

Let Them Eat Cake: Why the Inherent Bias in Professor Grading Should Change to Individual Performance Assessments

Kim Gower
University of Mary Washington

Self-performance and other appraisal is a ubiquitous part of the organizational landscape in the U.S. If you ask employees, however, it is not a popular one. In viewing the way people are assessed from the start of grade school through professional careers, final performance evaluation is almost always in the hands of somebody else telling us our worth, and often without full context or information. Therefore, when it comes to writing and rating from a personal perspective, a lack of tools, including practice and objective self-assessment, can cause a serious impediment to an employee's ability to achieve promotions, earn bonuses and raises, and further their professional progression. More importantly, assessments on individuals as done by others have brought about mounds of research on the inherent bias of such systems, at all levels. This paper examines that bias as it applies to professors grading their students' work, and outlines teaching students, and business professionals, how to properly evaluate their performance and advocate for themselves as a valuable lifelong skill.

Keywords: self-assessment, bias, grading biases, professional development, college grading, grade inflation

INTRODUCTION

As performance management and performance reviews have matured, self-evaluations have become commonplace. As such, self-evaluation is an important skill and, with the right approach, it is a valuable exercise no matter what company you work for (Adams, 2019). Despite this, college student "evaluations" (grades) are still mired in the old one way "supervisor to employee" mode, which restricts students from developing one of their most important mastery skills - professional self-assessment.

While there has been a slow decline in formal performance evaluations in organizations, in 2017 there was still a hearty 80% of companies that practiced the formal evaluations system (Smith, 2018). In the United States, Multisource Feedback (MSF) is the most widely used employee evaluation and development tool (Atwater, Waldman, Ostroff, Robie, & Johnson, 2005). The process of MSF - also known as multisource performance ratings (MSPRs) and 360 feedback - gathers and compares employee assessment ratings from multiple rater group observers including peers, subordinates, supervisors and, most importantly, self (Viswesvaran, Schmidt, & Ones, 2002). The results are used to determine significant corporate investments in training and development programs, promotions, and employee compensation for the targeted individual (Scullen, Mount, & Judge, 2003). And, while performance appraisals might be slowing down, the use of 360 performance appraisals is on the rise (Johnson, 2018).

This makes the process of being able to prepare an honest, professional self-assessment a critically valuable skill for both the employee and the organization, because along with opportunities for increasing pay and stature, MSF is also the predominant leadership development tool. This development is based on the premise that rater groups differ in their perceptions of the employee's performance (Hoffman, Lance, Bynum, & Gentry, 2010), and therefore each group's perceptual differences contribute unique information to the overall assessment of the employee, creating a more well-rounded analysis of their strengths and opportunities for improvement. Thus, the rater group results, when compared to the employee's own perception of performance, highlights areas where agreements and disparities exist between groups about performance and behaviors, and therefore helps with employee self-improvement to further develop their leadership skills and abilities (Hooijberg & Choi, 2000).

However, while this is decisively an important mastery skill for college students, it is rarely, if ever, taught in the classroom. The most popular form of assessment, to decide what grade a student has "earned", is written testing, including multiple choice, true and false, fill in the blank, and essay (Bruce, N.D.). As we watch the importance and use of 360 feedback rise when performance appraisals are used, sending graduates out with no idea of how to assess their performances, as well as the performance of others, leaves a large gap in their potential success.

COLLEGE PERFORMANCE ASSESSMENT

As college educators, we tend to take the power position of being the arbiter of what grades our students' "earned". This often opens us up to challenges by students that grading is biased, "unfair", subjective, etc. These critiques are even harsher if it is a project-based course based on team work, and students are learning other important skills like active learning, critical thinking, and becoming familiar with the ambiguity that usually surround professionals in the workplace – the intangibles that have invaluable impact, hence ambiguous grading. In fact, the entire integrity of performance appraisals/grades depends on the appraiser, who is responsible for providing objective evaluations (Bell, et al., 2020). Based on the student/professor relationship, integrity could be a pivotal point in how the student's feel about their grades, and when looking at grading "systems", an argument can be made for the integrity of these as well.

There are multiple studies on how to remove professor bias from the grading process because if we are honest, we are human and we are biased. Some peer review articles address the pitfalls of grading based on things like halo bias (Malouff, Stein, Bothma, Coulter, & Emmerton, 2014), student's predictions of their grades versus their actual grades (Saenz, Geraci, Miller, & Tirso, 2017), and even the idea of mistreatment and justice in a cross university empirical study (Al-Hussain, 2008). This latter research was conducted because of poor student perceptions in Jordan about the behavior of some teaching staff and the occurrence of different forms of student mistreatment and student mistrust of the grading system in the Jordan University of Science and Technology (JUST). In short, students and their concerns about grading also appear to be a multi-cultural issue, thus increasing the importance of studying this all important process at the university level and for future professional performance success.

Perhaps the most concise description about the overall failure of our current grading system came from Tomar in *The Quad Magazine*, December 12, 2019.

...the reality is, our five-letter grading system isn't necessarily that informative, it generally isn't very effective, and it certainly isn't educational. Bad grades can stigmatize and discourage those who need help. Good grades tend to reward and elevate those who already have all the intellectual and cognitive advantages. And ultimately, grades are a gross oversimplification of what students are capable of learning and doing...The grading system is inherently subjective. It's inherently punitive. And it's inherently reductive (Tomar, 2019).

This fairly harsh overview of the current grading system, however, is tough to argue against. If a "B" stands for "good", wherein does the difference lie between adding plusses or minuses to this, or changing

the grade to a higher (“A” for excellent) or lower (“C” for average)? This should cause educators to ponder how and why these three grades carry so much weight, and where do we draw the line? Also, how informative are these arbitrary grades, particularly in the era of grade inflation? After all, we have all had students who learned, yet performed poorly on multiple choice tests or essay tests, due to test anxiety. If these are the only form of assessing student performance, the punitive and reductive nature of assigning grades is self-evident.

Recent disturbing evidence supporting the stigma and subjectivity and teacher bias behind bad grades has been studied in Germany, where a research indicated that teachers gave lower grades to overweight students (Dian & Traventi, 2021). The total findings did not support teacher bias due to data analysis, but data results that did emerge add to human psychology of implicit theories and biases when it comes to judging others based on personal prejudices.

In fact, there is significant current research about the bias against perceptually overweight students, including a recent US study from 2020. Essays for students who were overweight were judged to be similar in structure quality, but they were still assigned lower grades compared to their perceived healthy weight counterparts. Additionally, teachers assumed that students who were overweight put forth more effort, needed more remedial assistance, and had lower overall grades in school (Finn, Seymour, & Phillips, 2020).

The Finn et al. article also provided an excellent review of a large amount of previous research, across ranges of education levels, showing a pattern of bias against overweight students, and even how that crosses over into gender (Branigan, 2017). However, the most disturbing comment in the article was that there exists so much empirical research about this bias in grading, yet nothing has been done about it (Finn, Seymour, & Phillips, 2020).

In a rather ambiguous dispute against grade inflation problems, or grading problems in general, are is Jephcote et al. 2021, whose research contends that it is not grade inflation, but rather a much more positive and proactive picture of a higher education system that is engaged in a process of continuous enhancement. The unexplained variables, rather than automatically being labelled as grade inflation, should instead point to a need to investigate further the local institutional contextual factors that inform grade distribution. While their conclusion proposes higher grades as a positive improvement in higher education, it also simply suggests more research is done rather than support their original premise (Jephcote, et al., 2021).

In sum, a strong argument about higher education grading bias comes from a meta-analysis published in the Australian Journal of Education in 2016, which examined 23 analyses from 20 studies. The results strongly supported that bias can occur in subjective grading when graders are aware of irrelevant information about the students.

WHY COLLEGE GRADING IS UNDER SUSPICION NOW

Much of this grading incongruity might have stayed in the research of psychology realm rather than evolve into the application reality until *Inside Higher Ed* got a hold of a story and published the fall-out from a professor at the University of Georgia who tried to offer a self-grading system to their students, saying “Emotional reactions to stressful situations can have profound consequences for all involved,” stated Rick Watson, an instructor at the Terry School of Business (Kabbany, 2017).

This highly publicized case was fraught with internal problems and drew the ire of enough outsiders that the school demanded the syllabus outlining the program be taken off the school website. The university’s response was swift: “A recent online report published a syllabus that a Terry College of Business professor had placed on his website,” Benjamin C. Ayers, Dean of the Terry College of Business, said in a statement to *Inside Higher Ed*. “The syllabus stated that his grading policy would allow students inappropriate input into the assignment of their own grades. I want you to know that the syllabus did not conform with the university’s rigorous expectations and policy regarding academic standards for grading” (Roll, 2017).

The CSC Media Group, which first reported on the professor’s policy, called it a “stunning but not-to-surprising example of the deteriorating quality of education and discipline in America’s universities.” This is a stunning statement in and of itself, since most people reading this article are sure to have had supervisors

give them poor assessments based on bias, not actual performance. It also reverts back to the days of old when professors held themselves in lofty esteem, and as such seemed to consider themselves unbiased authorities of student grades.

The official university comment brings up equally interesting and arguable points on its own. The Dean of the Terry School of Business swiftly claimed the professor's practice defied the "university's rigorous expectations and policy regarding academic standards for grading".

After using just a portion of the research findings that support bias in grading in this paper, comments that only the university is able to set rigorous expectations and policy regarding academic standards means what? That universities, despite volumes of research, do not allow bias or assumptions to enter into their student grading? That, unlike any other profession, educators are able to objectively measure student against student, scantron tests against scantron tests, writing skills against writing skills, and come up with a completely unbiased, five-point choice of grade to measure a student's learning? This would also be blatantly bypassing the concerns and results of research grading studies (and academics do love a good study), and the conscious and unconscious psychological prejudice inherent in any type of judgements we make.

HOW TO WORK ON CHANGING COLLEGE GRADING

As with any significant problem, there are answers that make sense, do not make sense, and are not feasible. One question that begs some answers is why the Instructor from UG was not asked, or perhaps not allowed to reply, to the Dean's comments, the news outlets comments, and the context around his particular system at the time. If all the student grades are posted on the learning management system, and he talks in class about students reviewing their work before emailing him, and the email needed to include solid documentation of why the student's perception of a certain grade is in dispute with the instructor's perception (if it is a perceptual grade), and of the "other" work the student had done outside of the task list, like organizational citizenship behaviors/extra-role behaviors/prosocial behaviors (numerous studies have shown that prosocial behavior impacts performance within organizations [Yaakobi & Weisberg, 2020])-than how can his system (let alone the premise of academic freedom), be questioned? Perhaps this was a practice in self-assessment analysis, a critical skill in the professional world as it leads to promotions, bonuses, raises, and job retention, but the lack of rationale (which to be fair would have been a huge failure on the part of the author if it was a research paper) leaves a gaping hole in what might have been an excellent step in an efficient and effective way to take bias out of grading, and teach students a life-long, necessary skill.

CONCLUSION

The existing data around grading bias, at all levels of education systems, and apparently world-wide, is of great concern. The amount of empirical research is extensive, yet bold ideas for how to begin solving this problem are not. Change, particularly when it takes power out of the hands of one group and redistributes it to others, is one of the most stress-inducing actions in personal and professional settings. However, giving negative appraisal has also been found to be one of the most stressful work interactions we encounter. The suggestions that individuals and organizations take action to increase the objectivity and accuracy of employee evaluations and employee feedback (grading), resulting in an assessment system that is better at meeting its objectives (learning goals), and ultimately will improve organizational productivity (learning) seem like a research supported place to start (Bell, et al., 2020). Even when, however, as our students say, they are only there for the grade.

REFERENCES

- Adams, T. (2019). *Performance Reviews: A Smart Guide to Self-Evaluating*. TinyPulse.com. Retrieved from <https://www.tinypulse.com/blog/performance-reviews-self-evaluating>
- Al-Hussain, S.M. (2008). Prevalence of mistreatment and justice of grading system in five health related faculties in Jordan University of Science and Technology. *Medical Teacher, 30*(3), 82–87.
- Atwater, L., Waldman, D., Ostroff, C., Robie, C., & Johnson, K.M. (2005). Self-other agreement: Comparing its relationship with performance in the U.S. and Europe. *International Journal of Selection and Assessment, 13*(1), 25–40.
- Bell, C.M., Goldman, A., Vogus, C., Zhang, I., & Merritt, S. (2020). Don't Shoot the Messenger: Causes and Levers of Manager Performance Appraisal Anxiety. *Journal of Organizational Psychology, 20*(6), 83–99.
- Branigan, A.R. (2017). (How) does obesity harm academic performance? Stratification at the intersection of race, sex, and body size in elementary and high school. *Sociology of Education, 90*, 25–46.
- Bruce, L. (n.d.). *Common Types of Tests in College*. College Success. Provided by: Lumen Learning. License: CC BY: Attribution, 4.0 International and Commercial. Retrieved from <https://courses.lumenlearning.com/waymaker-collegesuccess/chapter/text-common-types-of-tests-in-college/>
- Dian, M., & Triventi, M. (2021). *PLoS One, 16*(2). San Francisco.
- Finn, K.E., Seymour, C.M., & Phillips, A.E. (2020). Weight bias and grading among middle and high school teachers. *British Journal of Educational Psychology, 90*, 635–647.
- Hoffman, B., Lance, C.E., Bynum, B., & Gentry, W.A. (2010). Rater source effects are alive and well after all. *Personnel Psychology, 63*(1), 119–151.
- Hooijberg, R., & Choi, J. (2000). Which leadership roles matter to whom? An examination of rater effects on perceptions of effectiveness. *Leadership Quarterly, 11*(3), 341–365.
- Johnson, S. (2018). *Tips on Evaluating your Boss*. Chron. Retrieved July 1, 2018, from <https://work.chron.com/tips-evaluating-boss-7179.html>
- Jephcote, C., Medland, E., Lygo-Baker, S., Jephcote, C., Medland, E., & Lygo-Baker, S. (2021). Grade inflation versus grade improvement: Are our students getting more intelligent? *Assessment and Evaluation in Higher Education, 46*(4), 547–571.
- Kabbany, J. (2017). *A professor in Georgia created a stress-reduction policy that lets students choose their own grades*. BusinessInsider.com. Retrieved from <https://www.businessinsider.com/georgia-professor-lets-students-choose-their-own-grades-2017-8>
- López-Pastor, V., Fernández-Balboa, J.-M., Santos Pastor, M., & Fraile Aranda, A. (2012). Students' self-grading, professor's grading and negotiated final grading at three university programmes: Analysis of reliability and grade difference ranges and tendencies. *Assessment & Evaluation in Higher Education, 37*(4), 453–464. <https://doi.org/10.1080/02602938.2010.545868>
- Malouff, J., Stein, S., Bothma, L., Coulter, K., & Emmerton, A. (2014, December 14). Preventing halo bias in grading the work of university students. *Cogent Psychology, 1*(1).
- Malouff, J.M., & Thorsteinsson, E.B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education, 60*(3), 245–256.
- Roll, N. (2017). Professor pulls syllabus letting students set their own grades as 'stress reduction policy'. *Inside Higher Ed*. Retrieved from <https://www.insidehighered.com/news/2017/08/09/professor-pulls-syllabus-letting-students-set-own-grades-stress-reduction-policy>
- Saenz, G., Geraci, L., Miller, T., & Tirso, R. (2017). Metacognition in the classroom: The association between students' exam predictions and their desired grades. *Consciousness and Cognition, 51*, 125.
- Scullen, S.E., Mount, M.K., & Judge, T.A. (2003). Evidence of the construct validity of developmental ratings of managerial performance. *Journal of Applied Psychology, 88*(1), 50–66.

- Smith, A. (2018). More Employers Ditch Performance Appraisals. *SHRM*. Retrieved from <https://www.shrm.org/resourcesandtools/legal-and-compliance/employment-law/pages/more-employers-ditch-performance-appraisals.aspx>
- Tomar, D. (2019). Eliminating the Grading System in College: The Pros and Cons. *The Quad Magazine*. Retrieved from <https://thebestschools.org/magazine/eliminating-grading-system-college-pros-cons/>
- Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology*, *87*(2), 345–354.
- Yaakobi, E., & Weisberg, J. (2020). Organizational Citizenship Behavior Predicts Quality, Creativity, and Efficiency Performance: The Roles of Occupational and Collective Efficacies. *Frontiers in Psychology*, *11*, 758. Retrieved from <https://www.frontiersin.org/article/10.3389/fpsyg.2020.00758>