

TEACHER LEADERSHIP FOR FLEXIBILITY IN GRADING:
AN EXPLORATION AND ANALYSIS OF GRADING PRACTICES IN ONE URBAN
HIGH SCHOOL

By

Ryan Pivonka

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Education

Major: Educational Leadership

Under the Supervision of: Dr. Tamara Williams

Omaha, Nebraska

September, 2020

Supervisory Committee:

Kay Keiser, Ed.D.

C. Elliott Ostler, Ed.D.

Amanda Steiner, Ed.D.

Abstract

TEACHER LEADERSHIP FOR FLEXIBILITY IN GRADING: AN EXPLORATION AND ANALYSIS OF GRADING PRACTICES IN ONE URBAN HIGH SCHOOL

Ryan Pivonka, Ed.D.

University of Nebraska, 2020

Advisor: Dr. Tamara Williams

Grades are perhaps the most widely used and universally accepted representation of a student's academic achievement. More importantly, they can be a vehicle to higher education, specialty programs, and access to highly competitive scholarship dollars. Given the wide disparity of grading practices from teacher to teacher and from district to district, the potential impact it could have on final grades for students, and the ever-increasing importance of grades in post-secondary pursuits, it is essential that this is further studied to discover to what extent the variability of teacher grading practices impact student grade point average (GPA).

The focus of the study was around the question of *what makes a grade a grade?* This study examined the grading practices of one urban high school to determine what impact they had on students' final HSGPA. A survey was given to all teaching staff asking teachers to self-score their current grading practices. The results of the survey created individual teacher scores as well as departmental averages. These scores were analyzed along with trend data of student HSGPAs and ACT scores for the past five years. Outlier students were identified and the full student transcripts were evaluated to determine if the teacher grading practices had an impact on the incongruency of the

student's GPA compared to ACT score. This study will help to determine whether current high school grading practices are solid indicators of students' content mastery, or if they are artificially inflated (or deflated) by other factors.

Results of the study were inconclusive in determining that teacher grading contributed to an inflated GPA, therefore, causing the incongruency. The study did expose a wide variance in grading practices within departments and more so schoolwide. However, the study did point to the need for further research on the purpose of grading and the impact of entire schools or individual departments aligning with best practices.

Acknowledgements

The completion of this program and creation of this dissertation did not happen on its own. As I stop and reflect on the process and sit down to write the final pages, I realize that support and inspiration has come from every facet of my life – family, friends, colleagues, and professors. The journey is finished, and I have many to thank.

I could not have accomplished the completion of this program and dissertation without the unending support of my husband, Jeremy. His unwavering encouragement has been a constant throughout this process – from supporting me in finishing my masters, my taking additional course work in order to be accepted to the doctoral program, and then the doc program from start to finish. He allowed me to disappear into the basement office for days at a time and kept me on track at times when I wanted to hit pause and work on other pursuits or interests. Hence came the expression “we don’t take the stage until we cross the stage.” As with everything in our life together, it was a team effort.

Thank you to my chair, Dr. Tamara Williams who was always a text or zoom away. Thank you for guiding me through this process, helping me to frame my thinking and better articulate my thoughts, for keeping me focused, and not hesitating to tell me when aspects of my submissions were not quite there (or just plain bad). Thank you also to my committee and the entire Ed Leadership department at UNO. Every professor I had was always encouraging and supportive. Forever ringing in my head will be Kay Keiser’s words of wisdom: “when this is all done, we will call all of you Doctor; we will not, however, call you all great writers.” That saying stayed with me in a positive way.

While the dissertation is the culmination of the program, I must thank and acknowledge the cadre of peers in my classes. I looked forward to every class session because of the rich discussions, the support and encouragement given universally and the friendships that were formed. Special thanks to Jaclyn Gibbons, Alysha Collins and McKenzie White. Our text threads were a source of encouragement, reassurance and humor when needed most.

To Dr. Mark Adler - a huge thank you for believing in me. You took a chance on me and gave me my first administrative role and have been my constant champion and mentor ever since. Thank you to my colleagues for covering when I would miss an evening event so that I could attend classes. A very special thank you to my proofreaders who endured endless “special” drafts and sifted through the various incarnations of this final product. Beth McGrath and Kristin Haas, thank you for being kind with your edit suggestions and never making me feel like a bad writer.

Thank you to all of my friends who understood when I “had to write this weekend” and couldn’t join them for their invitations. Now that the dissertation is complete, I may have to come up with a new excuse or start a second dissertation...

Finally, great educators inspire others to go into education. I would like to acknowledge and thank two such teachers who had a profound impact on me. Cathy Pierson and the late Barbara Lebedz. I am certain that I would not have chosen the path that I did had it not been for their influence.

Now that this process is complete, I not only look forward to the new opportunities that this degree presents, but I also look forward to canceling my

Grammarly subscription, reading for pleasure and not worrying about highlighting, notating and citing, and to students calling me Dr. P. instead of Mr. P.

Table of Contents

Acknowledgements	i
List of Figures	vii
List of Tables	viii
Chapter 1: Introduction	1
<i>Conceptual Framework</i>	6
<i>Research Questions</i>	7
<i>Operational Definitions</i>	8
<i>Significance of the Study</i>	9
Chapter 2: Review of Literature	10
<i>Grades' role in Post-Secondary opportunity</i>	10
<i>College Admittance</i>	10
<i>GPA and Class Rank</i>	13
<i>Compound Impact of Student Grades</i>	18
<i>College Success Predictor</i>	19
<i>Inflated Grades and College Response</i>	22
<i>Employers and Other Uses</i>	23
<i>Test-Optional Universities</i>	25
<i>Summary</i>	26
<i>Purpose of High School Grading</i>	26
<i>Brief History of Grading</i>	28
<i>Perception of Grades to Stakeholders</i>	30
<i>Perception of Grades to Administrators</i>	32

<i>Perception of Grades to Students</i>	33
<i>Perception of Grades to Teachers</i>	35
<i>Grade Inflators and Deflators</i>	37
<i>Behaviors: Rewards and Punishments</i>	40
<i>Motivation and “Point Grabbing”</i>	42
<i>Attendance</i>	44
<i>Late Work</i>	45
<i>Extra Credit</i>	46
<i>Utilization of the Mean</i>	48
<i>Using Zeros and the Impact on Final Grades</i>	49
<i>Grading Homework and Formative Grades</i>	50
<i>Teacher Bias</i>	51
<i>Participation Points</i>	53
<i>Effort</i>	55
<i>Lack of Teacher Training</i>	56
<i>Other Factors</i>	57
<i>Summary</i>	60
<i>Alternative Practices</i>	60
<i>Standards-based Grading</i>	61
<i>Moving away from the 100-point scale</i>	62
<i>Multiple Grades</i>	63
<i>No Grades at All</i>	65
<i>Relationship Between Grades and Standardized Test Scores</i>	66

<i>Current (and Best) Practices</i>	68
<i>Conclusion</i>	78
Chapter 3 – Methodology	80
<i>Design</i>	80
<i>Research Questions</i>	80
<i>Participants</i>	81
<i>Instrument</i>	81
Chapter 4 – Analysis and Findings	87
<i>Response Rate</i>	87
<i>Demographics of Study Participants</i>	88
<i>Analysis of Research Questions</i>	88
Chapter 5 – Conclusions and Discussion	109
<i>Review of the Literature</i>	109
<i>Research Design</i>	110
<i>Conclusions</i>	110
<i>Considerations for Further Research</i>	113
<i>COVID-19 and Grading</i>	114
<i>COVID and Equity in Grading</i>	115
<i>Implications for Educational Leadership</i>	117
<i>Summary</i>	118
References	122

List of Figures

Figure 3.1	Teacher Self-evaluation of Current Grading Practices	86
Figure 4.1	Frequency of Students Who Earned a 19 ACT with Corresponding GPA	90
Figure 4.2	Scatterplot of Individual Students' ACT Scores and HSGPA	91
Figure 4.3	Core Subject A Survey Results	96
Figure 4.4	Core Subject B Survey Results	98
Figure 4.5	Core Subject C Survey Results	99
Figure 4.6	Core Subject D Survey Results	100
Figure 4.7	Elective A Survey Results	101
Figure 4.8	Elective B Survey Results	102
Figure 4.9	Single Teacher Response Survey Results	103
Figure 4.10	Scatterplot of Individual Students' ACT Scores and HSGPA	106
Figure 4.11	Identified ACT Score Outlier	107

List of Tables

Table 4.1	Pearson Correlation of ACT Score and HSGPA	89
Table 4.2	ACT and GPA Averages by Year	89
Table 4.3	Average Teacher Grading Score by Years of Experience	92
Table 4.4	Average Teacher Grading Score by Gender	93
Table 4.5	Average Teacher Grading Score by Course Designations	93
Table 4.6	Average Teacher Grading Score by Department	94
Table 4.7	Survey Question Abbreviations	95
Table 4.8	Breakdown of Survey Responses by Department	105
Table 4.9	Average Teacher Grading Score by Student Subgroup	108

Chapter I: Introduction

This study aims to determine the level of impact that high school grading practices have on final grades for students. When examining the components of a final grade, multiple artificial inflators and deflators of grades can be identified that are not necessarily indicators of the level of content mastery that the student has demonstrated. Given the wide disparity of grading practices from teacher to teacher and from district to district, the potential impact it could have on final grades for students, and the ever-increasing importance of grades in post-secondary pursuits, it is essential that this is further studied to discover to what extent the variability of teacher grading practices impact student grade point average (GPA).

Grades are perhaps the most widely used and universally accepted representation of a student's academic achievement. More importantly, they can be a vehicle to higher education, specialty programs, and access to highly competitive scholarship dollars. While parents and community members each have their own perceptions and interpretations of what grades communicate, the research shows that the components contributing to a student's final grade can vary greatly and that grading is subject to factors that can create incongruency between a student's level of content mastery and their final grade. As many colleges and universities move to test-optional admittance policies, it is imperative that student grades and GPA are an accurate and universal representation, and that grading practices be heavily scrutinized until their validity can be guaranteed and provide equity for all students.

As grading is examined, many common teacher practices emerge that do not wholly represent student mastery of content. These practices artificially inflate or deflate

the student's score creating situations where a student might receive a high mark or superior score, but in fact, did not demonstrate mastery of the content on summative assessments. Conversely, it can also create situations where a student fully understands and masters the standards, but because of punitive grading methods, they may receive lower final marks.

Grade inflation is not a new trend. Alfie Kohn (as cited in Smith & Fleischer, 2011) states that it first appeared in the 1890s. That trend continued, which results in the current situation where today, students are increasingly earning more A's in their courses, but not achieving the same increase in their standardized test scores. Recently, studies have begun to focus on determining the teachers' rationale and thinking about grading as opposed to just critiquing their practices (Guskey & Brookhart, 2019).

Grading needs to be an accurate and valid representation of a student's mastery of a particular set of content standards. Grades are reported to and used by many outside agencies and stakeholders; therefore, it is vital they are calibrated and universally interpreted. Unfortunately, that is not the case. Students who receive artificially higher grades than other students with similar ability, content knowledge, and environment may have an advantage in college admissions (Rauschenberg, 2014). When teacher grading practices vary, and their application is uneven or unfair, a systematic advantage (or disadvantage) is provided to one group of students over another (Carifio & Carey, 2013).

Being left unchecked may compound that advantage. Over the past twenty years, there has been more considerable attention and focus placed on applying and analyzing standardized testing. However, the entire practice of grading and grade composition has been left alone (Bowers, 2011). There is no way of knowing the level of rigor and the

validity of a grading system by looking at a student transcript. It is not reasonable to believe that grading is the same in every school district because of the inherently flawed practices across school systems. Each one is unique, and their courses, teachers, and students vary from one school to the next. So do their grading practices. (Westrick, et al., 2015). This is problematic as the consistently lenient grading policies and no system to document and address the frequency of which the grading practices are implemented compromise the validity of student GPA as a barometer of achievement (Marx, 2013).

The research supports that the problem exists. In a study by Cross and Frary (as cited by Marzano, 2011), 39 percent of teachers included effort and behavior in grading practices. In a similar study by Liu (as cited by Guskey and Brookhart, 2019), over 90 percent of the teachers reported using effort in grading, over 60 percent used student ability, over 60 percent used attendance and participation, and over 40 percent used classroom behavior.

Because of the many non-academic components of grades, researchers typically declare teacher-assigned grades as not of great use. This is because of the belief that grades are an inaccurate representation of academic achievement (Bowers, 2011). Grades established entirely by the teacher have become viewed as subjective and inconsistent. Because of this, they are typically not used in decision making and strategic planning by school officials and administrators (Bowers, 2011).

Despite the acknowledgment of its practice and the studies done to identify its use, grade inflation is experiencing a surge. This is causing colleges and universities to adapt and offer more remediation courses as incoming high school students start college with lower mastery despite their high GPA (Wormeli, 2018). One study found that the

proportion of students who reported an A average on their high school transcripts has grown more than 30% in the last 40 years (The Higher Education Research Institute, as cited by Cook, 2004). The same study stated that half of the students reported graduating with a GPA at or above 4.0.

This is a problem. Colleges are making adjustments for their students to combat their academic deficiencies, even though on their applications, they appear ready for college-level coursework. They admit students based on potentially flawed information, and that information plays a crucial role in their decision making. According to the National Association for College Admission Counseling (2016), grades are the most important factor that colleges and universities use in admitting students. Colleges are under the impression that grades are a representation of students' readiness for the rigor of college coursework and curriculum and grant admission partially based on those grades (Feldman, 2019).

Students have become savvy at manipulating the school system to achieve the highest GPA possible. They take the courses needed to get the extra point boost in their GPA. They find the teachers who have a reputation to "grade easy," or that accept tons of extra credit.

By focusing strictly on acquiring a high GPA for college admission, some students have become very talented at working the system, but consequently, less prepared academically (Vatterott, 2015). Grades are misleading as predictors of a student's ability to succeed in college because a high GPA might help a student be accepted to a college. However, it does not necessarily guarantee that they are prepared

academically to be successful in college. Because of this trend, the United States has one of the highest rates of college dropouts in the world (Canady, 2017).

One of the primary factors that affect the validity of high school grades is the use of grading practices that incorporate artificial inflators or deflators. School systems have used these commonplace systems for decades. They are methods that academically penalize students for non-academic infractions, or reward students academically for doing non-academic things. Examples of these practices include penalizing students for misbehaviors by deducting points, deducting points for absences and tardies, lowering a grade for an assignment being turned in late, offering extra credit, utilization of the mean, use of a zero for missing work, assigning points to formative work, awarding points for merely participating, and subjectively scoring a student's effort. Teacher bias and a lack of quality teacher training with ample emphasis on assessment and grading can also influence grades.

The application of these practices makes interpreting a grade's validity difficult. When outside factors that are not academic are included in final grades, the level of student mastery is potentially miscommunicated. A student who regularly scored C's on assessments throughout the course could be given A or B grades because of their adherence and compliance to the teacher's grading policies. For example, a daily participation grade can add a litany of entries in the grade book and provide an un-earned boost to the student's final score. This can also happen when extra credit is readily available for students. The result can potentially raise a student's final grade a full letter value.

Conversely, grade deflators can negatively skew a student's grade so that their final score is not indicative of their mastery level. They could have superior scores on all of their summative assessments. However, because of penalties imparted throughout the term, they may finish with a grade well below their level of content understanding. This could be caused by a teacher taking points off for talking in class each day. The previously mentioned daily participation grade may cause them to lose a large number of points for something that is not at all academic resulting in a final grade that is lower than their level of mastery. A student who regularly turns assignments in late may be deducted points for being late, even though the paper or project is of mastery-level. Grade deflators and practices that penalize students academically negatively skew grading.

Because conditions exist creating inconsistencies between grading practices of different teachers and different districts, and evidence suggests that the inclusion of formative exercises in final grades skews the calculation of that grade, and since grade inflators and deflators are utilized, it is important to study further the impact these practices have on students' final grades.

Conceptual Framework

This study examines the relationship between teachers' adherence to best practices in grading and student final GPA. A list of best practices in grading based on Ken O'Connor's work *Fifteen Fixes for Broken Grades* (and supported by other experts) will generate a scorecard. Teachers in a mid-sized, urban high school used this scorecard to self-assess their grading practices for each class taught during a predetermined period. The scores were compared between departments and by years of experience to analyze patterns in grading practices. The scorecard was formatted so that if a teacher utilizes a

particular practice, they get a point. The best practices suggest avoiding the use of the practices listed. Therefore, the lower the teacher score, the more aligned to best practices they would be.

Additionally, composite ACT scores were compared to students' final GPA over a five-year span to determine the relationship between the two scores and whether there is a strong correlation. The researcher examined a sample set of subgroups from the 2019-2020 senior class student roster. That is, students who have a high GPA but a low ACT score, students who have a low GPA and high test scores, and students who are closely correlated with high GPA and high test scores. Those students' course histories were further analyzed to determine the average teacher self-score and whether the grading practices impacted the GPA. For example, if a student had a high GPA, but a low ACT score, what was the teacher average for the teachers that taught that student?

Research Questions

The central research question for this study is what makes a grade a grade? The study will answer the following sub-questions:

1. What is the relationship of cumulative high school GPA at the time of graduation and that student's highest ACT score from 2015 to 2020?
2. What are the teachers' current grading practices? Does this vary by subject, course designation as AP/honors, or years of experience of the teacher?
3. What is the relationship between GPA/ACT alignment with teacher grading practices?

Operational Definitions

Teachers' grading practice: the methods a teacher uses to determine reporting marks for students. For the purposes of this study, these will be determined in how teachers respond to survey questions regarding their use of certain grading practices that inflate or deflate student grades. (Inflators - non-summative grades or points added that potentially add to a student's grade that does not reflect content mastery. Deflators - Practices that take points away from students for things that are not related to the curriculum, standards, or content of the class).

Highest ACT score: The ACT is an entrance exam used by most colleges and universities to make admissions decisions. It is a multiple-choice, pencil-and-paper test administered by ACT, Inc. The purpose of the ACT is to measure a high school student's readiness for college and provide colleges with one common data point that can be used to compare all applicants (Princeton review, 2020). For the purposes of this study, the students' highest ACT score will be used (for students who took the test multiple times).

GPA: A grade point average (GPA) is the sum of all course grades throughout a student's high school career divided by the total number of credits. Most high schools (and colleges) report grades on a 4.0 scale. The top grade, an A, equals a 4.0, a B equals 3.0, a C is 2.0, and a D is 1.0. In a weighted GPA system, such as the one used in this study, certain courses receive an additional point based on their designation as AP (advanced placement) or Honors. These courses carry a higher level of rigor than other courses of the same content.

Course designation: The determination by school officials of whether a course is a traditional course or labeled as one of higher rigor because of its content. Examples of these courses would be honors courses or AP (advanced placement).

Years of experience: the number of years of teaching experience the teacher had (in their current content area) when completing the survey, regardless of school district.

Significance/Purpose of the Study

This study is important because it examines the relationship between teacher grading practices and student GPA. By identifying the level of impact that the utilization of best practices in grading has on grades, a more persuasive argument can be made to implement a universal grading practice. When schools are consistently scrutinized and undergoing massive reform efforts, and students are doing whatever it takes to get into their top choices for college, it is increasingly important that students get top marks. However, what do those marks indicate? This study will help to determine whether current high school grading practices are solid indicators of students' content mastery, or if they are artificially inflated (or deflated) by other factors.

The idea to study grading practices came as the administrative team continually encountered students who earned high GPAs but struggled to earn an above-average ACT score. The team would continually see students who took top classes and earned A's in their honors and AP courses, but, because of their low ACT score, they would have a difficult time getting into colleges without having to take remedial college classes as a freshman, or even being accepted altogether. More and more students would earn above a 4.0 cumulative GPA and only get an 18 or 19 on the ACT. Something was wrong. The team began to examine factors such as test anxiety or a low frequency of test attempts.

However, it always came back to asking whether the curriculum was rigorous enough, or if students who played school really well could navigate the inflated grading system to earn high marks, but then eventually get caught when it came time for standardized tests.

Chapter 2: Research Literature Review

The purpose of the study is to examine what makes a grade a grade, and to what extent the variability of teacher grading practices impact student GPA. What factors contribute to a grade, and to what level is that grade indicative of the level of a student's academic achievement? This literature review will focus on six main components: grades' role in postsecondary opportunity, the purpose and history of high school grading, grade inflators and deflators, alternative grading practices, the relationship between grades and standardized tests, and best grading practices.

Grades' Role in Post-Secondary Opportunity

College Admittance

Perhaps the most prevalent use of high school grades is their use as a qualifier for admittance to colleges and universities. Brookhart (2015) calls them "the fundamental currency of our educational system; they signal academic achievement and non-cognitive skills to parents, employers, postsecondary gatekeepers" (p. 269). Schneider and Hutt (2014) stress their importance and long-lasting impact. They state that grades have real and lasting consequences and implications and play a role in determining a student's future, as schools, universities, and employers use them to measure the student's academic achievement.

According to the National Association for College Admission Counseling (2016), grades are the most important factor that colleges and universities use in admitting students. With students competing for coveted scholarship dollars and places in top collegiate programs, grades must be accurate representations of students' abilities and proficiency in content standards. Grades provide access to specialty programs and the award of scholarships (Harland, 2014). However, as Reeves et al., (2017) points out, "with consequences for scholarship opportunities and university admissions, grading remains the wild west of school improvement...anyone with a red pen can make decisions with devastating instructional consequences" (p. 42).

While the term "wild west" can be extreme, many researchers agree that there are a wide variance and inconsistency among schools' grading practices. Grades are calculated in so many ways and represent student achievement so inconsistently, many decisions based on grades may be wrong, potentially resulting in consequences for students. Colleges believe that students' grades reflect preparation for postsecondary work and offer admission based in part on those grades (Feldman, 2019), and a chance for the student to garner funds for college (Stanley, 2010). Grading practices do vary widely across the country, and researchers are calling for uniform grading methods so that all parties can easily interpret them to make informed predictions about the students' academic performance (Azeem, 2010).

With the wide variance in grading practices, how do colleges and universities trust the validity of the grades submitted, and use them to determine which student(s) they admit? As Rauschenberg (2014) claims, "high school grades play an important role

in college admissions. Many colleges make admissions decisions based almost exclusively on a student's GPA and SAT or ACT score" (p. 3).

However, with such a wide disparity in grading practices, likely, a student who receives an A from one school does not have the same level of mastery as a student who earned an A from another school, or even a neighboring school. Differences in the rigor of curriculum, grading scales, and teacher grading practices could cause two students to present the same grade when their level of achievement and actual content mastery is vastly different. Gonzales (2015) agrees that this is problematic, stating that inflation and interpretation are inherently what is wrong with grades. When colleges are looking at two students with similar applications, they have no way of knowing if one applicant's grades are inflated when they decide which student to accept into their program.

While increased attention has been paid to preparing for, proctoring, and analyzing results from standardized testing since the rise of the accountability movement over the past 20 years, the vast industry of assessing and assigning course grades has continued unabated (Bowers, 2011).

Research supports that high school grades are an accurate indicator of how students will perform in college, particularly in their first year. High school grades are a more accurate representation than scores on college admission tests when it comes to predicting first-year college GPA (Sawyer, 2013). While there are ongoing debates on the selection criteria for college admissions (Westrick et al. 2015), "findings reinforce the centrality of measures of standardized achievement and high school performance for understanding the readiness to complete college entry-level general education courses" (p. 43).

When selecting which candidates to admit, colleges need established criteria, and high school grades are a logical part of that process. Colleges and universities are looking to enroll students who are the most likely to succeed. They use their criteria when making course placement decisions and making highly competitive scholarship decisions. One primary reason for the utilization of these criteria is objectivity, that is including a component that can be interpreted the same way for all applicants (Sawyer, 2013). It is then worth asking if it is being interpreted the same by all? Each student goes through high school differently, and each family has their own memory and expectations of the school experience. Allen and Mattern (2019) call the high school experience "an amalgamation of the type and sequence of courses taken, intensity and difficulty of the courses, alignment to postsecondary expectations, quality of instruction, and level of student engagement and effort" (p. 6). That is a lot of variables.

GPA and Class Rank

The practice of compiling student grades into a holistic cumulative grade point average to convey the breadth of their high school experience is not new. High schools also use those GPAs to rank students in order of achievement. However, many colleges and universities have gone away from factoring class rank into their considerations. This is for a variety of reasons. For example, if a student attends a school where the majority of the students are high achieving, they may receive exemplary marks. However, their class rank could still be low because of the competition level at their particular school. Their cumulative GPA of 4.1, while certainly well above average, may give them a class rank far below a student from another school who had drastically less competition from their peers and earned a lower GPA and had a higher class rank. That student with the 4.1

could have a class rank below the top twenty percent of the class, preventing them from being eligible for specific scholarship opportunities with that requirement. Students who attend a smaller school have this mathematical problem compounded. A rural school district with a graduating class of twenty students might easily have eight students earn a 4.0 or above. Because of the small number of students in the class, the student with a class rank of eighth would be in the 60th percentile of their class. If colleges do not have protocols in place to screen for those instances, students could be rejected for a low percentile despite exemplary grades.

Students who receive artificially higher grades than other students with similar ability, content knowledge, and environment may have an advantage in college admissions (Rauschenberg, 2014). The side-by-side comparison of two GPAs or class ranks from students from different schools can be misleading and difficult to interpret without dissecting the class rank and looking at individual course grades. While there are a variety of factors that colleges and universities consider when selecting applicants, the process takes time and needs to be streamlined in order to complete it. Large universities may be dealing with 50,000 applications and cannot dig deep on each one to discover whether there is grade inflation. They are left to screen the application based on the GPA (Marcus, 2017).

In Wormeli's (2018) research, more colleges have stopped asking for class rank, stating they are finding that the data point is not a reliable predictor of how a student will perform in college. The elimination or omission of class rank from the application process places additional weight and importance on the student's grades, thus amplifying the importance of their accuracy. The importance of that accuracy is amplified since by

earning high grades, students get admitted to their choice colleges and universities and receive scholarships and tuition assistance since grades are a significant selection criterion in the college admission process (Allen, 2005).

There have been debates over two grading methods – norm-referenced and criterion-referenced. Norm-referenced grading compares students to each other, whereas criterion-referenced grading compares the student's work and achievement to a predetermined set of standards. The need for schools to rank students for college admissions causes high schools to utilize norm-referencing (Guskey & Brookhart, 2019). It is fair to ask whether the purpose of grading is to provide rich feedback to students in order for them to thrive and make informed decisions about their academic future, or if grades are merely a method to rank students against each other for future opportunities (Carifio & Carey, 2013). The ranking system that has been in place served a purpose that may be obsolete.

Vulperhorst (2017) argues that a GPA may not be a truly accurate depiction of a student's achievement because the classes the students take are not the same. Cumulative scores are being compared, but the factors used to create those scores are not necessarily equal. "Even though high school GPA may summarize all grades neatly in one variable that reflects the overall achievement score of a student, and can be seen as a variable with little measurement error, this variable is actually built upon different standards of achievement" (Vulperhorst, p. 400). For example, two students could both have a cumulative GPA of 3.85. However, the rigor of the courses that make up that GPA could vary drastically. One student could have taken the bare minimum core academic classes and a slew of easy elective credits. In contrast, the other student could have taken the

most challenging and advanced courses the school offered. While the two students earned the same final GPA, their scores are telling two different stories.

Colleges typically will ask for a weighted and an unweighted GPA. Since a weighted GPA can often have classes on a five-point scale, it can better reflect the rigor of the courses taken. However, a standard GPA can present a clearer picture of the student's academic record. Regardless of which GPA is submitted, colleges will ultimately take a deeper dig into the student's transcript to look at individual courses taken and the grades earned in those courses, therefore reinforcing the importance of accurate grading.

Because of this, colleges do look at more than just the GPA when considering their applicants. Along with courses completed and test scores, class rank and grade point average are among the top four determinants of whether a student will be accepted for admission. Nevertheless, despite the importance of class rank to this process, there is no agreed-upon procedure that high schools use (Lang, 2007). Lang's research found that 80% of schools surveyed apply some type of additional weight to advanced placement or honors courses to determine class rank. His main finding is that most districts use flawed procedures, suggesting that class ranking's robustness is questionable. The same study found that schools reported as many as thirty students tying for first place and all being named Valedictorian. Since there is no universal practice of how GPAs are derived, most colleges and universities ask that an unweighted GPA be reported separately. This helps create equal opportunity among applicants since not all schools provide the same level of advanced rank courses or award an extra point when calculating GPA.

When not only the grades but the class rank becomes increasingly important, students are impacted and sometimes forced to alter their course selection to game the system. Students and parents have become savvy about how the GPA calculation works in their schools, and this can often lead to students choosing not to take courses that do not carry additional weight for their GPA. A student may have a passion for instrumental music but may stop taking any performance classes because earning an A in the orchestra does not compute the same way that earning an A in an honors or advanced rank class does. Conversely, class rank pressure may influence a student to take a less rigorous course load for which they can be more confident they will achieve top marks, rather than pushing themselves to take the more difficult and rigorous classes where they may earn a lesser grade. Students are like most other people. They try to maximize their benefits at a minimum cost or effort (Lang, 2007).

Another factor that comes into play is the location and socioeconomic status of the school or school district. New research shows that students who are enrolled in private and suburban public high schools are being awarded higher grades than their urban public school counterparts with no less talent or potential. The research, conducted by the College Board, found that grade inflation is accelerating in the schools attended by higher-income Americans, who are also much more likely to be white (Marcus, 2017). The research concluded that the GPA of private school students taking the SAT rose between 1998 and 2016 from 3.25 to 3.51 (approximately 8%). In suburban public high schools, the change was from 3.25 to 3.36. However, in city schools, the scores barely moved, going from 3.26 to 3.28.

Some states have adopted policies where any student who completes high school in a specific top percentage range of their graduating class are admitted automatically to the state university. California, Texas, and Florida, three of the country's largest four states, have adopted this practice. This adds to the pressure for students to be top achievers and places additional importance on the validity and integrity of grading practices.

Compound Impact of Student Grades

The path for college-bound students can start relatively early, and grading practices can have a domino effect on a student. If a student does not get an A in sixth-grade math, then they may not be on track to be in the highest level of math in high school, which means they potentially will not get into their top college choice (Lukianoff & Haidt 2018). This can also happen in schools that require a specific grade as a prerequisite for an upper-level class. A student may be ineligible to take advanced math because they earned a D in the introductory course. However, if that D was the result of inadequate grading practices and not a reflection of the student's ability, they could potentially miss the opportunity in their own school. This narrow focus can also cause students to miss out on actual learning. By focusing strictly on acquiring a high grade point average for college admission, some students have become very talented at working the system, but consequently, less prepared academically (Vatterott, 2015).

Teachers' judgments of students' achievement may play a role in everyday classroom interactions. Through school grades, recommendations, and certificates, teachers' judgments can impact children's future educational paths and, more broadly, educational justice (Westphal et al. 2016). Another result can be grade disparity. When

different grading practices are applied unevenly or unfairly, they provide a systematic advantage (or disadvantage) to one group of students over another (Carifio & Carey, 2013). Some grading practices' subjective nature puts students in jeopardy of being given grades impacted by teacher perception and bias. This bias can compound, as a teacher may label a student as low-achieving based on previous grades or efforts, and that student may never get out of that image to the teacher.

Nitko (as cited by Allen, 2005), states that grades are used by students, parents, other teachers, guidance counselors, school officials, postsecondary educational institutions, and employers. Therefore, teachers must assign grades with the utmost care and maintain their validity (Allen, 2005). In addition to their postsecondary institutions' use, grades can impact students while they are in high school. Students must maintain certain GPAs in order to remain eligible to participate in sports and activities. They are afforded privileges such as reduced rates on car insurance when they achieve a certain GPA. Students can also face consequences at home for grades that their parents deem unacceptable. Grades can have an impact long after they are finalized.

College Success Predictor

A study done by ACT found that high school GPA was not a valid predictor of success at higher levels (ACT website). For example, even a 4.0 HS GPA corresponded to very low probabilities of success at the 3.25, 3.5, and 3.75 levels of first-year GPA. However, some would disagree. Bowers' (2011) research found that high school grades have shown to be *strong* predictors of college entrance exams and first-year college performance. These associations have remained consistently strong despite the reputation of grades as imperfect and subjective measures of academic knowledge.

When studying whether high school GPAs or ACT scores were more valid indicators of a student's likelihood to succeed in a college, Allensworth & Clark (2020) found that GPAs had a stronger correlation.

HSGPAs perform in a strong and consistent way across high schools as measures of college readiness, whereas ACT scores do not. HSGPAs are not equivalent measures of readiness across high schools, but they are strongly predictive in all schools, and the signal they provide is larger than the differences across schools. As measures of individual students' academic readiness, ACT scores show weak relationships and even negative relationships at the higher achievement levels (Allensworth & Clark, 2020, p. 209).

This could be the result of several factors. Some of the soft skills that are part of being a good student and earning high grades transfer well to a college setting. Students are more likely to have study skills, strong attendance, and to be engaged in the learning process. These traits would segue naturally to a college setting. The lack of those skills could explain the lower correlation of test scores to college readiness. Standardized tests do not measure soft skills.

In a 2013 study by Pattison, Grodsky, and Muller (as cited by Brookhart & Guskey, 2019), they found that since 1982 GPA has risen for high school students but dipped at four years colleges. The same study found that GPA and standardized test scores correlated at about .5. They cited a similar study done by Woodruff and Ziomek (2004), who stated that high school grades from 1991 to 2003 inflated by around .23 on a four-point grade scale. They also found a moderate correlation between standardized tests and GPAs.

Another study found that high school GPA and ACT's correlation was inconsistent and pointed to the variance in grading practices as a possible contributor. They found that the validities of ACT scores and high school GPA vary across institutions, stating that this finding should not be a surprise.

It is unreasonable to argue that they should be the same across institutions or to argue that the predictor variables are flawed because the validity coefficients vary across institutions. Every institution is unique. The departments, courses, instructors, and students differ from one institution to the next, as do their grading standards. (Westrick, et al., 2015, p. 39)

With the varying results on the validity and accuracy of test scores and grades as predictors of college success, it only reinforces the need for the grades to be more accurate.

If grades have consistently risen from year to year, it would make sense that the SAT has risen proportionately. This is not the case. SAT scores have gone down since the mid-2000s, primarily in writing and critical reading (Twenge, 2017). More students wanted to go to college sometimes attributed to previous dips in SAT and ACT scores; when more students take the test, the overall score tends to go down. This was not the case for the recent dips. Similar results occurred with the ACT. According to the National Center for Education Statistics (as cited by Blankstein and Noguera, 2015), results from the ACT showed that only 39% of the students who took the exam were deemed college-ready. The percentage decreased substantially when disaggregated into subgroups. Only 11% of African Americans and 18% of Latinos were deemed college-ready. With

standardized test scores going down, and grade point averages going up, colleges and universities face a challenge.

Inflated Grades and College Response

Inaccurate or inflated grades contribute to students getting accepted into programs that they may not be ready for. Universities and colleges have had to incorporate remediation programs to catch students up who, on paper, are ready for their program of study based on their high school transcripts. Grades are misleading in predicting success because high grades may help students get into college, but they do not necessarily prepare them academically to succeed in college (Vatterott 2015). This leads to the United States having one of the highest college dropout rates in the industrialized world (Canady, 2017).

Colleges continue to see a rise in overall GPA from their applicants from year to year. However, the same increase is not documented in the applicants' overall standardized test scores or success once they begin college studies. American School Board reports that even though SAT scores remain unchanged, college applicants are receiving more A's than ever as grade inflation reaches new heights in the nation's high schools. Harvard professors have complained that "in the present practice grades A and B are sometimes given too readily - grade A for work of not very high merit, and grade B for work not far above mediocrity" (Goodwin, p. 80). This is not a new discovery. Alfie Kohn, (as Cited by Smith and Fleischer, 2011), references that grade inflation first appeared at Harvard as early as 1894. If students are earning more A's at the high school level, then it would be logical to have the same increase in standardized test scores; however, that is not the case.

Wormeli (2018) fears that grade inflation is on an upward trend. Colleges and universities are forced to offer increased remediation classes because high school students are going to college with less and less mastered for their high grades. Teachers have to hold students accountable for the material so that those grades mean something. Some safeguards are in place to ensure that a high school classes' curriculum adequately prepares a student for a college course. For example, courses taken for college credit while at the high school are subject to adhering to college and university guidelines and must follow their syllabus and curriculum. This helps ensure that the grade the student received in that course is an accurate representation and can be used to place them in an appropriate college course. Another example is students who complete the AP exam for AP courses. While they receive a grade from the classroom teacher, they can also present a score from the AP test that is standardized and not subject to factors the way the classroom grade could be.

Students have some discretion over the courses they take in high school, and these courses can vary in terms of the level of rigor or difficulty. Colleges, universities, and researchers desire a means to quantify the rigor of a student's course load to understand better their level of academic preparation and readiness for postsecondary pursuits (Allen & Mattern, 2019).

Employers and Other uses

Students who attend college are not the only ones affected by their grades and grade point averages after high school. Grades inform decisions outside the educational world as well. Potential employers consider grades when hiring, and GPAs are often required for youth work permits and reductions in car insurance (Feldman, 2019).

School grading practices have long been the subject of heated discussion; yet, the debates concerning student assessment and teacher grading methods have grown more intense over the past decade. This increased attention comes as teachers, administrators, and parents realize that traditional grading schemes, in place and largely unchanged for over 100 years, are proving insufficient in meeting the demands of the 21st century. (Carifio & Carey, p. 19)

Conversely, a student can also suffer if their GPA is not indicative of their abilities. Invalid grades that understate the student's knowledge may prevent a student with the ability to pursue specific educational or career opportunities (Allen, 2005). An error is made when academically underprepared students are directed toward post-graduation educational options, careers, or jobs that exceed their ability or level of intellectual capacity because of a relatively high GPA (Steward et al. 2008).

Some students may not see any correlation between their grades and the workforce. While poor grades in high school may rule out the college option for students, it can potentially make the employment alternative more attractive. Students may know of others who are successfully employed despite doing poorly in school and use this to convince themselves that academic performance is not particularly important if they are not going to pursue their education. While many high school students value pursuing a college degree, those from lower socioeconomic classes may not believe that high school grades will lead to college. Parental education may also influence a student's beliefs and attitudes toward grades (Linnehan, 2001).

Test-Optional Universities

For various reasons, some colleges and universities have begun to alter their admissions processes and are no longer requiring students to submit scores from the ACT or SAT. With those universally understood and calibrated measures of assessment no longer being required, greater importance is placed on a student's high school transcript and grades, which increases the necessity that those grades be accurate and their interpretation standardized. The reasons for dropping the college entrance exam requirement are varied. One rationale is that the test highlights inequities between students. Lower-income students are at a disadvantage because of their lack of access to test preparation programs or even the means to pay for the test itself. Students of color are often at a disadvantage because many test questions contain inherent cultural bias (Koran, 2020).

In a study done by Hiss and Franks (2015), they found a statistically insignificant difference between college students who applied submitting standardized test scores and those who did not. The results were clear: "non-submitters performed just as well in college as submitters. Specifically, non-submitters earned cumulative GPAs that were only .05 points lower than submitters; the difference in their graduation rates was .6 percent. By any standard, these are trivial differences" (Hiss & Franks, p. 33).

Another reason the tests are no longer being required is because of the impact of COVID 19, which caused cancellations of the tests and critical disruption of learning due to school closures. With fewer opportunities for students to take the test, colleges and universities have adjusted their requirements during the pandemic and even for several years afterward (Vidgar, 2020). The University of California has waived the test for the

next four years and plans to consider phasing it out altogether. While they may potentially replace it with their own admissions test, the importance of grades and student scores now rises.

Summary

With increased importance and emphasis placed on high school grades and those grades being interpreted and used by many different entities, it raises concerns about their validity.

The real policy question is how to increase the rigor of the K-12 curriculum to increase college readiness and to understand better the motivational and academic behavior components of high school GPA to ensure that all students entering a highly demanding, unstructured postsecondary work environment are able to succeed. (Westrick et al. p. 43)

Whether students are leaving high school to pursue postsecondary education or enter the workforce, there is a need for a universally understood grading system where an A means the same level of achievement in all parts of the country and terms like class rank and grade point average are computed systematically. When various factors come into play and have to be analyzed and extrapolated, leaving the grade report subject to interpretation is problematic.

Purpose of High School Grading

While there is considerable debate about what factors should contribute to the creation of a grade, an equal variance as to what *purpose* grades should serve exists. At their base, grades refer to "the symbols assigned to individual pieces of student work or composite measures of student performance on student report cards" (Brookhart et al., p.

804). However, what to do with these symbols, and how to interpret them by various stakeholder groups is another matter. Students view grades differently than their parents, who view them through a different lens than school officials. Bailey and McTighe (as cited by Allen, 2005) say that "the primary purpose of secondary level grades and reports is to communicate student achievement so that informed decisions can be made about the student's future" (p. 219).

If grades directly impact a student's future, teachers must take great care to ascribe and report them. A grade should represent a clear and accurate indicator of what a student knows and can do. Grades document the progress of students and teaching, providing feedback to students and parents or guardians. They are also used to make instructional decisions regarding the students' learning (Wormeli, 2018).

Wormeli (2018) developed what he called essential tenets for grading.

- (1.) Teachers must be ethical. They cannot knowingly falsify a score or grade.
- (2.) Grades must be based on accurate evidence of students' performance related to standards; otherwise, they should not be used to make instructional decisions or document student progress
- (3.) Report cards should communicate performance related only to publicly declared standards/outcomes, and
- (4.) Any test format that does not provide accurate evidence of students' performance related to standards must be changed or replaced. (p. 12)

Grades can have a profound impact on students, whether it be leading to decisions made about entry into a college program, or placement in a remedial course. Students' grades must be accurate representations and universally understood by students, parents, and all stakeholders. "Grades must provide a fair, undistorted, and unbiased reporting of

student academic progress and achievement. When schools assign grades that are inflated, skewed, or simply based on uneven standards applied in an uneven manner, they cannot possibly fulfill these critical social responsibilities" (Carifio & Carey, p. 21).

Brief History

The methods and purpose of grading have changed significantly since its inception. During the 19th century, student progress reports were presented to parents orally by the teacher during a visit to the student's homes, with little standardization of content (Brookhart et al., 2016). Initially, the public charged schools with preparing students to meet the industrial and commercial world's needs. They could do so most efficiently if they matched each student with the appropriate curriculum based on their ability (Feldman, 2019).

Early American grading systems were primarily based on the European model. They were focused heavily on competition and the awarding of rank order and used mainly for pedagogical purposes. As the requirement for schooling expanded, the role of grades changed. The initial design for schools was a communication tool between the teacher and the family; they were now also being used to communicate externally (Schneider & Hutt, 2014).

Early in the 20th century, compulsory attendance laws changed the practice of K-12 education in the United States and schooling initially evolved for the purpose of sorting and ranking (Vatterott, 2015). This sorting of students worked and was beneficial at a time when the workforce required relatively low skill. The community tasked schools with deciding which students entered the labor force and which students went on to higher education. This system was effective and accepted when jobs were available for

workers without much formal education. The economy was built on low-skilled laborers, and the school system played an important role in sorting students for their next steps, whether it be upper-level education or as laborers (Canady, 2017). Additionally, Stiggins (2017) talks about what he called the "institutional mission of sorting the most able learners from the least able ones," where teachers graded on a curve designed to give a finite number of students an A, and a select few more a B. By following the curve, the majority of the students were labeled as average, and a group was labeled as low achieving, helping schools to complete their task of sorting students.

Eventually, grade reporting practices developed into what it looks like today. According to Feldman (2019), "no longer could educators use the clumsy "unscientific" narrative of reporting -- it was time-consuming and too unstandardized. Instead, there was pressure to identify a standardized system of communicating student achievement" (p. 23). This led to the commonly used 100-point system that yields a final letter or number grade. High schools favor this system because it was less time-consuming than descriptive feedback to parents. However, it streamlined student abilities into a percentage grade (Brookhart et al., 2016).

Grading has since evolved beyond the sorting and ranking of students and moved toward the idea that all students can and should master the necessary standards to be successful. The Every Student Succeeds Act (Pub.L. 114-95, 2015) tasked schools with ensuring that every student is ready for college or the workforce, not just those who were ranked accordingly (Stiggins, 2017).

Grades communicate a perceived level of academic achievement. They are supposed to inform parents how well their child did in a particular class. The parent

applies interpretation of what that grade means to them, given their context of schooling. Grades have moved beyond just being used in education. Entire rating and grade systems seek to rank products and services. Consumers are apt to read reviews and see how many stars a product earned and what the Yelp reviews say before buying. When ordering from Amazon, buyers typically read the comments before purchasing. How is a grade supposed to communicate the entirety of a student's content knowledge through one letter? When there are no comments or reviews to read, the parent is left to interpret the meaning of a grade.

In her podcast, *A Cult of Pedagogy*, Gonzales (2015) interviewed Starr Sackstein, who shared that notion. She lamented being limited to be able to enter only a handful of pre-coded comments. "I don't know if a report card is supposed to communicate learning to students and families, it seemed off that I didn't get more space to actually do that." Stiggins (2017) takes that idea further, stating how the student's grades are then all combined into a GPA, leaving out a great deal of information about the level of learning that took place and replacing it with a single number.

The perception of grades to stakeholders

Do parents trust that the teacher grade is an accurate representation of their child's academic achievement? To what level is parental buy-in to the process? In a study done by Marzano (2001), he found that when a representative sample of parents was polled about their opinions of teachers and administrators, one of the primary findings was that "educators need to rebuild public trust in a few important areas" (p. 2). One of those important areas was the ability to make judgments about individual students. A study

done by Baron (2000) (as cited by Allen, 2005) showed a lack of coherence in the beliefs about grades held by parents, students, and those held by the education community.

While stakeholders prefer that grades tell a universal set of information about a student, research says otherwise. Although measurement experts and professional developers may wish grades were unadulterated measures of what students have learned and can do, substantial evidence indicates that they are not (Brookhart, 2015).

Due to the non-cognitive behavioral aspects of grades, psychometricians and assessment researchers have historically maligned teacher-assigned grades as non-useful. This is due to the perception that grades are unreliable measures of academic knowledge, and they have urged teachers to work to align their grading practices to specific standards and procedures. (Bowers, p. 141)

In question is the validity of what learning is being assessed, and the validity of the communication of that assessment to others. Allen (2010) feels that the majority of teachers do not give valid grades. "Because grading is something that has been done to each of us during our many years as students, it is hard to change the invalid grading schema that has become embedded in our minds" (p. 218).

Guskey and Brookhart (2019) share this premise. The idea of what schooling is and what grades should be has become engrained in people's understanding of what is normal. While people may be quick to point out flaws and inconsistencies in the grading process, they still cling to the idea because it is familiar to them -- it is what they know and have experienced for themselves. They write:

If hodgepodge grading is so deplorable, why haven't students, parents, and administrators or the general public called for reform? It may well be that they

share a common understanding that grades often do, in fact, represent a hodgepodge of attitude, effort, conduct, growth, and achievement, and that is what they expect and endorse. (p. 33)

By establishing and actively enforcing achievement-only grading policies at the school level, administrators can create a school culture where all stakeholders (parents, teachers, and students) can trust the meaning of grades for all students (Randall, 2010).

Perception of Grades to Administrators

Administrators (both building level and district level) use student grades in many ways. They use the data provided by classroom grades to inform decisions about curriculum, staffing, textbook adoption, and more. However, some of those grades may be affected by subjectivity. "Teacher subjectivity blurs the meaning of grades and makes it challenging for school leaders to decide what are fair grades for students. Without explicit grading policies, school districts and leaders are vulnerable to litigation challenging the notion of fair grading" (Guskey & Brookhart, p. 161). It becomes a delicate balance of wanting to provide teachers with the autonomy to run their classrooms while ensuring that the grading process is fair and consistent throughout the building and district. Left unchecked, grades can be subject to bias and subjectivity.

Despite grades' vital role in the decision-making processes at various educational levels, many experts call their validity questionable, influenced by outside factors such as teacher subjectivity, and the lack of standardization (Feldman, 2019; Guskey & Brookhart, 2019; Allen, 2005)

Students must earn the grades to pass classes. Students who fail classes cause a ripple effect not only to themselves but to the school. Administrators are aware of not only the emotional costs but also the financial costs of student failure. Tally the costs to the district each time a student is forced to repeat a failed course, attend summer school to complete a failed course, or be enrolled in a credit recovery program to make up a failed course (Carifio & Carey, 2013). Student failure places not only an emotional impact on the student but also a financial impact on the district. Administrators must evaluate grading practices to ensure that students repeating courses are doing so because they genuinely did not master the content and require support and reteaching. Not because they were subject to subpar grading practices that did not reflect the level of content mastery of the student and instead was a formula of non-academic factors working against the student. Canady (2017) calls this process a "*take class, fail class, repeat class* instructional model" (p.19).

While it would be ideal and beneficial for districts to fully trust and embrace the data derived from their students' grades, some researchers state that this is not the case. Teacher-assigned grades have come to be thought of as subjective and unreliable, and therefore not used for systemic decision making by administrators, central offices, and state and federal policymakers (Bowers, 2011).

Perception of Grades to Students

For high school students, grades can mean any number of things. Grading is necessary to study because they have a presence and impact on all students' educational experience (Brookhart et al., 2016). All students, regardless of where they attend school, experience grading. How students perceive those grades, and what they take from them

can vary. Some students are motivated by the acquisition of top marks and see their benefit. Other students may only do the bare minimum to pass and move on and find no motivation or value in the grading process.

Teachers' grading introduces a distinction between real and perceived achievement. Perceived achievement is the students' indirect perception of their real achievement, as mediated through the grades given by their teachers. Under some circumstances, the student may care more about the perceived achievement than real achievement (Bonesronning, 2004). In *Deschooling Society*, Ivan Illich (as cited by Morrison, 2003) has argued that grades make students believe that only those things that are measured and documented are worth anything: "To them, what cannot be measured becomes secondary" (p. 21). Because of grading, many students go around feeling that nothing is worth doing unless they receive credit. "The students then become intellectually quiescent and dependent, waiting for teachers, or other authority figures, to feed them" (Illich, p. 21). Grading makes students think less creatively. When the fear of failure or the risk of a bad grade, potentially paralyzes students, they are unlikely to want to take risks and be original. Grading also makes students less apt to prefer challenging tasks over safe ones (Morrison, 2003).

Grading can be motivational, but it can also pit students against each other, creating a sense of competition. Some would argue that the very act of grading gets in the way of actual learning. "Grading causes students to focus almost solely on competition and performance, not on the joyful process of learning. In essence, then, grading violates our society's intellectual heritage. Is this what we want grading to do?" (Morrison, p. 20).

Students can feel the pressures of grades at home. Children and parents trust and accept grades as the measure of a student's success. To them, grades have come to represent the relative value of a learner -- regardless of the criteria used to arrive at the grade (Vatterott, 2015). Since students' familial support systems vary, as do the level of understanding that parents have of the grading system, grades must be equitable in their derivation.

The meaning that students give to grades is what is most important. "Teachers need to consult with students about how they interpret grades -- what message students receive, what this suggests about their learning, and what the effect of the grade is on motivation" (Guskey & Brookhart, p. 106). The research suggests that stakeholders often view grades as a form of feedback to students. However, it is worth examining whether the feedback is what the teacher intends for the student to hear.

Perception of Grades to Teachers

If one hundred different teachers were interviewed about their philosophy of grading, there would be one hundred different responses. Teachers can be very protective and territorial about their grading. More recent studies have sought to understand teachers' thinking about grading instead of merely criticizing their grading practices (Guskey & Brookhart, 2019). Some view questioning their policies as undermining their authority and questioning their very professionalism. It can become a delicate balance to try to align grading practices to create a universal policy while acknowledging and maintaining the teacher's voice.

Feldman's (2019) research found it ironic that most teachers detest the act of grading, calling it unpleasant, time-consuming, and anxiety-provoking. He also found

that teachers felt that grading was arguably the only aspect of schools in which the teacher and supervisor's power dynamic was inverted.

With administrators and policymakers defining nearly every aspect of a teacher's practice, we have one remaining "island of autonomy": our grades. Grades are entirely within our control -- the declaration of our professional judgment of student performance and the most concrete symbol of our authority and expertise. (Feldman, p. 4)

Brookhart (2015) found that many of the interviewed classroom teachers did not follow many recommended practices for grading, stating that teachers are often uncomfortable with grading. He cited three general reasons for the discrepancies between recommended and actual practice. "1. best practice may be a matter of opinion; 2. recommended practices do not take some of the practical aspects of teaching into account; 3. teachers lack training or expertise" (Brookhart, p. 123).

Some teachers believe that grading gets in the way of student learning. They would find that after spending a great deal of time reading and grading a lengthy essay, the student would skip over all the feedback and constructive comments given and merely look at the grade. Once they saw the score, very few would take the time to read and process the feedback, much less apply it. Once a grade is put on something, the learning stops (Gonzales, 2015). Real learning occurs when a student takes the opportunity to process, synthesize, and apply the feedback, but that is lost to the student when they only process the score.

Grade Inflators and Deflators

One of the primary factors that affect the validity and accuracy of teacher-assigned grades is the practice of using grading methods that artificially inflate or deflate the grade with non-academic items that add or take away points. This is problematic for many reasons. As Marx (2013) points out, "the usefulness of GPA as a standard of excellence has been compromised by increasingly forgiving academic policies coupled with no obvious means of assessing the frequency with which such policies are employed" (p. 40). This illustrates the need to further examine and scrutinize grading practices that artificially inflate student achievement.

Teachers are often left to their own devices when it comes to assessing work and assigning grades and have increased autonomy in their classrooms. This can lead to grading practices varying widely from teacher to teacher, not only between different districts but even within the same school. This is a widespread practice, and in many districts, the norm. It is not a malicious act; teachers use their professional judgment to create policies that they believe are in the students' best interests and promote learning. However, they are creating these policies in isolation and often do not align with their department's peers. While well-intentioned, they are inadvertently creating a flawed system of evaluating student performance (Feldman, 2019).

This phenomenon is not new. According to Brookhart, et al. (2016), the earliest study (1936) found that the students who earned high grades but had low test scores performed well in other areas including penmanship, attendance, punctuality, and effort. Teachers also scored the students high in industry, perseverance, dependability, cooperation, and ambition. While these qualities may contribute and aid in a student's

success, they should not receive a grade for them. If a student writes neatly, the teacher can more easily assess the content of their writing. When students are present, and on time each day, they undoubtedly have an advantage over the absent students. Students who try hard and cooperate with others surely have a better attitude about education and their learning. However, these traits and characteristics are challenging to measure without being inherently subjective, and students' grades should not include them. Grades should report only what students know and can do after a learning cycle, and not the routes they used to get there (Wormeli, 2018). When teachers use academic achievement as a grading criterion, they assign grades in a manner proportional to the amount of content students learn. If they learn a great deal of content, students receive a high grade; if they learn little content, they receive a low grade (Marzano, 2001).

Final grades are subject to a litany of practices that can potentially impact their validity. Examples include the use of grading as a form of classroom management, grading for attendance and punctuality, penalizing late work, the use of extra credit (or bonus points), utilization of the mean, using zeros for missing assignments, grading homework and practice work, and assigning points for participation. In addition to these practices, grading can also be influenced by teacher bias (despite best efforts) and the lack of or quality of training in teacher preparatory programs.

Ideally, teachers would not include effort, behavior, and attendance with academic achievement in an overall grade. Cross and Frary (1999) (as cited by Marzano, 2011) state that 81 percent of teachers and 70 percent of students agreed with that statement when asked. However, in the same study, 39 percent of teachers included effort and behavior in their grading practices. Despite agreeing with and acknowledging that

those factors should not be included in the final grade, almost forty percent of them still incorporated them.

Previous research indicates (Randall, 2010) that teachers consider many factors other than pure academic achievement when assigning final grades: homework, participation, improvement, ability, effort, and behavior. Additional research (Rauschenberg, 2014) include differences in teacher grading standards, district grading policies, teacher stereotypes, teacher quality, and curriculum adherence. Liu's 2008 article *Using the Teachers' Perceptions of Grading Practices Scale* (as cited by Guskey & Brookhart, 2019) states that over 90 percent of the teachers reported using effort in grading, over 60 percent used student ability, over 60 percent used attendance and participation, and over 40 percent used classroom behavior.

If schools expect that a grade is trusted as a true and authentic measure of a student's academic achievement, then it is paramount that all factors that comprise that grade must be measurable examples of achievement and nothing else (Sadler, 2010). When non-achievement categories are factored into grading, it creates a disparity. In some cases, whether a student passes or fails may depend simply on which teacher is assigning the grade (Carifio & Carey, 2013). Brookhart et al. (2016) suggest that most teacher's grades do not yield a pure achievement measure but are instead a hybrid of student learning and their behaviors in the classroom. There is little excuse for poor grading practices, and all systems have room for evaluation and improvement. A century of grading research confirms widely varying grading practices among teachers and in the reliability of final grades, both in their meaning and accuracy. With that amount and chronicled pattern of inflation, teachers should not rely on these practices.

While it is well-known and documented that grade inflation is taking place, educators are likely not doing it maliciously. Most would support their own methods. Reeves et al. (2017) recognize that education professionals have wide-ranging perspectives on grading and can often readily find research to support their alternative points of view. There are four areas where they state that there is clear evidence with grave consequences for inaction. "The use of the average or arithmetic mean to calculate a final grade; the grading of practice, or homework; the use of the zero on the 100-point scale; and the use of grading as punishment for misbehavior" (Reeves, et al. p. 42). When these practices are further analyzed for their potential inaccuracy, it could be difficult for educators to support their classroom use.

Behaviors: Rewards and Punishments

Some teachers have come to rely on grading as a form of classroom management, believing they can control students through the fear of losing points for misbehaviors or forgetting materials. Whether it be docking points for not having the necessary supplies, or a lowered grade on a presentation because the student was talking during their peers' presentations, these practices can result in final grades that are not indicative of what learning took place. While these practices are prevalent in schools, Reeves (2000) would advise against it; "When grades are used to punish poor behavior, the true meaning of the grade becomes unclear because it is now an uncertain mix of achievement and behavior" (p. 44). Rewarding certain behaviors are meant to support student learning. However, it forces students to fit within a set of behaviors anchored to the teacher's subjective implicitly biased idea of what a successful student is. These behaviors are more often

than not the behaviors that the teacher values, and are embedded within that teacher's specific culture, upbringing and learning style (Feldman, 2019).

This practice is doing little more than teaching students to "*play school*." As Brookhart (2015) states, "teachers' assessment of students' ability to negotiate the social processes of schooling represents much of the variance in grades that is unrelated to test scores" (p. 823). Additionally, Bowers (2011) claims that emerging literature indicates that grades may be a multidimensional assessment of both student academic knowledge and a student's ability to negotiate the social process of schooling, such as behavior, participation, and effort.

Guskey and Brookhart (2019) agree that using grading to assess behavioral factors, teachers may base grades on the students appearing to be fully engaged, but in actuality, the students are merely complying. They are not giving the level of legitimate effort and engagement that leads to increased learning. These practices can be damaging to students (Dueck, 2014). Teachers risk conveying that they are willing to sacrifice the intended purpose of grading in order to produce the desired behavior, thus placing compliance ahead of achievement.

Using grades to deter off-task behaviors could be useful for some students and in some environments. Immerwahr (2011) warns against it, recalling his fears as he began teaching, questioning how he would make the students do their work. "It is tempting to rely on the power of the grade as the solution, but you may degrade the environment you're trying to foster" (Immerwahr, p. 336). Grading, especially in low-income schools, may reflect a hidden curriculum of compliance and control, in which teachers use grades as carrots and sticks to keep students in line (Goodwin 2011). This practice can easily

mistake compliant behavior with learning, and off-task and disruptive behavior as a deficiency in learning (Guskey & Brookhart, 2019).

While there is likely to be an inherent connection between appropriate and positive behavior and a student doing well in school, it is essential that the student not be rewarded academically for good behaviors or punished academically for undesirable ones. The poor behaviors could stem from a multitude of sources, none of which should invalidate the student's grade by having points removed as a consequence.

Motivation and "Point Grabbing"

The concept of "playing school," as previously addressed, has created an environment where many students replace authentic learning with an elaborate game of acquiring as many points as possible. Vatterott (2015) calls it a form of extortion, and a developed barter system, where the students approach learning with the mindset of "I will work, you pay," and points are the currency. Students learn how to game the system and how to earn points from each teacher.

That concept is further explored in Feldman's (2019) work; the teacher motivates students to behave by creating a currency of points and awarding or withdrawing currency based on students' actions, thus becoming a full-blown economic system of incentives and penalties. Through this practice, the teacher is "essentially the Federal Reserve of the classroom who can 'print' more currency and inject more points into the classroom economy when needed" (Feldman, p. 35). Brookhart (2015) had similar findings, calling it an economic mechanism, supporting the idea that grades were a form of currency. "Among teachers, a more common image than achievement is that of grades as currency; this image is evident in teachers' frequent use of the words *earn*, *work*, and

perform" (Brookhart, p. 139). Clearly, the researchers suggest that grading systems promote learning and cultivating student thought as opposed to creating systems where students focus strictly on point gathering.

While many teachers see this method as an effective way to motivate students to learn, Feldman (2019) points to Guskey's (2009) research in opposition. Many students may be motivated by the acquisition of points and receiving high grades as an incentive to produce more and learn more. However, students become jaded and unmotivated by zeros and F's. No studies support the use of low grades as punishment. Instead of promoting more significant effort, low grades more often cause students to withdraw from learning. Wormeli (2018) would argue that a point system is altogether ineffective for student motivation, questioning whether teachers seek control and coercion, or students' learning and maturation. Teachers can coerce students into doing tasks, but real motivation and engagement come from within. Kanold's (2018) research supports the claim. Students are aware of when a teacher designs a grading method to keep score. Students also spend a large amount of effort, just trying to figure out the system. Kanold (2018) found that it was more important to figure out the teacher's grading system to some students than it was to actually learn the content.

When a student equates learning to nothing more than point acquisition, it becomes little more than a game. They are no longer interested in feedback, improvement, or actual learning. They become more focused on mere point collecting until they reach their desired grade. This system is not an accurate depiction of their learning – it becomes more of a carnival where they trade in their tickets for a prize.

Attendance

Grading students on their daily attendance can be considered an unfair practice primarily because students do not always have control over whether they attend school each day. Situations arise when a high school-aged student may have to stay home with a younger sibling to provide care while the parent(s) are away. Students are often beholden to others to provide transportation to school. If the family vehicle is out of commission, that student has few options or other opportunities to get to school. Taking points off because a student is not there for the same amount of seat time as their peers unfairly penalize a student academically. They are perhaps doubly penalized as they are already missing some of the presented content and will potentially not do as well on an assessment as a result (Bowers, 2011; Wormeli, 2018).

Another example of penalizing a student for absences are teachers who give a graded quiz at the start of each period or grade their bell work. If a student is habitually five minutes late, they miss the daily opportunity to demonstrate what they learned the previous day. Unless they are allowed to make up the missed daily quiz or formative assessment, they are academically penalized for their tardiness and not allowed to make it up. In some instances, missing daily points could result in a B-level student being given a C or D because of punctuality and attendance.

Some teachers make the argument that a student *should* lose points if they skip their class. That goes back to using grades as a classroom management tactic. A consequence of skipping should always be making up the time or some other restorative justice practice. The endgame should always be getting students to learn the content and show mastery.

Late Work

The practice of lowering scores for late papers is quite common. Proponents argue that by holding students accountable to deadlines, teachers are preparing them for the real world. However, the experts disagree with that mindset. While students must be regularly turning in work so that teachers can see their progress and collect that formative data as they move through the unit, is it equally important that the student turn an assignment in on a given day? It can be argued that the due date is arbitrary at best. If a student submits an essay that is high quality, but three days later, should that student be penalized because they had extra time to accomplish that quality? Teachers have different practices of handling such situations, ranging from a few points off to a lowered letter grade each day. The "A" quality essay could then potentially receive a "D" grade. That grade would be a misrepresentation of what the student was able to demonstrate and would be a convoluted mixture at best. It is then worth examining what is being graded, adherence to timelines, and the ability to follow a schedule or the mastery of the content (Vatterot, 2015).

This practice can have a negative impact on students. In Feldman's (2019) *Grading for Equity* he quotes a student who is frustrated with her teacher's grading procedures: "Teachers constantly and subjectively judge me, adding or subtracting points for everything I do: my punctuality to class, whether I meet deadlines, effort, politeness, what and when I say in class, my homework performance, and my test scores" (Feldman, p. 60). This mindset can lead students to become disenchanted with school or offer them perceived acceptable excuses as to why they are not achieving – pushing the blame and responsibility onto the teacher.

The learning process is ongoing, and each student has their respective timeline. Some may take more time than others to master a concept. While at times, hard deadlines are appropriate; more often, students perceive them as arbitrary. Dueck (2014) argues that it is natural for students to turn things in late. Procrastination is a human condition. Just because students procrastinate does not mean that they do not know the causes of World War II. Evidence of understanding is not dependent on a due date.

At times a firm deadline makes sense and could be part of the learning objective. Courses in journalism, where students are learning the process of putting together a newspaper or yearbook, often have actual deadlines they need to adhere to meet their print dates. In situations like these, the due date is part of the learning process and is an example of a grade that could justifiably be lowered for being late.

Extra Credit

The awarding of extra credit and bonus points is one of the most prominent and likely rampant examples of grading practices that distort and dilute final grade reports. When bonus points augment authentic well-designed assessment scores for knowing the instructor's cat's name or for extra credit from bringing tissues for the classroom, it undermines the integrity and validity of the course grade. When teachers offer extra credit, they reinforce that the class is not really about learning or mastery of standards, but about acquiring enough points, however possible (Feldman, 2019).

Extra credit supports the practice of point grabbing and working the system. The student is no longer concerned with how much they learn, but more so with gaming the grade book until they hit the requisite number of points needed to earn their desired grade. On Gonzales' 2015 podcast, Sackstein addressed the issue, calling it bothersome

that students simply wait until the end of the term and then scramble to do a litany of extra credit to try to make up their missed points.

Extra credit contributes to the idea of hodgepodge grading and skews the final letter grade. What that grade becomes is a conglomeration of artifacts of student learning, and a series of non-academic points "earned" from attending an event, participating in a class fundraiser, or performing any number of tasks that do not enhance learning or promote content mastery. In lieu of extra credit, real learning that would benefit the student would be to redo the assignments or allow for makeup exams. Feldman (2019) points out some of the inequities within the practice of extra credit. He calls it inequitable "because it reflects a student's environment over which she has no control. Extra credit, although it is optional and open to all students, doesn't allow all students to take equal advantage because it requires extra resources beyond the course requirements" (Feldman, p. 114). Since extra credit ranges widely, some students may not have the means to participate. For example, when students are given extra credit for bringing a prepared dish for ethnic food day, or for attending an event during the evening. These examples present barriers to students who may not have the financial means to participate or have access to transportation.

Extra credit gives students the message that they do not have to put forth their most significant effort because they know there will be a safety net. Instead of pushing themselves academically, they will often settle and then try to work the system to get a bump later on. This practice ultimately can result in grades that are augmented with non-academic fluff.

Utilization of the Mean

Teachers commonly average a series of scores together to arrive at a final grade. Grades throughout the entire grading period are averaged together, with each mark given equal importance. Researchers find this practice problematic as it penalizes students as they are learning and makes it difficult to recover from stumbles in the learning process. If the goal is mastery, then the process potentially should not matter.

To calculate a grading average across time is to engage in the fantasy that proficient individuals never make mistakes or, more likely, that their mistakes are counterproductive. Watch any toddler learning to walk, and it is clear that mistakes are the engine of success. To say the toddler should get a poor grade in walking because of her many spectacular failures along the way would be ridiculous. She eventually got there. She mastered the skill. (Reeves, et al., p. 43)

Kanold (2018) advises against the practice. To rely on a basic calculation to assess a student's content mastery and understanding is a pitfall that can create obstacles for students.

Wormeli (2018) argues that mode is more practical. Using the mode would be a better representation of a student's overall achievement over a period of time by demonstrating what score the student earned most frequently, as opposed to having one or two scores heavily skew the final grade. "We want our grades to have integrity, which means we need the curriculum to align with assessments, and we need to use mode (and in some situations, median) instead of the mean" (Wormeli, p. 189). While standard deviation is no longer considered an acceptable or appropriate grading practice (grading on the curve) when it is applied to a class of students, it has merit when applied to each

individual student. Their final grade should reflect where the bulk of their learning scored and not be skewed in either direction by extreme outliers. The use of the average can fall victim to such skewing.

Using Zeros and the Impact on Final Grades

The assigning of a zero for an incomplete assignment or assessment has a tremendous negative impact on a student's final grade. The damage of one zero can be catastrophic mathematically. A zero can misrepresent an absence of learning, and that the student did not in any way demonstrate content knowledge for the assessed standard.

Guskey and Brookhart (2019) state that a particularly egregious example of how grading is unfair is when a teacher calculates a missed assignment as a zero. This is because the zero assumes no knowledge of proficiency and unduly skews the overall score in a negative direction. Conversely, many educators believe that students should receive zeros as both a consequence and a motivator to do more. "Defenders of the zero claim that students need to have consequences for flouting the teacher's authority and failing to turn in work on time. They're right, but the appropriate consequence is not a zero; it's completing the work" (Reeves, as cited in Feldman, p. 77).

However, the desired motivation may not be the result. Receiving a zero is not a positive motivator for students who receive them. This averaging approach creates an environment where students give up. When the only way to overcome the deficit and earn a D at best is to accumulate several perfect scores, the student can feel defeated, and no longer try (Canady, 2017). Because of the adverse effects and impacts of zeros, many districts have adopted the practice of assigning fifty percent in place of zeros. Additional information on this theory will be found in later sections.

Grading Homework & Formative Grades

The decision of whether to grade homework is a contested one and ranges in teacher opinion. If the teacher does not grade homework, then what will prompt students to complete it? However, if homework *is* graded, how can a teacher know for sure that it was completed with fidelity? Additionally, if homework is meant for a student to practice and learn and grow from the practice and the feedback provided, why should the process be graded? Feldman (2019) argues that this is precisely when a grade should not be given because student mistakes are penalized during the very stage of learning when students should be making mistakes.

Dueck's (2014) research cites several reasons that homework grades should not be used. Grading penalties aimed at completion compel some students to cheat. Take a math class, for example, where the teacher assigns twenty problems. If a student knows that they need to get all of the questions completed and correct to get the points, they will likely use several methods to do so to ensure they get the points: copy from a peer, get the answers from an internet source or app that will complete the problems for them or have a parent, sibling, or tutor do the work. While the student will receive points for completion, there is no guarantee that learning has taken place. However, if teachers empower students to try the assignment and work through the problems without fear of being penalized for their learning process, they are more likely to work independently and complete the assignment with fidelity. If students feel pressured to copy work for the sake of completion, they do not realize that someone else has done the deeper thinking, problem-solving, and processing. Grading homework confuses *completion* with *understanding*. By removing data from the grade book that merely reflects completion

and compliance, teachers can see the aggregate data more accurately reflect student understanding.

Reeves (2017) presents another point. Grading homework leads to two types of adverse outcomes. The first is compliant students who work at skills that rarely matter, and the second is students who work at nothing, unable even to approach the task because they cannot do it independently or do not see the value in it. The use of first attempts and practice in a permanent grade penalizes students while they are still in the learning process. The teacher permanently records their mistakes, and the students have no opportunity for redemption. In conjunction with averaging grades, one bad grade can seal a student's fate (Vatterott, 2015).

Teacher Bias

Whether conscious or not, teachers are prone to fall victim to their own biases as they assess students and assign classroom grades. These biases can range from a teacher's perception of how students *should* act and perform in the classroom to inadvertently skewing a grade because of when it was graded. Anything subjective in nature is prone to fall subject to bias. According to Morrison (2003), grading in and of itself is anti-intellectual, irrational, and arbitrary. One of the biggest criticisms of grading is that it is so subjective. Allen (2005) supports the notion of the bias of subjectivity, stating that many factors such as effort, motivation, and student attitude are subjective measures made by a teacher. Their inclusion in a grade related to academic achievement increases the grade's chance to be biased, unreliable, and invalid.

Teachers also get a reputation for the rigor and severity of their grading practices. "So and so is the tough teacher." "I hope I get teacher X because they are so easy."

Students rarely state, "I hope I get her because she is very transparent about her grading practices, and all of our work is scored against a provided rubric after we receive rich and supportive feedback throughout the process." That level of rigor, as Brookhart (2015) supports, aids to variance in grading. In his research, teacher severity or leniency was found to be a source of variability in grades.

Bias ties in with other previously noted areas of grade inflation. In Guskey and Brookhart's (2019) work, they state that teachers may inappropriately identify how much effort or behavior to consider in a grade. They may confuse boisterous behavior with a lack of learning. They are illustrating this kind of bias in grading.

Wormeli's research (2018) found that *when* a teacher graded an assignment in relation to other students' assignments had a factor in what score it received. He noted that good work received a much higher score when it followed examples of imperfect work. If that paper followed two or more examples of low quality, the biased advantage was even more significant. If students were aware of this, they could theoretically place their paper beneath a low-performing classmate's to receive a subconscious, unmerited boost in score.

Feldman calls the bias and subjectivity a recipe for inequity. "White teachers can misinterpret African American students' behaviors, incorrectly believing them to be signs of disrespect or stemming from some evil intent" (Feldman, p. 43). The teachers judge it as inappropriate due to their differing cultural background where that type of behavior is not as familiar or accepted. Westphal (2016) also found that some teachers potentially judge lower SES students less favorably than their higher SES peers, despite their similar standardized tests.

Many well-intentioned teachers use their understanding of individual student circumstances, instructional experience, and perceptions of equity, consistency, accuracy, and fairness to make professional judgments instead of relying solely on a grading algorithm (Brookhart et al., 2016). Because of that, grades often become measures of how well students adhered to the teacher's expectations instead of measuring actual academic achievement and growth (Allen, 2005).

Racial, gender and other stereotypes of student performance may also influence how teachers issue grades. Beliefs that male students perform better in science and math or that female students are better writers can also have an impact. Ethnic and socioeconomic stereotypes can also have an impact on how some teachers view students. "These stereotypes may cause grade discrimination, where a teacher assigns a grade at least partially based upon stereotypes of a student's innate characteristics rather than solely based upon student performance" (Rauschenberg, p. 3). Rauschenberg cites a study by Hanna and Linden, where exams were blindly scored by two teacher panels to measure the validity of the scoring. In the first test, they were aligned. On the second batch of exams, one group was given a fact sheet with student demographic information (age, gender, and socioeconomic status). They found that teachers issued lower scores to the students labeled as lower-socioeconomic than those who had no fact sheets.

Participation Points

Participation points are simply points given for doing what is expected and contributing to the class. This practice has wide acceptance in schools, but it may be a sign of little more than compliance. The teacher may assign a participation grade weekly to motivate the student to insert questions into the discussion, bring the required

materials, or dress out for the activity. The practice is open to questions about what the teacher is measuring. Usually, there is no rubric provided to assess the quality of the questions or comments that the student contributes to the class discussion. There are simply tick marks that they participated three times that week. While it is true that it is usually necessary that a student has the appropriate clothing on in order to engage in a physical education class fully, it may not be an accurate assessment of a student's knowledge of or ability in the rudiments of a layup if they are given a 60% because they did not dress out for two of the days. This practice can alter the final product and give an inaccurate and murky picture of the student's actual mastery of the standard.

Feldman (2019) cautions against participation points, as it is an additional example of bias. Participation and effort categories are almost exclusively subjective, especially when the teacher is white, and the students are not. Wormeli (2018) goes on to state that teachers should be seeking *evidence*, not compliance. He calls participation the tool used to render that evidence.

Teachers must consider whether students' participation is a technique used to learn the standards, or if participation is the standard itself. If participation is merely the way a teacher helps students to arrive at mastery, then it is inappropriate to grade it. (Wormeli, p. 164)

Participation may be a valuable exercise or activity in the learning process. Students can learn skills from the day to day activities in which they participate. Whether it be games of badminton in a physical education class or rehearsing a symphony movement, daily student participation is part of the learning process. However, teachers

should give the student authentic feedback about their participation and not include it in formal final grades.

To grade participation, regardless of the criteria, focuses more on a student's conduct than what she has learned and is subjective and, therefore, a bias-infected judgment of a student's behavior (Feldman, 2019). Granting points for participating is not assessing what level the student performed or providing data about their deficiencies. Participation is practice time, and grading it gives potentially inaccurate data.

Effort

The incorporation of grading effort is perhaps the most inherently subjective practice. It may not be possible for a teacher to fairly assess the level of effort a student put into an assignment. Even if the teacher sat beside each student while they completed an assignment, they would still have no idea of the intellectual output level that the student invested in its completion. This creates a situation where a student who takes less time to complete something could be downgraded because of a perceived lack of effort or a hand-drawn map of the 50 states deemed less accurate because the coloring is sloppy. If North Dakota is still in the correct spot, but it appears that the student did not put in their best effort, it does not change North Dakota's geographic location, and it is not evidence of a student's lack of knowledge.

Teachers are told to ignore effort when assigning final grades. However, most stakeholders know, or suspect, that educators either explicitly (i.e., effort on a rubric) or implicitly (homework completed, class participation) assign grades, including the effort criteria (Randall, 2010). In Brookhart's (2016) research, it was found "teachers indicated the need for fairness and accuracy, not just accomplishment, saying that grades are fairer

if they are lowered for lack of effort or participation and that grading needs to be strict for high achievers" (p. 827). Surveys of grading practices have documented that when teachers combine grades from individual pieces of student work to arrive at composite (report card) grades, they take into account academic enablers like effort and ability, at least to some extent, raising issues of validity (Brookhart et al., 2016).

Lack of teacher training

Guskey and Brookhart (2019) point out that little research has been done on *why* teachers include factors other than achievement in grading. A critical area contributing to the variance of grade configurations and assessment methods could be the difference in the quality of training teachers receive in teacher preparatory programs. According to Allen (2005), fewer than half of the fifty states require specific coursework on assessment for their teachers' initial certification.

Additionally, teachers often make grading decisions with little school or district guidance (Brookhart et al., 2016). While most teachers are involved in some professional learning community or have colleagues to collaborate with on lesson planning and common assessments, the alignment does not find its way into grading practices.

Unfortunately, most teachers and professors today are not well trained, typically grade alone, and rarely seek help from colleagues to check their grading reliability. Thus, working toward more explicit criteria, collaborating among teachers, and involving students in the development of grading criteria appear to be promising approaches to enhancing grading reliability. (Brookhart et al., p. 836)

Stiggins (as cited by Allen, 2005) suggests that the concepts of reliability and validity related to classroom grading practices are not addressed in the courses which introduce these terms to our preservice teachers. Brookhart's (2015) research found that classroom teachers do not follow many of the recommended practices for grading because teachers are often uncomfortable with grading. He states three reasons for the discrepancies between recommended and actual practice: "1. best practice may be a matter of opinion; 2. recommended practices do not take some of the practical aspects of teaching into account; 3. teachers lack training or expertise" (p. 123).

Allen (2005) found that less than one-third of teacher education programs require an assessment course. This leaves new, inexperienced teachers left to rely on their own experiences as students, and their limited student teaching time as examples from which to form their grading practices. It was also found that "teachers-in-training have a difficult time accepting theoretical principles that do not match with their personal experience" (p. 221). For example, a teacher who, when they were a student, experienced a particular method of grading that they achieved high marks in, is generally accepting of that practice. If their teacher training program does not offer any contrasting style of grading, and their student teaching experience aligns with that practice, they will likely embrace the practice as universally accepted and appropriate.

Other Factors

Additionally, a few other factors contribute to inflated and deflated grades. Some can be attributed to teacher error (Guskey & Brookhart, 2019). Teachers are subject to fatigue and can make mistakes when scoring assessments. Depending on their classroom practices, students may not be able to review the test and appeal the mistake. Fatigue can

also cause teacher judgment to lapse when grading a long series of papers over an extended amount of time. If the teacher is not careful to recalibrate before each new paper, they risk subjecting the paper to a grade influenced by their fatigue.

Others can stem from the software or grading programs that teachers are using--the calculations that we commonly use to derive grades--and often embedded in our mathematically unsound grading software (Feldman, 2019). Even when grading programs are not in use, a teacher's mathematical error can result in a final grade that is inaccurate. Examples can also include teacher clerical error where they may simply mistype a score or enter it into the wrong field.

Teachers often have to move quickly to get massive amounts of grading completed in an appropriate timeframe so that they can provide feedback to students. In doing this, they are subject to clerical errors. If students do not self-audit their scores against the grade book (which with current technology and online grading platforms they can), then their grades are potentially subject to data entry and calculation errors.

Sometimes teachers are required to teach outside of their content area. "In college, out of field teaching does not exist. Yet in public schools, through the generous auspices of the "emergency certificate," almost anyone with a pulse and a degree can teach almost anything" (Stanley, p. 103). Stanley also notes that in some states, there is a fifty percent chance of having a teacher who is not endorsed in all content areas they are teaching. This may be due to budget restraints, or the inability to recruit and retain an endorsed teacher in a rural area. Not having the requisite endorsement can potentially lead to the teacher not aligning their grading practices for what is best for that particular subject.

McMillan (as cited by Duncan & Noonan, 2007) believes that the core subject plays a role in certain grading practices and that some subjects are more apt to use certain practices. "Mathematics teachers reported using academic enablers, extra credit, graded homework, and the use of zeros less frequently than English, science, and social studies teachers" (Duncan & Noonan, p. 3). They also suggest that the *size* of the class can impact the grading practices used. For example, larger class size may force the teacher to rely on penalizing for behaviors as a method of management and daily participation points to keep track of student engagement. In contrast, a smaller class would allow the teacher to lead discussions and naturally provide more opportunities for students to be engaged and eliminating their need to use penalties for behaviors. Smaller class size also allows teachers to take more time with the grading since by having fewer papers and tests to grade, they are less pressed for time and less likely to make an error.

Lastly, the variance in grade *scales* used can contribute to differences in final letter grades. It is common that one school district is using a seven-point scale while other districts in that area are using a ten-point scale. When students from these different districts are applying for the same scholarships and college acceptance, the admissions officer would have no way of knowing if their A in chemistry was 92% or 98%. If the student finished the course with a 92, they could potentially receive an A in one district and a B in another.

One possible reason schools may intentionally use grade inflation is to *attract* families to the district. Walsh's (2010) research suggests that schools inflate grades and soften their grading scales to promote the image of higher achievement, making them more desirable to families. As school choice gets increasingly competitive, Walsh

suggests that some schools are altering their grading practice to produce more straight-A students. The hope is that other parents talk up the school in their networks, saying, "our daughter is a straight-A student is surrounded by students who earn the same." This, in turn, creates an image for the school of high academic achievement. Conversely, Walsh suggests that some schools inflated grades and relaxed practices in response to No Child Left Behind to avoid the state labeling them as a school of improvement.

Summary

Regardless of which type or types of inflators and deflators are used, all of them contribute to the growing concern of grade validity and reliability. When something has an element of subjectivity to it, it is open to interpretation. In Brimi's (2011) study (as cited by Brookhart et al.), 73 English teachers all graded the same essay. The result was a range of scores that spanned 46 points on a 100-point scale. The teachers awarded the essay grades ranging from A to F. In Bailey's 2012 study, it was found that the gender of the teacher played a role in the type of grade inflators used. The study found that male teachers placed a higher emphasis on effort, whereas female teachers emphasized behavior.

Alternative Practices

While traditional grading practices have their inherent flaws and are potentially subject to practices that undermine their intent and impact their validity, other grading methods are gaining popularity and increased implementation. Some alternative practices prevent and even eliminate many of the inflators and deflators that traditional methods can be subject to. These practices range from standards-based grading, where grades are

reported strictly as to the level of mastery the student demonstrated on each content standard to abolishing grading altogether.

Standards-based Grading

Standards-based grading moves away from traditional grading practices and reports each standard separately, as opposed to an all-inclusive "omnibus" grade (coined by Marzano and Heflebower). Most standards-based scales are 0-4 or 0-5 and reflect students' increasing skills or mastery. For a 1-4 scale, a "1" indicates that students have little understanding of a concept, and therefore cannot demonstrate mastery of it. As students learn and progress, they can demonstrate partial mastery and score a "2." Once they meet a target, they score a "3". The "4's" are reserved for students who exceed the learning target (Davis, 2020). Another example of this utilizes letters as to where the student is in the learning process. The alternative system uses the following; NI; no information, A = approaching the standards, M = meeting the standards, and E = exceeding the standards (Gonzales, 2015). This would take some explaining and calibration of parent thinking, as it would be potentially confusing that an A means approaching the standard instead of the A, meaning what it does in current grading practices.

The intended benefit to this approach is that it provides much more information about the student and better identifies areas of strengths and weaknesses than a traditional letter grade potentially would. Some also assert that "standards-based grading can provide exceptionally high-quality information to parents, teachers, and students and, therefore, has the potential to bring about instructional improvements and larger educational reforms" (Brookhart et al., 2016).

To be useful for improvement, grades must provide information about what students, individually or collectively, have and have not learned; know and do not know; can and cannot do. Those who advocate for standard-based grading argue that their systems provide the necessary detail (Anderson, 2018). Standards-based grading has the potential to restore integrity to the grading process. It can and will change our students' futures (Vatterott, 2015). It would allow educators to base grades on common reference points (standards), making it possible to develop uniform grading policies and standards. This would result in less variability among teachers and schools and focus on content mastery (Canady, 2017).

A study by Swan, Guskey, and Jung (2014) found that both parents and teachers preferred standards-based grading over traditional report cards. The parent participants felt that they received higher-quality information, more information, and were clearer and easier to understand than traditional report cards. The teachers acknowledged that standards-based grading took more time, but they received more information from the process (Guskey & Brookhart, 2019).

Moving away from the 100-point scale

Researchers have called the 100 points grading scale unreliable and difficult to calibrate. They question how teachers can delineate between a paper given the grade of 88% and one given 85%. Grading practices with fewer, well-defined categories tend to have significantly higher consistency between teachers and are easier to understand. They are much like a rubric. Typically, rubrics do not have 100 levels of mastery; there are usually four or five. In the absence of a genuinely accurate measuring device, adding more gradations to the measurement scale offers only the illusion of precision. Studies

have been done that investigated whether or not changing a grading scale would positively impact its reliability. Previously, the most common scale was 0 to 100, which was very inaccurate. When teachers were given scales with fewer categories or increments, their grades were much more consistent. They were most consistent when there were five categories or fewer (Guskey & Brookhart, 2019).

Feldman (2019) thinks, "we value failure over success - why else would we want to describe failure in sixty different ways but proficiency (B or above) in only twenty ways?" (p. 81). The one-hundred point scale, compounded with the utilization of a zero, can be doubly negatively impactful to students. Reeves (2000) cites a system where teachers used the values of 4,3,2,1 and 0 to represent grades of A, B, C, D, and F. When the increments are equal, the impact of one failing grade entry would equal a single grade difference. However, when that failing entry receives a zero on a 100-point scale, it has a much more significant negative impact.

Multiple Grades

Another practice is giving multiple grades, for example, breaking the grade down into different elements, including academic, effort, and participation. This practice also aims to remove grade inflation elements by reporting them out separately, leaving an academic grade that is designed to be a better representation of the level of academic achievement, while also informing parents about how the student behaved in the classroom. This practice is primarily seen in elementary school settings, but it has many benefits and practicalities at the secondary level. One example is having two grades, academic and soft skills (Feldman, 2019). Other teachers may call the second grade a professionalism grade. This method lets teachers report on the students' study habits,

behaviors, and adherence to due dates without punishing them academically. All factors that matter in assessing a student should be reported separately, therefore elevating the importance of work habits (Wormeli, 2018).

Guskey and Brookhart (2019) call the process of assigning multiple grades a "more useful and meaningful description of students' performance includes multiple grades. At a minimum, it provides grades that distinguish *product*, *process*, and *progress* learning criteria" (p. 221). Tomlinson and McTighe (as cited by Carifio & Carey, 2013) recommend separate grades for the achievement of goals, progress toward goals, and work habits. Their recommendation comes with a caution. Utilizing multiple grades can increase the already complicated grading process. The separate marks can create extra work for the teacher and can potentially be confusing for parents to interpret.

Researchers acknowledge the role that some of the subjective factors (primarily effort, behavior, and attendance) play in the learning process. There is a remarkably high correlation between academic success and effort, behavior, and attendance. However, they should be reported out separately. Although ancillary information such as effort and attitude could be part of an overall student report, they should not be part of a grade that represents academic achievement. Grades should only reflect academic achievement if they are to be reliable information. If it is necessary to report other factors such as attitude, effort, and social behaviors, then schools should use an alternate form of reporting (Allen, 2005). When we mix ancillary criteria that are not meant to serve as indicators of mastery with assessments meant to serve as such, we cannot trust the results or make decisions based on such criteria (Wormeli, 2018).

No Grades at All

Some educational leaders and experts explore the concept of not having grades at all (Allen, 2005). One of the most well-known critics of having grades is Alfie Kohn. In a series of publications, Kohn asserts that almost all forms of grading should be abolished. In an article entitled "*From Grading to Degrading*," Kohn asserts that 1. Grades tend to reduce students' interest in learning itself, 2. Grades tend to reduce students' preference for challenging tasks; 3. Grades tend to reduce the quality of students' thinking (Marzano, 2011).

Mark Barnes (who writes about teachers who stopped grading their students), and Starr Sackstein (who asks educational leaders to focus on feedback), both advocate for the removal of grading systems from our schools (Wormeli, 2018). In her podcast, Gonzales talks about the group *Teachers throwing out grades*--a growing movement spearheaded by writer, Mark Barnes. The concept is gaining momentum as "teachers start to discover that the quality of students' learning improves when we shift the way we assess their work in our classrooms, getting rid of the letters and numbers that typically define how well a student performs" (Gonzales, 2015).

While the idea that students may feel more emboldened to explore more difficult material if they were not afraid of a lower grade may be appealing, the likelihood of a large-scale movement toward eliminating grades as we know them is low. Entire social networking educational groups are centered around the concept. One example is *Teachers Throwing Out Grades*, a Facebook public discussion group where educators share ideas and questions about running a gradeless classroom. The group currently has over ten thousand members. Examples of discussion items include labor-based grading contracts,

feedback only classrooms, and competency-based learning. Even as the new ideas gain momentum, mounting such a paradigm shift away from what parents and the community understand and accept takes a great deal of time.

Relationship Between Grades and Standardized Test Scores

If grades claim to be an accurate representation of a student's academic ability, one could assume that they would align and strongly correlate to the student's performance on a standardized test. However, much of the research done on this topic suggests otherwise. Guskey and Brookhart (2019) refer to the correlation as *consistently moderate*. "This consistent moderate relationship persists across a significant amount of time and studies and despite large shifts across the educational system" (p. 38). With multiple factors potentially contributing to the inflation of grades, many officials have questioned their validity. Grades have gone up while SAT scores have remained flat, and someone will catch on soon (Cook, 2004).

The Higher Education Research Institute conducted a study finding that the proportion of students who report an A average on their high school transcripts has grown more than 30% in the last 40 years. In the same survey, it was reported that half of the students reported graduating with a GPA at or above 4.0 (Cook, 2004). Some of this data can be misleading, as students often self-report their grades when taking the ACT and SAT.

If students are collectively earning higher grades in high school, then their average score on the ACT or SAT would theoretically rise as well. However, a study done by Woodruff and Ziomek (as cited by Brookhart, 2015), found only a moderate correlation. The study examined composite ACT scores and high school GPA for all

students who took the ACT between 1991 and 2003. Consistent correlations were discovered (ranging from .56 to .58). However, even though the consensus believed that the correlation between grades and standardized test scores was strong, the study concluded that it is consistently modest (within the .5 range).

Vatterott's (2015) research supports the idea that a higher ACT score generally yields a higher cumulative high school GPA. However, the reverse is not always the case. A higher GPA does not consistently yield a high ACT score, and GPAs are rising at a proportionately higher rate than ACT scores. He cites examples of strong performances on standardized tests from relatively average students, and conversely, high-achieving, straight-A students who receive low scores on the same test. Many students complete high school with superior grade point averages that encounter challenges academically in college.

In Goodwin's (2011) study, he analyzed the mathematics grades of high school students taking the math portion of the ACT between 1991 and 2003. The students' average grade-point average rose from 2.80 to 3.04, whereas their ACT average scores on the math portion of the ACT rose only slightly from 20.04 to 20.555. ACT concluded that the higher GPA's reflected grade inflation rather than an increase in achievement.

We have noted these differential strengths of grades and tests: Grades can represent broader content and reflect unique accomplishments, but tests can more easily emphasize the most important content. Tests can more readily assess cognitive skills, but grades can more readily assess motivational components of achievement. Grades can reflect progress on what each student is studying, but tests can reflect progress on more significant long-term educational objectives.

Test scores are more comparable from one school to another, but grade scales are more readily accommodated to local situations and programs. (Brookhart, 2015, p. 283)

Current (and Best) Practices

While there is undoubtedly debate and argument over what a grade should constitute, and what best practices are, there are some key concepts that most grading experts agree on. One of the foremost experts and prolific writers about grading is Ken O'Connor. In his work, *A Repair Kit For Grading: 15 Fixes for Broken Grades*, the author highlights the need for educators to align their practices to ensure the accuracy and validity of grades by suggesting the best practices for grading. Each of the concepts, or fixes as he calls them, centers around ensuring that the grade is a fair and accurate representation of student learning.

The list is certainly not all-inclusive, nor is it universally agreed upon. Grading can be very personal to teachers. When their practices are questioned or scrutinized, teachers can feel attacked as though their judgment is being questioned. The concepts outlined are primarily pulled from O'Connor's work, with support from other grading experts and researchers. Most of the items in this section are referenced earlier as grade inflators or deflators with additional rationale provided.

Do not include student behaviors in grades.

"Make grades reflect only student performance in mastering the public, published learning goals of the school" (O'Connor, 2011, p.16). Whether the grade is inflated due to the reward of the proper or desired behavior or deflated by lowering a student's grade because of undesired or off-task behaviors, the experts suggest that best practices would

be to avoid factoring any form of student behavior into grading. Additionally, if teachers elect not to penalize students for their negative behaviors, they should also not inflate grades to reward positive behaviors (Dueck, 2014).

Many teachers, especially younger, inexperienced teachers are tempted to use grading as a form (sometimes their only form) of classroom management. Feldman (2019) states, "It's a tempting incentive strategy to use points to manage behavior, whether it's to get our students to complete the homework or to be respectful, particularly when it gives us immediate results" (p. 36). Using grades as a behavior management tool is a belief held by many educators, particularly secondary teachers (Brookhart et al., 2016). Unfortunately, many teachers and administrators rely on grading as a classroom management system, seeing it as a powerful and effective way to manipulate students (Carifio & Carey, 2013). As tempting as the practice may be, and despite whatever short-term results, the practice may yield, it is ultimately discouraged, and best practice would be to not incorporate it into grading policies (Guskey, 2013; Dueck, 2014).

Grades should be a representation of a student's level of academic achievement in a particular subject and not a combination of achievement and behavior. A student who talks to their peers or is out of their seat or even talks back to the teacher could still be an A-level student. Their choices in behavior should not negate their achievement. Behavior issues should be dealt with through disciplinary outlets, and not through punitive grading practices.

Do not reduce marks on work submitted late; provide support for the learner.

Teachers range wildly in the practice of late work from taking a few points off, not accepting the assignment at all, and having no consequences. The best practice would

be to "not use penalties and set up support systems that reduce or eliminate the problem of late work" (O'Connor, 2011, p. 24).

The practice is problematic for several reasons. Firm due dates assume that all students have equal access to the necessary materials and tools needed to complete the assignment- primarily, *time*. High school students are stretched with the demands of their coursework, extracurricular activities, part-time employment, and social lives. Having rigid due dates with steep penalties imposed forces the students to make choices. Those choices are often out of the students' hands. If a child needs to work a part-time job in order to help support the family, and a teacher assigns a homework assignment that takes an hour to complete that must be turned in the next day, the student is forced to choose whether to stay up very late once they come home from work in order to complete the assignment, or not do it at all and face the consequences. They may be able to complete the assignment the next evening or over the upcoming weekend and spend the necessary amount of time to complete the assignment with fidelity and learn from it what was intended by the teacher. However, then the student is penalized for having learned the material a few days late -- regardless of the reason or the level of mastery shown on the assignment.

Deadlines are important in (some) capacity, but if the goal is mastery, it is not fair. The deadlines are arbitrary (Gonzales, 2015). As Dueck (2014) points out, "getting it done does not mean learning occurred." The practice of taking points off for late work is a form of using grading as a classroom management tactic (Immewahr, 2011). Penalizing late papers is an example of motivational grading since it imposes a consequence to the student if they submit an assignment late.

When best practices are used, and teachers accept the work as students complete it, research has shown an increase in the quality of work. Teachers are often surprised that when they discontinue the practice of penalizing late assignments, students turn in more work, and their work is of a higher quality. When students are granted flexible time to complete their work, they can take ownership of their schedules and prioritize assignments, which leads to increased pride in their work and less of an incentive to copy a peer's work to meet a deadline (Feldman, 2019) simply.

Do not give points for extra credit or use bonus points.

The use of extra credit points and bonus items creates a litany of problems when trying to determine students' actual proficiency by looking at their final letter grade. It creates situations where a C student can potentially "earn" an A in the course because of pervasive extra credit acquisition. It undermines the teacher's curriculum and instruction (Feldman, 2019). While extra credit is commonly used and has a long history of use, most experts agree that the best practice would be to avoid extra credit and bonus points. "Extra credit that simply allows students to compensate for low test scores or inadequate papers is not reasonable, especially if the extra work does not help them overcome demonstrated deficiencies" (Azeem, 2010, p. 590). Instead, create real, authentic ways for a student to earn points while reinforcing or reteaching the concepts where they need assistance.

Teachers will often utilize optional "bonus questions" on quizzes or exams. These are frequently more difficult questions that rely on higher-order thinking skills for the student to be successful. O'Connor (2011) states that teachers should give these types of questions to all students as part of the assessment. Teachers can then garner the evidence

and data needed to make instructional decisions. If the question is important for students to know and be able to answer, then it should be part of the assessment and not optional.

Do not punish academic dishonesty with reduced grades.

Plagiarism, cheating, or academic misconduct are usually met with the student receiving a zero on the test or assignment. However, "this uses the grading policy as a tool to discipline students for inappropriate behavior, thus distorting student achievement" (O' Connor, 2011, p. 38). If the goal is having the student demonstrate mastery of the topic, and the student (for whatever reason) utilizes dishonest methods, then the student should be met with a consequence that is not academic in nature. The student should also still be required to complete the assignment, test, or task so that evidence of learning can be documented (Sadler, 2010).

While it can be upsetting that a student cheats on a test or assignment, awarding a zero does not solve the problem of assessing the mastery level of the student. It is essential to discover *why* the student was cheating, and then to support them through the next steps, including reteaching and reassessing with an alternate assessment. (Dueck, 2014). Some teachers would say this is more work and that the student should not be given the opportunity, but if the goal is evidence of learning, then the student needs to demonstrate what they know. Awarding a zero, despite the student's misconduct, is not an accurate indicator of student knowledge level. Teachers and administrators can and should assign other non-academic consequences.

Do not consider attendance in grade determination.

It is generally accepted that students with higher attendance tend to do better in school. However, the practice of penalizing students academically is not educationally

sound and does not result in an accurate representation of student knowledge. Absences should be reported separately from grades, and grades should be determined only from the evidence of achievement (O'Connor, 2011).

The correlation between academic success and effort, behavior, and attendance are very high (Wormeli, 2018). However, while students are likely to do better in school when they have excellent attendance, attendance should be reported separately and not used in the calculation of a grade. Many circumstances contribute to a student missing school, whether it be illness, transportation issues, suspensions, or caring for a family member. Penalizing the student academically, when they have already missed content, places them in further jeopardy and makes it difficult for them to recover. The lowered score for attendance is also an unsound practice because whether students were there for an entire unit should not impact their score if they can demonstrate mastery despite being absent.

Do not include group scores in grades.

Too often, teachers incorporate a group score for a project where all members earn the same grade. They do this as a perceived method of fairness to the students, and to motivate and hold members accountable for their work. Group grades are the amalgamation of multiple factors and cannot be a valid representation of a single student's learning. There is no way to know which group member contributed what, if anything, to the final process or ascertain who did the deep thinking and heavy lifting of the project and who (potentially) sat and did nothing. "Grades are broken when they include group scores from work done in cooperative learning groups" (O'Connor, 2011, p. 52).

A common rationale and justification for assigning group work are to prepare students for the real world and to teach collaborative skills. Teachers are often quick to point out that students will need to work together with their peers to solve problems the way that they will in the future. There may be some truth to that. However, as Kagan (1995) points out, "group grades don't necessarily foster social skills. And even if they did, there are far more efficient ways to accomplish this goal. Group grades tell us nothing reliable about individual performance" (p. 68).

Feldman (2019) suggests that there may be harm done from group work and group grading. There are complex interpersonal dynamics at play when students collaborate. While it may seem logical that the more advanced students will elevate the weaker student's performance, that is not the case. They are more apt just to do more work because they do not want the lower-performing student to bring down their grade. This can cause resentment toward their peers and the teacher. Group grades also take away individual accountability and may allow students to do far less work. All of these factors can result in inaccurate grades. When the use of group grades is prominent, then it is difficult for a parent, scholarship committee, admissions officer, or potential employer to interpret grades when they partially reflect the work of other students (Kagan, 1995).

Do not assign grades based on a student's achievement compared to other students.

Grades should be based on preset achievement standards --to be criterion-referenced, not norm-referenced in assigning grades. It should be possible for all students to earn an A, or for all students to earn an F. There should not be a predetermined bell curve forcing students into categories against their peers. (O'Connor, 2011).

When grading is fair and comprised of well-designed assessments, students should be graded for their level of achievement of the predetermined standards. They should not be scored against their peers, and their grades should not be contingent on how well other students did. The practice of allocating a certain number of each grade for the class and then only allowing that number to earn it is unsound. For example, a teacher might arbitrarily decide that the top five students in each class would earn an A. This is problematic for two reasons; first, the students who happen to earn the top scores may not achieve A-level mastery; and, second, the students who score beyond the cut-off of the top five may have demonstrated mastery that is at the A-level, but be given a grade that is not indicative of their performance and is lower as a result of a flawed grading practice.

Do not include zeros in grade determination when evidence is missing.

The use of zeros in grading is controversial for several reasons. As O'Connor (2011) points out, "zeros give a numerical value to something that has never been assessed, and therefore has no basis in reality, they can have a counterproductive effect on student motivation, and they involve inappropriate mathematics" (p. 96).

Arguments can certainly be made that students should not receive credit for nothing. That belief pushes back against the idea of assigning a 50 percent instead of a zero. The concept is not awarding a student points for not completing any work; it is instead acknowledging that an assignment not turned in is not actual proof that there was no learning. Also, the use of a zero has an enormous mathematical impact on an overall grade. Using a 50 to represent a missing piece of evidence aligns with the traditional ten-point grade scale giving each grade level an equal range instead of having sixty levels of failure.

While teachers may view a zero as a motivator to students and parents to complete and submit the work, Dueck (2014) calls it the "academic death penalty." Guskey and Brookhart (2019) refer to the practice (when a missed assignment is calculated as a zero) as "a particularly egregious example of how grading is unfair is-- because the zero assumes no knowledge of proficiency and unduly skews the overall score in a negative direction" (p. 105). Since grading should be a measure of academic learning, best practice would be not to use a zero, but rather to use a 50 so that it bears an equal impact on a ten-point grading scale. Ideally, the student would be held accountable to complete all assignments and be given an incomplete until there is adequate evidence of learning.

Do not use information from formative assessments & homework and practice to determine grades.

A growing number of researchers support the claim that grades should come strictly from well-designed summative assessments or methods of demonstrating students' content mastery. The assignments and activities that help them reach mastery should not count against them as part of the process. What is ultimately important is where the student ended up, not the route they took. For example, would it be fair for an Olympic swimmer's daily practice times to be factored into their final race time? The practice activities are designed to stimulate thinking and help students learn and grow throughout the process. Best practice would be not to have them count as part of the final grade.

Learning is a process in which learners increase their knowledge, understanding, and skills as a result of effort, instruction feedback, and self-assessment and

adjustment. For this process to work well, it is important and worthwhile to try and that it is acceptable to take risks and make mistakes; it is not necessary to always get it the first time. (O'Connor, 2011, p. 107)

In order for students to take ownership of their learning, teachers need to realize that learning is not free of mistakes. Students will undoubtedly make errors in the learning process. Teachers need to instill the idea that mistakes are an acceptable part of the process and remove the practice of penalizing students during that process. Homework should be a place of practice and experimentation without the fear of failure. It can be graded for completion, but not penalized for incorrect answers when done with fidelity. Through checks for understanding in the learning process, students will be better prepared for summative assessments. Those scores will better reflect the student's mastery without being diluted by low marks in the learning process (Canady, 2017; Feldman, 2019; Vatterott, 2015).

Additional research questions whether homework is even beneficial to students and questions the level of impact it has on student learning. In a study by Eren and Henderson (2011), they found that the only subject where homework practice improved student achievement was mathematics. In the same study, they found that how the teachers treated the homework (whether it was graded) did not affect the amount of completed work done by the students.

Do not include non-academic entries or participation points.

Grading should represent a student's level of academic achievement and not be a combination of other factors. Therefore, items that have nothing to do with academic merit should not be graded or counted against the student. A prime example of this

practice is assigning points to the completion of a parent-signed syllabus. While some educators may argue that this is a method to ensure that students are communicating the course's requirements and expectations to their parents in order to help create a support system for the student, it is not a sound practice.

For one, it is inherently biased toward students who do not have a support system at home. Some students may not see their parents for days at a time because of a variety of situations, or have a parent who, for whatever reason, refuses to sign the syllabus. Secondly, and importantly, awarding students points for bringing back a signed piece of paper does not measure any learning and is not tied to any standard. It is subjectively grading a student's responsibility of bringing back a piece of paper, and the practice should be avoided. This would also be an example of grading behaviors as opposed to grading learning.

Additional examples of non-academic grading policies that should be avoided include having required materials, points for dressing out appropriately (in physical education classes, for example), and participation points. While it is undoubtedly necessary that a student has a pencil in class to actively take notes or complete practice problems, their mastery of the practice problems is not directly connected to the pencil (Brookhart, 2019; Feldman, 2019; Randall, 2010).

Conclusion

There is still debate about many aspects of grading. The practices vary heavily from school to school despite best efforts and intentions from teachers and school leaders to ensure their validity. However, there are some grading practices that a large number of grading experts have supported. While the list is certainly not all-inclusive, nor is it

universally accepted and agreed upon, it is a starting point in the discussion and movement for a universal grading system. As agreed upon by the experts, the school improvement committee incorporated these best practices, and designed the scoresheet and evaluation tool that will be used for the purpose of this study.

- Do not include student behaviors in grades
- Do not reduce marks on work submitted late
- Do not give points for extra credit or use bonus points
- Do not punish academic dishonesty with reduced grades
- Do not consider attendance in grade determination
- Do not include group scores in grades
- Do not assign grades based on a student's achievement compared to other students
- Do not include zeros in grade determination when evidence is missing
- Do not use information from formative assessments and practice to determine grades
- Do not use information from homework to determine grades
- Do not include non-academic entries
- Do not use participation points.

Chapter 3: Methods

Grades, regardless of their origin or initial intent, have become an incredibly important part of a high school student's life. They carry with them many external and internal implications, including self-worth and adolescent identity; students use them to compare themselves to peers and to their perception of what is average. Additionally, they are used for college admittance, access to scholarships and specialty programs, and overall opportunities afforded the student based on their grades. Because of the ever-expanding importance of grades, grading practices should be heavily scrutinized until we can guarantee their validity and provide equity for all students. Therefore, this quantitative explanatory study aims to determine the impact that teacher grading practices have on student GPA.

Design

This study is a quantitative modified exploratory study. There are two sets of data. The first set of data examines the correlation between a student's cumulative GPA (grade point average) and their ACT score. The data was collected over a five-year period and compared to analyze what correlation, if any, exists. The second set of data examines the use of a scoring rubric on a school-wide survey of teacher grading practices.

Research Questions

The central research question for this study is what makes a grade a grade? To answer this, the researcher addressed three specific questions:

1. What is the relationship of cumulative high school GPA at the time of graduation and that student's highest ACT score from 2015 to 2020? This will use data provided from five years of ACT scores (on a scale of 0-36), and five years of

weighted cumulative high school GPAs (on a scale of 0.0 – 5.0). It will be analyzed using a Pearson Correlation Coefficient.

2. What are the teachers' current grading practices? Does this vary by subject, course designation as AP/honors, or years of experience of the teacher? This will be measured using a teacher self-score checklist of 2019-2020 grading practices per course, with an administrator read-behind for accuracy check of every 10th course.
3. What is the relationship between GPA/ACT alignment with teacher grading practices? This research question will utilize data from questions one and two.

Participants

The study took place in a mid-sized, urban high school. The student's GPA and ACT results from the research high school from 2016 to 2020 were used. The teaching staff of the 2019-2020 school year were surveyed.

Instrument

The primary instrument was a teacher self-score survey. Each teacher was responsible to self-score their grading practices by answering a series of questions. The teachers needed to self-score because some of the areas and practices that the scorecard measured the researchers could not readily ascertain from merely viewing the final grade book. Each teacher completed an online Google survey with definitive questions about their practices and policies around grading. The Google survey format also allowed for a more accessible collection of demographic data, such as years of

experience. A small sample of the survey submissions was checked manually to verify the accuracy of teacher self-reporting. Provided below is a sample of the survey.

The scorecard assessed a score for each class taught. The researcher tabulated scores and compared them by various disaggregated subsets such as subject, core vs. elective, years of experience of the classroom teacher, etc. The scorecard consisted of twelve of the common-themed best practices within the existing research. Each class was given a score of 0-12 (with 12 being the highest or least aligned to best practices), and those scores were analyzed to evaluate patterns and correlations. The twelve research-based criteria were primarily based on O'Connor's (2011) work, specifically his "15 fixes for broken grades." Some of the criteria are not from O'Connor's work, but from other experts in the field with the rationale given in chapter two. For this study, the scorecard included; behaviors and attendance, late work, extra credit or bonus points, group projects, grades for non-academic items (for example a parent signing a class syllabus), utilization of the mean, use of zeros, formative grades used in the final calculation, participation points, homework (or practice work) graded for content, assigning a group grade for group projects, use of the curve, and penalization for academic dishonesty. Chapter two fully explained and justified each of these criteria.

Behaviors and attendance - The practice of penalizing students' through grading is one prime example of grade deflation. Too often, teachers use grading as a form of discipline instead of a more meaningful consequence that addresses the behavior. Lowering a grade due to attendance is another example of lowering the

amount of points a student receives due to something that has nothing to do with achievement or content mastery.

Late work - A common and often standard grading policy that lowers final grades is the practice of lowering an assignment grade because it is late. This practice is defended by many educators as a “real-world lesson” to teach students “real-life” consequences of their actions. However, should an A-level essay receive a C because it is turned in a week late? Is the point of the assignment for the student to produce high-quality work and grow from the experience and through the writing process, or to hit a deadline? There are times where a deadline might be realistic and gradable (for a school newspaper or yearbook, for example). However, the widespread practice of penalizing for late work often lowers a student's final grade.

Extra credit or bonus points - There is a great amount of research that supports the idea that extra credit and bonus points are rarely, if ever, a replacement for real learning. Too often, a student will scramble at the end of a term to try to recover points by doing superfluous worksheets, or their quiz score will be padded by answering a funny question about what the teacher's dog's name is. Other forms include five extra points for each box of Kleenex that the student's family provides. This teaches students to game the system and is not an accurate indicator of learning.

Group projects - This item is included because best practices, according to O'Connor and others, discourage the use of group project grades in a student's final marks. The group grade should not count for all of the students because there is no fair way to measure each student's actual contribution toward the project. Each student

should be graded on individual submissions of work and tangible contributions to the project, and students should each be graded separately and individually.

Assigning points for non-academic items - Examples of this practice would be assigning points for a signed syllabus. The points do not reflect anything academic or any learning. This category also includes points for showing up to some event. For example, attending an evening concert, or showing up for a movie night for a US History class. There is no concrete evidence that anything was learned, and the points are awarded for compliance passively.

Utilization of the mean - Several top grading researchers suggest that the use of the mean is not the best practice. O'Connor (2011) suggests using other measures of central tendency.

Use of zeros - When a teacher assigns a zero for a missing assignment, the mathematical impact on the final grade is massive. Furthermore, is a zero really saying that the student learned nothing for that standard, or are they being punished for not completing a task? The alternative would be to assign a 50 in place of the zero so that it still reflects an 'F' but is more aligned with the other grade increments and is a better representation. An even better practice would be to hold the student accountable and keep after them until they complete all work.

Formative grades used in the final calculation - The entire purpose of formative grades is for the student to try, and to get feedback and to improve. This process should not be included in their final marks if the grade is truly about mastery and not about the journey.

Participation points - Performance-based classes often see the practice of participation points. The teacher will list participation points for the week and mark that the student received them. However, unless there is a daily or weekly rubric provided to the student with a clear criterion of what the students must demonstrate each class or week, the points are arbitrary and not evidence of learning. Five points for dressing out for gym class. Five points for the student having their instrument ready and assembled. While these are important behaviors that are necessary so that instruction can occur and learning take place, they are not gradable as evidence of learning.

Homework (or practice work) graded for content - Many scholars attest that homework should only be for practice and not for a grade. Dweck's book *Grading Smarter Not Harder* (2014) goes into great detail as to why homework is not an accurate representation of student knowledge and content mastery. They could have copied it, a parent or friend could have completed it, or they could have Googled all of the answers.

SELF-EVALUATION OF CURRENT GRADING PRACTICES	
QUESTION	TYPE
1 Years of teaching experience	Demographic; short answer
2 Class name	Demographic; short answer
3 Is this class an AP or Honors class?	Demographic; yes/no
4 Which of the following departments would this course best fit into?	Demographic; multiple choice
5 Do you dock points for behavioral issues (i.e., talking in class, not having materials/required clothing, falling asleep, etc)?	Scorecard data; yes/no
6 Do you dock points or lower grades for work submitted late?	Scorecard data; yes/no
7 Do you award bonus points or allow for extra credit?	Scorecard data; yes/no
8 Do you punish academic dishonesty with reduced grades?	Scorecard data; yes/no
9 Does a student's attendance factor into any portion of your grading?	Scorecard data; yes/no
10 When assigning group work, do all members of the group receive the same grade?	Scorecard data; yes/no
11 Do you score tests/assignments against other students' scores? (i.e. "grading on the curve")	Scorecard data; yes/no
12 Do you use zeros (for missing work) in your final grade calculation?	Scorecard data; yes/no
13 Do you include formative assessment scores in your final grade calculation?	Scorecard data; yes/no
14 Do you include homework scores in your final grade calculation?	Scorecard data; yes/no
15 Do you assign points for non-academic entries such as a signed syllabus?	Scorecard data; yes/no
16 Do you assign participation points?	Scorecard data; yes/no

Figure 3.1 - Teacher Self-evaluation of Current Grading Practices

Chapter 4: Analysis and Findings

The purpose of the study was to examine what makes a grade a grade, and to what extent the variability of teacher grading practices impacts student GPA. What factors contribute to a grade, and to what level is that grade indicative of the level of a student's academic achievement? This study is important because the research examines the relationship between teacher grading practices and student GPA. By identifying the level of impact that the utilization of best practices in grading has on grades, a more persuasive argument can be made to implement a universal grading practice. When schools are consistently scrutinized and undergoing massive reform efforts, and students are doing whatever it takes to get into their top choices for college, it is increasingly important that students get top marks. As many colleges and universities move to test-optional admittance policies, it is imperative that student grades and GPA are an accurate and universal representation, and that grading practices be heavily scrutinized until their validity can be guaranteed and provide equity for all students. This study will help to determine whether current high school grading practices are solid indicators of students' content mastery, or if they are artificially inflated (or deflated) by other factors in teacher grading practices.

Response Rate

The school district utilized for this research study consists of one high school, and the study focused strictly on that school's practices. At the time of the study, the school employed 70 certified personnel in teaching roles. The teacher grading practices self-score survey was distributed to all 70 teachers. Of those distributed, 62 completed the survey. This resulted in a response rate of 88.5%.

Demographics of the Study Participants

The survey, distributed by the building leadership team as part of the school improvement plan, asked for four pieces of demographic information - years of teaching experience, the department in which the class being listed is part of, whether the class was offered as AP/AR (advanced placement, advanced rank), and gender. Respondents were asked to complete the survey for each class taught, because some teachers use different practices depending on the course and the level of student. For data analysis, each teacher's collective responses were averaged together to create the teacher score. The specific teacher score for each corresponding course was used when appropriate.

Analysis of Research Questions

The primary research question was what makes a grade a grade? This question will be answered with three sub-questions, each of which will be analyzed through the data acquired in the study. Each sub-question will be analyzed, and the results will all be considered when discussing the primary question.

Research Question 1. What is the relationship of cumulative high school GPA at the time of graduation and that student's highest ACT score from 2015-2020?

Results. Data was provided from the district's central office listing every student's cumulative high school GPA and the highest ACT score. Data was grouped by graduating class. Another member of the administrative team served as a secondary researcher for the gathering and disaggregation of the data. This was to protect the individual student scores from being seen by the researcher (author). The secondary researcher presented data that omitted individual student identification numbers. A Pearson Correlation test was done for each data set. A moderate positive correlation was

found for each of the five years of data. In a Pearson correlation, an absolute value of 1 indicates a perfect linear relationship. A correlation of 0 indicates no linear relationship between the variables (in this case, ACT and GPA). Table 4.1 shows the correlation.

Over the five-year span, the average Pearson correlation was .7214.

Table 4.1 Pearson Correlation of ACT Score and HSGPA

<i>Year</i>	<i>R Value</i>
2019-2020	0.7086
2018-2019	0.7022
2017-2018	0.7176
2016-2017	0.7373
2015-2016	0.7413
Five-year average	0.7214

The correlation and consistency of the Pearson score is also seen in the average ACT scores and average GPA's. Table 4.2 illustrates the yearly averages and five-year averages of each score.

Table 4.2 ACT and GPA Averages by Year

<i>n =</i>	<i>Graduation year</i>	<i>ACT average</i>	<i>GPA average</i>
224	2019-2020	19.27	2.95
263	2018-2019	18.95	3.08
237	2017-2018	18.85	3.01
227	2016-2017	19.36	2.99
204	2015-2016	18.94	2.87
	average	19.07	2.98

Examination of the data shows that the average student in this five-year window had an ACT of 19 and a cumulative GPA of approximately 3.0 (2.98). A deeper dig into the data reveals that all students who earned a 19 on the ACT were examined to see what the corresponding GPA was. There were 88 students who earned a 19. The average GPA was 3.17, which is significantly higher than the average of all students. Figure 4.1 illustrates the findings.

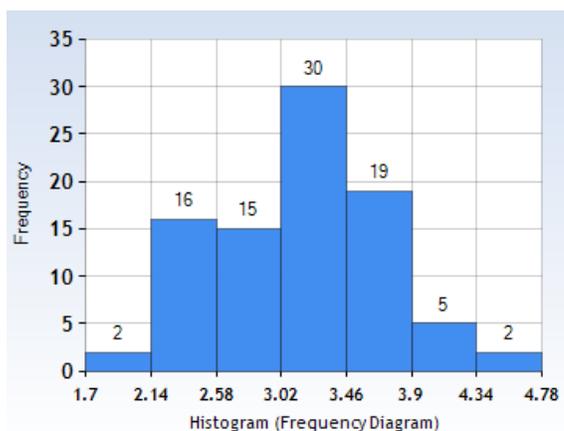


Figure 4.1 *Frequency of students who earned a 19 ACT with corresponding GPA*

ACT SCORE/ HSGPA

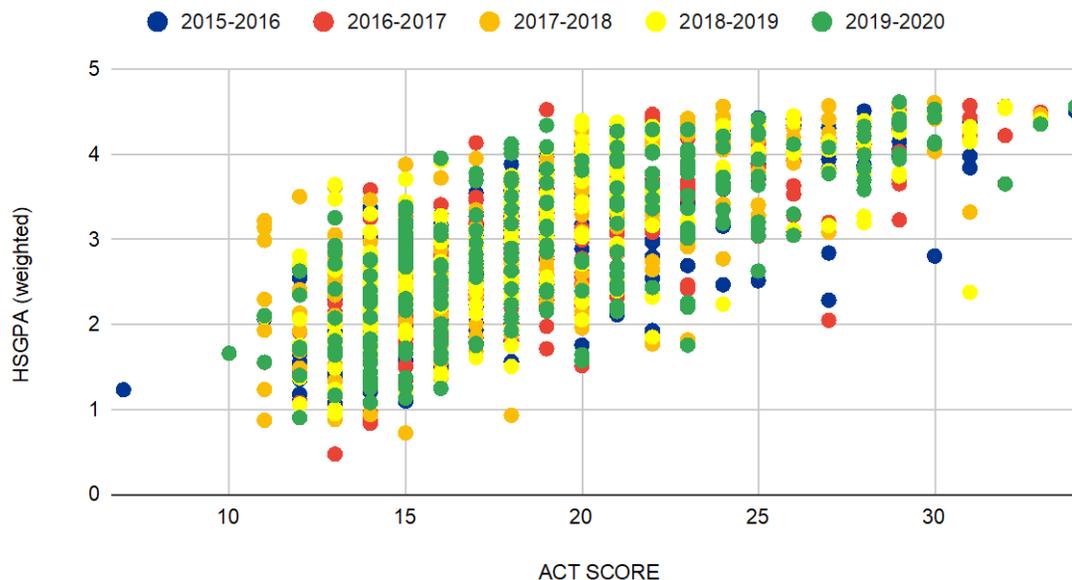


Figure 4.2 *Scatterplot of Individual Students' ACT Scores and HSGPA*

Figure 4.2 displays the trend data for the five years of ACT scores and corresponding HSGPAs. The trendline for each year is about the same. Also similar are the appearance of outliers in each quadrant. These outliers have scores that whose relationship are incongruous based on peers with similar GPAs. These outliers will be examined and discussed further in research question 3.

Research Question 2. What are the teachers' current grading practices? Does this vary by subject, course designation as AP/honors, or years of experience of the teacher?

Results. The survey asked each teacher a series of questions about current grading practices. If the teacher utilized the practice described, they answered yes and received a point. The higher the amount of points, the farther away from best practice the grading practice is. The highest possible number of points (score) is 12. A perfect score would be zero. Of the 62 teachers who completed the survey, no teacher received a 12.

One teacher scored a zero. The average score for all teachers was 6.24. This means that the average teacher acknowledged to utilizing an average of six grading practices that have been shown to artificially inflate or deflate a student's final grade.

The data was disaggregated into various subgroups. The first explored the years of experience that each teacher had and whether that had an impact on their grading score. Table 4.3 illustrates.

Table 4.3 Average Teacher Grading Score by Years of Experience

<i>Years of experience</i>	<i>Number of teachers</i>	<i>Average score</i>
1-3	10	6.95
4-10	17	6.73
11-20	22	6.2
20+	12	5.35

As the teacher gains more experience, the grading practices align more with best practices. It is interesting that the newest teachers, who have most recently completed teacher training, are the farthest from best practices. Veteran teachers who change and adapt over time, align the most with best grading practices.

Another subgroup that was looked at was the gender of the teacher. Some of the research in chapter 2 highlighted certain studies that found a teacher's gender could impact grading. Table 4.4 shows the results of teacher gender on grading practices.

Table 4.4 Average Teacher Grading Score by Gender

<i>Gender</i>	<i>Number of teachers</i>	<i>Average score</i>
Female	37	6.5
Male	25	5.86

While the male teachers had a slightly lower average score, it is not significantly different than the female teachers' average score. The next subgroup explored looks at whether there is a difference between classes listed as AP/Honors (advanced placement) and those that were not. It would be understandable that a teacher would have a higher level of rigor for an AP/Honors class and therefore have different grading policies and those scores could be different. Table 4.5 shows the results.

Table 4.5 Average Teacher Grading Score by Course Designation

<i>Course Designation</i>	<i>Teacher score average</i>
AP/Honors	6.17
Non-AP/Honors	6.3
All courses	6.24

There is virtually no difference in the grading practices of the teachers who teach AP/Honors courses and those who do not. All three averages are basically the same.

The next breakdown of subgroups looks at the different scores from each department. Another member of the administrative team served as a secondary researcher for the gathering and disaggregation of the data. This was to protect the individual teacher scores from being seen by the researcher (author). The secondary researcher presented data that omitted individual teacher names and identified departments only as “core” (A-D) or “elective” (A-B). The secondary researcher also provided all the final data for research question three so that the primary researcher did not see individual teacher scores when creating the data scores for the research question.

When looking at the individual department data, several things surface. Table 4.6 illustrates the department averages.

Table 4.6 Average Teacher Grading Score by Department

<i>Department</i>	<i>Average teacher score</i>
Core A	6.34
Core B	5.85
Core C	6.35
Core D	7.09
Elective A	6.98
Elective B	5.43

Each of the core areas' grading practice is different. Core A and Core C are statistically the same (although it's not clear as to which practice they align with, only the final average). Core D is slightly higher. What stands out is that Elective B has the lowest average. In the research presented in chapter 2, elective areas tended to most demonstrate practices that inflated grades, for example participation points and grading attendance.

A deeper dig into each of the departments' practices reveals the actual alignment of grading amongst department members. Figures 4.3 – 4.9 break down each survey question and the varied responses of the department. For ease of reading and interpreting these tables, the survey questions have been abbreviated. Table 4.7 highlights the abbreviation of each question as a reference.

Table 4.7 Survey Question Abbreviations

<i>Survey question</i>	<i>Abbreviation for figures 4.3 - 4.9</i>
Do you dock points for behavioral issues? (i.e. talking in class, not having materials/required clothing, falling asleep, etc.)	Behavior
Do you dock points for work submitted late?	Late Work
Do you allow bonus points or extra credit?	Extra Credit
Do you punish academic dishonesty with reduced grades?	Plagiarism

Does a student's attendance factor into any portion of your grading?	Attendance
When assigning group work, do all members receive the same grade?	Group work
Do you score tests/assignments against other students' scores? (i.e. "grading on the curve")	Use of the curve
Do you use zeros (for missing work) in your final grade calculation?	Use of zeros
Do you include homework scores in your final grade calculation?	Formative
Do you include formative assessment scores into your final grade calculation?	Homework
Do you assign points for non-academic entries such as a signed syllabus?	Non-academic
Do you assign participation points?	Participation

CORE SUBJECT A

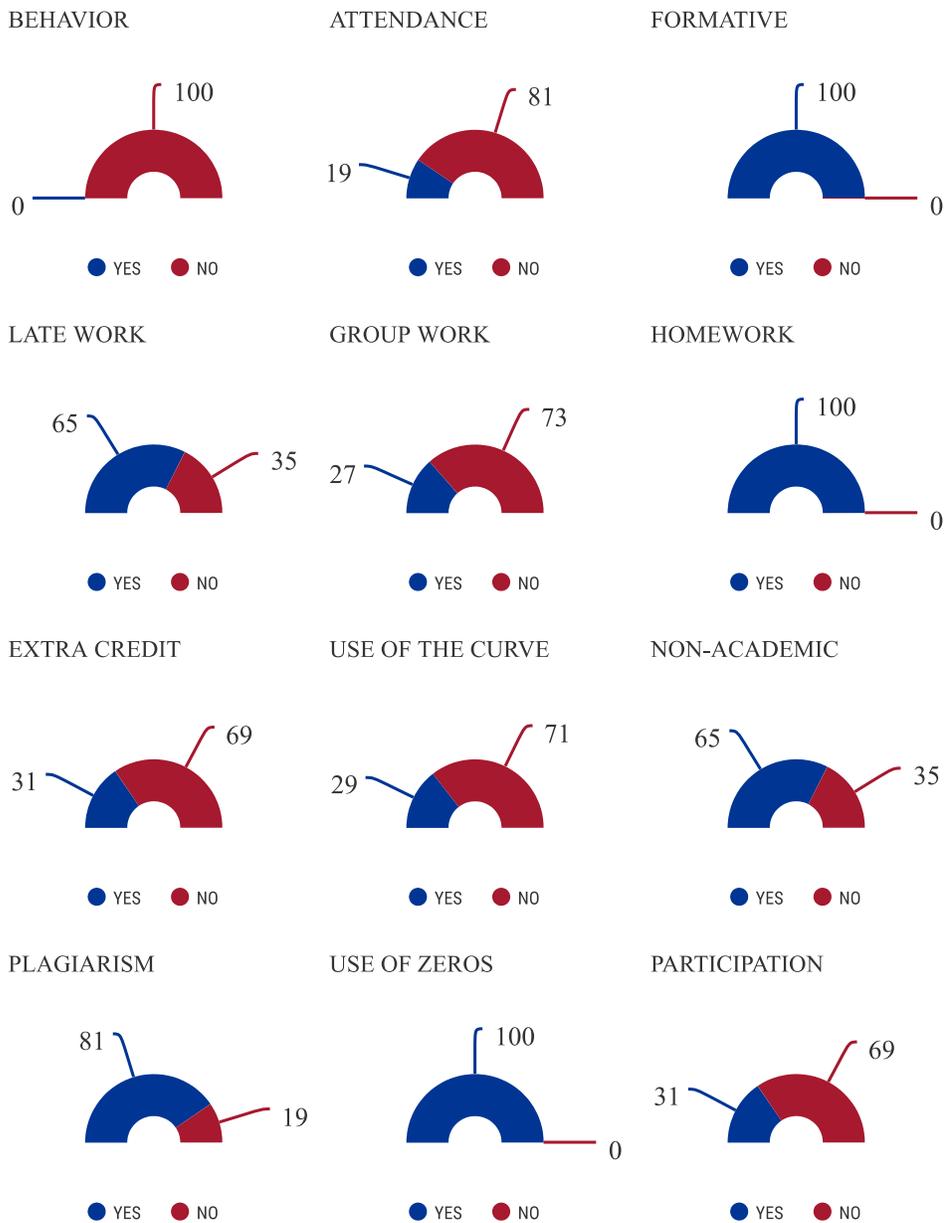


Figure 4.3 – Core Subject A

Some of the main things that stand out about Core Subject A's practices are the non-alignment in several areas. The first one being the acceptance of late work. This is an area where a student's grade could be negatively impacted because the student has one teacher as opposed to another. Two other practices that stand out are grading non-academic work and accepting extra credit. One third of the department allows for extra credit whereas the teacher across the hall may not. The same teacher may award points for a signed syllabus. Compounding the varied practices could result in the difference of a letter grade based solely on which teacher a student had and not necessarily mastery of content.

The department is completely aligned in some practices. They do not grade based on behavior. They all grade homework and use formative assessments in the final grade. They also all use zeros for missing work. While some of the practices are aligned, the majority are not.

Figures 4.4 – 4.8 outline the responses of each of the other departments and all appear below. The same themes appear in each department. There are some practices that the department agrees on, but the majority they do not. Elective B had no questions in which the teachers were in complete agreement. However, this can be somewhat expected, as it is not clear which electives make up the subgroup, and the instructors are likely different content areas, whereas a core department would tend to naturally align.

The next figure looks at individual questions and the breakdown of how individual teachers answered them. Figure 4.9 breaks down each grading practices and the percentage of teachers who utilized it.

CORE SUBJECT B

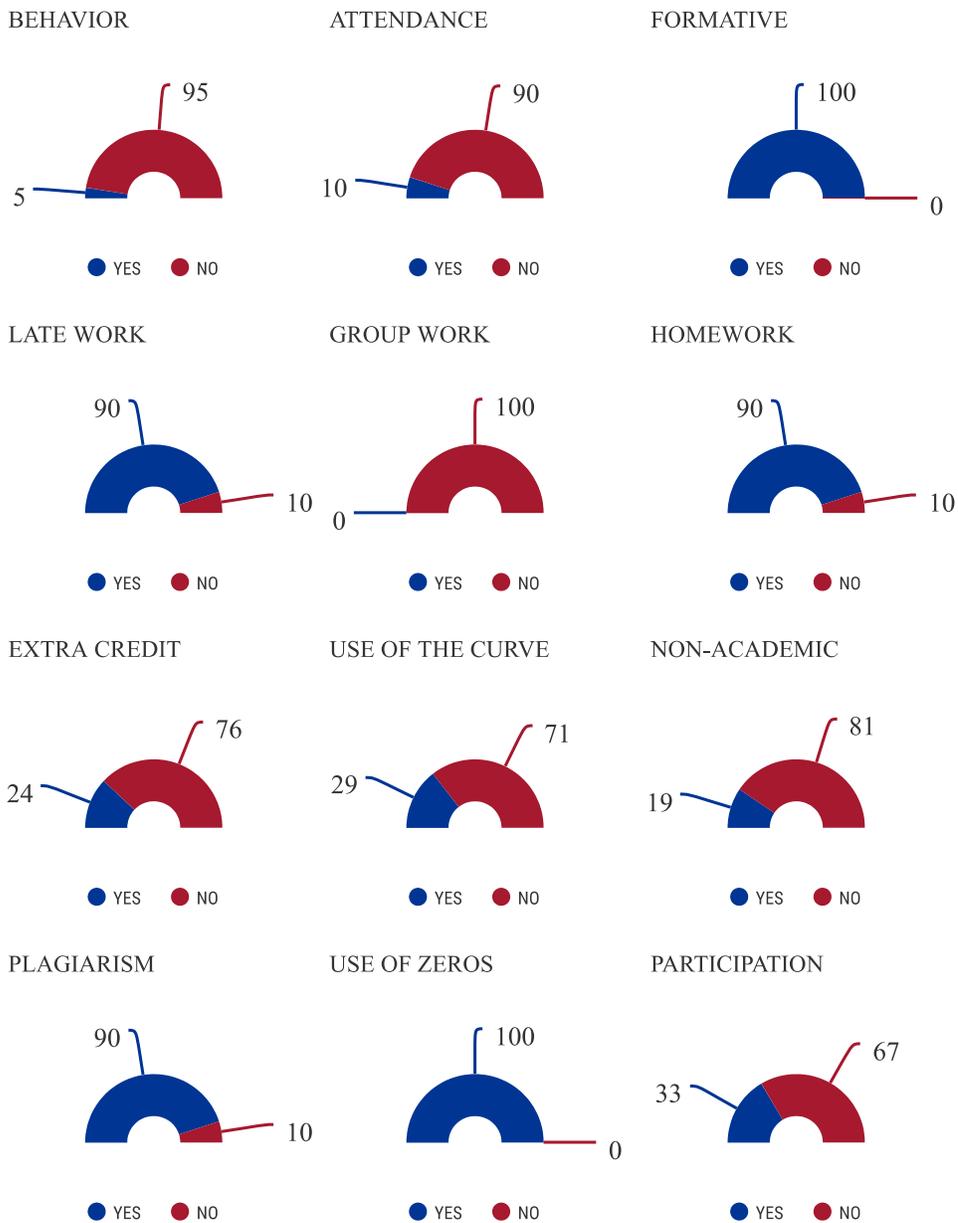


Figure 4.4 – Core Subject B

CORE SUBJECT C

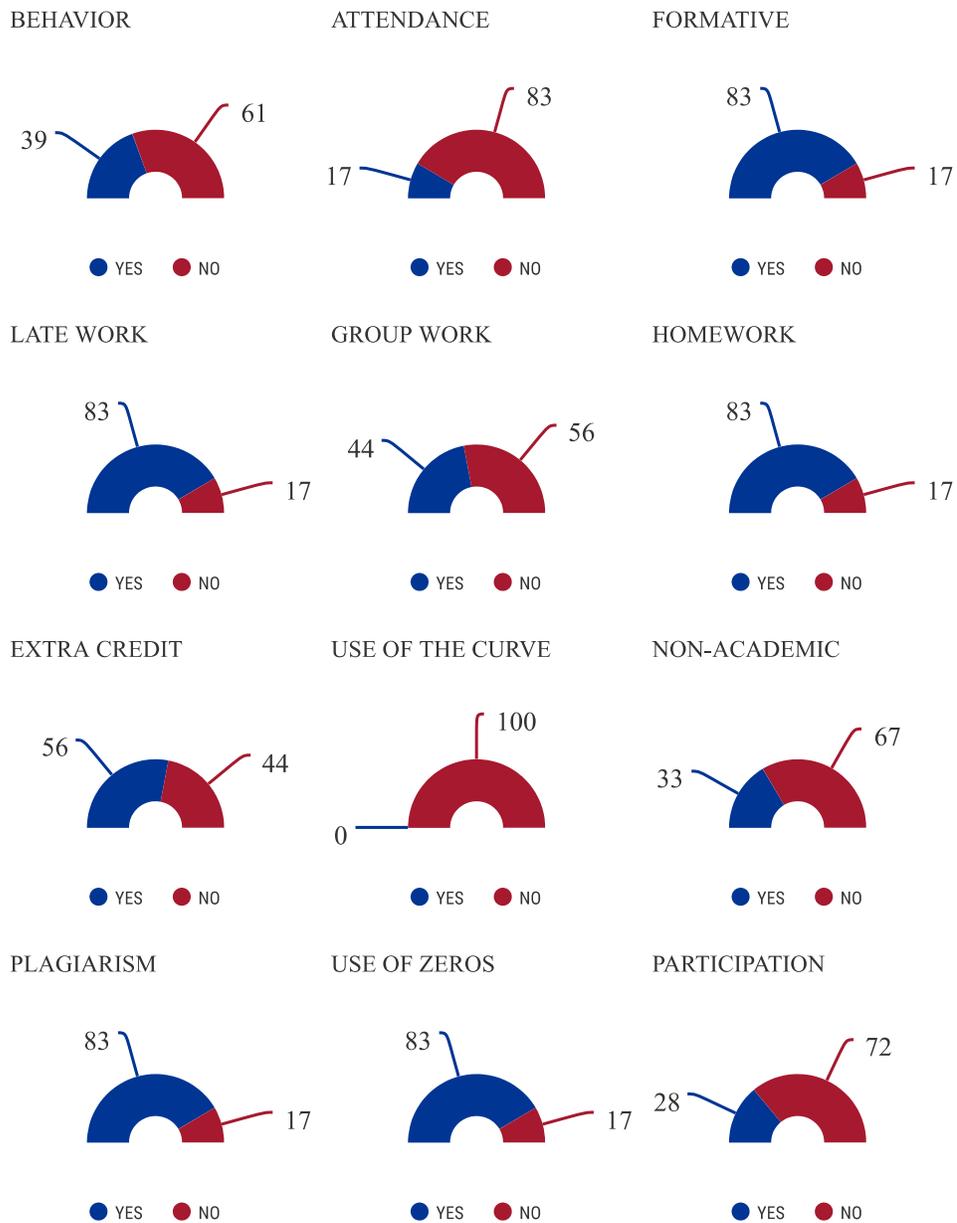


Figure 4.5 - Core Subject C

CORE SUBJECT D

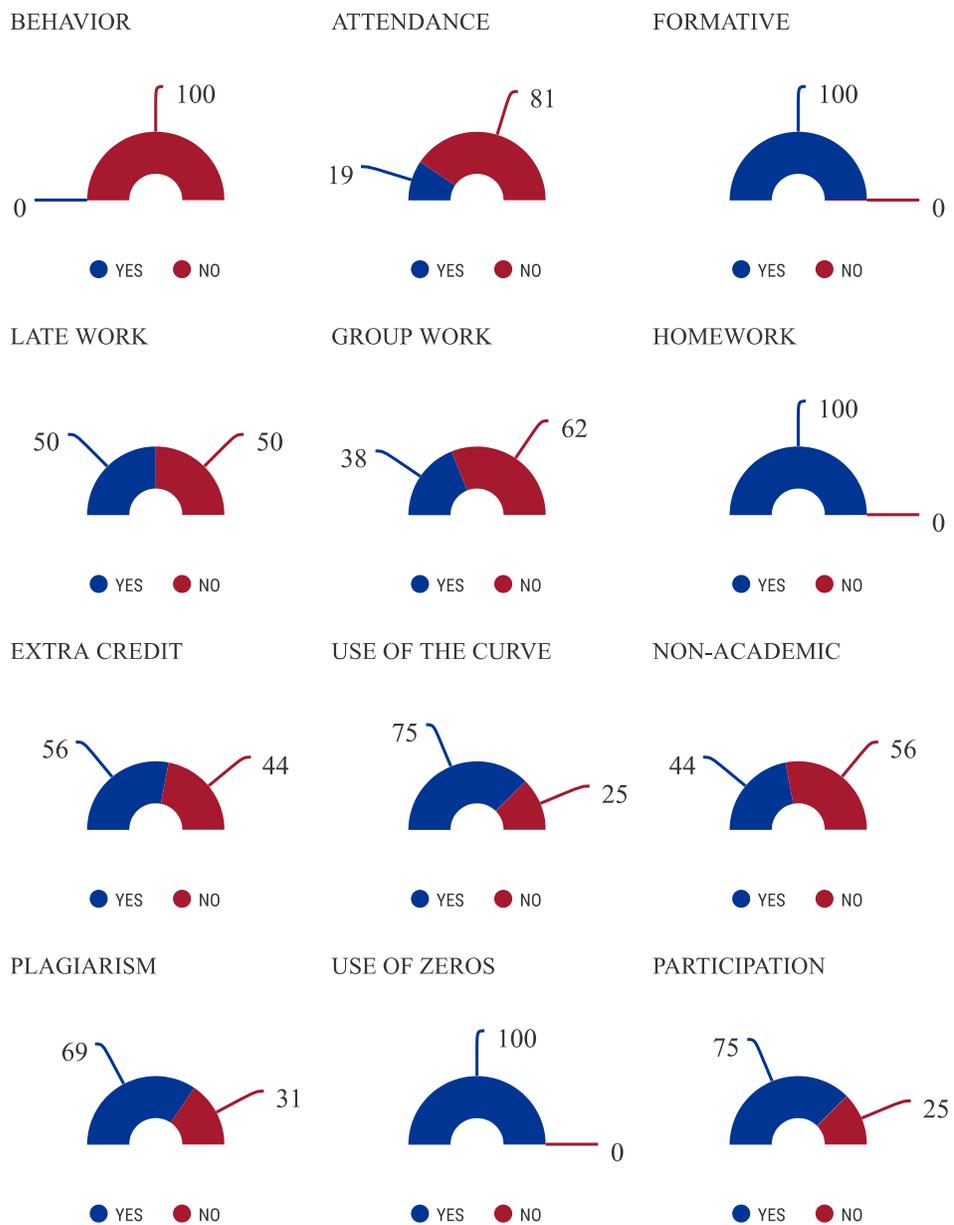


Figure 4.6 - Core Subject D

ELECTIVE A

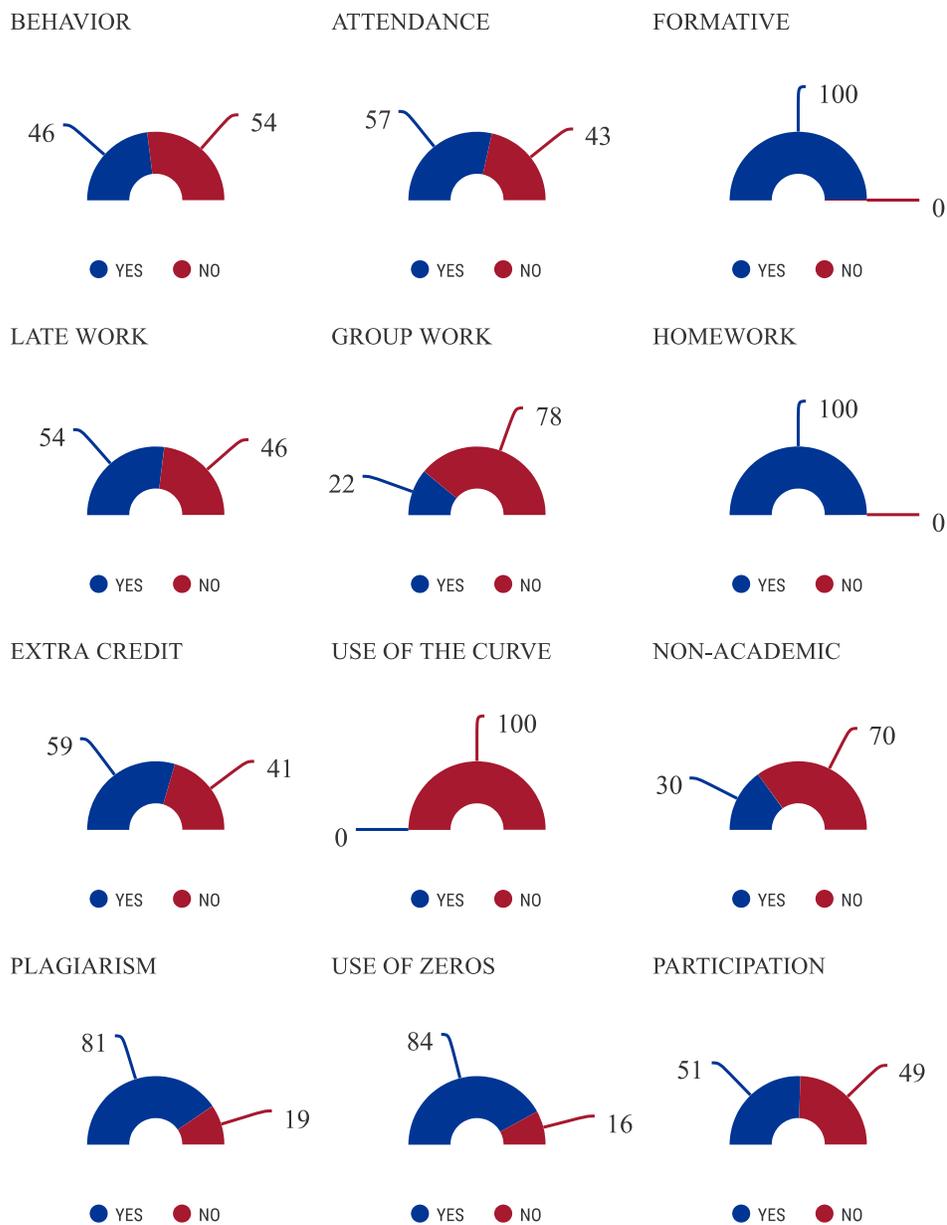


Figure 4.7 – Elective A

ELECTIVE B

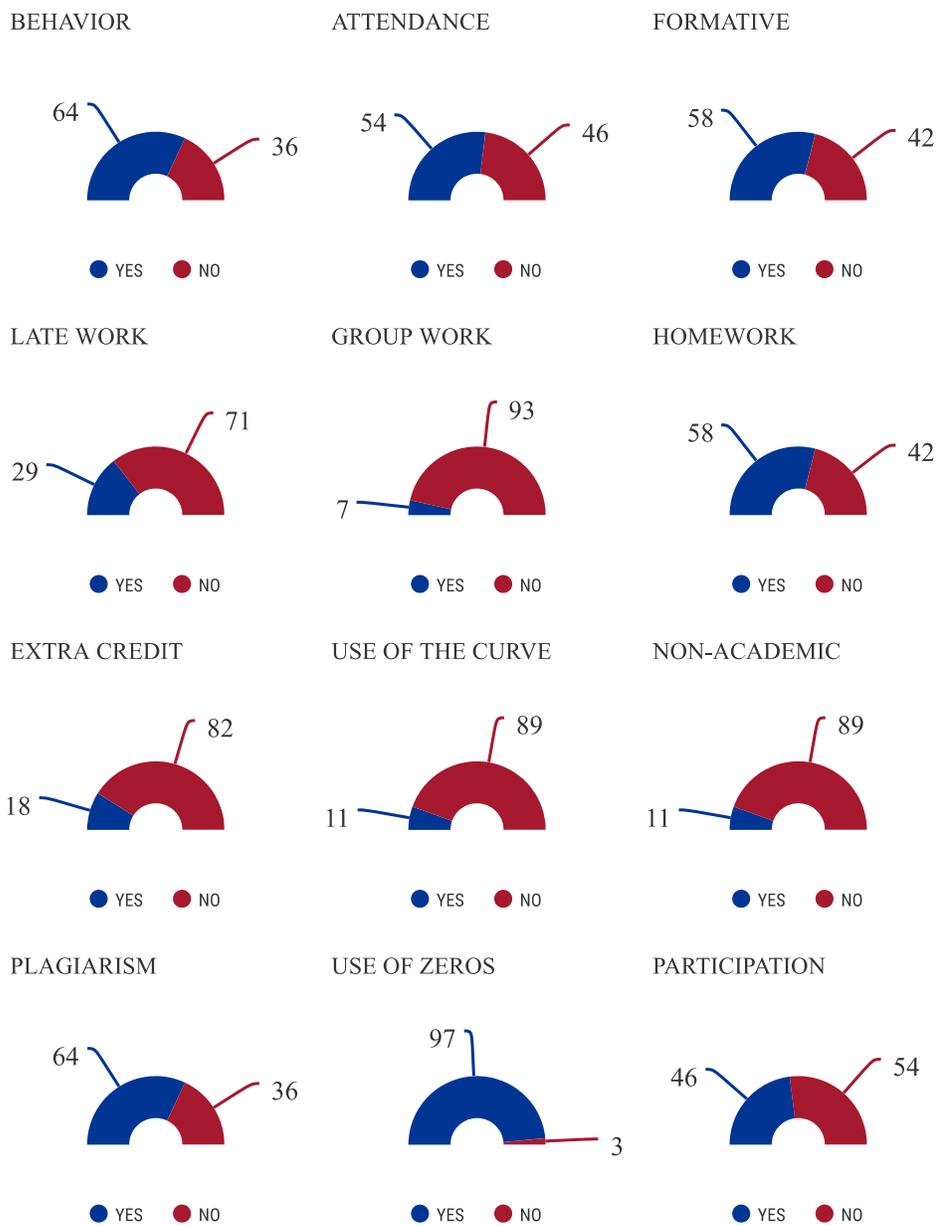


Figure 4.8 – Elective B

SINGLE TEACHER RESPONSE

n = 62 teachers

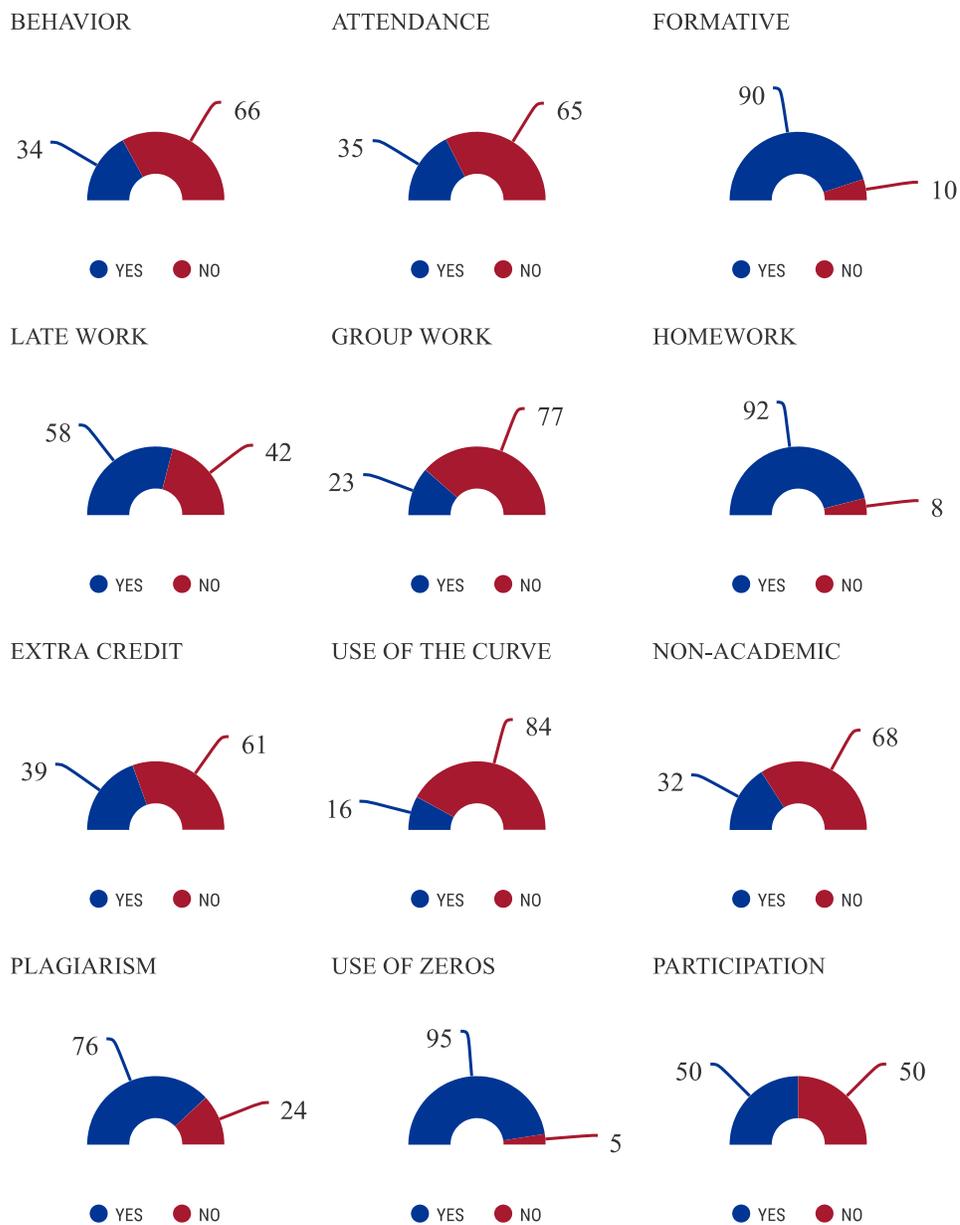


Figure 4.9 – Single Teacher Response

Figure 4.9 displays the percentage of teachers who answered each question in the survey. There is not a single practice that all the teachers agree upon. A few have a large majority; 95% percent of teachers use zeros for missing work, 90% include formative assessment in the final grades, 92% include homework in the final grades and 76% of teachers penalize academic dishonesty or plagiarism with a lowered grade. These are large majorities of teachers, all using what experts state, are not best practices.

There are areas where the majority of teachers were in line with best practices. For example, 84% of teachers do not grade students compared to other students, 77% of teachers do not give the same grade to all members of the group, and 66% of teachers do not penalize students academically for behavioral issues. Practices with a fairly even split include penalizing students for late work, and the use of participation points.

The next table (4.8) illustrates the breakdown of each department's responses side by side in one table. When examining and comparing the departments, definite differences in their grading practices emerge. The table mostly aligns with the results from the previous table. Overall, there are few areas where an entire department agrees on a particular practice. Four of six departments agree entirely that formative work should count toward a student's final grade, despite experts' opinion that it is not a best practice. With twelve grading practice categories, and six department groups, there are 72 opportunities where a department could be completely aligned on a particular practice. In this study, departments only aligned 15 times, or 21%. Which means that 79% of the time, teachers are not in alignment with peers.

Table 4.8 Breakdown of Responses by Department

		Core A		Core B		Core C		Core D		Elective A		Elective B	
		#	%	#	%	#	%	#	%	#	%	#	%
Behavior	Y	0	0	1	5	7	39	0	0	17	46	18	64
	N	26	100	20	95	11	61	16	100	20	54	10	36
Late work	Y	17	65	19	90	15	83	8	50	20	54	8	29
	N	9	35	2	10	3	17	8	50	17	46	20	71
Extra credit	Y	8	31	5	24	10	56	9	56	22	59	5	18
	N	18	69	16	76	8	44	7	44	15	41	23	82
Plagiarism	Y	21	81	19	90	15	83	11	69	30	81	18	64
	N	5	19	2	10	3	17	5	31	7	19	10	36
Attendance	Y	5	19	2	10	3	17	3	19	21	57	15	54
	N	21	81	19	90	15	83	13	81	16	43	13	46
Group work	Y	7	27	0	0	8	44	6	38	8	22	2	7
	N	19	73	21	100	10	56	10	62	29	78	26	93
Curve	Y	3	12	6	29	0	0	12	75	0	0	3	11
	N	23	88	15	71	18	100	4	25	37	100	25	89
Zeros	Y	26	100	21	100	15	83	16	100	31	84	37	97
	N	0	0	0	0	3	17	0	0	6	16	1	3
Formative	Y	26	100	21	100	15	83	16	100	37	100	22	58
	N	0	0	0	0	3	17	0	0	0	0	6	42
Homework	Y	26	100	19	90	15	83	16	100	37	100	22	58
	N	0	0	2	10	3	17	0	0	0	0	6	42
Non-academic	Y	17	65	4	19	6	33	7	44	11	30	3	11
	N	9	35	17	81	12	67	9	56	26	70	25	89
Participation	Y	8	31	7	33	5	28	12	75	19	51	13	46
	N	18	69	14	67	13	72	4	25	18	49	15	54

Research Question 3. What is the relationship between GPA/ACT alignment with teacher grading practices?

Results. This research question uses the combined data from RQ1 and RQ2 to determine the impact of teacher grading practices on student GPA. Outliers were identified from the data set in RQ1. That is, students who had a high GPA, but a lower than expected ACT score, and students who had a high ACT score, but a lower than expected GPA. A subset of students was also pulled who had an average ACT score and an average GPA.

ACT SCORE/ HSGPA

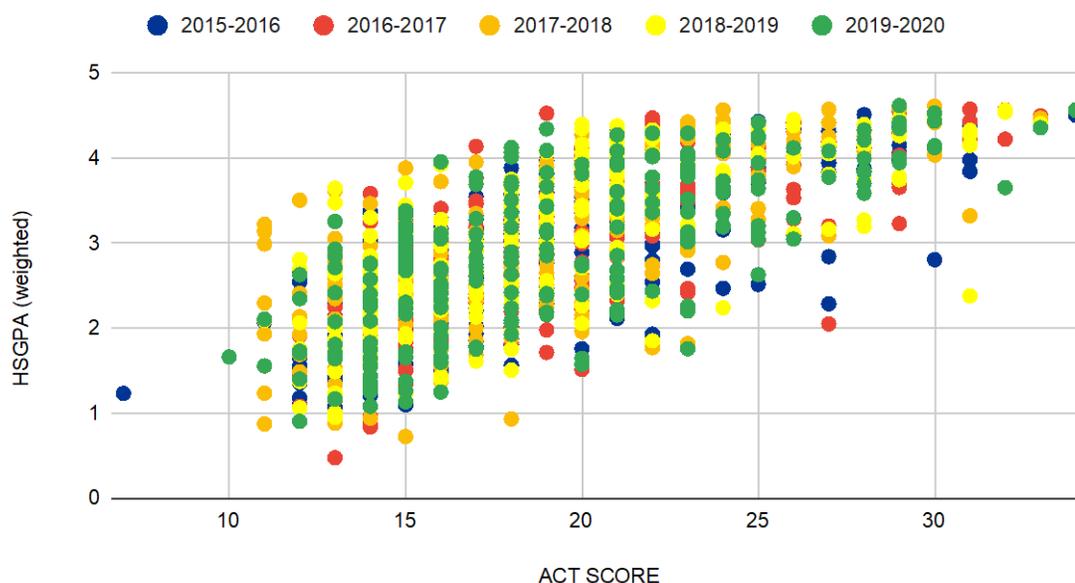


Figure 4.10 *Scatterplot of Individual Students' ACT Scores and HSGPA*

Figure 4.10 displays again the findings of research question one. For this portion of the study, subgroups of outliers were identified. For each student in the subgroup, high school transcripts were evaluated. The secondary researcher analyzed each transcript and identified all core subject area courses that were taught by a teacher who took the 2020

survey. The teacher grading scores for each teacher were averaged together to produce the average score for all core teachers the student had. This was done to determine if the outliers had teachers who utilized significantly different grading practices. Figure 4.11 illustrates a significant outlier from the normal trend. This student earned a cumulative HSGPA of 4.531 and scored a 19 on the ACT.

ACT SCORE/ HSGPA

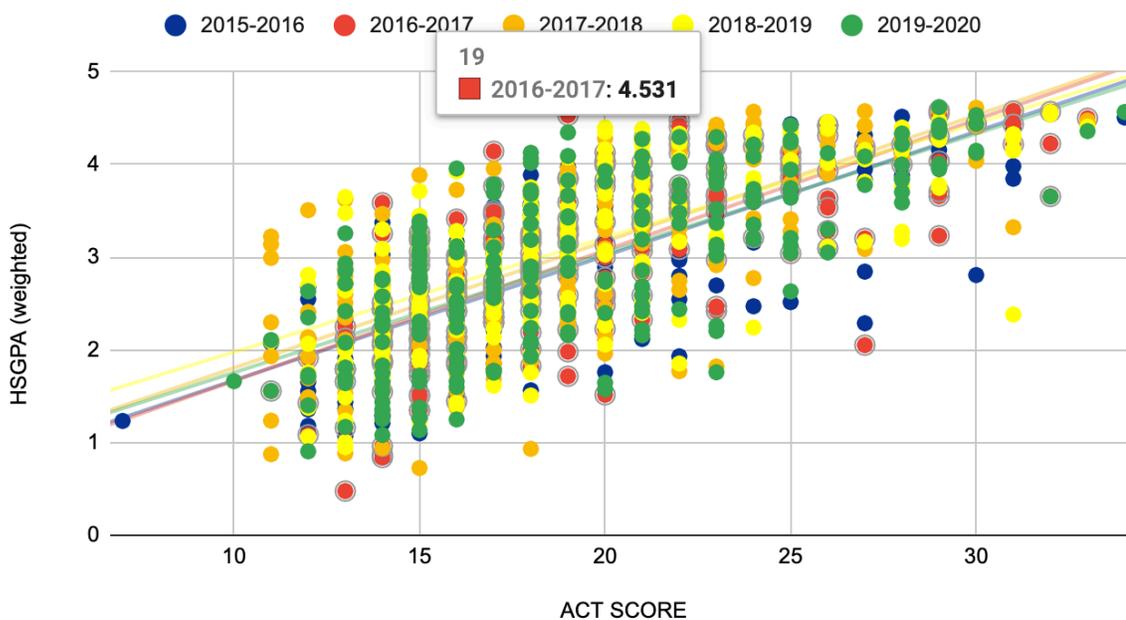


Figure 4.11 Identified ACT Score Outlier

In order to better understand why this incongruence exists, a sample of other students with similar GPA/ACT comparisons were pulled. Each of these students' transcripts were analyzed to determine an overall teacher grading score. This was done to determine whether students whose HSGPA was significantly higher than peers with the same ACT score had teachers with significantly higher teacher grading scores. However, as is seen in table 4.9, that was not the case.

Table 4.9 Average Teacher Grading Score by Student Subgroup

<i>Subgroup</i>	<i>Average score</i>
Average teacher score	6.29
High GPA (4.0-4.5); Low ACT (18-21)	6.75
Low GPA (2.8-3.0); High ACT (30-32)	6.73
Avg GPA (3.0-3.1); Avg ACT (19)	6.64

The data determined that the individual grading practices and adherence to best practices had no impact on the student's final GPA. In fact, all three subgroups' averages were only slightly above average.

Some of the limitations of the study that could have impacted the data are the small size of the school where the study took place. Most of the students in the subgroups tended to have the same teachers appear on transcripts, therefore yielding similar scores. The study was also limited to the teachers who were teaching in the 2019-2020 school year and who completed the survey. The average teacher score for the student transcript looked at a four-year period and had to omit many courses from being included in the average because that teacher was not part of the study.

Chapter 5: Conclusions and Discussions

Grades are perhaps the most widely used and universally accepted representation of a student's academic achievement and are a vehicle to higher education, specialty programs, and highly competitive scholarship dollars. The research shows that the components contributing to a student's final grade can vary greatly and that grading is subject to factors that can create incongruency between a student's level of content mastery and the final grade.

The goal of this study was to determine the extent of the impact that teacher grading practices had on a student's final GPA. While there are obvious conclusions that can be drawn between grading and grades, this study aimed to look deeper at the grading practices themselves and how the use inflates or deflates a final grade. While a direct correlation between adherence to best grading practices and student achievement could not be drawn, the study did reveal that grading practices in the study group were largely unaligned within departments, and even more so schoolwide.

Review of the Literature

The purpose and intent of grading have evolved over the years, but experts still believe there is room for improvement. "Letter grades do several different things, none of them well, and the result undermines students' learning" (Schneider, para. 13). While the study determined that the grading practices varied widely throughout the sample building, that is not an uncommon situation or unusual for school environments. Despite grading practice variance, even within departments, high school grades are still the best indicator of student success in college. Substantial research shows that high school grade point averages are stronger predictors than test scores of college outcomes (Allensworth &

Clark, 2020). High school grade point averages are five times stronger than ACT scores at predicting who will graduate from college (Jaschik, 2020). This could be due to a variety of factors. HSGPAs are comprised of data from a four-year span where students are assessed by many different teachers and grading systems and measured by different metrics and criterion for each course and subject. Conversely, the ACT can be viewed as a mere snapshot and a single lens view of a student's ability.

Research Design

This study was conducted using quantitative methods to analyze data generated through the test building's school improvement cycle. To shield individual teachers from being identified by the researcher, a secondary researcher assisted in preparing data for analysis. The author was provided with data sets where all individual names were omitted, and teacher groups were disaggregated in a manner where departments and subsets were still unidentifiable. The subject school's teaching staff was given a 16-question survey comprised of four demographic questions, and 12 questions about the individual grading practices. The grading questions were all yes/no responses. A response of no meant that the teacher aligned with best practices in grading. The lowest possible score was zero, and the highest possible score was 12.

Conclusions

Research Question 1. *What is the relationship of cumulative high school GPA at the time of graduation and that student's highest ACT score from 2015-2020?*

When looking at this correlation, it is important to acknowledge that ACT is a norm-referenced test, while grading and therefore, a GPA, is criterion-referenced. There

is an inherent mismatch. GPA's have the ability to rise year to year from a variety of factors.

The first ACT was administered in November 1959 to measure how well students would perform academically in the first year of college (ACT Website). Since that time, the test has grown in popularity and increased dramatically in the number of students who take it each year. With that increase also came the increase in the number of students who earn a perfect score. The percentage of students earning a perfect score (36) on the ACT is 17 times what it was just 20 years ago and continues to rise. ACT attributes this more to students gaming the test and prepping specifically for the test, as well as the number of students increasing who take the test (Findlay, 2017). Both of those are likely factors to the increase in perfect scores. As the number of states who adopt the ACT as a state test increases, the testing pool also increases. The high-stakes nature of the test increases the measures to which students and families will go to earn top marks, thus ushering in test-prep programs and students taking the test over and over again to achieve a perfect score.

The five-year data showed a consistent correlation between the students' GPA and ACT score. This is not surprising because it is logical that as students do better in school, corresponding test scores would reflect an increase. The study aimed to look at specific outliers who did not match that correlation. For this research sub-question, the class average was examined and reflected what was expected.

Research Question 2. *What are the teachers' current grading practices? Does this vary by subject, course designation as AP/honors, or years of experience of the teacher?*

The teacher survey data revealed that inconsistencies are evident in the teacher grading practices across the building and within departments. Veteran teachers aligned

more with best practices than newer, less experienced teachers. After analyzing the responses to grading practices, it is worth examining if it is not the *grading* that has the greatest impact, but rather the teaching *practices* in the classroom. The United States Department of Education conducted a study in which researchers examined 32 instructional practices and determined the four practices that had the most significant impact on student achievement -- fostering student engagement, having students participate in discussions, having fewer class period disruptions (classroom management), and developing a classroom climate conducive to instruction (Schaffhauser, 2019). While grading has its importance, it is worth acknowledging that the method and quality of instruction have a high impact.

Research Question 3. *What is the relationship between GPA/ACT alignment with teacher grading practices?*

The results of this question were the most surprising. Something is contributing to the incongruency of the subset of students who earn high GPAs, yet the performance on the ACT is well below the average of peers with the same GPA. The argument is often made that grades are "inflated" or "subjective" because of evidence that HSGPAs have increased over time, yet the same change has not been seen in test scores. However, each of those things measures different aspects of achievement. Therefore, one component can change without a change in the other (Allensworth & Clark, 2020).

The study determined that the grading practices of the teachers did not appear to factor into the incongruence. This raises a new question of whether grades were inflated, or were accurate representations of the skills of school, and not content mastery. While the ACT measures specific skill sets, a HSGPA is still a significant measurement of

learning of a different type of skills. "High school grades are powerful tools for gauging students' readiness for college, regardless of which high school attends, while ACT scores are not. GPAs measure a very wide variety of skills and behaviors that are needed for success in college, where students will encounter widely varying content and expectations" (Jaschik, para. 4). GPAs may be inflated by giving value and points for non-academic or curricular traits, but those traits and skills are beneficial overall to the student. While attendance, turning work in on time, and other soft-skills may not be a fair representation of content mastery, they *are* skills that will aid students for the rigor and demands of a college schedule, therefore being a solid predictor of college readiness.

Considerations for Further Research

The data for this study is inconclusive about the impact that teacher grading practices have on student GPA. However, the results of the study highlight the overall variance in grading practices, even within departments. If grading is to be universally understood and interpreted by all stakeholders, and if all students are to be given the same opportunities when it comes to policies regarding grading, it is imperative to align grading practices. When all teachers in a building adopt and adhere to the same, clearly identified, and published grading standards, stakeholders can understand and are not left to guess and interpret the meaning of a grade based on which teacher assigns it.

An additional study should be done focusing on the characteristics and demeanor of *students* and how it impacts the teacher's perception of them, and thereby impacting how the teacher grades them. Students who "play school" (as discussed in chapter two) tend to please the teacher. The teacher may subconsciously or even consciously grade them easier than other students because the teacher appreciates the positive behavior and

being a model student. It could also be examined in light of the demographic of the student. When disaggregating the data for this study, the secondary researcher noted that the majority of the students in the outlier group for research question three were Hispanic females. This raises additional questions and areas for potential further research. Are these students not performing well on the ACT because of test bias? Are the teachers grading them more leniently because they are rooting for them to do well and because they demonstrate all of the positive skills mentioned previously for the students who "play school?" The subjective element of grading is always present, and additional research is necessary to examine how student traits (demeanor, rapport with teachers) factor into it.

COVID-19 and Grading

COVID-19 most certainly impacted the educational system and will continue to have an impact long after the pandemic passes. Not only did COVID expose inequities between school districts across the nation, but it also brought to the forefront some questions about grading. When schools operated remotely, many schools switched to a pass/fail model or a standards-based grading model (outlined in detail in chapter 2). This raises questions about whether those methods are appropriate all the time. If educational leaders felt that switching to standards-based grading or a pass-fail model was the best option to measure learning when it was done remotely, why is it not the best option when students are present in the classroom? What differences are there in the two instructional models? What additional criterion might the students be graded on when they are learning in-person, vs. learning remotely? Perhaps participation, attendance, behavior, group work, and bringing a box of Kleenex?

In an online discussion forum about school closure and its impact on grading, the following practices were revealed. One district switched to pass/fail, with everything over 65% being a pass (Pierson, 2020). Another averaged the term grades from 1-3 to create a term 4 grade (Ryan, 2020). One district in Connecticut moved to a parent option of pass/fail or a letter grade, giving the family the choice of what method was most beneficial to them to interpret their student's level of success (Clifton, 2020). One district in New Mexico moved strictly to standards-based grading (Healy, 2020). While these examples do not necessarily address grading in a remote or in-person context, they do however bring up the question of looking holistically at a year's worth of evidence when assigning a grade instead of an arbitrary grading window of nine weeks.

COVID and Equity in Grading

As schools have dealt with the closures and challenges of distance learning plans, one thing that school leaders universally struggled with was grading. It was unclear how to accurately and fairly assess student learning in this new, virtual environment. It created a new set of obstacles for students and further highlighted inequities that play into students' education. Michelle Novak, a school board member in Ohio, stated in a meeting, "If we want to give our kids grades at this time, we're really going to be grading what their home life looks like. And I don't think that's fair to anyone." (Sawchuk, p. 3). What is a fair method of grading when some families have a top-speed internet connection, and another family relies on a borrowed signal while sitting outside a closed library or business? When some students have a parent at home with them to work through the online problems while another student is waiting for their turn while their

three siblings are working on their online learning, what is the best method to document and communicate their learning?

In an effort to address questions about the fairness of grading while schools are primarily using distance and virtual learning as the methods of instruction, many districts have switched to a pass/fail format. While many argue that pass/fail is the fairest and most equitable solution, some say that it is the low-income students who stand to benefit the most from earning letter grades. These students may not have the other resume boosters such as international travel, essay-writing tutors, or expensive summer classes. These can all help in setting a college application apart (Goldstein, 2020). Without those additional pieces on their college applications, low-income students are forced to rely even more on test scores and course grades. However, as was discussed at length in previous chapters, letter grades do several different things, none of them well, and the result undermines students learning (Schneider, 2020).

The experts agree that grading should not look the same during the pandemic. Feldman (2020) states three reasons not to grade during the pandemic:

1. COVID-19-related stresses will impact student academic performances in a negative way.
2. Racial, economic, and resource differences are more likely to be reflected in student academic performance during school closures
3. Most teachers do not have adequate training and preparation to provide high-quality remote instruction.

Colleges are going to have to adjust admission practices and make concessions in the requirements for students who are applying who are affected by COVID-19. Many

colleges nationwide appear to be favoring a pass-fail, or pass-no credit system for work completed during remote learning (Sawchuk, 2020). However, some experts say that pass/fail grading is not a long-term solution. “The problem can only be addressed at its root. Shaken from our complacency by a crisis, perhaps we can begin the conversation about what comes next” (Schneider, para. 22).

Schneider offers a different solution to grading altogether. In multiple articles and podcasts, Schneider (2020) suggests the use of portfolios as opposed to more traditional grading, a practice common in IB and AP programs. Creating portfolios of sample work allow students to present evidence of what they know and can do. The portfolio allows the student’s product to do the communicating instead of the transcript. The portfolio fixes many of the inherent problems with grading; the motivated students want it to be the best it possibly can; the non-motivated will do enough to pass, thereby de-weaponizing the system. “Americans may come to recognize by the end of this schooling crisis that we would all be better off without letter grades” (Schneider, para. 5). Once COVID-19 is passed and schools resume to normal, there will undoubtedly be educators calling for an overhaul of practices where inequities surfaced and better methods to assess student learning emerged.

Implications for Educational Leadership

While this study did not find a clear correlation between teacher grading practices and inflation of student GPA, it does not mean that the impact does not exist. What the study was able to illustrate is the largely apparent disparity in grading practices throughout the subject building and within each department. Not one department agreed on grading practices, and school-wide, there was not one single practice that all six

departments/subgroups agreed upon. While experts may never agree on one perfect grading system, continued research needs to be done in the area of grading.

The survey used was an adequate tool to measure the grading practices of the teachers in the study, but it was not perfect. The committee formed the survey with an endgame in mind - to expose flawed practices. The wording and verbiage of the questions are even almost accusatory. What was not taken into consideration was the inherent trait of teachers to *adapt* each situation to the needs of the students. Sometimes grading practices need to be fluid so that they can best address each situation.

Summary

When grading is too rigid and prescriptive, it can cause unintentional harm to some of the most at-risk students. Teachers are professionals and often use their expertise to slightly bend the rules a bit in favor of the student. This isn't meant maliciously, or as a means to falsify a document and misrepresent the student's mastery or ability. Sometimes students need an extra bit of encouragement, and a slightly inflated grade can help them to gain confidence in themselves. Sometimes a teacher needs to adjust scores because of an unforeseen circumstance that impacted the classes' performance on a test; the teacher is not doing this to cheat the system - they are not altering scores on high stakes tests, but rather making professional judgments to score students in a way that best reflects the performance.

Should schools switch to a standard grading system? Before the study, the researcher would have without a doubt said "yes." However, through all of the research and analysis of data, a clear answer was not revealed. Grading is incredibly important, and it is essential that all stakeholders understand and are able to intelligently interpret

the meaning of them. However, grading practices cannot and should not be so rigid as to not allow for teachers, who are experts in their field, to make decisions about grading that are in the best interest of the students. Some of the best practices identified in this study would benefit all students and should be reviewed and possibly implemented as common practice for all teachers. In order to provide equity, teachers who teach the same course would ideally have very similar grading practices so that a student's grade isn't subject to chance because of which teacher is assigned. But throughout the study, what became clear is that grading will likely never be a one-size-fits-all system. There are drawbacks and limitations to implementing a universal grading system just as there are consequences for not implementing one.

Grading is also incredibly complex and made up of many components. Some professionals argue that grades should *only* be a representation of the level of content mastery for a student, meaning that a final grade is nothing more than a series of summative assessments. However, that score would not take into account a myriad of other factors that are also key representations of what the student learned - working with others, managing time, adhering to due dates, actively participating in class discussions, and being prepared with materials. While the combination of summative and assessments and "other" factors may dilute the data and artificially inflate the final score, the student who mastered all of the soft skills is likely better prepared for the rigor of college coursework.

The research revealed that high school grades are five times more accurate than ACT scores in predicting student success in college (Jaschik, 2020). This is because of all of the immeasurable traits that grades are secretly measuring. While grading systems

are somewhat inherently flawed in their design and subject to inflation by certain widely-accepted practices, they are also the best method to serve a multi-faceted set of needs; as a communication tool, a motivator, a ranking system, a barometer of mastery, and as a rite of passage.

Yes, grades are inflated. Yes, there are students whose HSGPA is padded by extra credit and other inflated grading practices. But that “inflated” HSGPA still tells a story about the student who earned it; that they understand the schooling system and expectations and rise to meet them, that they likely have excellent attendance and punctuality, study habits, and homework completion record. These “immeasurable” soft skills may be just as important, if not more important, than the test scores they are inflating.

Having a fluid grading system tells teachers that they are trusted as professionals to make decisions about students. It acknowledges that teachers know what works for one student will not work for every student and that adjustments need to be made. There are aspects of the grading survey that are best practices for learning and should be incorporated by all teachers; but what school leader is so bold as to say they know more about the needs of each individual classroom better than the expert in the room with the students each day?

Grades are more important than ever. *Because of the ever-expanding importance of grades, grading practices should be heavily scrutinized until we can guarantee their validity and provide equity for all students.* This statement was made on the first page of this dissertation and it still holds true on the last page. However, what was discovered in the pages in between is that good teaching is more important than ever. Good teachers are

harder to come by, and *all* teachers need to be given the trust and tools needed to do the best job possible. A grading system is perhaps one of the most important tools a teacher has. It should be left to teachers to collaborate with their peers to create the system that works for them in order to serve students best.

References

- ACT at Sixty*. (n.d.). ACT. www.act.org
- Allen, J. D. (2005). Grades as Valid measures of academic achievement of classroom learning. *Clearing House*, 78(5), 218–223.
- Allen, J., Mattern, K., & Ndem, E. (2019). An empirically derived index of high school academic rigor. *Educational Measurement: Issues & Practice*, 38(1), 6–15.
- Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, 49(3), 198–211.
<https://doi.org/10.3102/0013189X20902110>
- Anderson, L. W. (2018). A critique of grading: Policies, practices, and technical matters. *Education Policy Analysis Archives*, 26 (49), 1-21.
<https://doi.org/10.14507/epaa.26.3814>
- Azeem, M., Afzal, M. T., & Majoka, M. I. (2010). How and why grading. *International Journal of Learning*, 17(3), 579–596.
- Berkshire, J., & Schneider, J. (2020). *Have you Heard?* (No. 93; Making the grade). (Podcast) Retrieved from <https://www.stitcher.com/podcast/have-you-heard/e/76393086>
- Blankstein, A. M., Noguera, P., Kelly, L., & Tutu, D. (2015). *Excellence through equity: five principles of courageous leadership to guide achievement for every student*. Corwin.
- Bonesrønning, H. (2004). Do the teachers' grading practices affect student achievement? *Education Economics*, 12(2), 151–167.
<https://doi.org/10.1080/0964529042000239168>
- Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research & Evaluation*, 17(3), 141–159.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30, 123–142.

- Brookhart, S. M. (2015). Graded achievement, tested achievement, and validity. *Educational Assessment, 20*(4), 268–296.
<https://doi.org/10.1080/10627197.2015.1093928>
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & 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>
- Canady, R. L., Canady, C. E., & Meek, A. (2017). *Beyond the grade: Refining practices that boost student achievement*. Solution Tree Press.
- Carifio, J., & Carey, T. (2013). The arguments and data in favor of minimum grading. *Mid-Western Educational Researcher, 25*(4), 19–30.
- Dueck, M. (2014). *Grading smarter, not harder: Assessment strategies that motivate kids and help them learn*. ASCD.
- Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research, 53*(1), 1–21.
- Eren, O., & Henderson, D. J. (2011). Are we wasting our children's time by giving them more homework? *Economics of Education Review, 30*(5), 950–961.
<https://doi.org/10.1016/j.econedurev.2011.03.011>
- Feldman, J. (2019). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms* (1st edition). Corwin.
- Feldman, J. (2020). To grade or not to grade? *Educational Leadership, 77*, 43–46.
- Goldstein, D. (2020, April 30). Should the virus mean straight A's for everyone? *New York Times*.
- Gonzalez, J. (Host). (2015) *Could you teach without grades?* (Audio podcast). Retrieved from <https://www.cultofpedagogy.com/episode-13/>
- Goodwin, B. (2011). Grade inflation: Killing with kindness? *Educational Leadership, 69*(3), 80-81.
- Guskey, T. R. (2013). The case against percentage grades. *Educational Leadership, 71*(1), 68-72.

- Guskey, T. R. (2015). *On your mark: Challenging the conventions of grading and reporting*. Solution Tree Press.
- Guskey, T. R., & Brookhart, S. M. (Eds.). (2019). *What we know about grading: What works, what doesn't, and what's next*. ASCD.
- Haley, M. R., Johnson, M. F., & McGee, M. K. (2010). A framework for reconsidering the Lake Wobegon effect. *The Journal of Economic Education*, 41(2), 95–109. <https://doi.org/10.1080/00220481003617228>
- Harland, T., McLean, A., Wass, R., Miller, E., & Sim, K. N. (2015). An assessment arms race and its fallout: High-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*, 40(4), 528–541. <https://doi.org/10.1080/02602938.2014.931927>
- Hiss, W. C., & Franks, V. W. (2015). Brave new world: How test-optional policies could (and should) change college admission. *Journal of College Admission*, 228, 32–34.
- Immerwahr, J. (2011). The case for motivational grading. *Teaching Philosophy*, 34(4), 335–346.
- Kagan, S. (1995). Group grades miss the mark. *Educational Leadership*, 52(8), 68–71.
- Kanold, T. D. (2018). *Mathematics homework & grading in a PLC at work*. Solution Tree Press.
- Koran, M. (2020, May 2). University of California to drop SAT and ACT as admission requirements. *The Guardian*.
- Koretz, D., & Langi, M. (2018a). Predicting freshman grade-point average from test scores: Effects of variation within and between high schools. *Educational Measurement: Issues & Practice*, 37(2), 9–19.
- Lalley, J. P., & Gentile, J. R. (2009). Classroom assessment and grading to assure mastery. *Theory into Practice*, 48(1), 28–35. <https://doi.org/10.1080/00405840802577577>
- Lang, D. M. (2007). Class rank, gpa, and valedictorians: how high schools rank students. *American Secondary Education*, 35(2), 36–48.
- Laurie, R. (2009). Raising the bar: A data-driven discussion on grade inflation. *Education Canada*, 49(4), 32–37.

- Linnehan, F. (2001). Examining racial and family educational background differences in high school student beliefs and attitudes toward academic performance. *Social Psychology of Education, 5*(1), 31–48.
- Lukianoff, G., & Haidt, J. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*. Penguin Press.
- Marcus, J. (2018). The newest advantage of being rich in America? Higher grades. *Education Digest, 83*(6), 12-16.
- Marx, J., & Meeler, D. (2013). Strike four! Do-over policies institutionalize GPA distortion. *Quality Assurance in Education, 21*(1), 39–53.
<https://doi.org/10.1108/09684881311293052>
- Marzano, R. J. (2000). *Transforming classroom grading*. Association for Supervision and Curriculum Development.
- Marzano, R. J., & Heflebower, T. (2011). Grades that show what students know. *Educational Leadership, 69*(3), 34-39.
- Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Association for Supervision and Curriculum Development.
- Morrison, K. (2003). Is grading doing what we want it to do? *Paths of Learning, 15*, 20-24.
- O'Connor, K. (2011). *A repair kit for grading: 15 fixes for broken grades* (2nd ed). Pearson.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372–1380.
<https://doi.org/10.1016/j.tate.2010.03.008>
- Rauschenberg, S. (2014). How consistent are course grades? An examination of differential grading. *Education Policy Analysis Archives, 22*(92). 1-39
- Reeves, D., Lee, A. J., & O'Connor, K. (2017). What's worth fighting against in grading? *Educational Leadership, 74*(8), 42-45.
- Reeves, D. B. (2000). Standards are not enough: essential transformations for school success. *Nassp Bulletin, 84*(620), 5–19.
<https://doi.org/10.1177/019263650008462002>

- Sadler, D. R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education*, 35(6), 727–743. <https://doi.org/10.1080/02602930902977756>
- Sawchuk, S. (2020). Grading students during the coronavirus crisis: What's the right call? *Education Week*, 39(30), 6.
- Sawyer, R. (2013). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education*, 26(2), 89–112. <https://doi.org/10.1080/08957347.2013.765433>
- Schaffhauser, D. (2019). 4 instructional practices with impact on student achievement. *THE Journal*. THEjournal.com
- Schneider, J., & Hutt, E. (2014). Making the grade: A history of the A-F marking scheme. *Journal of Curriculum Studies*, 46(2), 201–224.
- Schoology. (n.d.). Schoology. Schoology.com
- Smith, D. E., & Fleisher, S. (2011). The implications of grade inflation: Faculty integrity versus the pressure to succeed. *Journal of Research in Innovative Teaching*, 4(1), 32–38.
- Stackstein, S., Pierson, A., Ryan, M., Clifton, M., & Healy, C. (2020, July 26). *Teachers throwing out grades*, Facebook. <https://www.facebook.com/groups/teachersthrowingoutgrades>
- Stanley, G., & Baines, L. (2001). No more shopping for grades at b-mart: re-establishing grades as indicators of academic performance. *Clearing House*, 74(4), 227–30.
- Steward, R. J., Hill, M. F., Neil, D. M., Pritchett, T., & Wabaunsee, A.-S.-N. (2008). What does GPA in an urban high school actually mean? *Educational Considerations*, 36(1), 11–16.
- Stiggins, R. J. (2017). *The perfect assessment system*. ASCD.
- Twenge, J. M. (2017). *iGEN: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy-- and completely unprepared for adulthood and (what this means for the rest of us)*. Atria Books.
- Vatterott, C. (2015). *Rethinking grading: Meaningful assessment for standards-based learning*. ASCD.

- Vulperhorst, J., Lutz, C., de Kleijn, R., & van Tartwijk, J. (2018). Disentangling the predictive validity of high school grades for academic success in university. *Assessment & Evaluation in Higher Education, 43*(3), 399–414.
<https://doi.org/10.1080/02602938.2017.1353586>
- Walsh, P. (2010). Does competition among schools encourage grade inflation? *Journal of School Choice, 4*(2), 149–173.
- Westphal, A., Becker, M., Vock, M., Maaz, K., Neumann, M., & McElvany, N. (2016). The link between teacher-assigned grades and classroom socioeconomic composition: The role of classroom behavior, motivation, and teacher characteristics. *Contemporary Educational Psychology, 46*, 218–227.
<https://doi.org/10.1016/j.cedpsych.2016.06.004>
- Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT[®] scores, high school grades, and SES. *Educational Assessment, 20*(1), 23–45. <https://doi.org/10.1080/10627197.2015.997614>
- Wormeli, R. (2018). *Fair isn't always equal: Assessing and grading in the differentiated classroom* (2nd ed). Stenhouse Publishers.