Grading Policies in Education

Amber Kiesel

Department of Education, University of Washington Tacoma

TEDUC 599

Dr. Matthew Weinstein

June 2023

Abstract

Grading is one of the most pivotal aspects of education, and grading results profoundly impact students' future in grade school, higher education, and employment. This research explores grading policies in education. Three themes were identified: effective grading policies, equitable grading policies, and challenges or obstacles in grading policies. The paper examines how practices in schools and districts could be aligned with research and then explores implications for future research and transformed practice.

Keywords: grading policy, standards-based grading, minimum grading, grading implementation, equitable grading

Grading Policies in Education

The focus of this paper is the research on grading policies in education. My experience in education has led me to the following problem: school or even district-wide grading policies are rare. Most decisions we make as educators are based on grading and assessment, so why would there not be a policy for it? During my initial research, I started to think about what makes a grading policy equitable, especially in our world now, where classrooms are becoming increasingly inclusive and differentiated.

At one of my first meetings with my grade-level team, grading policies were discussed. No one could agree on what our team grading policy should be. The discussion became intense, and I quickly realized this was quite a contentious topic among educators. Everyone had their ideas about how we should go about grading, and no one could seem to agree. I decided to research some of the grading policies suggested by my colleagues. This topic is also fascinating in my specific situation because my district follows Marzano's (2017) Guaranteed and Viable Curriculum theory. This means that every student in each grade level and subject gets the same lessons and assessments regardless of which school or teacher they have within our district. Our grade books should be identical to every other sixth-grade math and science teacher. However, the grade books will never be aligned without an official grading policy. Therefore, there is inconsistency for our students. In every classroom, they must learn and remember an entirely different grading scheme, which is precisely what we are trying to avoid. Each teacher has their preferred way, and without any oversight, there is no guarantee that we are all doing the same thing.

Regionally, this means that there needs to be more consistency between districts or even schools. I have many friends who work in districts that border mine whose grading policies are

entirely different. When students move from one teacher, school, or district to another, they and their families must adapt to a wholly new system. In school, grading is like currency; it would be frustrating if you went to the next town over and had to learn an entirely new way to pay for goods. The same can be said for grading as a teacher, parent, or student.

Nationally, the picture becomes even more confusing and disjointed. The National Education Association has published countless articles about grading policies and practices. The one consensus is that there is no uniting behind a single approach but that educators recognize, especially in our post-online schooling era, that there is a need for change (Walker, 2021). Searching the National Association of Secondary School Principals (NASSP) archives will garner comparable results. However, the articles all agree that when developing a new policy, fairness and consistency are of the utmost importance (NASSP, 2019). Finally, the U.S. Department of Education states on its website that there is no nationally mandated grading system in the U.S. and that all structures for grading and assessment are regulated by the local systems (U.S. Department of Education, 2008). Despite no recommended approach, the major trend in grading policies nationally is standards-based grading using the common core standards, which will be discussed later in this paper.

This paper aims to examine data and literature on different grading policies.

Focal Questions

- What makes for an equitable grading policy?
- What grading policies have the most compelling research support?
- What are the challenges or obstacles to different grading policies?

Literature Review

This literature review examines research on grading policies in education. In this section, I will examine research on the effectiveness of grading policies, the challenges and obstacles of grading policies based on research, and the equitability of grading policies.

For almost 100 years, it was standard practice in education to grade on a 0-100 scale, where nothing below 70% was a passing grade. In recent years, especially since No Child Left Behind in 2001 and the introduction of Common Core standards, there has been a shift to move away from that practice (Brookhart et al., 2016). There are many different options when deciding on a grading policy. The following literature will show what policies are being chosen, why they might be selected, possible barriers that might be present in policies, and the equitability of grading policies.

Setting a baseline for what actual practices are happening in classrooms and their effect on students is an important starting point. The first section will contain eight studies to help understand the structures of grading policies currently in place. Major themes discussed in this section include the consistency of grading policies across schools, factors teachers consider, how those choices affect students and their performance, and some important guidelines when creating grading policies.

The second section of this literature review explores what makes a grading policy effective. Eleven different pieces of literature are considered in this section. The themes of the research articles are data surrounding specific grading policies (mastery-based grading, minimum grading, grading on a curve, and standards-based grading) and general practices in creating grading policies with fidelity.

The third section describes the challenges and obstacles identified in the literature. This section contains twelve articles. The major themes include a need for more consistency in grading, clarity in policies and training, and teacher buy-in to policy change.

The final section includes nine studies that examine what characteristics make grading policies equitable and how they might need to be changed or adjusted. Three aspects of equitability are looked at in this section: gender, race, and ability level.

Consistency and Structures of Current Grading Practices

The first section of this literature review examines a wide array of current grading practices. The studies reveal much information about what works and what does not. The major themes discussed in this section are consistency in grading, teacher accountability, student experiences of grading policies, and some guidance on what some research suggests are best practices.

Research by Dardanoni et al. (2009) indicates little consistency in grading policies. The researchers wanted to establish how consistent grading policies were across schools in sixteen different countries. In a 2003 survey done by the Organization for Economic Co-operation and Development (OECD) and overseen by their Program for International Student Assessment (PISA), data was collected from schools and students which reported information on school grades, namely whether students passed mathematics on their most recent report cards and the grading policies of the schools. Dardanoni and their colleagues then did a statistical analysis of the data. During this analysis, it was found that standard grading practices in schools across all countries were diverse; the only pattern to emerge was that those schools that fell in the middle of the data sets for students passing mathematics had a higher percentage of grading policies that used grading on a curve.

With the understanding that across larger entities, like countries, there is inconsistency in grading policies, what the literature says about consistency in individual teachers' practices should come next. Guskey and Link (2019) are two significant figures in the discourse surrounding grading practices; they will be seen a few times throughout this literature review. In this first study, they sought to investigate what factors teachers consider when judging students' grades. They intended to determine the characteristics and see if they varied between teachers at different grade levels and with differing years of experience in the classroom. To gather data, a questionnaire comprised demographic, grading, and recording practices. The questionnaire was sent to 2,233 teachers across five school districts in a Southeastern state, and 943 teachers responded. The data showed that about 10-20% of the weight of students' assigned grades had to do with non-cognitive factors such as participation, effort, and punctuality across all grade levels. Teachers at the high school level gave more weight to exams, projects, and homework than non-cognitive factors. At the same time, those at the primary levels gave more weight to observations, formative assessments, and displays of student work. Experience level did not show a significant statistical difference in grading factors.

While considering how teachers use a policy to grade their students, Simon et al. (2010) took an in-depth look at the practices of one teacher and the tension present between her practice and the grading policy she was working under. The teacher's interview took over three months while preparing final grades for two tenth-grade math classes. Overall, twelve teachers were interviewed for the more extensive study. However, Anne was selected out of the group because the problems she identified encompassed most of what the other teachers did, the fundamental tensions she picked out were articulated plainly, and she provided a wide range of artifacts to support her claims. The study results showed that Anne's interpretation of the policies her district

and her administrator presented vastly differed, which caused her to adjust her grading practices in the middle of the second semester. The policies seemed conflicting and inconsistent to varying levels within the education system. The Education Department wanted to see the students' progress from semester to semester. Still, it required 35% of the final grade from semester one, limiting how much of a difference the most recent evidence could make.

In contrast to the previous study, which showed the teacher's experience when policies are being unclearly presented, Puhani and Yang (2020) looked at increased teacher accountability and how it affects student grading outcomes. This empirical study took data from Hesse, Germany, over five years after a significant educational reform where students exiting high school were required to take a centralized written exam graded by an external source. The analysis was based on 364,445 exit exams from 72,889 students. This policy change aimed to decrease subjectivity in exit exams they had previously given. The statistical analysis showed that the overall scores for the written exam increased by 1% (0.046 standard deviations); this trend was seen in STEM subjects and social studies but not in language arts. The researchers concluded that this data shows that the increased accountability for teachers has led to a decrease in grading leniency that had been present previously.

With accountability in mind, Winters and Cowen (2012) conducted a study from the largest school district in the country, New York, looking into how keeping schools accountable for their students' progress affected their outcomes over the immediate years following. Their study showed that only the schools that received the lowest scores showed marked improvement over the following two years. Those schools ranked second lowest grade and higher showed little to no significant changes over the following years. The comparison of these two studies shows

how teacher accountability versus school-wide accountability measures can affect student outcomes and which had the most significant impact overall.

The effect on student outcomes was precisely what Klapp (2015) was trying to answer in her study, where she took data from the Swedish school system to try and answer questions about the long-term effects of grading on students' achievement and if grading affected different subsets of students divided by concerning cognitive ability and gender. She used 50 years of data from 1948-1998, consisting of 430 classes in over 29 municipalities and 8558 students. The data was taken from Sweden during this period because the curriculum and grading were regulated and overseen by the government, eliminating many of the variables in collecting empirical data elsewhere. The comparison of graded versus ungraded students was made because half of the schools started receiving grades at grade six while the other half started at grade seven. Overall, the results showed that grading in the traditional sense had an overall negative effect on later achievement. For those students with lower cognitive abilities, the results showed that those who were graded performed worse throughout their secondary schooling than those who were not. The gender differences in grades increased as students advanced through school, where girls achieved higher rates than boys through grades 7-9. Lastly, the results showed that graded students were less likely to finish secondary school compared to those who were ungraded coming into seventh grade.

While student motivation is important when examining how students are affected by grading policies, it is also important to consider what information they are getting from their grades and how they can be more accountable for their growth. Wormeli (2006) discussed this in his article about the importance of feedback when giving assessments and emphasized that grading is supposed to be about where a student is in their learning journey, not used as a reward,

motivation, or a reflection of their behavior. If those things are included in the grading process, then feedback and demonstration of progress will not be accurate. He stresses the importance of formative assessment and how much feedback it can provide for the student. By reducing our emphasis on high-stakes summative assessments and letter grades and providing more feedback to the students about their work, there will be significantly more growth over time and, as a result, highly accountable students.

Guskey (2020) takes this idea of informative grading even further by suggesting that progress reports and report cards should have multiple grades for each subject. He indicates that grades be split into, at the very least, product, progress, and process criteria. Product criteria follow a more traditional grading system as an amalgamation of tests, projects, or other culminating activities based on grade-level standards. Progress criteria show how much the student has grown and improved their learning. Process criteria include the day-to-day activities like homework, formative assessments, and social-emotional skills that help facilitate that student's growth. He claims that having multiple grades in the same subject will better help families to understand their students' performance and where they might need additional support. Guskey also argues that this does not result in more work for teachers because this is all information they are already collecting data on; it is simply a matter of reporting it differently.

Effective Grading Policies

The next set of research will establish what the literature says about specific grading policies and their effectiveness in the classroom. The first specific grading policy being examined is mastery-based grading. In this policy, students are graded on their holistic achievement during a course, sometimes through a portfolio or end-of-course exams. Then the minimum grading policy will be examined, which does not allow teachers to assign a zero grade

to students. Finally, research surrounding standards-based grading and its effectiveness will be discussed.

Research done by Alex (2022) on the effects on overall student achievement while using mastery-based grading was examined. The question posed was, is this grading policy the answer to modernizing our grading system? To answer this question, Alex interviewed Joe Feldman, one of the foremost proponents of this policy and author of *Grading for Equity*, educators, and students. She also took data from school districts in Nevada and California and an individual school in Virginia. In her research, she found that one of the districts in California, Placer Union, had a significant reduction in the number of Ds and Fs and narrowed the achievement gap between white students and students of color after mastery-based grading was implemented. The other district in California, L.A. Unified School District, had 15,000 grades go up because of the change.

In contrast to the previous study, the intent of Nowacki (2013) was to investigate the impact of portfolio-based grading, which is similar to mastery-based in that it is an assessment based on the student's entire bulk of work from a course. The central question this research was attempting to answer is: Are the needs of students being met within a portfolio system alone, or is there value (positive or negative) in adding a letter grade as supplementation? The study was conducted in a college-level course for Introductory Biostatistics, where 33 students' portfolios were submitted to a committee for assessment. The committee evaluates the work in the portfolios based on subject mastery. Of those 33 students, eight students also had to receive letter grades for the Masters-level work. At the end of the course, all students did a survey that gathered information about their perception of how the assessment method they received, just a portfolio or portfolio plus a letter grade, impacted the course for them. The study results showed

that overall, there was no significant difference in scores for the course; students with just portfolio grading averaged 72%, while students with both portfolio and letter grades received an average of 76%. However, in the survey about perception, students with both portfolio and letter grades reported more preparation for the course (84% versus 66%), increased willingness to ask for help (83% versus 68%), a higher sense of personal achievement (81% versus 62%), and more confidence that they learned what they needed from the course (85% versus 68%).

Once there is a clear establishment of the grading policy, and how it might impact students, it is important to look into implementation. Dingelman-Parente (2011) has developed a working guide to implementation, which she tested in an Organic Chemistry course which she claims increases student participation and an assessment guide with a mastery-based approach in mind. She states that for mastery-based learning to be successful, course content must first be broken up into key concepts. The author then describes the importance of group instruction while keeping in mind that not all students will learn at the same pace, and therefore some may need additional time to reach mastery of each key concept. Within each key concept, students can do their work over again and extend their learning when their mastery goal has been reached. This idea of teaching together but learning independently is central to Dingelman-Parente's pedagogy and what she has found to be most effective through trials of classes versus control classes.
The following two studies will look at data surrounding a minimum grading policy. The articles were written by the same authors using the same data but focused on finding answers to different questions.

Carey and Carifio (2012) sought to answer questions raised in their 2009 article, *A critical examination of current minimum grading policy recommendations,* about the effectiveness of minimum grading (more commonly referred to as a No Zero Policy), where

there is a minimum threshold on grades given to students (most typically 50% on a 100-point scale). The questions they sought to answer were: How often are minimum grades assigned, and how often does a passing course grade follow the assigning of a minimum grade? They accomplished this by doing a quantitative study on grading data from a high school from 2003-2010 that had implemented a minimum grading policy in the late 1990s. The results showed that of the 343,425 grades, only 29,187 (8.5%) began with a minimum grade of 50% after the first term. Of those 29,187, only 1,159 ended with a passing grade in the course, which is 0.3% of all assigned grades. This data showed that assigning minimum grades did not result in passing many courses that would have otherwise failed, and critics claimed that students would unfairly benefit from the policy.

A year later, Carifio and Carey (2013) continued their exploration of the effects of minimum grading by testing the data they had gathered in the previous study against the significant reasons critics are against the policy. The first argument they address is that minimum grading would give students grades beyond what they have rightfully earned and contribute to grade inflation. The empirical data they gathered in the previous study showed that those students who were having challenges in school were still receiving lower grades or failing. In contrast, their better-performing classmates were not affected by the policy. The students who actively benefited from the policy even outperformed their peers regarding growth in state-mandated exams. The results directly contradict the claims of critics. The authors say the minimum grading policy does not prevent students from failing. Instead, its purpose is to prevent students from "early catastrophic failure" (Carifio & Carey, 2013, p. 26) and give them a way to turn their grades around given effort. They conclude by stating that this policy is a low-cost change that helps to remove the inherent inequities in the traditional grading system. Wormeli

(2006) pointed out that when zeroes are turned to 50 or 60% in the grade book, students are not getting something if there is no effort. Instead, the grades are being adjusted to have a comparable impact on the overall grade. Therefore, the grade is a more accurate measure of the student's proficiency.

Many of the arguments against minimum grading are focused on the part of the policy which allows students to receive a grade on assignments regardless of completion (Caneva, 2014). They fear that this will allow students to pass classes without effort. However, effort has to be made for students to pass, as 50% is still a failing grade. The effort put in by a student is important, and research surrounds the importance of including effort within the grading policy. In 2004, Benedict College implemented a new grading policy for its first and second-year students, which took effort into account (Swinton, 2010). The justification behind this change was that if a professor only grades on effort, then that would, in turn, maximize the effort given by students; alternatively, if the professor only grades on knowledge, which may cause those students with a lower ability level to give up entirely. Therefore, balancing effort and knowledge will increase both the effort made by students and, in turn, the knowledge gained by them. This theory proved to be true when Benedict made the change; the added grade on effort had a significantly positive effect on the knowledge grades of students. Paredes (2017) supports this by statistically showing that lower-ability students exert higher effort and perform better when grades are relative to both effort and knowledge. However, the data also shows that some high-ability students reduce their effort when their knowledge performance differs from their grades' absolute determiner.

One of the most popular new grading policies in education is standards-based grading. The next set of articles will focus on what some research says is effective about this type of policy.

In their article about the impact of standards-based grading, Huey et al. (2022) define standards-based grading as "a method of grading students based upon a set of defined learning objectives with separate grades provided for work habits (process), improvement (progress), and learning (product)" (p. 2). This definition is an extension of Guskey's (2011) work which was examined earlier in this section. This study's data was taken from a school transitioning to standards-based grading and is currently at a hybrid implementation stage. The researchers wanted to evaluate the changes in student performance during this first year of implementation and have the students self-report any changes to their habits and possible explanations for why they changed their behavior regarding class work. For this, they divided the students into two separate cohorts. Cohort 1 grading was weighted 85% of the final product grades and 15% for their practice work. While Cohort 2 was graded 100% on their final product grades, and their practice work did not impact their final grade. The findings showed that Cohort 1 averaged about 6% higher final grades than Cohort 2, ranging from 2-17% through the units. Overall, Cohort 1 averaged higher for every unit of the class. In the survey sent to the students in Cohort 2, 23% reported doing less practice work than the previous year, 27% did more practice work, and 50% did about the same amount. When asked if they would complete more work if it impacted their grade, 75% answered they would. The researchers concluded from this information that having a more balanced system between product, process, and progress will incentivize students to improve cohesively.

Link and Guskey's (2022) analysis aimed to establish the critical criteria for creating a standards-based grading policy and determine its effectiveness. The ultimate question they attempted to answer was whether standards-based grading is effective. To do so, they propose that three steps need to be taken:

1. Standards-based grading needs to be precisely defined.
2. Explicit criteria for rating the effectiveness of any grading system need to be established.
3. Standards-based grading needs to be judged by those criteria.

The authors did a literature review of research on standards-based grading to define it and assess if it improved student achievement. They then used this same analysis method to determine what makes an effective grading system. They used all this data to determine if the definitive version of standards-based grading would fit the criteria established in the second step.

The three criteria by which grading policies should be judged come from over 100 years of research evidence and are as follows: (1) student performance should be rated based on key grade-level standards (not an entire content area), (2) student progress should be reported using a limited amount (about 3 to 5) of performance categories (3) academic achievement should be reported separately from behavioral factors. For the effective implementation of a standards-based grading policy, all of these need to be met. To meet the criteria means having a clear starting definition of what the grading policy will look like in that school or district, having thorough communication and training on what it is and will look like, and why it is effective.

Zimmerman (2017) argued that using a standards-based grading scale alone is not a significant enough shift from the traditional grading system to be effective for students' growth. Along with the grading scale, detailed learning objectives (standards) must be assigned to problems to give students enough feedback. Zimmermann argues that if the learning objectives

are too broad, you will have fewer objectives, and students will only receive general feedback. However, if the learning objectives are focused, multiple can be applied to a single problem, and students can get more accurate and focused feedback. Toledo and Dubas (2017) agree with Zimmermann but take it a step further by saying that not only is meaningful feedback important, but providing multiple avenues to and opportunities to achieve proficiency will help students reach academic success. Implementing this provided the students and educators with more detailed data and targeted feedback. They argued that this implementation method gives clearly defined scaffolding necessary for end-of-year standards achievement.

Challenges and Obstacles of Grading Policies

Now that it has been established what some research says makes for effective grading policies, seeing what the challenges and obstacles are to grading policies naturally follows. Twelve studies will be analyzed in this section; some contain empirical data, while others offer insight into the individual experiences of teachers and students.

The research conducted by Alex (2022), as mentioned in the previous section, also found some obstacles relating to those experiencing a change in grading policy. An interviewee from the district in Nevada (comprising 360,000 students) claimed that the policy was rushed out after returning to in-person learning and resulted in students being continually late on assignments because they felt their expectations of them were lowered. The high school in Virginia had a different experience. Educators at that school rejected the policy before it could be implemented, arguing that it would reduce expectations, rigor, and accountability.

Many of the proposed grading policy changes in education involve content mastery. While the research shows that high-stakes exams are not a good indicator of a student's abilities, they are still required to obtain higher education (such as the SAT or ACT). Senko and Miles

(2008) suggest that grading students with mastery- or standards-based policies may harm their success on these high-stakes exams. To determine if this was true, they took data from a psychology course where students' achievement goals were measured at the beginning of the course, and at the end of the course, their study strategies and final grades were used to assess if they achieved those goals. Some of the students were mastery-oriented, meaning their goals had to do with deep learning of the course content, and others were performance-oriented, meaning they wanted to do whatever it took to get a good grade. The authors hypothesized that those students with a mastery focus would have a lower overall achievement because their studying would become more interest-based, meaning they would pursue more profound knowledge on those topics that interested them and neglect other vital topics in the course content. The results supported the researchers' theory. Students with mastery goals found more specific interests than those with performance goals but could have performed better overall. Those students with performance goals, though, needed more interest in the topics presented in the course. The conclusion is that there has to be a balance between topic mastery and performance for actual student achievement.

The following articles concern the challenges presented by grading on a curve. This grading policy is often seen in science-based classes and is still widespread in grade school and higher education.

Bowen and Cooper (2022) were looking for a change in traditional grading methods in science education, such as grading on a curve. As stated above, it is widespread but needs to be more equitable, according to these authors. They argue that the notion that grading on a curve is more efficient than other grading methods is false. This grading method is used to sort students rather than measure their ability, encourage students to compete instead of learning, and

intensifies social inequities. They examine the history of grading policies in general as well as the rise in popularity of grading on a curve that occurred in STEM. They critique specific parts of grading on a curve: it is a comparison of what is considered normal in our population, alignment with culturally responsive teaching, consistency between educators, efficiency in grading, promoting competition over learning, and lack of informative feedback. Grading on a curve is a comparison of what is considered normal. The authors found that this practice stemmed directly from eugenics and therefore does not include the diverse population in classrooms today. Grading on a curve aligns differently than culturally responsive pedagogy because it automatically means some students will fail and, as a result, have relatively high failure rates. Consistency between educators is not promoted because while they may be grading on the same bell curve, they need to teach in the same way, influencing students' learning. While the practice may be more efficient in theory, it is a way of sorting students rather than grading their performance. Grading on a curve encourages competition rather than learning, which gives an unfair advantage to those students with access to more resources and opportunities. It also does not communicate how they can improve but shows students how they are doing relative to their peers. This qualitative data was gathered from several studies, with input from educators, administrators, and students who have experienced this method of grading policy.

In addition to creating competition among students, grading on a curve can also be incredibly unfair when there is an average of higher or lower-achieving students. Calsamiglia and Loviglio (2019) show through data from Catalonia that students' grades are drastically impacted by their peers' average quality rather than their achievement and abilities. This difference is enhanced when grading on a curve is present, not necessarily because of the student's actual abilities but because of the perception the curve creates for the teacher grading

them; the curve is augmenting the teacher's unconscious biases. These conclusions were made after comparing students in school grades to the high-stakes nationwide testing that Catalonia requires for college admissions (much like the SAT or ACT).

The subsequent two studies discuss the challenges present in implementing standards-based grading. The first article comes from interviews with principals who have or are planning to change over to the policy. The second article discusses the obstacles to communicating with students about their performance using standards-based grading.

Many of the studies presented in this review have been looking at grading policies from the perspective of urban or suburban schools and districts. On the other hand, Buckmiller et al. (2020) specifically gathered information from rural schools in the midwestern United States. This approach was chosen for two reasons: first because there needs to be more data about rural schools, and second, policy changes tend to move slower in these schools due to a lack of support. The administrators tend to be stretched thin. This study was done as a survey completed by 85 schools that consisted of questions about switching to standards-based grading and if they felt they had the content knowledge and support to implement the change correctly. Five-point scales were used to quantify the data gathered. This data was then compared to the answers given by principals at urban and suburban schools. When the principals were asked if they felt they had sufficient resources to implement the change, the average answer was 3.74, much lower than the urban/suburban average of 4.13. When asked if they felt a thoughtful plan was in place for transition, the mean answer was 3.81, again lower than the urban/suburban average of 4.20. The principals were then asked if their teachers supported the change 69% of the principals answered that they somewhat agreed, indicating that teacher resistance may be a significant barrier to implementation.

Getting an educator's perspective on standards-based grading is important to understand how the policy works. A professor chose to implement standards-based grading into his college course to see how it would impact his communication with his students about their progress in the course, where they needed work, and how they could improve their grades (Scarlett, 2018). The positive effects of this change were significant. Scarlett reported that his students were much more focused on learning the content than their score, and he felt their final grade was a much closer fit to their actual abilities than when he used traditional grading. However, implementation came with its drawbacks as well. The first that he encountered, and which is a common concern among K-12 teachers, were students not completing practice work like homework because it would not be included in their final grade. To combat this issue, he made standards for work completion and coming to class prepared, which fall under the process grade. He also admits that setting the system up was very time-consuming, and there was more effort into grading assignments because he was thinking more about the level of understanding shown by the students rather than just checking whether an answer was right or wrong.

Implementing any new system will be time-consuming and require much training. Is the absence of sufficient training a major downfall for standards-based grading systems? Wisch et al. (2018) wanted to see if this was true and sought to discover if the problem with implementation needed to be more knowledgeable or a difference of opinion on the part of the teachers. They surveyed 551 teachers for their opinions on the policy, its implementation, and its perception. The results showed that of the teachers who said their school had implemented a standards-based grade policy, 65% of them still altered grades based on classroom participation and behavior. The teachers indicated that they felt they did not have ownership over the policy change and were not empowered to voice their opinions or suggestions for the change. Based on this

information, the authors conclude that teacher buy-in and training are essential to successful policy implementation.

The rest of the articles in this section discuss obstacles present in the grading policies and the implementation process. Challenges in these studies include focusing grading policies around high-stakes exams, rigor of grading, and inconsistency in grading.

Typically, high-stakes exams in the U.S. are considered extensive college admissions tests like the SATs or ACTs. However, for students' high-stakes exams are any that determine the majority of their grades or moving on to the next level of their education. The problem of practice this study focuses on is whether it is possible to align teachers' grading practices more closely by choosing a national grading policy that does not rely solely on high-stakes national testing (Jönsson et al., 2021). The researchers looked at two different grading approaches, analytic grading and holistic grading. Analytic grading assessment takes individual grades for student mechanics, grammar, organization, and writing and averages them together for the student's final grade. Holistic grading simultaneously considers all these criteria, effort, and overall performance to create a student's grade. Participants included 74 teacher volunteers randomly assigned to one of the two grading policies. These teachers received writing from four students four times over a semester. The teachers using analytic grading were then asked to grade the assignments individually, and the student's overall score was determined based on the average at the end of the semester. The teachers asked to grade holistically were given all four assignments simultaneously and asked to provide an overall grade for each student. The results showed that the teachers using analytic grading agreed on student grades more than those grading holistically. This data led the researchers to believe that analytic grading may reduce the

complexity and increase agreement between teachers when grading student work. A major limitation of this study is the sample size and the objectivity involved in both grading policies.

While looking at grading through an analytic/holistic approach, some research also looks at strict versus lenient grading. The purpose of Elikai and Schuhmann (2010) was to examine the impact of grading on student performance in a course. Two different grading policies were compared, one more lenient in what overall grade is needed to pass the course and one with a stricter grading scale. The more lenient grading policy follows a traditional 0-100 scale: A = 100-90, B = 89-80, C= 79-70, D = 69-60, and F < 60. Depending on the university, a C or D is sometimes needed to pass the course. The stricter grading scale uses the same letter grades but represents different percentages: A = 100-93, B = 92-85, C= 75-84, D = 74-65, and F < 65. On this grading scale, a minimum of a C or 75% is required to pass the course, which is by professional examinations required for a professional license in accounting. The two major arguments about grading policies versus student achievement are that a strict grading policy will motivate students to put in more effort, and the other is that it will cause unnecessary stress to the student resulting in poorer performance. This study was conducted at a university (with over 12,000 students) over two semesters during the 2004 academic year. Two sections of undergraduate classes were examined each semester; the first section was the control group (91 students), graded on the traditional scale, and the second was the treatment group (95 students), graded on the stricter scale. The course was an important requirement for all students in the study; class sizes and gender demographics were similar in each course, and all courses met in the morning with the same instructor, materials, and teaching techniques. The student's GPA before the study began was used to measure current overall ability and as a comparison tool to rate student achievement in the course. Four exams were given over the semester to all groups,

and the results were used to determine student achievement. The study results showed that 66% of students in the treatment group received average grades of As and Bs on their exams, compared to the control group, which only received 55% overall. About 8% of students in the control group received an overall grade of Ds, while only about 3% received that grade in the treatment group. These results show that, statistically, more rigorous grading resulted in higher levels of achievement by the students.

Similarly, Lu et al. (2021) tried to find how different levels of leniency would affect student outcomes. Data sets were collected from twelve courses at a university in Taiwan. Each set includes individual students' digital activity and final scores from different learning environments. The environments include massive open online courses (MOOCs), eBook software, and a typical learning management system (LMS). One thousand two hundred sixty students participated in these courses, ranging from six to 18 weeks. Researchers found that they could divide the results into three groups. Group one, labeled "grading on discrimination," which was graded on a more traditional scale, had a majority of the students receiving average scores, with those on the low and high end being a relatively low percentage. The second group, labeled "grading on stringency," which was graded with more strict criteria, had overall lower grades than the other groups. The third group, labeled "grading on leniency," was where educators used high grades to motivate and build self-confidence during the course and received high or almost perfect scores on their final exams.

Figlio and Lucas (2003) agreed that leniency in grading policies does not benefit students' success. This conclusion was reached after an in-depth empirical study of detailed data from a large school district over three years. Data from third, fourth, and fifth graders were taken year to year and analyzed using the students' state test scores, report cards, demographics, and teacher

information. The results showed that of those teachers with more lenient grading standards, only nine percent of their students who received As achieved a level four (above grade standard) in their state tests. Furthermore, of those students that received Cs in class, which would be considered at grade level, only fourteen percent tested at grade level in the state test. However, the teachers who graded on higher standards in the classroom had 65 percent of their A student's scores at level four. Thus, those teachers who graded on stricter and higher standards in the classroom were more accurately measuring their students' abilities with their letter grades.

There are other reasons that students' grades can change drastically from class to class or even student to student; sometimes, it has to do with the teacher themselves. The purpose of Rauschenberg's (2014) article was to establish the major causes of differential grading. Differential grading is where students receive inconsistent grades in courses that use the same content and curriculum depending on their district, school, or teacher. The factors being assessed for the cause of this are teacher grading standards, district grading policies, student behavior, teacher stereotypes, teacher quality, and adherence to the curriculum. For this research, course grades and end-of-course test scores from 2007-2010 throughout North Carolina were analyzed. Two courses were used for the study, English I and Algebra I. It examines student, teacher, school, and district-level patterns across the state. The data gathered from this research showed that student characteristics were a higher predictor of differential grading than teacher, school, or district characteristics. Students who statistically earned higher grades in the courses were female, English language learners, and in 12th grade. Low-income students earned overall lower grades than other demographic examinations. Black students earned higher Algebra I grades but lower English I grades than white or Asian students, even when they earned the same end-of-course score. The best determiner of these statistics was the comparison of course grades to end-

of-course scores, which is how the author determined the primary causation as being the students' demographics.

Equitability of Grading Policies

Now that research has established what is effective in grading and what might prohibit progress, equitability must be the next topic because things can work and not be equitable and vice versa.

Grading for Equity by Feldman (2019b), one of the leading researchers in equity-based grading, consists of qualitative and quantitative data compiled from interviews and different studies. It addresses the how of grading practices, why they have stayed the same over the past century, and why they need to. According to Feldman, these antiquated grading methods do not allow for effective teaching or learning and perpetuate inequitable outcomes in our schools. It is a meta-analysis of articles and data from 1909-2016 about the history and evolution of grading in education. Feldman evaluates the equitability of this data and uses it to put together what he calls the Driving Principles—mathematically accurate grading practices, value knowledge over behavior, support hope and growth mindset, lifting the veil, and build soft skills. His use of data and focus on historically underserved communities lends to the validity of this research by fighting against implicit bias. By examining prior research, Feldman established that equitable grading is more accurate, motivational, and bias-resistant, especially for historically underserved communities. With this information, Feldman proposes that educators only use grading policies that uphold these five principles: mathematically accurate grading policies, value knowledge over behavior, support hope and a growth mindset, which lift the veil behind the whys of education, and that build on soft skills. These policies should include practices such as avoiding

zeros, weighting more recent performance of the student, being related to required content, being based on summative assessments and not formative, and allowing for retakes and do-overs.

Feldman (2019a) then took the research from his book, *Grading for Equity*, and compared it to a standards-based grading method called FAST (Fair, Accurate, Specific, and Timely). This policy takes away grades based on student behavior (late work), uses various assessment methods, and only bases course grades on summative assessments. He argues that while this policy is better than the traditional grading method, it is not equitable because it still can perpetuate race and socioeconomic inequities that have historically been apparent in our education system. He also notes that while most schools in the U.S. incorporate restorative justice discipline, culturally responsive teaching, and representation policies into our official practices, we need to catch up when it comes to updating grading policies. Feldman briefly examines the history of grading policies and how the continued use of these results in achievement gaps that disproportionately affect students of color. He then describes how implicit bias has affected many educational policies, especially regarding discipline. He uses this method of change-making and applies it to grading policies. Feldman argues that while grading policies may seem impartial, they are still promoting disparities between students who have resources, those who come to the classroom with none, or those who have a quiet place to do homework and those who do not. The evidence he has gathered shows that when equitable grading practices are implemented, it results in less failing grades and less A's, closing the achievement gap. In a district in California, the number of D and F grades was reduced by a third. Research from his non-profit, Equitable Grading Initiative, also shows that schools that implement an equitable grading policy have a higher correlation between course grades and standardized test scores.

Grading policy changes might happen at a school level, but they could also happen at higher levels, such as districts, states, or nationwide. The question then becomes, are the policies still able to be equitable? One study done in Sweden examined policy changes related to political movements during their time of implementation and whether they were intended to benefit the lowest-performing student population (Arensmeier, 2022). The questions they were trying to answer through this analysis are: What political issues have grading reforms aimed to solve, and how have the lowest performing students been considered in justifying those reforms? Unlike in the U.S., the Swedish school system has always controlled grading policies. It began with a norm-referenced numerical grading system from the 1960s to the late 1990s, shifted to a criterion-referenced system in the late 1990s, and then that system was overhauled in 2012. The author used empirical material, which consisted of policy documents spanning the entire time, and used them as discourse analysis to map the political arguments behind the grading reforms (28 documents, about 4000 pages). They took these findings and used a qualitative interpretive method to conceptualize and create a structure for their analysis. The grading reforms began to try to solve the problem of unequal access to secondary education. This initial reform aimed to take potential admission to a secondary school from standardized testing to admission based on primary school grades. It changed the grading system to a one-to-four scale, using the national population of students as a scale of point reference. Criticism of this process was that teachers were not using the national averages and instead ranking students compared to their peers (much like grading on a curve), and a shift was made to criterion-based grading. The political shift behind being focused on representation started in the 1980s. Criterion-based grading was adopted because it was based on a student's abilities rather than ranking them and created more equity in grading. The last reform came in reaction to declining results coming out of schools.

The solution was to begin grading earlier, with the argument being made that earlier grading would lead to earlier intervention and work as a motivator for educators and students.

When discussing equitable grading, different demographics of students, such as ability levels, gender identity, sexuality, religion, or race, could be examined. The intent of Cotner and Ballen (2017) was to examine whether there were gendered differences in STEM courses that graded solely on high-stakes exams (i.e., midterms and finals) and those that used mixed assessment methods, which included more low-stakes assessments (i.e., lab work, written assignments, small quizzes). The method of this research was to do two studies. The first took course data from nine Introduction to Biology courses whose final grades were based mainly on exam scores (41%-52%). A third-party individual then took that data and matched the student grades to the student's gender, age, and academic preparedness. The second study looked at courses for which the final grade was based mainly on more low-stakes work (exam grades accounted for 22%-30% of the final grade). The two studies were compared to see if women were disadvantaged when exam scores heavily weighed the course grade. The results showed that when final grades were weighted towards high-stakes exams, women performed worse than their male cohorts. When the final grades shifted towards more low-stakes assessments, women's performance on exams increased significantly compared to their male counterparts. Therefore, performance gaps between genders are more significant when course grades are more heavily weighted toward exams. The researchers took the necessary steps to ensure the validity of this study and thoroughly explained their methods, mathematics, and examination of their results.

Racial biases are also a significant concern when it comes to grading policies. Quinn (2021) proposes standardizing grading rubrics is the best way to combat this problem. Quinn wanted to prove that rubrics would combat racial bias in grading, so an experiment was

performed where teachers were given two sets of students' work; they were told one was from a black student and the other from a white student; this was done randomly to see how it would affect the grades received. As predicted, the white student received higher marks. This was true except when a grading rubric with specific criteria was presented to the teachers; then, the bias seemed to be an issue no longer, thus proving Quinn's theory. Racial bias was present when the teachers were asked to grade the student's work with the vague direction of asking if the work met specific grade-level standards. Whereas when teachers were given a rubric with specific criteria, their racial bias was eliminated. The results indicate that having standards alone is not enough to mitigate the issue of racial bias; the grading policy has to go further than that.

The work of Green (2022) continues this line of thinking on a more general level. They argue that not only is the traditional grading scale inequitable, but reducing a student to a single number, GPA, is even more inequitable as it is just another form of tracking students. Students who receive good grades early on in school are set on track for gifted/talented programs. At the same time, those students who start with poor or average grades are tracked and placed in such a way that cuts them off from opportunities in the future, including becoming one of those high-achieving students. Tracking students and assigning a GPA to them removes all nuance from grades and perpetuates the inequities in the traditional grading system. Many of these inequities stem from students' resources and socioeconomic backgrounds. Those who come to school with more get higher grades, those who come without those resources are already disadvantaged, and the grading system perpetuates that. Green concluded that student achievement could not be reduced to a single number; nuance and context are essential to assessing students' actual learning, such as those seen in standards-based grading.

The following articles will address how grading policies relate to students with different ability levels. They address how to make grading policies equitable for students with IEPs, grading in inclusive classrooms, and how grading policies are related to tracking and effort. Salend and Duhaney (2002) did a literature review combining many schools of thought for grading with an inclusive mindset. The authors look at the dichotomy between having rigorous standards and LRE causing more general education classrooms to be inclusive and how that affects grading. The authors reviewed research on grading in inclusive classrooms. They made a comprehensive list of those practices from the research which reflected best practices and followed legal guidelines for students with disabilities but would also improve grading for all students. The article lays out guidelines for determining if a grading practice is equitable, responsible, and practical while balancing the needs of all students. They did this by first establishing the purpose of grading, then examining data and literature on ten different grading policies, and finally taking the most effective practices from each and combining them into one comprehensive checklist for establishing an equitable grading policy. The author's findings show that implementing an entirely new policy is not always necessary and that needs-assessments are important if done by diverse committees. They also found that knowledge of and following legal guidelines was problematic in some studies. Overall, their focus is on consistency and adaptability of the system and the teacher to make progress through a class or school, in general, easier for the students and community, especially those from a special education background. With guidelines in place for implementing a new policy in an inclusive setting, it is important to establish what a specific policy might look like in a classroom. The purpose of Jung and Guskey (2007) was to examine how a school or classroom that used standards-based grading could adapt its grading policy to students who qualify for an Individual Education Plan (IEP) but are in an

inclusive setting. In this instance, an inclusive setting means that the students are in a general education classroom for the majority of the day, their grades are given to them by their general education teacher, and the special education teacher only reports on the progress of IEP goals. The authors establish that to reach this step, a school must first have an effective grading method in place. In this case, they are examining a standards-based grading system in which teachers consider three learning criteria: product criteria, which relate to proficiency levels; process criteria, which relate to effort and behavior; and progress criteria, which relate to how far students have come. The authors examined schools that have successfully implemented a system that meets these criteria and what they would need to do to grade students with IEPs under that system fairly. They looked at how teachers developed policies and practices for inclusive classrooms and devised an inclusive grading model. The inclusive grading model that the authors suggest, based on research and evidence from schools, has five steps. Step one is determining if accommodations or modifications are needed for each grade-level standard. The second is then to establish what those modified standards will be. The third is to outline any additional goals related to the student's academic progress. The fourth is to implement the modified policy. The fifth is to communicate the grades to other educators, therapists, families, and students in a way that communicates the meaning behind the grades.

Many students placed in intervention or even special education programs get there by tracking their abilities starting before they even enter school. Grading and tracking go hand in hand, but research shows that separating lower-performance students from their higher-performance peers is not beneficial to their growth. Ehlers and Schwager (2020) conducted a statistical analysis of two types of schools, the first which separates students based on tracking and the second which attempts to balance classrooms with low, middle, and high-achieving

students. Their results showed that mixing classrooms (what they call comprehensive schools) positively affected effort and incentive for those low and middle-achieving students. While there was a slightly negative effect on those high-achieving students, it was unimportant. Costrell and Betts (2001) research supports the idea of mixed classrooms through their meta-analysis of literature about creating high grading standards and found that raising standards would raise the effort of most students. Those students who would not rise to the occasion were more likely to have not graduated regardless, and therefore, high school attrition would have remained the same based on the policy change.

Summary

Research indicates that the traditional grading system is not working, and changes must be made. That fact has been examined through this literature review. Some researchers support standards-based grading, some minimum grading, and some amalgamating different grading policies. All agree, though, that specific criteria will make our grading policies more equitable for our students: grading policies need to reflect a more nuanced assessment of the student's abilities, high-stakes exams are not an accurate representation of academic achievement, and ranking students does not result in more effort universally. In the next section, I will apply this research to my experience working in a school.

Action Plan

Through my research on grading policies in education, I have been attempting to figure out what research says makes an equitable grading policy, which grading policies have the most research support, and what challenges or obstacles are commonly seen with the use of different grading policies. The major themes I found through this research were the importance of properly implementing a new policy, the traditional 0-100 grading scale being inequitable, and

each proposed new policy's positives and negatives. The recommendations I have given in the tables are based on these themes which emerged from the research.

The site of practice I am examining is a small school district in Puget Sound (PSD). This district has eight schools ranging from Pre-Kindergarten-12th grade. The only set grading guidance in the district is that formative work is weighted at 40% and summative work at 60%.

Based on the themes found in my literature review, I have developed three different tables of recommendations for PSD. The first is recommendations for implementing a grading policy based on known challenges and obstacles. The second is how the current grading practices in PSD might be adjusted based on what the research supports in grading policies. The third is recommendations for implementing new policies in PSD based on what some research suggest are most equitable for their student population.

Putting a Grading Policy in Place

This first table addresses the development and implementation process of a new grading policy.

Table 1

Putting a Grading Policy in Place

What research says:	What PSD is doing:	What I recommend:
A clear grading policy provides educators, students, and families with consistency and less subjectivity in the students' grades (Alex, 2022; Guskey & Link, 2019; Simon et al., 2010; Wisch et al., 2018).	There is currently no grading policy in place.	I recommend gathering a committee of educators from all grade levels, administrators, and families to decide on a formal policy and implementation process.

Providing adequate time and training through Professional Development (PD) helps ensure grading consistency (Arensmeier, 2022; Buckmiller et al., 2020; Link & Guskey, 2022; Rauschenberg, 2014).	There currently needs to be PD provided for grading policies or practices.	PD about the new grading policy should be included in summer staff week, which already requires attendance from the entire district staff. There should also be more opportunities for PD during the school year to keep teachers fresh. This training should clearly and thoroughly explain why the policy is changing, how it is changing, and why this is important. As well as time to practice practical application with feedback from the instructor.
Keeping Teachers accountable to policies increases the fidelity of practice (Puhani & Yang, 2020; Salend & Duhaney, 2002; Simon et al., 2010; Winter & Cowen, 2012).	There is no oversight for grading practices in PSD.	The individual schools with the district should organize grade-level oversight committees. Having a grade-level oversight committee to do biannual audits of grade books. This will keep teachers accountable and provide feedback for educators and policymakers. It will ensure that a policy is not just being implemented but evaluated and adjusted throughout the year.

Changes to Current Practices

This second table describes how current grading practices in PSD can be adjusted to reflect best practices.

Table 2

Changes to Current Practices

What research says:	What PSD is doing:	What I recommend:
Heavily weighting high-stakes exams creates inequitable performance gaps (Cotner & Ballen, 2017; Jönsson et al., 2021).	Grades are currently weighted 60% for assessment and 40% for classwork, homework, projects, etc.	Have all work equally weighted in the grade book.
The standard 0-100 grading scale is weighted towards failure and needs to provide students with more information or room for growth (Alex, 2022; Feldman, 2019; Toledo & Dubas, 2017).	Grades are currently based on the 0-100 scale.	I recommend moving to a minimum grading policy or a 1-4 scale, which would allow for more student growth and not be weighted toward failure.

New Grading Practices

This third table describes new grading practices that I recommend PSD to include in their grading policy.

Table 3

New Grading Practices

What research says:	What PSD is doing:	What I recommend:

Minimum grading (giving 50% as the lowest grade available, regardless of completion or lateness) provides students a better opportunity to pass a class and motivates them to try harder (Carey & Carifio, 2012; Carey & Carifio, 2013; Swinton, 2010).	PSD currently has no minimum grading policy.	I recommend implementing a minimum grading policy at the beginning of the 23-24 school year. This would allow students and families to start a new year with the new policy rather than changing it during quarter 4 of this year.
Grading that provides informative feedback improves students' mastery of a topic (Bowen & Cooper, 2022, Guskey, 2020; Wormeli, 2006).	The only feedback currently given with grades are comments selected from a prewritten list twice a year during Semester grades.	Allow teachers to provide more detailed information through the grading system. This could mean the ability to attach notes to grades on assignments through our digital platform, which provides students with more information than just a percentage grade. Teachers also need more training on giving effective feedback and what that should look like. I recommend a summer PD about this topic specifically.
Applying the theory of Standards-Based Grading (assigning clear standards that students need to be proficient at) allows students to focus on mastering the skills and standards they need rather than on their letter grade (Guskey, 2020; Huey et al., 2022; Toledo & Dubas, 2017).	There is no way to assign a clear learning outcome in the grade book for PSD.	Have learning outcomes or skills listed in the grade book to match assignments so students know what they need to work on (i.e., maybe their problem-solving process was correct but not their computational skills)? This transition would require technical support since the grade book must be adjusted. Having a program that supports this change and training to use the new program is critical for a successful transition.

Summary

In this section, I made recommendations with support from the literature about grading policies and analyzed a school district to see how their practices reflect those recommendations. I will now examine the literature more closely in the discussion section.

Discussion

This project aimed to discover what research says about grading policies in education. Through my literature review, I found that there are many different styles of grading policies with a wide range of research support. The major finding through the research, though, was that the traditional grading system with a 0-100 scale needed to be more equitable for students. While there are many different possibilities for grading policies, empirical data on each can take time to come by. Certain subjects did emerge, though educators are concerned about the equitability of grading, and this is the driving force behind wanting policies to be changed. Two grading policies stood out in that discussion of equitability: standards-based grading and minimum grading policies. The biggest obstacle that the research revealed was consistency in grading, even after a policy is in place. These topics are consistent with what I have seen throughout my teaching career. The first school I taught at had a standards-based grading system which was followed some of the time, and the school I teach at now has no grading policy and little consistency between classrooms.

In this section, I will attempt to come to some conclusions about my initial questions and discuss the possible successes and limitations of grading policies at different levels of implementation (school, district, state, etc.). I will also posit what the implications of this

research are for educators and policymakers, as well as speculate on future research in this area of study and discuss the limitations of my project.

Discussion of Findings

In this section, I will return to the three questions I asked in the rationale at the beginning of this paper. Those questions were:

- What makes for an equitable grading policy?
- What grading policies have the strongest research support?
- What are the challenges or obstacles to different grading policies?

In addition, I will discuss the benefits and limits of grading policies at different levels of implementation (i.e., school, district, state, or nationwide).

Equitable Grading Policies

My first question was, what makes for an equitable grading policy? At the very start of this paper, I mentioned that through my initial research, the one glaring thing that stood out to me was that the traditional grading system was not working. Almost every article I encountered mentioned that the 0-100 grading scale was skewed toward failure, making it naturally inequitable (Alex, 2022; Brookhart et al., 2016; Ehlers & Schwager, 2020; Feldman, 2019b). This led me to my first research question and to figure out what research says makes for an equitable grading policy. To try and find an answer to this question, I had to decide what equity in grading meant to me, and one would hope that we could say that if you take the subjectivity out of grading, it will become more equitable. However, when you add in factors like Individual Education Plans (IEPs), 504 plans, and other differentiation, there has to be some level of subjectivity in grading. For example, I currently have a student named Laurie (pseudonym), who has an IEP that says she only has to finish half of her work because she struggles with stamina.

She will often set up all her homework problems at once, then go back and start working through them. Eventually, she runs out of steam and starts to make silly mistakes in computation. I typically will not lower her grade for these mistakes since I can see she understands the process, and that is really what I am looking for out of the assignment. Without subjectivity in grading, Laurie's grade would not reflect her true abilities. Feldman (2019b), one of the leading researchers in equitable grading, said, "Grades should be based on valid evidence of a student's content knowledge…How we grade should motivate students to achieve academic success, support a growth mindset, and give students opportunities for redemption" p.101). This implies that grading should be used as a way to motivate and help students grow rather than punish or pit students against each other. This quote became the lens through which I was dissecting equity in grading policies. There are other aspects of Feldman's philosophy that I agree with also. For example, he believes that for a grading policy to be equitable, zeros should be eliminated (much like the minimum grading policy); he also states that switching to a 1-4 scale instead of 0-100 makes grades more equitable and easier to understand and that student's grades should be made in such a way as to give good feedback. However, Feldman also believes that students' behavior should not be included in the grading process. When talking about behavior, he specifically refers to things like participation, lateness, effort, or even attendance. I disagree with this idea and am more inclined to agree with Guskey (2020) on this topic. He believes that student behavior should be a part of a student's grade since they are an important part of their success in school. Where behavior items need to be incorporated into standards is through the process criteria, which refer to the day-to-day goings on in a classroom. I also believe that included in this part of the grading criteria should be collaboration with peers, as this is such an important social-emotional skill that students learn in school. From the perspective of a teacher, a lot of the

time, when something is not compulsory, the students will choose not to do it. While some may argue that it is a choice that only will cause that student to suffer, I would disagree. Any teacher will tell you that if one student is off task, others will follow, which becomes an even bigger distraction to those trying to learn. It follows that this distracting environment is inequitable for students who need a calm, controlled environment to learn and grow. While I do not think these soft skills and social-emotional skills should be a majority, or frankly even a large percentage, of a student's grade, they are important and will help students be better citizens of the world after they leave school.

Another major topic emerging from the research was that for grading policies to be equitable, they must be adaptable and never be set in stone. Salend and Duhaney (2002) wrote a brief guide on selecting and implementing a new grading policy. Included are steps I mentioned in the Action Plan, like creating a committee to identify the needs and wants of the students, teachers, and families. The last step of their process stuck with me, though, "Evaluate the impact of grading policies and practices on students, family members, educators, the community, and the curriculum continuously and revise accordingly" (Salend & Duhaney, 2002, p. 14). That very last part is the most important to me, evaluation of the grading policy should be continuous and adapted when needed. This could mean the system as a whole, a group of students, or even just one individual student. We adapt and modify assignments and lessons based on the needs of our students. Why would we not extend that work into our grading policies? If there is anything that becomes glaringly obvious when you become an educator, it is this: one size will never fit all! Last year I would teach the same lesson 5 times a day, and every single day, every class was different. I would have to adjust what I would emphasize in the lesson, how I would manage the classroom, and what expectations I would have for successful lesson completion. Adaptability is

one of our greatest strengths as educators, and it needs to extend into our grading policies and practices. As a teacher, I have never had anyone check up on my grading process, and the only help I have ever received was when I specifically sought it out. The example I gave of my student Laurie at the beginning of this section was one of those times. Her IEP said, “Laurie will complete 50% of assigned work tasks.” There was no further direction than that, and since I was not a part of the IEP team when the goal was written, I had no background knowledge of the student to go on. This meant I had to seek information from counselors, administrators, and her previous teachers, all of whom had different suggestions on how to proceed. I was being adaptable, as suggested above, and decided I would wait until I had a solid amount of work from Laurie before I decided how to proceed. I knew that it was important for me to realize that being adaptable did not mean lowering standards for Laurie. I held her to the same standards as her peers but reduced the work she had to do completely; that is what adapting and making modifications meant to me.

Knowing the community, the needs of the students, and the feelings of the parents and educators are vitally important to the success of a new grading policy and whether it impacts the students. Jung and Guskey (2007) suggested using small needs-assessments for the grading standards themselves to make them more equitable. They proposed assessing the standards and what modifications or accommodations might be necessary to make them more inclusive and sufficiently communicate the meaning behind the grades students are getting. The major issue with assessing standards is that more work must be done ahead of time, and it should not be done on an individual teacher level. This is where a committee of different stakeholders (teachers, administrators, families) is so important. There must be some oversight to conduct needs-assessments and adjust if the grades prove to be inequitable for the students. This requires a

commitment to be made on, at the very least, a schoolwide level. In their 2007 article, Jung and Guskey said, “Educators at all levels desperately need clear and specific guidance in developing grading and reporting policies and practices…They also need concise and meaningful data on the effects of such policies and practices” (p. 52). For a grading policy to be equitable, it has to be meaningful for all students. This is precisely what the committee and needs-assessments are for, to gather meaningful data and assess what is working and what needs to change. This is important for any grading policy, new or old, and frankly, not something I have ever seen as a teacher.

Grading Policies with Research Support

The second question I asked at the beginning of this research was, what grading policies have the strongest research support? There are two grading systems that seem to have research support. The first is standards-based grading and the second is minimum grading.

Standards-based grading is one of the grading systems with research support. Toledo and Dubas (2017) discussion of standards-based grading explained that by showing a student the exact things you are grading on and looking for, they would understand what they did well and where they need to improve to align with where they should be at their grade level. It is all about transparency in grading. Link and Guskey (2022) agree with this assessment and say that the what makes standards-based grading effective is that student grades are reported using specific performance categories based on standards surrounding the particular skill the students are trying to master. Zimmerman (2017) expanded on this expressing that standards alone are not specific enough but should have focused learning objectives attached to them. This way students receive meaningful feedback and know where they need to improve to achieve proficiency. Having focused learning objectives makes sense if you are familiar with common core standards, which

standards-based grades are usually derived from. They tend to be too broad and therefore do not give the students enough feedback about where they need to improve. Providing educators with explicit data and giving student explicit feedback are important aspects of this grading system which are not sufficiently provided by common core standards alone. This means that in order for standards-based grading to be successful more work needs to be done on a district or even school wide level to hone in on the exact objectives students need to achieve for academic success.

The other grading policy currently gaining popularity is called minimum grading, sometimes known as the "No Zero Policy." Much like its name implies, this policy uses the traditional grading scale, but instead of grading students from 0-100, it shifts the scale to 50-100. This means that the lowest grade a student can earn, regardless of completeness or effort, the lowest score a student can receive is 50%, which I would like to point out is still a failing grade. Carey and Carifio (2009; 2012; 2013) are the leading researchers on this policy. They argue that the grading policy eliminates outlying scores that might negatively affect a student's future in a class and cause a student's motivation and effort to diminish if they feel there is no digging themselves out of a hole. They also point out that when deciding on any grading policy, it is key to remember that grades are there to "reflect fair and accurate assessments of student accomplishment and achievements" (Carey & Carifio, 2009, p.37). If a school implements a policy in which students will be assigned one final grade, is allowing their lowest points in the year to affect their entire outcome fair and accurate? I do not believe it is. However, many disagree, and this is a hotly debated educational policy (Caneva, 2014). In my personal experience, the biggest complaint I hear from teachers is that if students put in zero effort, they should not receive anything more than a zero grade. While I understand this argument, I would

point out two things. First, if a student does want to turn their grade around and zeros bog down their overall grade, there is a strong chance they will never catch up and, therefore, might think, what is the point of trying? I never want my students to think there is no hope for them. Second, if students do not want to put in the effort and receive a 50% on every assignment, they will still fail. A minimum grading policy is not a making wand suddenly passing every student; it is simply a way to make the traditional grading scale fairer.

Challenges or Obstacles of Grading Policies

Throughout reading the literature selected for this project, the authors made numerous comments that could have been labeled challenges or obstacles of grading, creating grading policies, or implementation. However, my job was to boil it down to the most significant obstacles. I found a couple of running themes through writing this section and trying to answer this question. First, some grading policies make grades into a competition detrimental to student success (Arensmeier, 2022; Bowen & Cooper, 2022; Calsamiglia & Loviglio, 2019). Second, basing grades on high-stakes testing does not accurately represent students' mastery of a topic (Cotner & Ballen, 2017). Lastly, even with the same student, inconsistent grading among educators shows that having a policy alone does not mean everyone is doing the same thing or is equitable (Rauschenberg, 2014).

Grading policies that result in ranking students against each other are not beneficial for student growth, do not motivate most students, and can easily be thrown off by one outlier. Students with access to more resources than their peers, whether tools, time, or outside support, have an advantage over their peers (Bowen & Cooper, 2022). Grading on a curve or any ranked grading gives those students an unfair advantage over their peers without access to the same resources. The research is clear that students can be ranked against themselves or national

averages, but the competition created by ranking them against each other is not equitable (Arensmeier, 2022). I have seen this happen in a classroom where one student scores 100% on a test and throws off the entire curve for everyone else. Calsamiglia and Loviglio (2019) discuss this and argue that this practice will never work unless you separate classes by ability level, which is also inequitable. I agree and would gladly state that the negatives far outweigh the positives regarding grading on a curve.

The second consistency I saw in the research when discussing challenges was grading policies that base a student's overall grade on high-stakes exams. Cotner & Ballen (2017) were trying to determine if high-stakes exams disproportionately affected female students in a biology course. However, their results can be applied to all students. Their results suggest that the more impact an exam will have on a student's overall grade, the bigger the gaps in the total scores are. Students tend to have similar grades if the exam has less impact on overall grades. This speaks to the impact of high-stakes exams and how they do not measure a student's true abilities. I have seen this in my teaching; when I call an assessment a test or a quiz, there is panic, even among my higher-performing students. However, when I express that the assessment is a mini-quiz or only intended for my information, my students are much more relaxed and get better scores overall.

The last major obstacle I found through the research was inconsistency in grading, even when a policy is in place. Rauschenberg (2014) specifically addressed this in the article about differential grading. Their research found that a student's demographics were the biggest determiner of differential grading, which is a considerable problem of practice. Link and Guskey (2022) suggest that to combat differential grading and remove subjectivity, clear guidelines must

be in place. Knowing precisely what is being graded, whether the student's process, structure, or computation, will help educators remove some of the bias from their grading.

Grading Policy Implementation Impact

When a grading policy is established, it can happen at different levels: school-wide, district-wide, state-wide, or nationwide. As mentioned in the introduction to this paper, there is no national grading policy in the United States. However, there are other countries where grading policies are overseen and regulated by the national government. Klapp (2015) did a study about this in Sweden. In it, she was trying to determine the long-term effects of the grading policy. She discovered a direct negative impact on those students that began receiving grades at the primary level (before 7th grade). Since the grading policies were being set nationally, districts could not adjust the policy for their communities. The inability to adjust might also be an issue at a state level, although no research supports that, as no states have a regulated grading policy. In the United States, the biggest level grading policies are implemented are districts. However, the same problem of not being able to adjust for the school community could also be true for larger school districts. I have worked in two school districts while teaching, one larger and one smaller. The larger district had a grading policy but needed more regulation and educator training. While they were trying to implement a tool for educators and have consistency, the lack of training made the policy fail to be successful. Most educators adjusted their grading style to fit the policy. In the smaller district I worked in, there was no grading policy, and while there were active talks about aligning grade books, when there was no agreement on what the policy should be aligning fell by the wayside. Salend and Duhaney (2002) express the importance of balancing the needs of students within the grading policy and the importance of knowing your school communities and their needs. This is the precise reason I suggested in my Action Plan to create a committee

that involves all stakeholders so that an agreement could be made. There could be an oversight committee on the grading policy to ensure proper implementation and consistency in the grading process.

Implications for Education

Districts need to have a clear and consistent grading policy in place. Successful implementation of the policy requires the establishment of a committee to decide on the policy, put it in place, and ensure it is being enacted with fidelity. Research reveals implications that affect both educators and policymakers. Those responsible for putting new policies in place will first have to agree on a set of guidelines to follow, garner educator support, implement the policy, including accounting for proper training time, and then ensure fidelity of the new policy once it is in place. Educators will have to use their training and learn a new way of grading and thinking about their student's grades. For a major shift like this to be successful and accepted, the educators enacting the change must have buy-in. They have to want it to work, and while this is a major implication for the policymakers, it is also on the educators to check in on themselves.

Recommendations for Future Research

As stated in the introduction to this section, the lack of empirical research on different grading styles was startling. This is one of the most important aspects of education. It is how we place students, how they get into higher education or apprenticeships, and possibly even can affect their job prospects. There has to be data to back up the changes we want to make. Otherwise, it is just a bunch of words.

Future research on this topic will hopefully also contain more longitudinal data, especially regarding standards-based and minimum grading policies. I also think that future

research should include grading policies at different levels of implementation as it will show the strengths and weaknesses those policies pose.

Limitations of the Project

When conducting my research for this project, I decided to limit myself to not using any literature written before 2001. I tried my best to even limit myself further to anything from 2010 to 2022. I wanted to have the most recent data possible. I started with my second parameter of 2010-2022. However, the data needed to be improved, so I had to expand my search dates. I did not choose 2001 arbitrarily, but it was a focused decision based on the enacting of No Child Left Behind. This law was a major shift in educational thinking and policymaking; therefore, I wanted to ensure that the research I analyzed considered that shift.

The search term I began with for this paper was "grading policies in education." However, I soon found that I would have to be more specific to get more empirical data. I made a list of the different grading policies mentioned in the research and started using those as search terms. For example, standards-based grading, minimum grading policy, and grading on a curve, as those were the policies I saw discussed the most. For an article to be included in this paper, I wanted to make sure it was drawing information from data or prior research with included data or (in a couple of cases) from the experiences of educators in schools implementing grading policies. The research I excluded was theoretical primarily as I felt there was insufficient information without data to support their claims or about grading policies I did not see mentioned elsewhere.

The most significant limitation I encountered in this project was the need for more empirical evidence regarding grading policies. I wanted to find longitudinal studies that showed the impact of having one specific policy versus another—alternatively, even data from schools

that had changed grading policies. Recent and relevant data took much work, and I had to do some digging to find valuable sources. I even contacted districts in my area to see if they had unofficial data for me. I was disappointed that it did not exist in the capacity I sought.

Conclusion

Grading has a considerable impact on the lives of our students. In part, grading defines their confidence in their academic abilities, it can put them on specific tracks which play a part in determining their future, and it impacts where they can go to college or if colleges will even consider them for admissions. These are huge, life-altering things that grading can impact. Choosing a grading policy is a complex problem that educators, administrators, and policymakers need help to complete successfully. Educators must ensure that students learn the content and are motivated to continue their learning journey through assessment and feedback. Schools need to have equitable, research-supported policies in place that help students continue to grow and thrive. Students and educators need consistent, clear, and meaningful grading policies and practices that allow them to do so. Ultimately, that is what I have explored through this project, the best ways to help our students through grading.

References

Alex, P. (2022). Time to pull the plug on traditional grading? Supporters of mastery-based grading say it could promote equity. *Education Next, 22*(4), 38–44.

Arensmeier, C. (2022). Institutionalizing school failure: From abandoning to reintroducing a failing grade—the rationales behind Swedish grading reforms. *Journal of Educational Change, 23*(2), 221–252. https://doi.org/10.1007/s10833-021-09421-7

Bowen, R. S., & Cooper, M. M. (2022). Grading on a curve as a systemic issue of equity in chemistry education. *Journal of Chemical Education*, *99*(1), 185–194. https://doi.org/10.1021/acs.jchemed.1c00369

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 Education Research*, *86*(4), 803–848. https://doi.org/10.3102/0034654316672069

Buckmiller, T., Townsley, M., & Cooper, R. (2020). Rural high school principals and the challenge of standards-based grading. *Theory & Practice in Rural Education*, *10*(1), Article 1. https://doi.org/10.3776/tpre.2020.v10n1p92-102

Calsamiglia, C., & Loviglio, A. (2019). Grading on a curve: When having good peers is not good. *Economics of Education Review*, *73*, 101916. https://doi.org/10.1016/j.econedurev.2019.101916

Caneva, G. (2014). For students' sake, say no to "No-Zero Policy" on grading. *Education Digest*, *79*(7), 52–54.

Carey, T., & Carifio, J. (2012). The minimum grading controversy: Results of a quantitative study of seven years of grading data from an urban high school. *Educational Researcher*, *41*(6), 201–208. https://doi.org/10.3102/0013189X12453309

Carifio, J., & Carey, T. (2009). A critical examination of current minimum grading policy recommendations. *High School Journal*, *93*(1), 23–38.

Carifio, J., & Carey, T. (2013). The arguments and data in favor of minimum grading. *Mid-Western Educational Researcher*, *25*(4), 19–30.

Costrell, R. M., & Betts, J. R. (2001). Incentives and equity under standards-based reform. *Brookings Papers on Education Policy*, *2001*(1), 9–74. https://doi.org/10.1353/pep.2001.0001

Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology classes more equitable? *PloS One*, *12*(12). https://doi.org/10.1371/journal.pone.0189610

Dardanoni, V., Modica, S., & Pennsi, A. (2009). Grading across schools. *The B.E. Journal of Economic Analysis & Policy*, *9*(1). https://doi.org/10.2202/1935-1682.2065

Diegelman-Parente, A. (2011). The use of mastery learning with competency-based grading in an organic chemistry course. *Journal of College Science Teaching*, *40*(5), 50–50.

Ehlers, T., & Schwager, R. (2020). Academic achievement and tracking – A theory based on grading standards. *Education Economics*, *28*(6), 587–600. https://doi.org/10.1080/09645292.2020.1808594

Elikai, F., & Schuhmann, P. W. (2010). An examination of the impact of grading policies on students' achievement. *Issues in Accounting Education*, *25*(4), 677–693. https://doi.org/10.2308/iace.2010.25.4.677

Feldman, J. (2019a). Beyond standards-based grading: Why equity must be part of grading reform. *The Phi Delta Kappan*, *100*(8), 52–55.

Feldman, J. (2019b). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms.* Corwin, a SAGE Company.

Figlio, D. N., & Lucas, M. E. (2004). Do high grading standards affect student performance? *Journal of Public Economics*, *88*(9), 1815–1834. https://doi.org/10.1016/S0047-2727(03)00039-2

Green, L. (2022). [GPA] in, [GPA] out: Uncovering inequity and flaws in grading policies. *AASA Journal of Scholarship Practice*, *18*(4), 40–50.

Guskey, T. R. (2020). Breaking up the grade: To make grading more meaningful, course grades should reflect various criteria that makeup student learning. *Educational Leadership*, *78*(1), 40–46.

Guskey, T. R., & Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, *26*(3), 303–320. https://doi.org/10.1080/0969594X.2018.1555515

Harbour, R., & Miller, J. (2001). A new system for grading recommendations in evidence-based guidelines. *BMJ: British Medical Journal*, *323*(7308), 334–336.

Huey, M. E., Silvey, P. R., Vaughan, A. G., & Fisher, A. L. (2022). Assessing the impact of standards-based grading policy changes on student performance and practice work completion in secondary mathematics. *Studies in Educational Evaluation*, *75*, 101211. https://doi.org/10.1016/j.stueduc.2022.101211

Jönsson, A., Balan, A., & Hartell, E. (2021). Analytic or holistic? A study about how to increase the agreement in teachers' grading. *Assessment in Education: Principles, Policy & Practice*, *28*(3), 212–227. https://doi.org/10.1080/0969594X.2021.1884041

Jung, L. A., & Guskey, T. R. (2007). Standards-based grading and reporting; a model for Special Education. *Teaching Exceptional Children*, *40*(2), 48–53. https://doi.org/10.1177/004005990704000206

Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy & Practice*, *22*(3), 302–323. https://doi.org/10.1080/0969594X.2014.988121

Link, L. J., & Guskey, T. R. (2022). Is standards-based grading effective? *Theory into Practice*, *61*(4), 406–417.

Lu, O. H. T., Huang, A. Y. Q., & Yang, S. J. H. (2021). Impact of teachers' grading policy on identifying at-risk students in learning analytics. *Computers & Education*, *163*, 104109. https://doi.org/10.1016/j.compedu.2020.104109

Marzano, R. J. (2017). The new art and science of teaching. Solution Tree Press.

National Association of Secondary School Principals. (2019). Class rank, GPA, and grading. *NASSP*. https://www.nassp.org/class-rank-gpa-and-grading/

Nowacki, A. (2013). Making the grade in a portfolio-based system: Student performance and the student perspective. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00155

Paredes, V. (2017). Grading system and student effort. *Education Finance and Policy*, *12*(1), 107–128.

Puhani, P. A., & Yang, P. (2020). Does increased teacher accountability decrease leniency in grading? *Journal of Economic Behavior & Organization, 171*, 333–341. https://doi.org/10.1016/j.jebo.2019.12.017

Quinn, D. M. (2021). How to reduce racial bias in grading: New research supports a simple, low-cost teaching tool. *Education Next, 21*(1), 72–79.

Rauschenberg, S. (2014). How consistent are course grades? An examination of differential grading. *Education Policy Analysis Archives, 22*, 92–92. https://doi.org/10.14507/epaa.v22n92.2014

Salend, S. J., & Duhaney, L. M. G. (2002). Grading students in inclusive settings. *Exceptional Children, 34*(3), 8–15. https://doi.org/10.1177/004005990203400301

Scarlett, M. H. (2018). "Why did I get a C?": Communicating student performance using standards-based grading. *Insight, 13*, 59–75. https://doi.org/10.46504/14201804sc

Senko, C., & Miles, K. M. (2008). Pursuing their own learning agenda: How mastery-oriented students jeopardize their class performance. *Contemporary Educational Psychology, 33*(4), 561–583. https://doi.org/10.1016/j.cedpsych.2007.12.001

Simon, M., Tierney, R., Forgette-Giroux, R., Charland, J., Noonan, B., & Duncan, R. (2010). A secondary school teacher's description of the process of determining report card grades. *McGill Journal of Education / Revue Des Sciences de l'éducation de McGill, 45*(3), 535–554. https://doi.org/10.7202/1003576ar

Swinton, O. H. (2010). The effect of effort grading on learning. *Economics of Education Review, 29*(6), 1176–1182. https://doi.org/10.1016/j.econedurev.2010.06.014

Toledo, S., & Dubas, J. M. (2017). A learner-centered grading method focused on reaching proficiency with course learning outcomes. *Journal of Chemical Education*, *94*(8), 1043–1050. https://doi.org/10.1021/acs.jchemed.6b00651

U.S. Department of Education. (2008). *Structure of U.S. Education: Evaluation and Assessment*. Retrieved May 5, 2023, from https://www2.ed.gov/about/offices/list/ous/international/usnei/us/edlite-evaluation.html

Winters, M. A., & Cowen, J. M. (2012). Grading New York: Accountability and student proficiency in America's largest school district. *Educational Evaluation and Policy Analysis*, *34*(3), 313–327.

Wisch, J. K., Ousterhout, B. H., Carter, V., & Orr, B. (2018). The grading gradient: Teacher motivations for varied redo and retake policies. *Studies in Educational Evaluation*, *58*, 145–155. https://doi.org/10.1016/j.stueduc.2018.06.005

Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education*, *34*(3), 14–27.

Zimmerman, T. (2017). Grading for understanding – Standards-based grading. *The Physics Teacher*, *55*(1), 47–50. https://doi.org/10.1119/1.4972500