Checked Your Bias Lately?
Reasons and Strategies for Rural Teachers to Self-Assess for Grading Bias

Patricia L. Hardré
University of Oklahoma

Multiple factors influence teachers’ grading and scoring of students’ class work, homework, projects and tests. Put simply, bias in grading is giving different grades on student work of essentially equal quality, based on factors irrelevant to the scope and criteria for that work. Grading is to a degree subjective, but it need not be biased. Intentional bias is a common criticism of teachers and has been the topic of numerous studies and reports. However, less attention has been given to unintentional sources of grading bias. For teachers in any school, bias can creep into grading despite a teacher’s best efforts to be fair and impartial. Teachers in rural schools may face more challenges to preventing and reducing grading bias. This paper identifies some of the issues relevant to teachers’ unintentional grading biases and discusses both conventional and innovative ways to reduce it.

Keywords: Grading bias; school accountability; student achievement and motivation; teacher grading strategies; assessment standards

Schools are being scrutinized, challenged and called to account for every element of the work they do. Even matters traditionally left to teachers’ expertise and judgment are now susceptible to close inspection. More than ever before, in the face of recent governmental increases in school accountability, rural schools need effective strategies to check and remediate issues of equity and quality. One such issue is intentional or unintentional grading bias, which can leave schools open to accusations from grade inflation to more extreme grade manipulation. Teachers and administrators in small and rural schools need defensible accountability and quality-of-practice strategies that do not require huge resources to implement and monitor. Systematically checking for grading bias is one such strategy.

Bias is a personal or unreasoned distortion of judgment (Webster’s Collegiate Dictionary, 2000). In assessment, bias as a technical term most often refers to a characteristic of tests that present advantage or disadvantage to a particular subgroup (e.g., by gender or ethnicity) (Nitko, 2004; Popham, 2005). However, as a more general term, bias refers to any differential in grading across learners that is not caused by completeness or quality of work on the assigned task. The Student Evaluation Standards specify that grades should be free from bias, that is, “free from influence by factors unrelated to the purpose of the assessment” (Joint Committee on Standards for Educational Evaluation, 2003, p. A7).

Teachers in rural and small schools are often deeply embedded in the community and connected to knowledge regarding generations of families. Because teachers’ knowledge of student characteristics is a significant predictor of grading bias (Rauschenberg, 2012), rural teachers may face more danger of grading bias than their urban and suburban peers, nationally. Many non-achievement factors are embedded into student grading (Howley, Kussimo & Parrott, 2000), both intentionally and unintentionally. All teachers are prone to elements of subjectivity in grading that takes scores and grades off-target from their original criteria and standards (Brookhart, 2004). To address these tendencies requires checking for instances of bias, and recalibrating grading practice to fit its original purpose and targets, to realign it with standards of accuracy, reliability and validity (Malouff, 2008). While urban and suburban teachers, or those in larger districts, may have more accessible resources to use in this checking and recalibration process, rural teachers, more isolated in small and geographically-distant schools, often have less immediate access to in-place and on-site resources to do so. Teachers and administrators at a rural conference expressed universal concern that the strategies often recommended to maintain accuracy, reliability and validity of assessments (to reduce bias) are particularly difficult for rural schools, given their small size and remoteness (personal communication, discussion group, NREA conference, 2013). When one teacher is the school’s whole math department or the whole first grade, it is harder to find a colleague with the right expertise to check grading. Rural schools and administrators have to innovate to maintain high-quality assessment standards like grading equity and accuracy. For these reasons, rural teachers and
administrators need strategies to identify and address the potential for grading bias.

What is Unintentional Grading Bias, and Where Does It Come From?

Unintentional bias in grading or scoring student work is giving what amount to different grades on work of essentially equal quality, based on factors irrelevant to the criteria for judging that work (Banks, 2005; Nitko, 2004). Most teachers introduce unintentional grading bias into their judgment at one time or another, usually without being aware of doing it. Grading bias comes from two major sources: design factors and personal factors, both of which are influenced by environmental factors.

Sources of Grading Bias

One source of grading bias arises from design and implementation of assessment, influenced by teacher beliefs about the nature of knowledge in the discipline and the multiple purposes of grades (Nitko, 2004; Popham, 2005), embodied in teachers’ choice of assessments and how they are used (Wylie, et al, 2012). Common design factors that may influence teachers’ perceptions and accuracy during grading (and produce scoring bias) include mismatch and misalignment of the assessment with its purpose. Teachers may inadvertently include performance expectations and components in their assessments that are not entirely aligned with the actual goal of an assessment, and if additional parameters are included without accounting for them in the scoring model, and without making students aware of them, these may constitute grading bias.

A second source, relational and interpersonal bias factors, is related to the identities and interpersonal relationships of teachers and students. This type of bias is caused by what teachers know about students’ past academic performances and experiences, past and present ability assessments, out-of-school and family circumstances, and affect or attitudes toward the teacher or class (Banks, 2005; Brookhart, 2004) Teacher knowledge and beliefs about individual students tend to produce most individual differences in assessment and grading (Rauschenberg, 2012). Research has demonstrated that factors such as emotional connectedness and relatedness, knowledge of at-home challenges students had to overcome, and interest in motivating students have caused teachers to adjust grades upward from strictly criterion-based grades or scores (e.g., Hardré, 2008; Hardré & Sullivan, 2008). These shifts may be intentional and explicit or unintentional and implicit, and implicit action can become a slippery slope, increasing subjectivity over time.

Systematic strategies for checking bias can expose such tendencies and allow teachers to make explicit choices. This is not to say that there is no place for using grades for developmental and motivational purposes, but such choices need to be explicit, made with full awareness and clearly justified, rather than loosely subjective or accidental. Teachers using “objective tests” sometimes assume their objectivity to be absolute, but wherever judgment comes into play, on an essay response, points for method in solving math problems, or perspectives that explain alternate responses, bias may creep in. Periodically checking and reflecting on assessment and grading practice is a key component of active, reflective teaching (Brookhart, 2004; Zeichner & Liston, 1996).

Mismatch between Target Knowledge or Skill and Scope or Method of Assessment

Freedom from bias means that if the purpose of a test is to assess science knowledge and skill, nothing else should have a significant impact on grades for that assignment. If the way that learners are being assessed gets in the way of their demonstrating the target knowledge and skill, it constitutes a response-type bias, as well as an underlying scope bias. An historic example of response-type bias is the use of essays or text-based items for math knowledge, in which language skills influenced how clearly knowledge was communicated, and therefore biased grades (Joint Committee on Standards for Educational Evaluation, 2003). It is perfectly legitimate to include the skills of communicating relevant knowledge in essay form and correctly using relevant vocabulary as part of a math or science class assessment; it just needs to be communicated as part of the expectations for the assessment and included in the instruction. If these things are assumed and not articulated, and the student who is a good writer gets a higher grade than the student who is not, it can produce frustration, demotivation and accusations of unfairness or bias.

Misalignment of Scoring with Performance Expectations

This type of bias overlaps with scope or method mismatch, but is broader, as it includes any gap between the explicit expectations placed on an assessment and how it is scored. Factors that bias teachers’ scoring of assignments include simple things such as fatigue or anxiety, or environmental factors such as interruption, that skew scores (Arends & Kilchner, 2010). Teachers’ personal approaches to scoring, whether derived from optimistic or pessimistic
philosophical underpinnings, can result in differences. For example, a teacher who approaches an assignment assuming passing quality and subtracts for errors, and a teacher who approaches the same assignment assuming a zero baseline and adds score for earned achievement can produce very different scores for the same assessment from the same rubric. These differences are infrequently addressed, yet they introduce bias into grading because the source of the difference is not contained in the student’s actual knowledge as demonstrated on the assessment. This kind of differential grading bias can be illuminated through systematic bias checking and reduction strategies.

Knowing Students, Their Families and Circumstances

Of course, teachers try to handle interpersonal factors objectively, but many admit they are sometimes unsuccessful. Positive connections and interpersonal relatedness can produce more lenient grading, and negative emotionality can produce stricter grading; all at the level of the subconscious (Malouff, 2008). Knowledge about students’ special challenges at home can promote more lenient grading. Knowing students’ families as neighbors or friends, community leaders or local outlaws; having had older siblings as previous students; and having their parents as professional colleagues can produce grading bias. Conversely, some teachers may actually grade harder in an effort to compensate for potential positive interpersonal bias (Cole, 2008).

Knowing Students’ Past Academic Achievements

There is a tendency, especially for less experienced or efficacious teachers, to be biased by past teachers’ grading (Brookhart, 2004). If a past teacher’s grading is perceived as a more expert evaluation of the student’s performance, a teacher may unconsciously (or consciously) align present grades with that past performance evaluation, without regard to how the student may have improved in the interim. When teachers feel less expert in making grading judgments, they may engage in mental tracking of students, such as thinking of one student as “an A student” and another as “a C student” and subconsciously align current grades with those mental tracks.

Subtle elements of communication that happen in schools may actually create bias regarding individual students. To reduce the probability of novice teachers being biased toward (or against) particular students, veteran teachers should avoid making sweeping judgments (or sharing gossip) about individual students. The new teacher who hears three experienced peers make the statement that a certain student “never works” or “does only average work” may be biased toward giving only average grades, despite what the student produces.

Knowing Students’ Relative Aptitudes and Special Needs

Teachers may also be biased by students’ tested aptitudes and special needs. That is, if one student has high scores on math tests and another has low scores, the teacher may be more critical of, or optimistic about, the first student’s math computation and strategy use. The teacher may unconsciously be using the model of aptitude (an “objective” predictor of probable achievement) to guide scoring judgments when the immediate judgment seems difficult to discriminate. Knowing students’ abilities may cause teachers to grade them relative to what they can achieve or produce instead of simply based on the objective criteria for what they have achieved or produced. Another issue is teachers knowing if students have identified disabilities or special needs. Even if they are high-functioning and in mainstream classrooms, research has demonstrated that teachers often tend to score students with special needs differently from their peers (Mastergeorge & Martinez, 2010).

Knowing Students’ Attitudes Toward the Teacher or Class

Teachers may tend to score assignments by students who are engaged and positive about the class more generously than those who are negative and critical about the class, even though their performance by objective criteria may be comparable.

Why is Grading Bias a Problem?

If teachers and administrators consider grading a tool for motivation or communication apart from claims to objectivity and equity, or if they consider subjective and affective judgments unrelated to the assignment standards acceptable in influencing students’ performance and assessment in school, then grading bias may not be a problem at all. However, if they believe that all students in a classroom should be assessed on an equal, fair and relatively objective basis, then unintentional bias is a serious problem. If they believe that grades on assignments (and in classes) should reflect the objective performance of each individual as measured against the relevant instruction and criteria for a given assignment, then they will be concerned with how teachers can self-monitor and check for personal grading bias.
It is also important to consider possible negative effects of bias—even positive bias—on students. Making grading criteria as transparent as possible and maintaining perceptions of fair, unbiased grading promotes healthy student attitudes and motivation, while perceptions of unfair, arbitrary or personalized grading practices can cause students to feel like their effort to learn has no effect on their achievement. If students have no evidence of teachers’ objectivity (or think they see bias), they may tend to attribute their lack of achievement to teacher bias or favoritism (factors outside their control) rather than to their actual effort and performance (factors within their control). Students’ attributions of what controls their performance outcomes are critical to their motivation and achievement (Banks, 2005; Brookhart, 2004; Brophy, 1998).

Students sometimes recognize and manipulate teachers’ grading. Older students admit that they can identify (and may manipulate) teachers’ grading tendencies (Banks, 2005). Aside from these extreme examples, teachers may ask, “Isn’t some consideration of students’ special circumstances okay?” Certainly, teachers should show students understanding and give them extra time or opportunity if doing so facilitates instructional goals (e.g., learning, motivation, management). However, special consideration is counterproductive to learning if it teaches students that teachers are manipulable or inconsistent. For rural teachers or administrators interested in checking for and addressing possible bias in assessment and grading, the following sections describe some effective practical strategies, with particular attention to how they can be implemented in rural and small school and districts.

How Can Teachers Self-assess and Adjust Their Existing Grading Biases?

A useful means to check and adjust for grading bias related to personal knowledge of students is to use multiple graders. If grading is done by both the teacher and at least one other qualified person and those grades compared, it can illuminate unrecognized bias. If multiple graders’ scores are averaged, the result may be a more balanced, objective assessment than one teacher’s grade alone. If an additional grader is not available, there are ways that one teacher can self-check for grading bias.

Exchange Grading

One way to check for grading bias is to share scoring tasks with colleagues. If teachers are in a school large enough to have several people teaching sections of the same class, they can give the same assignment and exchange student papers. Without looking at the colleagues’ grades, they can grade each other’s students’ papers and then compare. Grading with a fellow teacher teaching the same content to the same level of students and comparing scores is a good way to identify grading bias that may exist as well as to control for it. If several teachers of the same subject are not available, scoring by two independent raters is still a good check; however, the use of multiple graders needs to be attentive to their professional qualifications (e.g., only use a person well qualified in the discipline). In a small school, with only one teacher of Art or Geometry, for example, the teacher should find the colleague best qualified in the subject area (e.g., someone who minored in it or has other specialized expertise), or seek out a qualified peer teacher in a different school.

Collaborative Grading

Another excellent check is to have a group of teachers grade the same blinded student assignments, using the same rubric, and then compare grades. If teachers independently assign very different grades to the same assignment, it can indicate differential biases that they need to examine. Based on any different grades assigned to the same students’ work, they can discuss and try to identify what caused those differences. Telling students about collaborative grading can enhance their perceptions of a broader audience and encourage more objective reasoning.

Whenever teachers use shared grading strategies, it is important to pre-check for decision consistency with “test” papers similar to the actual student assignments, to make sure that all teachers are applying the rubric in the same way. While it is worth the investment, organizing and implementing shared grading strategies takes time, so teachers (and administrators) should select the most important opportunities to use it. Using multiple graders for particular assignments, but not for every assignment, respects teachers’ valuable and limited time. Teachers can also individually reciprocate with colleagues to share grading, with the potential residual benefits of sharing ideas, and infusing their teaching and assessment thinking with new energy.

Spaced Repeat Scoring

If multiple raters are not available or not practical in a given instance, a single teacher can use spaced repeat scoring. This is an effective tool for periodic self-monitoring of grading consistency. The teacher works through grading the assignments once using the appropriate scoring or grading model and rubric, but records grades in a separate place (not on the
assignments themselves or in the grade book). Then the teacher sets aside the whole set of assignments for several days or a week. After that time (or longer depending on how acute long-term memory is) the teacher takes out the assignments and independently scores them again using the same rubric and under the same conditions. The same conditions are important because bias can be introduced by the context and circumstances under which grading is done (e.g., location, noise, distractions, fatigue, time of day). After the second scoring, the teacher takes out the original grades and compares the two sets for consistency.

For any type of bias checking (multiple raters or one), some variability is expected, but there is a degree of difference that constitutes “inconsistency”. The teacher needs to define, in advance, what degree of disagreement indicates inconsistency, versus a tolerable amount of variability. General rules for “inconsistency” are a full letter grade, or more than 10-20% of a numeric score range (e.g., 2 points on a 10-point assignment). If two sets of grades on the same assignment are inconsistent, then the teacher looks back at the assignments and tries to determine why the inconsistency exists.

Similar patterns of inconsistency may exist between two repeat scorings as between two graders’ evaluations (e.g., more attention to small technical errors). If such patterns exist, the teacher can decide whether the difference is due to bias or to better discrimination on one occasion than on the other. If it seems to be bias (characteristics unrelated to the criteria for quality on that assignment) then actual grades assigned to the student should be on the less biased scores, those which focused more clearly on the established criteria. However, if the different scores seem to represent appropriate discrimination the two sets can be averaged.

In practical terms, implementing collaborative, shared or repeated grading strategies, wait time is generally a consideration. If the purpose of the practice is primarily as a check for teachers (self-monitoring), then assignments can be copied (graded and returned) and the more complicated process completed at leisure without pressure to get grades or papers back to students. If the multi-grader strategy will actually generate a collaborative or averaged grade for students, then it may help to make students aware that grading and return will take more time. Of course, such decisions and communication must be developmentally appropriate.

**Increasing Reach with Digital Technologies**

All collaborative and shared grading strategies require access to peer teachers, and rural and small school teachers may lack on-site access. Some distributed schools are using digital technology tools for connection and collaboration, and they can be utilized with the strategies above, with attention to their additional unique challenges. Compatibility and connectivity of systems are one set of hurdles to be addressed, and systems are improving that enable one site or individual user to host synchronous and asynchronous collaborations. In many cases, the school can purchase a single license that multiple teachers can use. Security is another critical issue in transmitting student grade information over digital systems. One strategy to add a layer of confidentiality is the use of codes or pseudonyms for transmission and communication, with the additional *caveat* that the most secure system possible should always be used to share any student information. Keys to retranslate identities can be retained by the host teacher. Rural teachers in solo roles can use digital media to recruit colleagues with similar expertise and to share rubrics to collaborate.

**What if these Checks Reveal Grading Bias?**

If teachers engage in bias checks and find reason to believe unintentional bias exists, the next steps are problem-solving. Based on the data from the bias check, consider why it exists. Which of the common causes in the list above may explain it? Does the pattern appear to be a design flaw in the rubric, an unrealized philosophical assumption, or perhaps an interpersonal connection to certain students? Once it is identified, it can be addressed. Sometimes awareness is enough, and cognitively monitoring enables the teacher to reduce and control that source of bias. If the issue seems to be environmental, like interruptions, teachers can find strategies to reduce those factors in the environment. If it seems to be due to personal fatigue or anxiety, teachers can brainstorm ways to reduce them, like taking frequent breaks while grading piles of student work. If the bias issue is more complex, teachers can identify strategies to support ongoing monitoring, strategies like including written reminders in personal grading instructions and notes.

**How Can Teachers Prevent Grading Bias?**

A number of effective strategies exist for reducing the probability of individual grading bias, including blind scoring assignments, and using precise scoring rubrics.

**Blind Score Assignments**

One well-established way to control for grading bias is to blind score assignments. This simply
involves masking the students’ identities from the teacher during grading. If grading is “blind” (without names or identifying characteristics), then identity is masked and bias linked to individual identities is reduced. Strategies for blind scoring include assigning students code numbers, masking names on papers during grading, and requiring work to be typed to reduce handwriting recognition.

Some teachers use the same code numbers during grading and when students’ grades are posted or read aloud, and often over a period of time such as a whole semester. The problem with this approach is that teachers memorize students’ code numbers, and so do other students (by watching their peers’ reactions when grades are made public). Thus, long-term code numbers can become essentially ineffective for reducing bias and even psychologically detrimental to students. If student code numbers are used to mask identity, then they must be changed frequently and kept secure, or they become essentially useless.

Use Precise Scoring Rubrics

Another strategy to avoid bias is to use precise, appropriate scoring rubrics when grading assignments. The more specific scoring criteria are and the more consistently they are applied, the less scoring bias is likely to occur. Teachers can also enhance students’ perceptions of grading objectivity and consistency by sharing rubrics with them up front, so they see the target performance standard, instead of possibly imagining a hidden agenda.

Bias Checking as Professional Development

Regular checks for grading bias can be effective professional development tools for individual teachers, for departments and for whole schools. They validate the claim that teachers and schools are grading on criteria central to the discipline and appropriate to students’ developmental levels. They can improve assessment consistency across grade levels and content areas. They can improve teachers’ confidence in their own grading methods and policies, and can reduce students’ perceptions that teachers’ grades are individualistic and subjective or arbitrary.

Some residual benefits beyond grade calibration and bias reduction exist for using professional development and faculty meetings to address grading needs and practice with strategies such as those described above. It can open up surprisingly productive conversations about the philosophies, purposes and meaningfulness of grading. Stepping back to consider, refine, and examine foundational issues such as grading holds promise to improve teaching across and between schools and districts, and among colleagues distributed over geographic distances.

Administrative and policy support for collaborative bias checks can promote ongoing collaboration within and between grade levels and departments as well. It can help build educational community in the school and help reduce teachers’ perceptions of professional isolation. Administrators can support this practice by enabling grade collaboration as an in-service option (so it is not an additional time requirement on top of all the teachers’ other development requisites). An admonition to administrators is that it is not generally effective to mandate bias checks. Teachers (just like students) can perceive mandated activities as “busy work” and do them poorly or mechanically, generating useless information and developing negative attitudes toward the practice.

School Policy and Culture Prevention Strategies

Beyond teachers taking individual responsibility for reducing grading bias, school administrators can encourage a school culture that creates a level ground for evaluating students’ performance. By supporting and encouraging the practice of periodically checking for grading bias, school policy makers can support consistency and fairness, and encourage collaboration and community among school staff. Grades communicate important information to students about their progress and competence (Nitko, 2005), and they exert influence on motivation, learning, and achievement (Brookhart, 2004; Brophy, 1998; Reeve, 1996). They are also important vehicles for communication with parents and other community stakeholders, so that investing in monitoring for grading bias is an investment in the school and in the broader impacts of education. Grading necessarily involves judgment and is subject to human error (Popham, 2005), but teachers and administrators can take steps to minimize unintentional grading bias.

Conclusions and Recommendations

Grading matters for its value in teaching: conveying messages to students and families about progress toward goals and performance standards, as well as contributing to student motivation, validation of effort and individual competence feedback. Grading matters because it impacts students’ motivation, attitudes and learning. Students’ achievement, reflected in grades, influences their self-beliefs, competence perceptions, goals and success expectations, all of which impact engagement, effort and investment in learning and skill development. Grades also contribute to larger information sets that feed into student grade point averages, achievement
records, program eligibility; and to larger-scale evaluations of teachers and schools. Grades have many uses and purposes, and for all of those uses they need to be accurate and meaningful, not arbitrary or biased.

All teachers and schools can benefit from implementing bias checks, along with bias reduction and prevention strategies. Rural teachers and schools, in particular, can use innovative ways to adapt these strategies to work in small and remote educational sites. Options such as using bias checks as professional development options, utilizing digital networks to reach and facilitate work with peers, sharing assessment tools, and engaging in collaborative and exchange grading practices can support teachers in ensuring and improving grading equity, and provide administrators with evidence of ongoing efforts to maintain quality of practice in assessment and grading, to answer accountability needs.

References


About the Author:

Patricia L. Hardré, Ph.D., is a Professor of Instructional Psychology and Technology at the University of Oklahoma. Her doctorate is in Educational Psychology is from the University of Iowa, specializing in Instructional Design, Motivation, and Measurement & Evaluation. She has worked in design, training, development and program evaluation for schools, businesses and non-profit organizations. Her research and professional interests include: rural education, applied professional expertise & competence development, and designing formal and informal learning environments. Email: hardre@ou.edu.