

Exploring Factors Affecting Peer Evaluation Scores: A Many-Facets Rasch Perspective

Jinjin Huang, University of Denver <jjinjin.huang@du.edu>

Donald Bacon, University of Denver <dbacon@du.edu>

Teams are widely used in marketing education as a means of developing team skills among our students and to encourage collaborative learning. To facilitate both objectives, peer evaluations have long been used to monitor and encourage good team behaviors when students work in teams (see LeRosen, 1976; Clark, 1989; or Strong & Anderson, 1990, for early examples). While early approaches to peer evaluation relied on paper-and-pencil technology, more recent approaches to peer evaluation have included online software, such as CATME (Loughry, Ohland, & Woehr, 2014). Importantly, there is some evidence that the repeated use of online peer evaluations can improve team skills (Brutus & Donia, 2010).

Another reason for interest in peer evaluation within student groups is to assess team skills. With the advent of assessment in higher education and especially with mandates from the AACSB to assess learning outcomes, business educators have become more interested in finding ways to assess team skills. Loughry et al. (2014) note that peer evaluations are acceptable to the AACSB as a direct measure of team skills. They go on to describe the CATME online system and to review the literature supporting the reliability and validity of the CATME peer evaluation scales.

To date, however, little is known about how other factors beyond student team skill ability level, may affect peer evaluation scores. For example, it seems reasonable that a team project completed in class with very little coordination outside of class would be easier for a team than a project that required substantial amounts of time outside of class. Further, a group task involving learning textbook material may be easier and less stressful than a group project involving working with a live client. Thus, the difficulty of the team task will affect peer evaluation scores just as the difficulty of an exam will affect student scores on an exam. Other factors that affect peer evaluation scores are likely to include the size of the student team (larger teams will have a higher chance of social loafing, see Comer, 1995), self v. peer ratings, the maturity of the raters (graduate v. undergraduate students), and the leniency of the students doing the rating. To best understand student team skills, we need to understand how all of these factors affect team skill scores.

Fortunately, the many-facets Rasch model is a measurement approach designed to address challenges just like this one. The single-facet Rasch model has been used to simultaneously estimate the ability of a student and the difficulty of a test, even when not all students take exactly the same version of the test (see Bacon & Stewart, 2006, for an example in the marketing education literature). Extending this model, the many-facets model can simultaneously estimate the ability of a student and the difficulty of each group project, the different challenge presented by each team size, the different leniency of each student rater, etc., even if not all students experienced all of these conditions. The many-facets Rasch model has been used before to explore rater effects (Myford & Wolfe, 2003, 2004), but it has not yet been applied to understanding the factors that affect team peer evaluation scores among marketing students.

The contribution of this research is to apply the many-facets Rasch model to marketing student peer evaluation data to demonstrate how each of several factors affects peer evaluation scores. A large data set has been collected to analyze the model. The data set includes data on each of the five CATME scales, and from marketing courses spanning 4 different terms, 18 different instructors, team sizes from 2 through 7, 1,008 students, graduates and undergraduates, and student self-ratings and peer ratings.

The preliminary results indicate that substantial differences in peer evaluations may be observed across several of the variables studied, including the course/group project itself, the group size, and the leniency of the student raters within a group. Other variables will be explored before the proposed presentation in April. In presenting these findings at MEA, we will also provide recommendations to educators on how to collect and use peer evaluation data in the most meaningful way.

References

- Bacon, D.R., & Stewart, K.A. (2006). How fast do students forget what they learn in consumer behavior? A longitudinal study. *Journal of Marketing Education* 28(3), 181-192.
- Brutus, S., & Donia, M. (2010). Improving the effectiveness of student groups with a centralized peer evaluation system. *Academy of Management Learning and Education*, 9(4), 652-662.
- Clark, G.L. (1989). Peer evaluations: An empirical test of their validity and reliability. *Journal of Marketing Education*, 11, 41-58.
- Comer, D.R. (1995). A model of social loafing in real work groups. *Human Relations*, 48(6), 647-667.
- LeRosen, R.G. (1976). Group projects, peer rating scales: Management education tools. *Journal of Business Education*, 15, 317-318.
- Loughry, M.L., Ohland, M.W., & Woehr, D.J. (2014). Assessing teamwork skills for assurance of learning using CATME Team Tools. *Journal of Marketing Education*, 36(1), 5-19.
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422.
- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5(2), 189-227.
- Strong, J.T., & Anderson, R.E. (1990). Free-riding in group projects: Control mechanisms and preliminary data. *Journal of Marketing Education*, 12, 61-67.