

Promoting Equity and Mitigating Bias in Online Grading Feedback Processes

Jennifer Schneider
The Community College of Philadelphia
USA
jschneider@ccp.edu

Abstract: It is no secret that grading feedback is a powerful influence on how well and whether one learns (Hattie, 2012). However, the topic of equity and grading feedback is less often discussed. Issues of bias persistently arise when evaluating grading feedback. Instructors often bring biases, some unconscious, some conscious, to the grading process. Bias can result from a variety of factors, including prior knowledge of student grades and scores, race, class, ethnicity, gender, and other factors (Malouff et al., 2014). This presentation seeks to raise awareness and infuse a more intentional reflection on bias, equity, and inclusion into the online feedback process. Original checklists are shared as tools to promote more reflection and inclusivity in the online grading feedback process.

Introduction

Grading feedback, in both face to face and online teaching environments, has been simultaneously characterized as one of the most important but most challenging aspects of teaching, learning, and education (Boud & Molloy, 2013; Hattie, 2012; Nilson, 2015; Tierney, 2013). Literature documents a variety of complex and multi-faceted challenges, including time, stress, variability, and ambiguity as ever present in grading and feedback contexts. (Hattie & Clarke, 2019; Tierney, 2013). There are questions of equity and bias in grading, as well (Schinske & Tanner, 2014; van Ewijk, 2011). By way of example, research has consistently found that grading practices vary greatly across and within schools and programs (Feldman, 2018; Kohn, 1999). Perceptions of feedback are also fraught with stress, negative feelings, and variability. For example, Guskey (2006) explored educators' recollections of their own student experiences with grading. Data collected through questionnaires revealed that nearly 70% of all educators recollected negative experiences in college level courses (Guskey, 2006). Further, in a significant majority of studied cases, recollections conveyed perceptions “of unfair treatment or personal bias on the part of their professors or instructors” (p. 1).

Research has also identified grading as a subjective and inconsistent evaluation process, where an individual student can, and often does, receive dramatically different grades and feedback for the same work, depending upon the time and place of grading and the identity of the grader (Schinske & Tanner, 2014, p. 163; van Ewijk, 2011). Tierney (2013) writes that “[a]n essay that earns a B+ at one moment might earn a B- the next day. It shouldn’t be that way, but any honest teacher will admit it’s true” (p. 2).

TCC 2021 Conference Papers

Challenges in grading reliability present, as well (Meadows and Billington, 2005; Schinske & Tanner, 2014). In a study where 53 professionals were asked to grade 300 freshman essays using a 1-9 scale, not a single essay received fewer than five different grades (Diederich, 1974; Joyce & Joyce, 2017). Further, as researchers for The Academic Senate for California Community Colleges acknowledge, everyone suffers stress and additional tension when an instructor, department, or institution develops a reputation for either less or more rigor in grading than another (Walton et. al, 2008).

The variety of factors impacting the consistency and reliability of grading are broad and far-reaching. Traits and author characteristics such as penmanship (Bull & Stevens, 1979), sex (Spear, 1984), ethnicity (Fajardo, 1985), likeability (Cardy & Dobbins, 1986), and attractiveness (Bull & Stevens, 1979; Landy & Sigall, 1974) all have the potential to impact the way an instructor interacts with and scores student work (Schinske & Tanner, 2014). Instructor experience levels (Weigle, 1999) and the order in which student papers are scored (Farrell & Gilbert, 1960; Spear, 1996) can also impact an instructor's grading and feedback process (Schinske & Tanner, 2014). van Ewijk (2011), citing earlier studies, highlights a range of factors, including group stereotypes, student attractiveness, and interpersonal relationships, that may influence instructor grading. For example, studies have also identified variations in scoring and evaluation based on name (first and surname) (Erwin & Caley, 1984; Lebuda & Karowski, 2013) and accents (Na, 2016). Research also suggests "that biases in teachers' grading practices may harm certain groups of students, depending on their sex, ethnicity, or socio-economic status (Dee, 2004; Dee, 2007; Figlio, 2005; Lavy, 2004; Lindahl, 2007; Ouazad, 2008)" (van Ewijk, 2011, p. 1). Arter & McTighe (2001) highlight related doubts common across teachers and teaching contexts. For example, teachers are often asked to assess and score criteria for which they are not comfortable. Further, questions such as "Maybe I'm not being consistent between students" and/or "Maybe Mrs. Jones next door wouldn't agree with my grades" are common (Arter & McTighe, 2001, p. 9).

Biases associated with grading raise issues of equity, where some students are unfairly harmed (Malouff et al., 2014). However, an inevitable reality of human instruction is that even well-intentioned individuals have biases that impact their actions, can promote discriminatory behaviors, and can hinder the desire to provide feedback in personalized and equitable ways (Fridell, 2017). As an example, Dee (2004) found that students of ethnic minorities received lower test scores when their teacher belonged to the ethnic majority than when their teacher belonged to the students' ethnic group. While van Ewijk (2011) noted that Dee's (2004) research did not determine whether the noted differences were a result of biased grading and/or some other factors, the irregularity and variability of grading persists as an ongoing challenge.

Research has also identified situations involving incomplete information, time constraints, and fatigue as those especially susceptible to implicit bias (Staats, 2016). Staats (2016) writes that "[g]iven that teachers encounter many, if not all, of these conditions through the course of a school day, it is unsurprising that implicit biases may be contributing to teachers' actions and decisions" (p. 30), including grading feedback. As another example, a 2014 study evaluated how confirmation bias (mental shortcuts that

TCC 2021 Conference Papers

support the activity seeking of information that affirms pre-existing beliefs, can unconsciously impact the evaluation of reviewed work (Reeves, 2014). In this study, researchers designed a fictitious legal memo that contained almost two dozen intentional spelling, grammar, analytical and technical writing errors. Identical memos were shared with law firm partners for evaluation. For memos with an author that was identified as African American, reviewers found more of the embedded errors and rated the memo as lower quality than for memos where the author was identified as white (Reeves, 2014). Although this study focused on the evaluation of a legal memo, “it is not a stretch of the imagination to consider the activation of this implicit dynamic in grading student essays or evaluating other forms of subjective student performance” (Staats, 2016, p. 31). Studies have also uncovered evaluative biases against a wide range of students, including female students (Spear, 1984), male students (Martin, 1972), black students (Piche, Michellin, Rubin, & Sullivan, 1977), and white students (Fajardo, 1985) (Malouff et al., 2014). Similarly, researchers have identified biases in pre-service teachers and their evaluation of students with migrant backgrounds (Bonefeld & Dickhäuser, 2018).

While the presence of implicit bias in grading and other contexts is well-documented, less is known about what steps and strategies can most successfully mitigate such bias. Research suggests that professional development training can be an effective way to mitigate implicit bias in a variety of contexts (Chamberlain, 2016; re:Work, n.d.b; Vianden, 2018; Woolf & Dacre, 2011). Research also suggests online diversity training has the potential to encourage behavior change (Chang et. al, 2019). Some argue that “[t]he starting point for any debiasing intervention is implicit bias training” (Race Equity Project, n.d.). Relatedly, the Implicit Association Test (the “IAT”) is a widely used resource that is designed to help identify and elicit awareness of implicit biases (Greenwald, Nosek, & Banaji, , 2003; Kang, Gray, & Dovidio, 2014; Project Implicit, 2011; Sukhera, Wodzinski, Rehman, & Gonzalez, 2019). For example, in a 2014 study, Kang, Gray, & Dovidio examined the potential impact of lovingkindness meditation on improving implicit negative attitudes toward members of culturally stigmatized groups. Participants completed measures of implicit attitudes toward Blacks and homeless people, using the IAT, at the beginning and end of the study period. The study found that decreases in implicit bias against stigmatized outgroups were observed only in the lovingkindness practice condition (Kang, Gray, & Dovidio, 2014).

Although there remains much to learn about utility and impact of different types of training, most notably in connection with the grading feedback process, we know that in our diverse classrooms, intentionality and awareness are critical components of anti-bias in grading work and instruction. In particular, our feedback must acknowledge the experiences and identities of all persons and students. Awareness is a helpful strategy. Early Childhood Education Assembly (2020) writes, if we do not “counter anti-Blackness every single day in classrooms through our attitudes, curriculum, policies, and practices, we are complicit in perpetuating a racist society” (para. 5). Checklists have proven useful in a variety of contexts, including medical care, medical education, medicine, hiring and courts (American Society of Trial Consultants, 2011; Burgess, Van Ryn, Dovidio, & Saha, 2007; re:Work, n.d.a).

TCC 2021 Conference Papers

This remainder of this article presents a collection of anti-bias in grading checklists that can be used when evaluating student work. The goal of the shared checklists is to raise awareness, counter bias in grading feedback, and infuse more intentional reflection on equity and inclusion in the feedback process. Individual faculty might use the sample checklists for reflection and to identify personal strengths as well as areas with room for improvement. Reflection might occur at the beginning of any individual grading session and/or in a regular and sustained fashion. Additionally, organizations might structure training for faculty in ways that both highlight the challenges of bias and grading and demonstrate how these resources might be integrated into faculty workflows and practice. As you review the resources, reflect: Do I do this consistently? How might I do better? Use one, use them all. Tailor as needed.

Anti-Bias Online Grading Feedback Checklists

Feedback Personalization Checklist

- Do I address students by their preferred names?
- Do I use students' preferred pronouns?
- Do I sign my comments with my name?
- Have I encouraged questions?
- Have I included my contact information (email, phone, etc.)?
- Do I adopt a professional, respectful tone?
- Do I sound genuine?
- Is my tone supportive?
- Is my tone positive, not punitive, in nature?
- Does my feedback initiate a dialogue between myself and the student/author?

Instructional Presence Checklist

- Have I included scholarly resources to clarify assignment topics and further student learning?
- Have I modeled proper formatting for any shared resources?
- Do I propose and/or acknowledge opposing arguments in connection with shared perspectives?
- Do I pose thoughtful questions to promote student reflection and critical thinking?
- Do I use grammatically correct sentences?
- Is my writing clear?
- Do I alert students when grades are posted?
- Do I remind students how to access and review grading feedback?

Intellectual Presence Grading Checklist

- Does my feedback promote ongoing learning and knowledge-building opportunities for students?
- Do I incorporate open-ended, forward-focused critical thinking questions to promote ongoing learning?
- Do I include supplemental resources to support further learning for students at all levels?
- Do I model proper citation format in my own writing?

TCC 2021 Conference Papers

Motivational Feedback Checklist

- Have I considered the tone and the phrasing of my feedback?
- Rather than focusing only what students didn't do, have I provided positive reinforcement and encouragement for what they did?
- Do I highlight and acknowledge specific points raised in student work?
- Does the language of my feedback demonstrate cultural sensitivity and support for all students of all backgrounds?
- Is my language and choice of words supportive, rather than accusatory?
- Does my language and choice of words emphasize achievements, rather than only deficits?
- Have I shared specific recommendations for what students could and/or should do to improve?
- Does my feedback focus on and emphasize the contents of student work and how the work-product can be improved going forward?

Inclusiveness Checklist

- Do I avoid an accusatory tone?
- Do I avoid overuse of "you" in my comments?
- Do I encourage questions and reiterate my availability?
- Do I express empathy in my comments?
- Does my feedback vary in style so as to support a variety of learning preferences?
- If I share video feedback, do I include a transcript?
- Are all shared resources accessible?
- Are shared resources authored by a wide range of diverse authors?
- Do recommended follow-up resources include works by authors that reflect the diversity of my students?
- Does my grading feedback support and consider a variety of diverse perspectives?
- Do I encourage cross student collaboration, reflection, and knowledge building in my feedback?

Unbiased and Equitable Feedback Checklist

- Have I provided all students consistent feedback?
- Have I reviewed my feedback for logical errors?
- Have I considered alternative perspectives?
- Do I clarify why points are deducted?
- Do I maintain high expectations for all students?
- Do I score work anonymously?
- Do I refer to rubric elements when scoring each student's work?
- If I offer resubmissions opportunities, do I extend those opportunities to all students?
- If I permit late submissions, do I permit that for all students?
- Do I approach grading with a positive and open state of mind?
- Do I approach grading in ways that minimize my own fatigue?
- Do I grade in comfortable conditions?
- Do I grade under similar (as similar as possible) conditions for all students?

TCC 2021 Conference Papers

Anti-Bias Feedback Checklist

- Do I make consistent efforts to minimize assumptions when grading?
- Do I score work without identifying the student author?
- Do I work to avoid letting my personal values impact my grading?
- Do I review all rubric and assignment elements before grading?
- Do I reflect on all shared comments to eliminate stereotypes?
- Do I attempt to score multiple-question assessments by question (thereby minimizing the potential for scores on later questions to be influenced by student performance on an earlier question)?
- Do I pause and consider whether and how my surroundings might positive and/or negatively impact the objectivity of my feedback?

Online Grading Feedback Process Acronyms

PASTE

Before copying and “PASTE”-ing prewritten comments, ask: Is my feedback: PASTE?

Personalized

Actionable

Specific

Timely

Encouraging

PRIMED

Before finalizing grading, ask: Is my feedback PRIMED?

Positive

Reflective (of student work)

Inclusive

Motivating

Equitable

Differentiated (by student work and student need)

Additional Resources

[Implicit Association Test](#)

[Community Toolbox, Building Cultural Competence](#)

[Race Equity Project, Debiasing Techniques](#)

[Before You Make That Big Decision](#)

[Rework and Google, Unbiasing Checklists](#)

[Document/Collection of Checklists](#)

[Promoting Inclusive and Unbiased Grading Feedback](#)

[Cultural Sensitivity and Online Discussion Board Rubric Language](#)

TCC 2021 Conference Papers

References

- American Society of Trial Consultants (2011). Strategies to reduce the influence of implicit bias. Retrieved from https://cdn.ymaws.com/www.napaba.org/resource/resmgr/2015_NAPABA_Con/CLE/400s/404_NAPABA2015CLE.pdf
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin Press.
- Bonefeld, M., & Dickhäuser, O. (2018). (Biased) grading of students' performance: Students' names, performance level, and implicit attitudes. *Frontiers in Psychology, 9*. <https://doi-org.ezproxy.snhu.edu/10.3389/fpsyg.2018.00481>
- Boud, D. & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education 38*(6): 698-712. 10.1080/02602938.2012.691462
- Bull, R., & Stevens, J. (1979). The effects of attractiveness of writer and penmanship on essay grades. *Journal of Occupational Psychology, 52*(1), 53–59. <https://doi-org.ezproxy.snhu.edu/10.1111/j.2044-8325.1979.tb00440.x>
- Burgess, D., van Ryn, M., Dovidio, J., & Saha, S. (2007). Reducing racial bias among health care providers: Lessons from social-cognitive psychology. *Journal of General Internal Medicine, 22*(6), 882–887. <https://doi-org.ezproxy.snhu.edu/10.1007/s11606-007-0160-1>
- Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology, 71*, 672–678. doi:10.1037/0021-9010.71.4.672
- Chamberlain, R. (2016). Five steps toward recognizing and mitigating bias in the interview and hiring process. *Strategic HR Review, 15*(5), 199–203.
- Chang, E. H., Milkman, K. L., Gromet, D. M., Rebele, R. W., Massey, C., Duckworth, A. L., & Grant, A. M. (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences of the United States, 16*, 7778.
- Dee, T. S. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment, *Review of Economics and Statistics, 86*(1), 195–210.
- Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources, 42*(3): 528-554.

TCC 2021 Conference Papers

- Erwin, P. G., & Caley, A. (1984). The influence of Christian name stereotypes on the marking of children's essays. *British Journal of Educational Psychology*, 54, 223–227. doi:10.1111/j.2044-8279.1984.tb02583.x
- Fajardo, D. M. (1985). Author race, essay quality and reverse discrimination. *Journal of Applied Social Psychology*, 15, 255–268. doi:10.1111/j.1559-1816.1985.tb00900.x
- Farrell M. J., & Gilbert N. (1960). A type of bias in marking examination scripts. *Br J Educ Psychol.* 30:47–52.
- Feldman, J. (2018). School grading policies are failing children. A call to action for more equitable grading. Retrieved from <https://crescendoedgroup.org/wp-content/uploads/2018/10/Call-to-Action-for-Equitable-Grading-Oct-2018.pdf>
- Figlio, D. (2005), Names, Expectations and the Black-White Test Score Gap. NBER working papers, No. 11195, National Bureau of Economic Research.
- FitzGerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: A systematic review. *BMC Psychology*, 7.
- Fridell, L. A. (2017). *Producing bias-free policing: a science-based approach*. Springer.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197–216. doi:10.1037/0022-3514.85.2.197
- Guskey, T. R. (2006). “It Wasn’t Fair!” Educators’ Recollections of Their Experiences as Students with Grading. In Online Submission. Online Submission.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.
- Hattie, J., & Clarke, S. (2019). *Visible learning feedback*. Routledge. New York, NY.
- Kaihlanen, A.-M., Hietapakka, L., & Heponiemi, T. (n.d.). Increasing cultural awareness: qualitative study of nurses’ perceptions about cultural competence training. *BMC NURSING*, 18(1).
- Kang, Y., Gray, J. R., & Dovidio, J. F. (2014). The Nondiscriminating Heart: Lovingkindness Meditation Training Decreases Implicit Intergroup Bias. *Journal of Experimental Psychology. General*, 143(3), 1306–1313.

TCC 2021 Conference Papers

- Kohn, A. (1999). From degrading to de-grading. Retrieved from <https://www.alfiekohn.org/article/degrading-de-grading/?print=pdf>
- Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology*, 29, 299–304. doi:10.1037/h0036018
- Lavy, V. (2004). Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment, NBER working papers, No. 10678, National Bureau of Economic Research.
- Lebuda, I., & Karwowski, M. (2013). Tell me your name and I'll tell you how creative your work is: Author's name and gender as factors influencing assessment of products' creativity in four different domains. *Creativity Research Journal*, 25, 137–142. doi:10.1080/10400419.2013.752297
- Lindahl, E. (2007) Comparing teachers' assessments and national test results – evidence from Sweden, IFAU Working Paper 2007:24.
- Malouff, J. M., Stein, S. J., Bothma, L. N., Coulter, K., & Emmerton, A. J. (2014). Preventing Halo Bias in Grading the Work of University Students. *Cogent Psychology* 1: 988937
- Meadows, M., & Billington, L. (2005). A review of the literature on marking reliability. Unpublished AQA report produced for the National Assessment Agency. Retrieved from http://archive.teachfnd.com/qcda/orderline.qcda.gov.uk/gempdf/184962531X/QCDA104983_review_of_the_literature_on_marking_reliability.pdf
- Na, E. (2016). Do College Instructors Have Implicit Bias toward Latino-Accented English Speakers? *Commission for International Adult Education*.
- Nilson, L. B. (2015). *Specifications grading: Restoring rigor, motivating students, and saving faculty time*. Sterling, VA: Stylus Publishing.
- Project Implicit (2011). Implicit association test. Retrieved from <https://implicit.harvard.edu/implicit/takeatest.html>
- Ouazad, A. (2008), Assessed by a Teacher Like Me: Race, Gender, and Subjective Evaluations, INSEAD Working Paper.
- Race Equity Project (n.d.). Debiasing techniques. Retrieved from <https://www.courts.ca.gov/documents/ab1058-implicit-bias-handout1.pdf>

TCC 2021 Conference Papers

- Re:Work (n.d.a). Using unbiassing checklists. Retrieved from <https://rework.withgoogle.com/guides/unbiasing-use-structure-and-criteria/steps/use-unbiasing-checklists/#:~:text=Unbiasing%20checklists%20are%20integrated%20into,and%20others'%2C%20unconscious%20biases.>
- Re:Work (n.d.b). Read Google's checklist research. Retrieved from <https://rework.withgoogle.com/guides/unbiasing-use-structure-and-criteria/steps/read-googles-checklist-research/>
- Reeves, A. (2014). Written in black & white exploring confirmation bias in racialized perceptions of writing skill. Retrieved from <http://nextions.com/wp-content/uploads/2017/05/written-in-black-and-white-yellow-paper-series.pdf>
- Schinske, J., & Tanner, K. (2014). Teaching More by Grading Less (or Differently). *CBE—Life Sciences Education*, 13(2), 159–166. <https://doi.org/10.1187/cbe.cbe-14-03-0054>
- Spear, M. G. (1984). The biasing influence of pupil sex in a science marking exercise. *Research in Science & Technological Education*, 2, 55–60. doi:10.1080/0263514840020107
- Spear, M. (1996). The influence of halo effects upon teachers' assessments of written work. *Res Educ.*, 56, 85–86.
- Staats, C. (2016). Understanding implicit bias: What educators should know. *American Educator*, 39(4), 29–33.
- Sukhera, J., Wodzinski, M., Rehman, M., & Gonzalez, C. M. (2019). The Implicit Association Test in health professions education: A meta-narrative review. *Perspectives on Medical Education*, 8(5), 267–275.
- Tierney, J. (2013). Why teachers secretly hate grading papers. Retrieved from <https://www.theatlantic.com/national/archive/2013/01/why-teachers-secretly-hate-grading-papers/266931/>
- van Ewijk, R. (2011). Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Economics of Education Review*, 30(5), 1045–1058.
- Vianden, J. (2018). “In All Honesty, You Don't Learn Much”: White College Men's Perceptions of Diversity Courses and Instructors. *International Journal of Teaching & Learning in Higher Education*, 30(3), 465.
- Weigle, S.C. (1999). Investigating rater/prompt interactions in writing assessment: quantitative and qualitative approaches. *Assessing Writing*. 6:145–178.

TCC 2021 Conference Papers

Woolf, K., & Dacre, J. (2011). Reducing bias in decision making improves care and influences medical student education. *Medical Education*, 45(8), 762–764.