[GPA] in, [GPA] out: Uncovering Inequity and Flaws in Grading Policies

Luke Green, MA
Instructor
Liberal Arts and Sciences
St. Cloud Technical and Community College
Sartell, MN

Abstract

This commentary explores how the omnipresence of letter grades and grade point average (GPA) as metrics in American education encourage uncritical acceptance of current grading practices despite the inherent inequity and flaws that harm students and institutions. The reduction of the student experience to a GPA launders both meaningful nuance of how course grades are assigned as well as potential evidence of inequitable grade distribution. The lack of consistency in how course grades are calculated erodes metric reliability and validity. Systems over reliant on GPA will continue to overlook the presence of inequity in grading when setting future policy, so developing diverse datasets is advised in an effort to promote equity within our schools.

Keywords

GPA, grading policy, equity, assessment

Education exists to support the proposition that individual growth and learning are possible. Additionally, evidence of intellectual growth and learning are observable and therefore believed to be measurable. These are not tremendously controversial claims; however, controversy can arise when deciding what metric best provides educators evidence of learning and academic attainment. Standardized tests and grading systems are two of the most prominent choices.

The reputation and industry surrounding standardized tests arguably are coming under increased scrutiny following 20 years of being regarded by policymakers as an effective way to create accountability in schools (Strauss, 2020). In places where standardized tests have waned, grades and grade point averages (GPA) have begun to reaffirm the influence grades have had on the American education system for the past 200 years (Brookhart et al., 2016; Durm, 1993).

The symbolic representations of student achievement by way of a letter grade and GPA are relatively easy to understand: An A (4.0) is most desirable. An F (0.0) is least desirable. There are several variations to symbolize student achievement (e.g., E, I, NP, O, P, S, etc.), but they all share the core idea that marking students with a singular symbol (letter or numeric based) is a suitable way to differentiate our students.

GPA is generally considered a heuristic that accurately represents the entirety of the academic experience in a quantifiable way that can be communicated in near-universal fashion within and between schools nationwide. The acceptance of this perspective results in using GPA to advise students and policy. It is through the retelling of this narrative that GPA has become a "proverb of education" (Souja, 2020), allowing it to keep its heralded dogmatic status without much criticism.

The continued use of GPA as a symbolic representation of our students has the potential to cause much harm to our students, and our systems, if the current shared understanding and comprehension of GPA amongst education stakeholders remains unchallenged.

The uncritical acceptance of reducing students to a number misrepresents student achievement due to problems with validity and reliability. The effect hampers the learning environment and exacerbates inequity (Blum, 2020; Brimi, 2011; Delgado & Stefanic, 2017; Farr, 2000; Kohn, 2018; Lipnevich et al., 2020; McMillan, 2001; Reeves, 2004; Solomon & Piggott, 2018). Systemic harm is a potential byproduct of over-confidently using GPA data to inform what will become ineffective policy (Bahr et al., 2019; Beatty et al., 2015; Brimi, 2011; Brookhart et al., 2016; Farr, 2000; Geiser & Santelices, 2007).

Grades Reduce Nuance, GPA Obliterates It

Producing a GPA is a commonly understood process: Individual letter grades are assigned at the completion of a course, translated into a four-point scale, then combined and averaged with other grades that have been received to generate a GPA. Producing this quantitative distillation of a student's academic history in the form of GPA allows the data gleaned to be used in guiding educational decision-making ranging from individual student advising to measuring and shaping federal education policy (Beatty et al., 2015; Brookhart et al., 2016; Ravitch, 2016).

Due to the significant impact these data sets can have on the decision-making process, understanding where these numbers come from, and what the symbols represent, serve as helpful reminders of what is being communicated by a letter grade or GPA.

The practice of measuring students on an A-F (and eventual 4.0-scale) emerged and evolved throughout most of the 19th century to replace the charting of student development via lengthy written narratives (Durm, 1993). The time and labor-intensive narratives were perceived by many to be cumbersome and made it difficult to transfer and compare students across time and institutions (Brookhart et al., 2016).

The lack of standardization within the narratives led to concerns about the potential of subjectivity to tarnish the validity of the metric. The innovative letter grades and GPA seem to solve many of these problems by providing an ordinal metric that could be understood in a seemingly universal way (Brookhart et al., 2016).

Elements of quantification, standardization, and universality of student data make GPA particularly well-liked by many in the post-*No Child Left Behind* era of data-driven decision-making (Strauss, 2020; Ravitch, 2016).

Although the process of calculating a GPA is well understood, there is uncertainty to be found in terms of what the course grades mean and how confidently we can trust what a GPA represents. What a letter grade on the A-F scale is purported to represent versus what it actually represents are influenced by grade level and a variety of classroom policies (which are influenced by many things which include teaching philosophies, content area, and institution policy).

The purpose of assigning grades and what the grades represent potentially shift throughout a student's career (Guskey, 2009). Although their form often deviates from a strict A-F scale, elementary teachers primarily use grades to start a conversation between educators, students, and parents, regardless of what letters are used (e.g., "Your son is doing great with reading, hence the O for outstanding, but we should spend a little more time helping him with math where he has an E for emerging").

Secondary education teachers can use the awarding—or withholding—of good grades as a compliance device to assist classroom management under the guise of preparing students for work or higher study (e.g., "Your content is great, but you will get a bad grade for not following formatting rules"). Post-secondary instructors report viewing grades as a determinate of whether future study in the discipline should continue and to weed out future applicants from selective programs (e.g., "This is the definitive measure of your academic potential").

Student experiences may vary from the findings of Guskey (2009); however, the research highlights the diverse criteria that determine a grade, thus affecting what a grade or GPA represents. Is the grade exclusively representative of content competency (e.g., understanding how to multiply fractions) or is it influenced by items unrelated to the material addressed in the learning outcomes whose influence comes about because of classroom policy (e.g., being a "good" student)? The answers to these types of questions provide tremendously relevant nuance that is rarely acknowledged when making sense of or comparing grades. This presents a big problem for the generalizability of what an individual grade or GPA is based on.

The Effect of the Status Quo

The current status quo of being comfortable with trading off nuance for ease of grade computability affects our ability to understand how students and our systems are performing.

Being grade centric affords the convenience of only having to look at a number. This potentially breeds complacency which prevents policymakers from remaining vigilant of what other stories are being told within our schools in ways not easily visible by looking at GPA. These blinders that prioritize uncertain data and grading policies potentially harm our students, curtail our ability to make sense of curriculum and instruction efficacy, and hamper achievement of institutional missions.

During the initial weeks of the Covid-19 pandemic, many acknowledged that grades received during the spring of 2020 might not be representative of true scholastic achievement but marred by myriad other factors.

The discussion of how pandemic-related disruptions would negatively impact the academic records of students caught in the maelstrom led to acceptance of the need to "hold students harmless" when grading (Castro et al., 2020).

These calls for benevolence reaffirm an unspoken reality: grades can be used to harm students. The timing of these messages imply we are comfortable harming students with grades as long as a global pandemic is not raging. When all students had to weather a life altering disruption, our ironclad grading policies softened, and we found a way to make it work.

Unfortunately, when equally life altering disruptions happen on an individual level, the willingness of our policies to acknowledge individual hardship are often less kind and less equitable.

Uncovering the negative effects that inequity in grading policy and GPA have on our students can frequently be uncovered by walking the halls of our schools.

Simply ask any student (as most have been burned by grading policy at some point) an example of what they feel is unfair or unhelpful about the ways that they are assessed. The willingness to fully listen to their experiences unfortunately does not always materialize, as legitimate grievances are quickly dismissed by administrators and faculty.

The predictable ad hominem retort, "of course you would say that, you are a student," prevents acknowledgment of the lived experiences of our students and dismisses worthwhile data.

The importance that grades will have on a student's future has been made abundantly clear to every pupil, which contributes to why it hurts so much when students experience what they believe to be unjust grading practice.

The introduction of grades into the learning environment introduces an external motivator that takes the pursuit of knowledge and mastery in a given subject and can turn learning into a performative game that rewards and punishes its players (Kohn, 2018). Increased emphasis on letter grades perpetuates motivation to "play the game of school" and encourages students to select a path of least resistance academically, as the reward for positive marks can supersede whether or not one was challenged and learned everything they could during their time in school (Kohn, 2018; Solomon & Piggott, 2018; Warner, 2020).

An additional impact of those approaching education as a game involves the potential to attach self-worth to the grades they receive (despite the, at times, arbitrary nature of what grades truly represent). It creates a meritocracy myth where it is to be believed that GPA is capable of definitively and accurately ranking a student's value. The perceived importance of grades is bolstered by the reification of the metric by our institutions in the form of valedictorian-adjacent awards/praise which foster self-fulfilling prophecies and drive schools farther from providing equity.

The reception of good grades early in one's academic career often opens doors for access into gifted and talented programs and advanced placement courses. Alternatively, those who received poor grades early on are likely to be set onto a track that makes the opportunity to become a high achiever much less likely.

 Stripping nuance from grades also strips awareness and acknowledgment of inequity amongst students. Exclusively attributing good grades and high GPA to academic prowess prevents critical inquiry into what else might be at play.

Whether or not students are harmed by grades often boils down to one's amount of privilege. Students whose families have stable housing, access to food, and present, supportive caregivers are fortunate in their ability to be more likely to focus primarily on school and extracurricular activities during their school experience.

On the other hand, students who need to work to support their families, care for their younger siblings, and lack parental support are likely more apt to struggle with academic due dates, grammar expectations, and completing assignments on a rigid schedule. These salient variables are often not going to be considered or valued when looking at a transcript.

A letter grade, in its current form, cannot begin to explain the performance of students in an equitable and meaningful way. The current system treats work not completed due to an obstinate and apathetic, but otherwise privileged, student the same as a student who would love nothing more in life than to be able to sit down and be selfish enough to take a half-hour for themself after school to better their understanding of their studies and brighten their future.

Arguably a better solution could be found for both students. However, in the current setting take a guess which one of the students (or parent) is going to have the ability to successfully litigate an opportunity for a second chance?

The truly gross nature of GPA is that the privileged students, who are already recipients of increased opportunity, are additionally rewarded by being able to brandish their high marks to interested colleges whereas the less advantaged, are burdened with a millstone of a bad GPA that makes an already challenging life more difficult going forward in a way that is devoid of any alignment with the core elements of what education should provide our students.

[GPA] In, [GPA] Out

Classroom policy is influenced by pedagogy specific to the content area, teaching philosophy of the instructor, educational dogma, systemwide/schoolwide grading policy, and other items (Brookhart et al., 2016; Warner, 2020). An overall course grade is often the result of a complex matrix of formative and summative assignments given

different weight and influence which vary greatly course to course. Final tests may influence 40% of the course grade in one class and 5% in another. Deductions for grammar, timeliness, formatting, and classroom management violations are not consistent either (Brookhart et al., 2016).

Many classroom policies in place do not incorporate sound pedagogy. Existence of these policies is attributable to "teaching folklore" (Warner, 2020, p. 206) in which classroom rules are largely shaped by policies the instructor had when they were students and endure, unquestioned, due to the inertia of tradition rather than sound best practice.

It is encouraging that, as systems begin to address inequity at a systemic level, a variety of safeguards (e.g., accepting late work, retake policies, etc.) have been put in place to minimize wholesale misrepresentation of course grades. Though a step in the right direction, these policies are still rare and often relate only to summative assessment.

Even if inequitable criteria for grades were resolved, the variability between instructors and teaching philosophies can severely hamper the descriptive and predictive value of grades due to issues with interrater reliability. Brimi (2011) looked at how 73 different high school English instructors independently evaluated the same essay. The results yielded assigned grades that spanned all five of the letter grades with a total range of 46 percentage points amongst the grades given.

One student essay is only a piece of a puzzle in what becomes the overall course grade. The lack of agreement between instructors compounds as more pieces are added. This is not proof of faculty being at fault; rather, it lays bare the impact of diverse expectations and approaches in the classroom.

The important takeaway is the potential for variability to exist within a singular assessment, which is folded in with the additional variability of other assessments, processed through the individual course/institution grading policies, and emerges as a course grade. The result being that the same student, progressing through the same course outcomes, taken with different instructors or at different institutions will potentially yield two different grades.

Despite this imprecision, and the inequitable grading criteria, there is little to no hesitation sending grades into a stream, that flows into the river of GPA. Once there, the GPA enters the ocean of institutional transfer where all GPA are assumed equal, and a 0.01 deviation in GPA can make or break a student being admitted to a receiving institution. A system that operates on the flawed premise that GPA from one school equals a GPA from another (Imose & Barber, 2015) is going to be operating on flawed interpretation of the data. The impact of this system creates unequal competition in the education marketplace and misrepresents interinstitutional comparisons as being equal when they are not.

Resolving apples-to-oranges comparisons by way of achieving a universal consensus of what grades should represent and how coursework is assessed for the purpose of a nationwide standardized grading policy is tremendously ambitious and borderline impossible. Before one can try to have any understanding of GPA use across schools, there is work to be done in fully understanding GPA in-house.

Consider these three students:

> Student 1 enters high school struggling academically, necessitating a tremendous amount of effort from their educational support team to end the year with a C (2.0) average, which is viewed as success relative to where the student began. The following year they build upon the foundation and earn a B (3.0) average for the year. During junior and senior year, the student excels in all of the most challenging electives the school has to offer earning an A (4.0) in every class both years.

> Student 2 enters freshman year not particularly interested in the school experience. The student is well-mannered, but not eager to go above and beyond in the classroom. The student does the work that is expected of them and is consistent in earning just above a B average (3.25) each of their four years.

> Student 3 enters high school as a graduate of the middle school gifted and talented program. They coast on their already established academic talent to straight A's (4.0) freshman and sophomore year. Junior year the student continues to not apply themselves and enrolls in easy electives, but the diminishing rate of return of their middle-school-talent drops their average for the year to a B+ (3.5). Their final year is rough, but they can still collect their diploma as their classes needed for graduation have been satisfied despite closing senior year with a D+ (1.5) average.

Arguably, Student 1 is the poster child for the transformative power of what is possible when effective policy, committed educators, and students unite; Student 2 represents those systems in place worked well enough to maintain and cultivate the competencies to graduate with an above average GPA; Student 3 represents several failed opportunities for intervention to take place.

The unifying relationship of these students is that each is going to graduate with a cumulative GPA of 3.25. The complexities of the three different student experiences have been reduced to a singular numeric representation that symbolizes their time in high school. Individual course grades are messy, but the longitudinal nature of GPA has laundered the different trajectories of the students making it difficult to know the true story without parsing over entire transcripts.

From an education leader standpoint any goal measured only by GPA without consideration of the deeper context misses the chance to best understand, and therefore serve, one's schools and one's students. Even if GPA was an accurate measure, when presented as a cell on a spreadsheet understanding what a certain number of students within a certain GPA range means is quite subjective.

How certain can you attribute your graduation rate to the value added of your schools (e.g., Student 1) versus students who otherwise would have succeeded, doing just enough to clear your graduation hurdle (e.g., Student 3) despite your ineffective policies and systems? Some would look at a graduating class that has 40 students graduating with a 4.0+ GPA and point to it as a sign of success, whereas others would look at it frustrated that more challenging opportunities for coursework were not available to these students who experienced a ceiling effect that limited their growth potential.

We are very quick to take a victory lap when simplistic statistics make us look good, but we cannot be lulled into a false sense of confidence. We should be mindful of the limitations that wholesale GPA data provide due to lack of qualitative context.

Ideally, assessments are structured to yield helpful and nuanced data that provides schools insight on when and how to respond in order to advance our institutional missions. GPA does not provide this.

Exploring and acknowledging the inherent shortcomings of grades and the GPA model should be a primary concern for those trying to achieve equitable solutions to student assessment. Being mindful of the shortcomings encourages development of metrics and measures that are more finely tuned to yield nuanced results. This work is not done alone and opens a dialogue amongst administrators, faculty, students, and stakeholders how student development is best measured within individual classes, buildings, and systems.

These efforts have the ability to refocus the educational experience into one that reaffirms the humanity and empathy that are at times lacking in current practice and achieves it in a way that reminds students and educators of the purpose, value, and mission of our schools.

Author Biography

Luke Green is a doctoral candidate in the educational leadership program at Minnesota State University-Mankato. His dissertation explores the impact of using GPA to guide admission decisions on student learning, programs, and institutional missions. E-mail: luke.green@mnsu.edu

References

Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R., & Koch, A. J. (2015). The reliability of college grades. *Educational Measurement: Issues and Practices*, *34*, 31-40. https://doi.org/10.1111/emip.12096

Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research & Evaluation*, *16*(17).

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., . . . Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, *86(*4), 803-848. https://doi.org/10.3102/0034654316672069

Castro, M., Choi, L., Knudson, J., & O'Day, J. (2020). *Grading policy in the time of Covid-19: Considerations and applications for equity* [Brief]. California Collaborative on District Reform. https://cacollaborative.org/sites/default/files/CA_Collaborative_COVID_Grading.pdf

Delgado, R., & Stefanic, J. (2017). *Critical race theory: An introduction*. New York: New York University Press.

Durm, M. W. (1993). An A is not an A is not an A: A history of grading. *The Educational Forum*, *57*.

Guskey, T. R. (2009). *Bound by tradition: Teachers' view of crucial grading and reporting issues*. American Educational Research Association. San Francisco, CA.

Imose, R., & Barber, L. (2015). Using undergraduate grade point average as a selection tool: A synthesis of the literature. *The Psychologist-Manager Journal*, *18*(1), 1-11. https://doi.org/10.1037/mgr0000025

Kohn, A. (2018). *Punished by rewards, twenty-fifth anniversary edition: The trouble with gold stars, incentive plans, A's, praise, and other bribes*. HMH Books.

Ravitch, D. (2016). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.

Solomon, T., & Piggott, A. (2018, June 15). *GPAs don't really show what students learned. Here's why*. The Washington Post. Retrieved from https://www.washingtonpost.com/news/grade-point/wp/2018/06/15/gpas-dont-really-show-what-students-learned-heres-why/

Souja, S. R. (2020). Effects of time metrics on student learning. *AASA Journal of Scholarship and Practice*, *17*(2), 55-66. https://www.aasa.org/uploadedFiles/Publications/JSPSummer2020.FINAL.v2.pdf

Strauss, V. (2020, June 21). *It looks like the beginning of the end of America's obsession with student standardized tests*. The Washington Post. Retrieved from https://www.washingtonpost.com/education/2020/06/21/it-looks-like-beginning-end-americas-obsession-with-student-standardized-tests/

Warner, J. (2020). Wile E. Coyote, the hero of ungrading. In S. D. Blum (Ed.), *Ungrading: Why rating students undermines learning (and what to do instead)* (pp. 204-218). West Virginia University Press.