 2 compare the rating and HLm (in this case $m = 2$) processes for a six-person class. Table 1 shows the traditional rating system. Each column consists of the grades given by that student to their peers (rows). The scores are then averaged and found in the rightmost column. Table 2 shows the same class rating each other with the HL2 system. Now the columns consist of the two Hs and Ls each student assigned to their peers. The nominations are then tallied across each row and then the net difference is calculated.

In the following section, we illustrate the technique applied to class participation in an undergraduate-level marketing management class. For comparative purposes, we first begin with an illustration of peer evaluation via a traditional rating task.

Two sections of principles of marketing recently taught by one of the authors during the same semester were randomly assigned to an assessment method. The principles of marketing class is required for all business majors and typically has an enrolment of around 30 students, almost all of them juniors. Due to the breadth of the university distribution requirements, the students spend their freshman and sophomore years in a wide variety of classes and typically have had little or no contact with each other prior to this semester. In both classes, participation counted for 25% of the students' total grade.

Table 1. Rating process.

		Student assessment of peers						
		Alice	Charlie	Elliot	Fred	Grace	Julian	Average rating
Students assessed by peer	Alice	—	48	42	50	45	47	46.4
	Charlie	44	—	50	50	49	47	47.2
	Elliot	48	50	—	50	48	47	48.6
	Fred	45	50	42	—	45	49	46.2
	Grace	44	49	47	50	—	49	47.8
	Julian	48	49	50	50	45	—	48.4

Table 2. HL2 process.

HL2		Student assessment of peers						Total number of highs	Total number of lows	Net number of highs – number of lows
		Alice	Charlie	Elliot	Fred	Grace	Julian			
Students assessed by peer	Alice	—	L	L	—	L	—	0	3	–3
	Charlie	L	—	H	L	H	L	2	3	–1
	Elliot	H	H	—	L	H	L	3	2	1
	Fred	—	H	L	—	L	H	2	2	0
	Grace	L	—	—	H	—	H	2	1	1
	Julian	H	L	H	H	—	—	3	1	2

Case 1: ratings in peer assessment of class participation

Method

The class assigned to the traditional rating task had 30 students and peer assessments were held during the semester's penultimate week. The class was 93% Caucasian, 3% African and 3% Asian. All enroled students completed the grading sheets. Before starting the assessment they were reminded of some ways to meaningfully participate (e.g. ask questions, answer other students' questions, relate their personal experiences to the course concepts, offer analysis of case studies, bring in examples from the newspaper and spot 'red flag' contradictions or 'green thread' linkages across the semester's topics) which had been listed in the syllabus and discussed during the first week of class; this instruction was explicit that both quality and quantity of contribution should enter into their assessment. A seating chart, used all semester, was distributed to the students to facilitate their grading by allowing them to identify their pairs based on facial recognition in addition to name recall. Students were asked to rate each of their classmates on a scale that ranged from 0 to 50.[3] The entire task took less than 10 minutes.

Results

The 30 students in attendance provided ratings for all class members, one student missed rating one classmate, resulting in 899 total ratings. These ratings were averaged for each student. Table 3 contains the frequency of these averages, rounded to the nearest integer, as well as the frequency of ratings given by the instructor. Although significantly correlated with those assigned by the instructor ($r = 71$, $p < .001$), student generated ratings ($M = 46.1$) were distinctly higher than those assigned by the instructor ($M = 40.6$, $t(29) = 8.70$, $p < .001$). This combination was driven by the relatively restricted range of the student assigned ratings.

While the student generated averages range from only 42 to 50, representing approximately 18% of the available scale, the instructor generated ratings ranged from 35 to 50. Looking at the student generated ratings, 23 students – 77% of the class – received an average grade of 45 (90%) or higher. While the average peer ratings suggest most students deserved an A with just a smaller number of Bs and nothing lower, the instructor's evaluations of participation spanned 60% of the grading scale, with one student who never contributed and rarely attended meriting a failing participation grade.

As noteworthy as the distribution of evaluations is the limited range of ratings used by the students in assigning their marks. The ratings given by individual students spanned only 15 out of a possible 50 points. Almost a quarter of the class (23%) used only 1 or 2 scale points (i.e. they gave either everyone the same grade or one of two adjacent point values) and 57% used a 10-point range in their ratings. Even the student who used the widest range of 15 points actually used a nine-point range for all but one of their ratings (apparently wishing to be especially harsh with a single classmate).

Discussion

Using ratings as a peer evaluation tool exhibited the issues identified in the literature as problematic with the technique. Peer ratings tend to be inflated, with too many students rating their peers with excessively high marks. If the participation grades were based solely on the peer ratings then too many students would have received the

Table 3. Frequency of received ratings.

Rating value	Frequency from	
Value	Students	Instructors
50	1	1
49	4	1
48	4	1
47	3	1
46	5	1
45	6	2
44	3	1
43	2	2
42	2	3
41	—	—
40	—	2
39	—	4
38	—	3
37	—	2
36	—	3
35	—	2
34	—	—
33	—	—
32	—	—
31	—	1

same participation rating despite the wide variance in their actual performance as judged by the instructor. While the process was easy for the students to complete this benefit came at the cost of insufficient diagnosticity.

Correcting for the substantial grade inflation would require, for example, translating a peer-generated B into a D, which would have tended to reduce student acceptance of the peer assessment process and results. It seems that even an explicit explanation of grade inflation would have prevented the students from accepting the legitimacy of the exercise and the fairness of the professor.

Based on this case, we conclude that rating, while widely used, is not the best method for students to use when assessing their peers' classroom participation. Although the process was well received by the students in that it was familiar and easy to understand and complete, it generated neither meaningful feedback for the students nor diagnostic information for the instructor. Can we do better? We suggest, and will demonstrate in the following case, that the peer nomination technique described earlier represents a significant improvement over rating.

Case 2: peer nominations of class participation

Faced with rating's endemic issues we now illustrate the use of peer nomination in the context of evaluating class participation. Our hope in applying the method was that it would deliver the diagnostic benefits of ranking but with the ease and clarity of rating. In this particular example, students identified the five of their peers they

thought contributed most to class discussion (high contributor) and the five who contributed least (low contributor), most and least again being broadly defined to reflect both quality and quantity of contribution. We will hereafter refer to this method as HL5.

We chose $m = 5$ as a number that balanced the reduction of required cognitive effort with the need for variation in the results. In our experience using the peer nomination technique in a variety of contexts, having the total number of nominations, highs and lows, equal to 1/3 the total objects (in other words, $m = 1/6$ the total) balances the two goals. Much larger and the task becomes onerous for nominators and much smaller and there is too little variability in the aggregated results.

Method

The second class had 33 students, 30 of whom completed the peer evaluations that were held during the penultimate week of the semester. Compared to the first class this one was unremarkable; there were no statistically significant differences across demographic characteristics, grades on individual assignment, the instructor's perception of level of class participation, quality of final projects and final grades. Students were given the same background explanation of the assessment process as described in Case 1. Again, the seating chart in use all semester was passed out to each student to prevent name familiarity – or the lack thereof – interfering with the assessment process. Students were told to use an H to mark the five classmates they thought did the best job of participating and to mark with an L their five classmates whom they felt participated the least. All submissions were anonymous[4] and again the entire evolution took somewhat less than 10 minutes.

Results

Each student's score was calculated by subtracting the total number of lows received from their total number of highs. This resulting participation net score ranged from −14 to +16, more than three times the range of average ratings and equal to 50% of the theoretical maximum range given the number of students submitting. Table 4 contains the number of high and low nominations received by each student and the corresponding instructor's assessment on the 50-point scale. The instructor's rating of this class' participation correlated with the results of the HL5 method ($r = .85$, $p < .001$). Consistent with the instructor's perception, a few students were recognised for regular, active participation, another small group for seemingly never participating and a large group in the middle that were infrequent contributors. There is another possible student profile that occurs not infrequently, the student who receives a relatively high number of both high and low contributor nominations.

For example, one student in this case received 11 high contributor and 5 low contributor nominations, yielding a net score of 6. This was a student who participated daily in ways that would probably be of interest to most instructors but not to many of her classmates; a typical comment from her would concern such things as the role of normative ethics in the Red Cross' market segmentation strategy. Easy identification of such a student is an important aspect of the HL method. It is thus trivial to distinguish between moderately involved innocuous students and the highly involved yet controversial students.

Table 4. Peer and instructor assessments.

Number of highs	Number of lows	Net difference	Instructor rating
16	0	16	50
14	0	14	48
12	0	12	45
11	1	10	46
9	0	9	46
8	1	7	45
9	2	7	46
11	5	6	47
5	1	4	41
3	0	3	43
5	2	3	41
3	1	2	44
6	4	2	45
4	3	1	40
1	1	0	41
2	2	0	45
4	5	−1	43
0	1	−1	41
3	4	−1	40
2	3	−1	42
1	3	−2	43
2	5	−3	44
0	4	−4	43
3	7	−4	44
1	6	−5	38
2	8	−6	40
3	9	−6	40
2	10	−8	38
3	11	−8	35
0	9	−9	37
0	11	−11	34
0	12	−12	34
0	14	−14	30

Discussion

The advantages of the high–low method of peer assessments in this class were both intuitive and subtle. The class indeed had talkative students, silent students, slight variations of each and a majority in the middle. The students were at first puzzled by the process only because of its novelty but were entirely accepting of the results because they appeared sensible and familiar – normal in both statistical and experiential respects. The strong consensus on the high performers is no surprise given the attention they drew to themselves during the semester. Rather it would have been surprising if they had not been particularly salient when it came to the assessment.

But what was surprising was the high agreement on who were the low contributor students. One might reasonably expect that a quiet student would simply have avoided nomination altogether. And yet this was not the case. Regardless of where the student sat, be it in the front row or the back of the room, the non-contributing students were called out by their peers. The salience of the low contributors was one of the true eye openers for us as instructors. Explicitly identifying low contributors provides additional discriminatory power of the HLm method over peer nominations that solicit only high contributors, what we would call Hm, as discussed in Kane and Lawler (1978).

General discussion

In this paper, we examined two methods of conducting peer assessment of class participation in a marketing class. The first was the traditional approach in which students rate their peers on a fixed scale. Ratings have the advantage that they are familiar, easy for the students to complete and easy for instructors to analyse with standard parametric techniques. However, peer ratings suffered from a tendency to produce inflated results whose lack of variance limited their diagnosticity. Ratings often have these issues and these can be exacerbated by social factors inherent in classroom environments. Participation grades based on peer ratings inevitably involve inflationary corrections by the professor that reduce student acceptance and deviate sharply from capturing the voice of the class, which was one of the reasons for using peer assessment in the first place.

To address these shortcomings, we illustrated a method of peer nomination in which students identified their five classmates who participated the most and the five who participated the least. This method consisted solely of each student identifying these two groups of students and implicitly a third group of students whose participation placed them in the middle. Within these groups, there was no further ranking or rating done. Following the sociometric literature reviewed earlier, a student's final participation score was calculated by subtracting the number of times students were identified as a low contributor from the number of times they were considered to be a high contributor. That this method has been effectively used with very young children (Cole, Cornell, and Sheras 2006) suggests that it is easy to use, can differentiate levels of behaviour that may be vaguely defined and involving topics such as peer performance that can be socially awkward to assess.

Unlike ratings, the HLm does not allow the indiscriminate assigning of grades nor does it accommodate any leniency in grading fellow students both of which were disadvantages of ratings. Unfortunately, it does not solve the lazy-grader problem, as there may still be graders who wish to complete the exercise with the minimum possible amount of effort just as there may be raters who give everyone the same score. Statistically, however, their effect will be less pernicious than in ratings as they now introduce noise rather than inflationary bias. It is possible that there might have also been a marginal effect on potential lazy graders because the task seems less formidable but, due to the process, this would be an unobservable effect.

The final advantage is that it strikes a balance between rating and ranking on the issue of forced delineation of similar assessments. Unlike ratings it forcibly reduces the number of possible ties but unlike ranking it does not proscribe them entirely. The

difference can be seen in the number of delineation points where a grader must explicitly say that student *X* is better than student *Y*. When ranking *N* students, graders must deal with $N - 1$ such boundaries between the final sorted elements and this is what makes ranking so difficult. However, HLm imposes only two points of distinction, in the current case with 30 students with $m = 5$ there are boundaries only between the fifth and sixth students and another between the 24th and 25th students.

However, while each assessor only has two such boundaries to consider in the HL method, each will have their own perception of the class participation of their peers. It is this variance in perceptions of participation that leads to differences in the nominations each respondent makes and these differences which lead to an aggregate ranking that has many gradations. A natural question concerns what leads to such different perceptions and the respective nominations perceptions? One factor will be the abilities and needs of the nominator. For example, class participation comments will vary along Bloom's (1956) taxonomy and students will evaluate those comments according to their own level of domain-specific knowledge and cognitive skills (Bean and Peterson 1998); contributions reflecting a level of the taxonomy near one's own are likely to be valued more than contributions either far above or below. Different cognitive preferences and learning styles (DiTiberio 1996) may also drive different evaluations of class participation. Student nominators may place different emphasis on quantity of contribution versus quality, and may evaluate and weigh aspects of quality in different ways (Mainkar 2008). Each of these factors, by driving individual difference in the nomination process, results in an aggregate nomination-based ranking of performance that ensures broad coverage of the entire class.

In this paper, we focussed on a net measure that is computed in the same manner as the social preference measure in the sociometric literature. We did not choose to specifically address the viability of a net measure analogous to the social salience measure from that domain. While a potentially interesting approach, we question the utility of the social salience measure in the context of peer nominations of class participation. In the domains where it has been applied, for example, understanding the peer relations among children (Bowker 2004; Cole, Cornell, and Sheras 2006) negative nominations are driven by active behaviour that other children dislike such as bullying. With respect to class participation, we believe negative nominations are more likely to be driven by lack of behaviour, that is, participation, rather than by making unhelpful contributions to class. A student who contributes to discussion but does so in ways that are tangential to the points previously raised by the instructor or other students is likely to receive some quantity of positive nominations, rather than solely negative nominations. In other words, we expect that negative nominations with respect to class participation will reflect disengagement, not misdirected engagement, making the social salience measure uninformative in this particular domain. However, this remains an area for future research.

One might wonder whether traditional ratings could simply be transformed into data resembling data generated by the HL method. Generally, the answer is no for the simple reason that there are too many ties given. When a given student rater gives 10 classmates a five on a five-point scale, how does the instructor select which of these 10 will be the *m* high nominations (assuming $m < 10$)? The same situation exists with respect to the low end of scale points awarded and identifying the low nominations. We note that this issue does not simply occur when a relatively small scale is employed; in our Case 1, we used a 51-point scale yet almost a quarter of the class

only used two values when rating their peers. Obviously these ratings could not be converted to HL5-like data. While conceptually it seems as if ratings should be easily transformed into HL nominations, in practice no mapping can generate information when the original source lacks sufficient variation.

Although unfamiliar to most students, the HL method of assessment is no more difficult than the ratings it replaces. Beyond the issues of familiarly and difficulty we note that regardless of the process used for peer assessment, students may question its fairness and their own ability to accurately assess their peers. This could affect how they complete the assessment. However, peer nominations are used extensively with young children and this should indicate that college students are likely to feel competent with the technique.

For instructors the analysis is also new but, as shown earlier, it is actually much simpler than parametric methods. The benefit of this method is that it better captures the class' actual performance while remaining true to the tenet that classroom participation is best judged by those who most experience it. The benefits were seen in the distributional properties of the results, the accurate identification of high and low contributors and detecting students with whom the graders had, for lack of a better phrase, mixed feelings. In short, the HL5 method delivered the benefits of ranking and rating while avoiding most of their attendant costs.

Notes

1. Some even argue that participation should not be assessed (Gilson 1994). While an interesting question, this paper focuses on the process of assessment rather than its propriety.
2. One might expect that rating peers on multiple dimensions might attenuate some of the end piling; however, data from Falchikov and Goldfinch (2000) suggest that this would not be the case.
3. A scale of such width is certainly not the norm although there is no particular reason to believe this would bias the final results (Dawes 2008). However, we wished to give students the most flexibility in terms of avoiding end piling.
4. This procedure favoured anonymity over preventing students from grading themselves. However, methods such as coded sheets or online surveys can also be used to achieve these dual goals.

Notes on contributors

James E. Heyman is an assistant professor in the Department of Marketing, Opus College of Business, University of St. Thomas. His research interests are behavioural economics, survey methodologies and judgement and decision-making.

John Sailors is an assistant professor in the Department of Marketing, Opus College of Business, University of St. Thomas. His research interests are consumer behaviour and marketing research.

References

Alwin, D.F., and J.A. Krosnick. 1985. The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly* 49: 535–52.

Antonioni, D. 1996. Designing an effective 360-degree appraisal feedback process. *Organizational Dynamics* 25, no. 2: 24–38.

Ballantyne, R., K. Hughes, and A. Mylonas. 2002. Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education* 27, no. 5: 427–41.

Bean, J.C., and D. Peterson. 1998. Grading classroom participation. *New Directions for Teaching & Learning* 74: 33–40.
Beatty, S.E., W.K. Martin, M.H. Yoon, and L.R. Kahle. 1996. Improving value rankings: An assessment of the rank-then-rate approach. Paper presented at the 1996 Developments in Marketing Science Conference, May, in Miami.
Bloom, B.S. 1956. *Taxonomy of educational objectives, book 1: Cognitive domain.* New York: McKay.
Bowker, A. 2004. Predicting friendship stability during early adolescence. *Journal of Early Adolescence* 24, no. 2: 85–112.
Brindley, C., and S. Scoffield. 1998. Peer assessment in undergraduate programmes. *Teaching in Higher Education* 3, no. 1: 79–89.
Cillessen, A.H.N., and W.M. Bukowski, eds. 2000. Conceptualizing and measuring peer acceptance and rejection. In *New directions for child and adolescent development.* Vol. 88, 3–10. San Francisco, CA: Jossey-Bass.
Coie, J.D., K.A. Dodge, and H. Coppotelli. 1982. Dimensions and types of social status: A cross-age perspective. *Developmental Psychology* 18, no. 4: 557–70.
Cole, J.C.M., D.G. Cornell, and P. Sheras. 2006. Identification of school bullies by survey methods. *Professional School Counseling* 9, no. 4: 305–13.
Dancer, D., and P. Kamvounias. 2005. Student involvement in assessment: A project designed to assess class participation fairly and reliably. *Assessment & Evaluation in Higher Education* 30, no. 4: 445–54.
Dawes, J. 2008. Do data characteristics change according to the number of scale points used? An experiments using 5-point, 7-point and 10-point scales. *International Journal of Market Research* 20, no. 1: 61–77.
DiTiberio, J.K. 1996. Education, learning styles, and cognitive styles. In *MBTI applications: A decade of research on the Myers-Briggs type indicator*, ed. A.L. Hammer, 123–66. Palo Alto, CA: Consulting Psychologists Press.
Durbin, J. 1951. Incomplete blocks in ranking experiments. *British Journal of Psychology* 4: 85–90.
Elliot, L.B. 1993. Using debates to teach the psychology of women. *Teaching of Psychology* 20, no. 1: 35–8.
Falchikov, N. 1995. Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International* 32, no. 2: 175–87.
Falchikov, N., and J. Goldfinch. 2000. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research* 70, no. 3: 287–322.
French, L.G. 1992. Teaching organic synthesis. *Journal of Chemical Education* 69, no. 4: 287–9.
Gagne, F. 1998. The prevalence of gifted, talented, and multitalented individuals: Estimates from peer and teacher nominations. In *Talent in context: Historical and social perspectives on giftedness*, ed. R.C. Friedman and K.B. Rogers, 101–26. Washington, DC: American Psychological Association.
Gilson, C. 1994. Of dinosaurs and sacred cows: The grading of classroom participation. *Journal of Management Education* 18, no. 2: 227–36.
Gopinath, C. 1999. Alternatives to instructor assessment of class participation. *Journal of Education for Business* 75, no. 1: 10–4.
Hayes, S.L., and R.A. Hayes. 1973. Towards objective assessment of class participation. *Journal of the Society of the Public Teachers of Law* 12: 323–32.
Henry, D.B. 2006. Associations between peer nominations, teacher ratings, self-reports, and observations of malicious and disruptive behavior. *Assessment* 13, no. 3: 241–52.
Hodge, G.K., and N.H. Nelson. 1991. Demonstrating differential reinforcement by shaping classroom participation. *Teaching of Psychology* 18, no. 4: 239–41.
Hollander, E.P. 1965. Validity of peer nominations in predicting a distant performance criterion. *Journal of Applied Psychology* 19, no. 6: 434–8.
Jacobs, L.C., and C.I. Chase. 1992. *Developing and using tests effectively: A guide for faculty.* San Francisco, CA: Jossey-Bass.
Jansen, A. 2006. Seventh graders' motivations for participating in two discussion-oriented mathematics classrooms. *Elementary School Journal* 106, no. 5: 409–28.

Jones, R.C. 2008. The 'why' of class participation. *College Teaching* 56, no. 1: 59–62.
Kamvounias, P. 1996. Assessment of class participation in commercial law courses. *Current Commercial Law* 4, no. 3: 128–31.
Kane, J.S., and E.E. Lawler III. 1978. Methods of peer assessment. *Psychological Bulletin* 85, no. 3: 555–86.
Liu, N.-F., and D. Carless. 2006. Peer feedback: The learning element of peer assessment. *Teaching in Higher Education* 11, no. 3: 279–90.
Lord, A.T., and K.B. Melvin. 1994. The prof/peer method: A technique to rate class participation. *Issues in Accounting Education* 9, no. 1: 109–17.
Love, K.G. 1981. Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction. *Journal of Applied Psychology* 66, no. 4: 451–7.
Magel, R.C. 1996. Increasing student participation in large introductory statistics classes. *American Statistician* 50, no. 1: 51–6.
Magin, D., and P. Helmore. 2001. Peer and teacher assessments of oral presentation skills: How reliable are they? *Studies in Higher Education* 26, no. 3: 287–98.
Mainkar, A.V. 2008. A student-empowered system for measuring and weighing participation in class discussion. *Journal of Management Education* 32, no. 1: 23–37.
Maio, G.R., N.J. Roese, C. Seligman, and A. Katz. 1996. Rankings, ratings, and the measurement of values: Evidence for the superior validity of ratings. *Basic and Applied Social Psychology* 18, no. 2: 171–81.
Mamola, K. 2005. Class participation on trial. *Physics Teacher* 43, no. 6: 330.
McCarty, J.A., and L.J. Shrum. 1997. Measuring the importance of positive constructs: A test of alternative rating procedures. *Marketing Letters* 8, no. 2: 239–50.
McCarty, J.A., and L.J. Shrum. 2000. The measurement of personal values in survey research. *Public Opinion Quarterly* 64, no. 3: 271–98.
Mehan, H. 1979. *Learning lessons: Social organization in the classroom.* Cambridge, MA: Harvard University Press.
Melvin, K.B. 1988. Rating class participation: The prof/peer method. *Teaching in Psychology* 15, no. 3: 137–9.
Moguel, D. 2004. What does it mean to participate in class? Integrity and inconsistency in classroom interaction. *Journal of Classroom Interaction* 39, no. 1: 19–29.
Mowl, G., and R. Pain. 1995. Using self and peer assessment to improve students' essay writing: A case study from geography. *Innovations in Education and Teaching International* 32, no. 4: 324–35.
Munson, J.M., and S.H. McIntyre. 1979. Developing practical procedures for the measurement of personal values in cross-cultural marketing. *Journal of Marketing Research* 16, no. 1: 48–52.
Newcomb, A.F., and W.M. Bukowski. 1983. Social impact and social preference as determinants of children's peer group status. *Developmental Psychology* 19, no. 6: 856–67.
Orsmond, P., and S. Merry. 1996. The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education* 21, no. 3: 239–50.
Ovadia, S. 2004. Ratings and rankings: Reconsidering the structure of values and their measurement. *International Journal of Social Research Methodology* 7, no. 5: 403–14.
Peery, J.C. 1979. Popular, amiable, isolated, rejected: A reconceptualization of sociometric status is preschool children. *Child Development* 50, no. 4: 1231–4.
Peterson, R.M. 2001. Course participation: An active learning approach employing student documentation. *Journal of Marketing Education* 23, no. 3: 187–94.
Petress, K. 2006. An operational definition of class participation. *College Student Journal* 40, no. 4: 821–3.
Russell, R. 2006. Mandatory peer review gains momentum, support. *Accounting Today* 20: 3, 37.
Russell, P.A., and C.D. Gray. 1994. Ranking or rating? Some data and their implications for the measurement of evaluative response. *British Journal of Psychology* 85, no. 1: 79–92.
Schwarzwald, J., M. Koslowsky, and T. Mager-Bibi. 1999. Peer ratings versus peer nominations during training as predictors of actual performance criteria. *Journal of Applied Behavioral Science* 35, no. 3: 360–72.
Sherrard, W.R., and F. Raafat. 1994. An empirical study of peer bias in evaluations: Students rating students. *Journal of Education for Business* 70, no. 1: 43–7.

Topping, K. 1998. Peer assessment between students in colleges and universities. *Review of Educational Research* 68, no. 3: 249–76.
Turner, J.C., and H. Patrick. 2004. Motivational influences on student participation in classroom learning activities. *Teachers College Record* 106, no. 9: 1759–85.
Tyler, K. 2007. Designing the review process. *HR Magazine* 52: 70–1.
Urch Druskat, V., and S.B. Wolff. 1999. Effects and timing of development peer appraisals in self-managing work groups. *Journal of Applied Psychology* 84, no. 1: 58–74.
Wooldridge, B.R. 2008. Golden duck awards: An interactive game to facilitate class participation. *Marketing Education Review* 18, no 1: 15–17.