




# Family Science Review

ISSN: 2331-6780

[www.familyscienceassociation.org/journal](http://www.familyscienceassociation.org/journal)

**FSA**  
family science  
association

## Using Grade Distributions for Assessment of a Child and Family Science Program

Kathleen D. Dyer, Child and Family Science, California State University Fresno 

---

**ABSTRACT.** Outcomes assessment in an academic family science program led to the accidental discovery of grade inflation that was causing impaction problems in upper-division major courses. The current analysis evaluates the effectiveness of a policy intervention designed to improve academic rigor in previously grade-inflated classes. The new policy required an annual grade distribution report placed in faculty personnel files in order to raise awareness of grading practices. Both before and after the policy change (2011 and 2018), we analyzed three data sources: course grade distributions, an objective, comprehensive exam for graduating seniors, and anonymous student surveys in classes. After the intervention, the percentage of As earned in classes was dramatically reduced, while performance on the comprehensive exam was dramatically improved. The number of failing grades remained constant, and student perceptions of classes in the program remained unchanged.

*Keywords:* grade inflation, grade distributions, program assessment, higher education, grading

---

### Cite this article (APA 7):

Dyer, K. D. (2023). Using grade distributions for assessment of a Child and Family Science program. *Family Science Review*, 27(2). <https://doi.org/10.26536/ZVHK6341>

---

Published online: August 2023

[Submit your article to this journal](#)

Direct correspondence to: Kathleen D. Dyer, Department of Child and Family Science, California State University, Fresno, 5300 N. Campus Drive, FFS 214, Fresno, CA 93740; e-mail: [kdyer@mail.fresnostate.edu](mailto:kdyer@mail.fresnostate.edu)

### **Using Grade Distributions for Assessment of a Child and Family Science Program**

Outcomes assessment in higher education is a fairly recent innovation, one that is about both accountability to outsiders and promoting reflective practice within academic programs (Ewell & Cumming, 2017). Programs generally self-assess by using objective measures of student learning (Kinzie, 2019), but other aspects of a program may be used to inform the assessment of its effectiveness. In the current study, analysis of objective measures of learning, subjective measures of perception, and grade distributions were used to first identify a problem with grade inflation in courses in a Child and Family Science program and then to evaluate the result of an intervention designed to address the problem.

### **Grading in College Classes**

The A-F system of letter grades has been in place in U.S. colleges as far back as the 1880s. It appears to have replaced a numeric grading scale (0 to 100) because of concerns about the reliability of such a fine-grained scale (Schinske & Tanner, 2014; Brookhart et al., 2016).

Schools first used grades to communicate with other schools about the adequacy of a student's work (Schinske & Tanner, 2014). Those grading systems needed to be reliable and somewhat comparable between institutions (Smallwood, 1935). This function of identifying the quality of performance to a third party remains a primary purpose of grading (Schinske & Tanner, 2014). For example, grades are used to facilitate the sequencing of college classes by using the grade from a prerequisite class to indicate readiness for a more advanced class. In addition, grades are used by graduate schools to consider prospective students and by employers to judge the credentials of job applicants (Johnson, 2003; Milton, Pollio, & Eison, 1986).

Grades do not appear to be an effective form of feedback if the goal is to help students improve their performance; descriptive feedback can improve performance in the rare circumstance that students read it and consider it, but evaluation of work, either written or in the form of a grade, does not (Schinske & Tanner, 2014). Neither do grades motivate most students to work harder for good grades, although they may stimulate some fear-based efforts to avoid bad grades (Schinske & Tanner, 2014; Chamberlin, Yasue, & Chiang, 2018). Therefore, the utility of grading practices for pedagogy is not considered in this analysis. Instead, grades are considered here to be indicators of academic performance for third parties (educators and employers).

### **Grading Philosophies**

There are two basic approaches to the assignment of grades for academic coursework: norm-referenced and criterion-referenced. Norm-referenced grading (Finkelstein, 1913) requires that grades be distributed by a normal bell curve or by predetermined quotas. It is sometimes called grading "on a curve" for this reason. With a norm-referenced grading system, grades are determined by how students compare to their classmates, independent of their mastery of course content. Presumably, this standardizes grading such that when students compete for entry to graduate school or a job, their grades are a reasonable basis for comparing one to the other. It also helps mitigate the effects of poor-quality grading tools by normalizing outcomes to overcome some types of measurement bias.

Norm-referenced grading is problematic for many reasons, not the least of which is that it creates competitiveness rather than cooperativeness among classmates, thereby actively impeding learning (Dubey & Geanakoplos, 2010). It is also associated with a belief in meritocracy and ability as something fixed and immutable (Terwilliger, 1993). The conception of academic ability as a fixed trait is now largely rejected by those who study learning (e.g., Dweck, 2007). These beliefs associated with norm-referenced grading can lead to an exclusion from the academy of those who have been traditionally disenfranchised (Terwilliger, 1993). Due to the distribution of grades over the

bell curve, norm-referenced grading in college classes is associated with higher numbers of low grades (Geisinger & Rabinowitz, 1980).

The primary alternative is criterion-referenced grading (Glaser & Klaus, 1962), whereby learning goals and grading standards are explicitly delineated in advance. Students are graded according to their mastery of knowledge and skills that are identified as objectives of the course. By this system, all students can earn an A if they all demonstrate mastery of the learning objectives identified for the course. It is equally true, however, that no students might earn an A if none demonstrate excellent mastery of the content. The National Grading Survey (Milton, Pollio, & Eison, 1986) found that most students, professors, parents of students, and employers prefer the criterion-referenced system for the assignment of grades.

Some versions of criterion-referenced grading demand that learning objectives are the only criterion on which grades should be assigned (Brookhart, 2016). However, research on grading reveals that this is rarely the case in practice; instructors generally use both cognitive performance and effort/participation to assign grades (Brookhart, 2016). Either way, a class can have a high or a low percentage of As with this grading system, depending on the performance of the students in the class.

Criterion-referenced grading is sometimes (but not always) accompanied by a mastery model, whereby students are allowed multiple attempts to achieve mastery of the learning objectives (Block, 1971). With a mastery model, all students might achieve an A grade if they are adequately supported in their efforts to learn, and they all put in adequate effort. The rationale that a mastery model necessarily produces all A grades has been criticized because “mastery” may or may not mean “excellence” (Block, 1971). It might be more reasonable to believe that, with multiple attempts, all students may achieve a basic level of competence rather than to believe that all students may achieve excellence. If so, grades A, B, and C can still reflect different levels of mastery even when there are no failing grades as a result of using the mastery model (Ebel, 1974).

There is considerable variability in grading practices in educational settings (Geisinger & Rabinowitz, 1980; Brookhart et al., 2016). Practically speaking, grading typically reflects some combination of norm-referenced and criterion-referenced standards (Lok, McNaught, & Young, 2016).

### **Grade Inflation in College Classes**

A classic definition of grade inflation is a “grade rise without increased achievement” (Bejar & Blew, 1981). By this definition, there is strong evidence that grade inflation has been occurring in U.S. colleges and universities since the late 1960s. While college faculty claim that they regard C as an average grade, B as above average, and A as exceptional (Johnson, 2003), grades given in college classes seem to imply that virtually all students are above average. A review of grades awarded in college classrooms based on data from more than 9 million students attending over 1500 accredited colleges and universities found that the number of As awarded grew dramatically in the 1960s and again, but more gradually, in the 1990s (Astin, 1998). Grades have continued to go up since then in U.S. colleges and universities (Rojstaczer, 2016). A historical analysis of data from the College Board concluded that grade inflation could not be accounted for by student demographics, patterns of course-taking, or anything other than the year they entered college, suggesting that grading leniency is the best explanation for better grades (Kostal, Kuncel, & Sackett, 2016).

A comprehensive historical review of grade distributions in U.S. universities (Rojstaczer & Healy, 2012) revealed that only 15% of grades were As in 1940, but by 2008, 45% of grades were As. Furthermore, the percentage of failing grades (Ds and Fs) has remained fairly constant over time, but the number of Cs has dramatically declined as Cs have become Bs, and Bs have become

As (Rojstaczer & Healy, 2012). It appears, therefore, that there is a compression of grades at the high end of the scale.

There are changes in the habits of students that would suggest that it is unlikely that they are becoming more homogeneously excellent. An analysis of time use data from 1961 to 2004 indicates that full-time college students spend far less time in educational pursuits (in class and studying) than they used to: 40 hours per week in 1961, compared to 27 hours per week in 2004 (Babcock & Marks, 2010). This drop in study hours happens whether or not students are employed outside of school (Babcock & Marks, 2010). By 2016, the number had dropped further to 19 hours per week (Reim, Amselem, & Hall, 2016). Therefore, it is hard to imagine how their learning could be so drastically improved over the years to justify the increasing number of them earning above average grades in their classes.

This inflation of grades in colleges and universities in the United States has been called a “crisis in college education” (Johnson, 2003). Grade inflation is likely to produce a fundamental inequity in that students’ career prospects are determined partly by the degree of grade inflation present in their discipline and the particular sections of classes they enroll in rather than by their mastery of course content. Furthermore, as students use grading rigor as a basis for choosing classes where choice is possible, instructors who demand rigor see lower enrollments and poorer student ratings. These outcomes have consequences for faculty members’ career prospects and therefore serve to force faculty to abandon rigorous teaching, which brings down the level of education for students and the level of scholarship in the academy overall (Johnson, 2003).

Rojstaczer and Healy (2012) assert that student grades are so inflated that they have become meaningless as tools by which employers and graduate schools can evaluate students. If grades only serve the purpose of signaling to students which classes are the easiest ones and nothing else, then they are worse than useless (Johnson, 2003; Rojstaczer & Healy, 2012).

Perhaps the most compelling argument against grade inflation is that it produces students who are ill-prepared for their next challenge, be it a higher level course or application of the knowledge in a job setting. One analysis of students who took a prerequisite algebra class at a community college versus a four-year university found that the community college students had higher grades in the algebra class but subsequently performed worse in the next level math class (Friedl, Pittenger, & Sherman, 2012) thus setting them up for an inevitable failure. When grades are high but do not reflect knowledge or skills, graduate programs, and employers cannot determine which candidates are better prepared. As a result, students and graduates are put into positions for which they are fundamentally unprepared, which in turn damages the reputation of the academic program and the discipline responsible for training them.

### **Implications of Grade Inflation in Family Science**

In higher education, in general, instructors tend to rely on process variables (such as effort) to assign grades (Lipnevich et al., 2020; Zinn et al., 2011). Qualitative research with college professors has identified a troubling trend: professors believe that grade inflation is a problem, but they do not believe that it exists in their own classes or department. In fact, they estimate themselves to be “tough” graders and seem not to be aware that their own grade distributions appear quite inflated (McCabe & Powell, 2004). Scholars have identified grade inflation in social work, a field related to family science, in part by comparing students’ earned grades in social work classes to the same students’ earned grades in other coursework (Miller, 2014).

We have no record of published empirical investigations of grading practices in family science programs. However, it is reasonable to assume that this pattern applies within our discipline. Grade inflation in family science programs may be especially likely because faculty in family science programs are acutely aware of equity, social justice, and systemic disenfranchisement of vulnerable populations, as this is a focus of research within the Family Science discipline (e.g.,

Blaisure, 2003; Letiecq, 2019).

Therefore, faculty in family science programs may be inclined to be compassionate and supportive, even to the point of exhaustion from the amount of emotion-work we do with our students (Mahoney et al., 2011). Even while the faculty may be inclined toward grade inflation as a result of a deep commitment to social justice, grade inflation is especially problematic for a discipline such as family science. Hamon and Smith (2014) argue that family science is at a historical moment that requires self-evaluation and innovation to develop the identity of our discipline. Perhaps due to the lack of a common nomenclature to describe our discipline, academic family scientists have the sense that our discipline is not widely recognized as an academic discipline (Hamon & Smith, 2014). Teaching classes that are known for the “easy A” may not help our reputation within the academy or outside of it. If students, and then our academic colleagues and administrators, and then our community partners see our discipline as trivial or expendable, then we may be less able to do our important work.

Furthermore, what starts as sensitivity and kindness to students can actually result in a disservice to them and potentially the community. When grades are inflated in lower-level academic coursework, there may be students who go into upper division coursework unprepared and unable to succeed. Upon graduation, some of those students may then enter into jobs for which they are fundamentally unprepared. These are jobs that provide service to and advocacy for the most vulnerable members of our society (Walker & Blankemeyer, 2013). Sending unprepared students to these jobs is a disservice to the students and a disservice to the children, families, and communities that they serve. Thus, the purpose of this study was to explore the impact of a department-level intervention on both the amount of grade inflation and the academic preparedness of students in a Child and Family Science undergraduate program.

### **Suspicion of Grade Inflation in My Program**

In 2010, I was unaware of the national trends concerning grade inflation. I had not carefully considered my grading practices, much less those of the family science discipline as a whole. I was a newly tenured faculty member in the Child and Family Science Department at a campus of California State University. We had approximately 350 majors, which was more than we could accommodate. At the time, we had six full-time and five part-time faculty members.

My newly tenured status meant that I assumed new responsibilities with regard to the curriculum, and I assumed the role of assessment coordinator for my department. Recently released from the bubble of the tenure track, I discovered our problem of impaction (having more enrolled students than can be accommodated in available classes) was worsened by high fail rates in select upper-division classes. In the spirit of program assessment, I set out to find out why so many of our students were not succeeding in core major classes. These students typically had strong GPAs, 3.0 or better, and had performed well in our lower-division classes.

The department implemented a comprehensive exam requirement to measure the foundational knowledge of our graduating seniors to identify the knowledge gaps in the hope that we could fill the holes. All faculty in the program collectively created the exam, which is described more fully below in the methods section. Comprehensive exam scores were abysmal. The first year we required the exam, in 2011, the average score was 59%; only 14% of our graduating seniors that year earned a passing grade on the exam. The high grades earned in lower-division classes (in many classes, 70-95% of students earned A grades) simply did not reflect students' level of mastery of that content. We wondered how the students earned those high grades when they clearly had not learned the content of the courses.

A review of syllabi revealed that many lower-division introductory courses were padded with attendance points and personal reflection activities that did not require any demonstration of mastery

of content. In the most egregious cases, syllabi revealed that students could pass classes by showing up even if they failed every single objective measure of knowledge because attendance and participation points alone were enough to produce a passing grade in the class. It became obvious that students could not succeed in upper-division coursework because they lacked foundational knowledge on which that advanced coursework was based. Because almost every student had earned As in those classes, we could not identify which students might need remedial help or which did not.

In the spirit of program assessment, we investigated. Then we intervened. Years later, we replicated the original investigation to determine if there had been any improvement since our intervention.

### Methods

At both the initial investigation in 2011 and the follow-up in 2018, the research was verified as a minimal risk by the university's institutional review board, and I obtained informed consent from both instructors and students who participated. The intervention was a change in department policy; only the evaluation of this intervention involved scientific study. Therefore, we did not seek informed consent from instructors to participate in the new policy.

### Measures

Archival records (the university's record system, PeopleSoft) were used to identify grade distributions for all Child and Family Science (CFS) classes. There is no institutional grading policy, but most CFS classes use a scale whereby 90-100% = A, 80-89% = B, etc. Institutional data included the course level (introductory, core major, or capstone), the rank of the instructor (tenure-track or part-time), class size, and the number of students to whom each grade was assigned at the end of the semester.

Short student surveys were created for the purpose of this research. Questions included self-report of their interest in the subject matter of the class, their typical study time for the class, how many classes they missed during the semester, how much of the reading they did at the time it was due, the perceived difficulty of the class, and whether their grade was an accurate reflection of their effort or learning. These questions were asked to identify potential explanations for high-class grades other than grade leniency.

Degree of interest in the subject matter is sometimes associated with better performance (Harackiewicz et al., 2016). Study time, reading, and attendance are measures of engagement, also associated with better grades (Bakker, Sanz Vergel, & Kuntze, 2015). Students were presumably not accurate reporters of their behaviors, so while their time-reporting was not considered to be objective, self-report was the only practical means by which to collect information like this about behaviors that could not be directly observed. We treated these self-reports as a reflection of the student's degree of engagement. We also included questions about the student's perception of class difficulty and accuracy of grading because such perceptions are associated with grades earned (England et al., 2019).

Surveys were administered in the final month of classes by myself and a student assistant. I scheduled classroom visits in coordination with the instructor. Students completed the surveys on paper in class during class time. Students did not put their names on the scantron forms, and no course instructor ever handled the completed forms from their class. Therefore, the surveys were completely anonymous. Students who were enrolled in more than one class were asked to complete the short survey for each CFS class they were enrolled in.

A comprehensive exam of foundational knowledge in developmental science and family science was required of all graduating seniors in this academic program. It is a multiple-choice exam consisting of items provided by the instructors of all core required classes in the degree program. It

was constructed by a department committee that solicited test items from faculty teaching core classes. They asked for items reflecting foundational conceptual knowledge that one would hope a student would retain after the class had ended. The committee reviewed items submitted and curated them to produce an exam that faculty believed accurately reflected the core curriculum of the department. It was approximately 35% developmental science, 35% family science, 15% theory, and 15% research methods. The validity of the exam was determined through convergence with other indicators of academic achievement, including grades and performance on specific measures in designated courses. A score of 70% or better was considered passing. Comprehensive exam scores presented here are not tied to specific courses or students. They are offered as general indicators of the health of the academic program at two different points in time.

### Sample

The unit of analysis was the class. The department offered 25 classes in the spring of 2011 with the CFS (Child and Family Science) prefix. I administered student surveys in 76% of those classes, with an average number of 31 surveys per class. This represented an 81% response rate for students in the participating classes in 2011. One instructor who did not participate later explained that she had simply forgotten to respond. A few classes were excluded because their instructor abruptly left mid-semester, and the classes were finished by part-time instructors who taught the class for less than a month.

We offered 37 classes in spring 2018 with the CFS prefix. I administered student surveys in 81% of those classes, with an average number of 27 surveys per class. This represented a 76% response rate for students in the participating classes in 2018. Virtually all instructors who did not participate reported that they simply forgot to sign the request for informed consent. One instructor, teaching a class for the first time, did not want to participate because she felt it was not a typical semester. Descriptive information about participating classes is offered in Table 1.

We aggregated student surveys by class instead of doing student-level analyses. Each class was assigned a score for each variable (e.g., student interest, attendance, reading frequency, etc.) that reflected the average student ratings on that item for all students enrolled in that class. We compared potentially grade-inflated classes (those with more than 45% of students earning an A grade) to those that were probably not grade-inflated (less than 45% of students earning an A grade) and compared them by ANOVA to look for explanations other than grade inflation. See details in the Results section.

### Intervention

The results of the 2011 data presented below convinced faculty in the program that our grades were inflated as a result of grading leniency and that this was detrimental to our program and students. We were unsure, however, how to solve the problem. To increase the academic rigor of our program, we looked at high-profile efforts to reduce grade inflation, such as the grade quotas required by Princeton University (Wofford, 2014). In 2004, Princeton adopted a grading policy that departments should award no more than 35% of students with a grade of A for coursework. The policy was later abandoned in 2014, but in 2011 it was still in force, so we considered following the example of Princeton University.

One problem with the Princeton model of grade quotas, however, was that it was essentially a norm-referenced system. Norm-referencing would not work for us because we had no way of knowing what the “correct” percent of A grades would be. It seemed reasonable that it would vary. Grades might be higher, for instance, in our senior capstone classes, as those students had already succeeded (passed) in all prior classes in the major, and only the best, most invested students remained. Grades might be lower, however, in introductory and general education classes. Even if

there was a different standard for different types of classes, we did not know how we would set that standard.

We also rejected the inherently competitive model of norm-referenced grading. We prefer a model of education that democratizes access to education and would celebrate the success of all of our students. Therefore, we were collectively more comfortable with a criterion-referenced system, one that is heavily weighted toward objective mastery of content but that might still consider student effort.

Finally, we were cognizant of protecting our academic freedom. We collectively agreed that academic freedom includes the right to choose one's grading practices, and we did not want to mandate a system that would prohibit any individual instructor from exploring alternative grading practices.

Still, we were alarmed and worried about the academic health of our program. After much consideration, we decided not to require any particular grading system. Instead, we enacted a policy that grade distributions would be monitored and discussed. We created an annual grade distribution report that was provided to each instructor and also placed into their personnel files for record-keeping. The report contained the percentage of grades earned in each of the instructor's classes for the year and also reminded the faculty member that they should consider whether the class grades were based on demonstrations of mastery of content.

This decision was made collectively, sealed with a unanimous vote, by the department's faculty. No faculty have reported feeling that the system is punitive, perhaps because we explicitly designed it not to be so. Faculty on the tenure track in the California State University system have probationary plans that determine criteria for retention, tenure, and promotion. Those criteria do not include characteristics of grade distributions, so the new grade distribution reports were not used for personnel decisions for tenure-track faculty, despite their placement in the personnel file. Since we intentionally did not identify acceptable and unacceptable benchmarks for grade distributions, not even the part-time faculty were subject to personnel decisions based on the grade distribution reports. We used the reports only as a system for raising awareness of grading standards and identifying classes that may be grade inflated, which then triggered an investigation of the grading standards and practices in those classes.

The department chair adopted the practice of reviewing these reports and investigating those classes that had either a high proportion of A grades or a high DFW rate. This practice is consistent with university policy requiring department chairs to review grading practices with faculty members on an annual basis. The chair reviewed the syllabi of courses that had remarkable grade distributions, with either high A rates or high DFW rates, and then met with their instructors to discuss the review. Very often, classes in which more than 45% of students earned an A grade included many assignments that were not graded for mastery of content, such as personal reflections and attendance points. Faculty were not required to change their grading practices as a result of the investigation, but most chose to do so after discussion.

These review procedures became standard operating procedures and continue to the present day. For brand-new instructors, the chair reviews grade distributions after their first semester and sits down to discuss the results with them. New part-time instructors are typically quite receptive to guidance about how to maintain rigor in the classroom. As professionals with deep regard for the importance of the work we are preparing students to do, it is not difficult to convince them that rigor is essential. Once faculty know that they have the support of the department, they can face the fear of bad student ratings that they assume will be the result of maintaining high grading standards.

By 2018, we wondered if this policy of raising awareness without imposing strict guidelines or any disciplinary procedures was successful at producing a sustained change in grade distributions

in our coursework. Therefore, we replicated the methods used in 2011 to see if there was any change. We compared the self-report data from students, as well as the grade distributions, from the two years (2011 and 2018) using ANOVA.

### Results

While the course list and descriptions were the same in 2018 as in 2011, the nature of our program evolved. In 2016, changes were made to the curriculum as a whole, mostly in prerequisites, to encourage some course sequencing to ensure that only adequately prepared students got into upper division classes. If the curriculum restructuring worked, then we would have prevented unprepared students from getting into upper division classes, and therefore we might have expected that grades in upper division classes would be higher as a result of a selection effect.

Despite the changes to the structure of our curriculum, by 2018, most students were still subject to the old catalog requirements and unaffected by the changes. We had a slightly bigger student body (approximately 400 majors compared to 350 in 2011). The high school GPAs of our students were virtually identical at the two points in time: 3.22 in 2011 and 3.32 in 2018. Cumulative GPA at the beginning of the term was also virtually identical: 2.91 in 2011 and 2.94 in 2018. On average, students were enrolled in 13 units both in 2011 and 2018. This suggests that our student body had not changed in any way that might explain differences in subsequent academic performance.

The participating classes are described in Table 1. Class sizes remained the same, but we offered more sections of courses and improved our tenure density.

**Table 1**

#### *Participating Classes*

	2011 (n=25)	2018 (n=37)
Total number of classes	25	37
Taught by tenure-track faculty	28%	46%
Average class size	38	38
Classes in which surveys were administered	76%	81%
Response rate in participating classes	81%	76%

By 2018, we had experienced a nearly complete turnover of tenure-track faculty as a result of normally timed retirements; only one from 2011 was still on the faculty. However, there was continuity in the part-time instructors; all five from 2011 were still teaching for our program in 2018. In our second round of grade inflation evaluation, we had seven full-time faculty and eight part-time faculty (compared to six and five in 2011).

#### **Potentially Grade-Inflated Classes**

In 2011, more than half of students (54%) in CFS classes earned an A grade. A whopping 83% earned an A or B, considered above-average grades. The data were dichotomized into classes that were potentially grade inflated (45% or more of students in the class earned an A grade) and those that were likely not grade inflated (less than 45% of students earned an A grade). The 45% indicator was chosen because it was the national average of As awarded in college classes at the time (Rojstaczer & Healy, 2012). In 2011, 14 of 25 classes (56%) met this standard and were considered potentially grade inflated. The 2011 data did not reveal any differences between the potentially grade

inflated classes and other others that would suggest any reason other than grade leniency for the high level of A grades. In 2018, 9 of 37 classes (24%) met the standard and were considered potentially grade inflated. Again, there were no differences that suggested any reason other than grade leniency.

For simplicity of presentation, the 2011 and 2018 data are combined in Table 2 because the results were identical in both years. Students in the potentially grade inflated classes were no more interested, engaged, or generally high-performing than students in other classes. Furthermore, they were no more likely to do all the reading or devote adequate study time. All of these alternative explanations for high grades are ruled out.

**Table 2**

*Student Self-Report in Potentially Grade-Inflated Classes*

	> 45% As (n=38)	< 45% As (n=23)	ANOVA
Class Size	38	38	F=.00, p=.994
Interest in the class (1=not at all interested, 2=a little interested, 3=moderately interested, 4=very interested)	3.4	3.4	F=.12, p=.731
Missed Class (1=never, 2=once or twice, 3=occasionally, 4=often)	1.9	1.9	F=.27, p=.606
Reading: Assigned reading completed when due (1=never, 2=less than half, 3=half, 4=more than half, 5=all or most)	3.5	3.5	F=.08, p=.784
Study Time: Hours per week studying or doing homework for this class (1=0-1hr, 2=2-3 hrs, 3=4-5 hrs, 4=6-7 hrs, 5=8+ hrs)	2.2	2.5	F=1.70, p=.199
Difficulty (1=lot easier than most, 2=little easier, 3=same, 4=little harder, 5=lot harder than most)	2.5	3.1	F=8.70, p=.005
Grade is accurate reflection of my effort and learning (1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree)	3.3	3.0	F=17.84, p<.0001

The only difference between the potentially grade inflated classes and the others is that students recognize that the grade inflated classes are easier than their other classes, and they think they deserve the good grades they earn in them. We also explored whether grade distributions varied according to the type of course or by the rank of the instructor. They did not. Introductory general education classes had roughly the same rate of A grades as did core major classes and senior capstone classes. Tenure-track faculty had roughly the same rate of A grades as part-time faculty.

### Changes from 2011 to 2018

Grade distributions in CFS classes drastically changed from 2011 to 2018; student perceptions did not change with them, as depicted in Table 3. The percentage of As earned in CFS classes reduced from 54% to 34%. The number of Bs and Cs both increased commensurately, thereby releasing the earlier compression of grades. The number of failing grades increased only slightly, from 4% to 7%. Using the standard that more than 45% As is probably grade inflated, the proportion of grade inflated classes in our program declined from 56% down to 24%.

**Table 3***Grade Distributions, Student Perceptions, and Comprehensive Exams 2011-2018*

	2011 (n=25)	2018 (n=37)	ANOVA
Percent A grades	54%	34%	F=9.71, p=.003
Percent B grades	29%	41%	F=7.05, p=.010
Percent C grades	10%	19%	F=7.25, p=.009
DFW rate	4%	7%	F=4.05, p=.05
Potentially grade inflated (>45% A)	56%	24%	F=7.86, p=.007
Study Time	2.3	2.5	F=1.86, p=.179
Reading	3.4	3.5	F=.97, p=.330
Missed Class	1.9	1.9	F=.08, p=.786
Difficulty	2.6	3.0	F=4.72, p=.04
Grade Accurate	3.2	3.1	F=2.72, p=.106
Comprehensive Exam Average Score	59%	71%	
Comprehensive Exam Pass Rate	14%	64%	

With regard to self-reported student behaviors related to their learning, average study time, amount of reading completed on time, and number of missed classes were not significantly different in 2018 as compared to 2011. Students in 2018 rated their classes as more difficult than students had done in 2011, but on average, they said that their CFS classes were about the same level of difficulty as other classes. They were equally likely to say that their grade was an accurate reflection of their work and learning.

During this timeframe, comprehensive exam scores increased from an average score of 59% in 2011 to 71% in 2018. In 2018, 64% of our seniors passed the exam, compared to only 14% in 2011.

### Discussion

This analysis of grade distributions, combined with student surveys and objective, comprehensive exam scores, suggests that grade inflation, without commensurate achievement, was present in our program in 2011. We intervened by regularly monitoring grade distributions in department classes, placing reports annually in faculty personnel files, and following up with individual faculty whose grade distributions appeared to be potentially inflated. By 2018, the compression of passing grades was released, and students were learning more content as measured by the departmental comprehensive exam. This change in grade distributions and improvement in comprehensive exam scores was not accompanied by a significant perception of the burden placed on students. A relatively minor and non-intrusive intervention likely produced this change.

## Grade Inflation

The department faculty were convinced by the 2011 data that grade inflation, produced by leniency in grading, was present in our coursework. We were convinced by the evidence. The percentage of A grades in our classes was higher than we had realized, higher than the national average. This was true despite the fact that very few of our students reported doing all of the assigned readings or spending as much time as recommended on their studies for the credit units earned by the class and despite the fact that our comprehensive exam suggested that our students were not mastering the content of their classes. In fact, our students reported that they knew our classes were easier than their other college classes, but they still believed their inflated grades were deserved.

We want students to succeed; we recognize their struggles and value their desire to succeed in college and beyond. This mindset may lead us to demonstrate our values through points: points for showing up because we value active participation, points for effort because we want to recognize their struggles, or points for extra credit to acknowledge the value of a second chance. These points determine student grades, and so we may end up giving grades based on our values rather than the student's demonstration of mastery of course content. This desire to grade based on values could easily result in a compression of grades, whereby students who do not show up and demonstrate effort still fail our classes, but everyone else gets an A. We want to reward them for their hard work and keep them moving toward academic success. This thought process produces a textbook case of grade inflation.

Our experience is probably typical. Most faculty resist believing that their own courses are grade inflated, even when they believe this happens elsewhere (McCabe & Powell, 2004). Furthermore, college students typically believe that high grades are their due. Adams (2005) reported that 70% of students believe they should earn a grade of at least C if they put in significant effort, even if they did not perform well in the class. Zinn et al. (2011) discovered that students are far more likely than faculty to believe that grades should be assigned based on effort, even when that effort does not produce high performance.

## Intervention to Reduce Grade Inflation

The intervention described here was actually a years-long process of discovery within one academic department. We landed on an annual report provided to all faculty with follow-up discussions for some, but before that, we collectively had many discussions raising our consciousness about grade inflation. The process of collecting data from all classes meant that even part-time instructors were aware that we were studying this issue. Therefore, the final decision and the changed practices that resulted from it were only a culmination of a long process of significant self-reflection and problem-solving. The outcomes of this intervention, however, suggest a significant release of grade compression without much impact at all on student perceptions of their coursework. Furthermore, the outcome has been sustained even through substantial turnover of faculty in the program.

This process has been reported, albeit rarely, elsewhere in the research on grades in higher education. In fact, our results are similar to the results of a similar study documenting a grading standards policy change in a department of education on another campus (Ridley et al., 2003). In that study, the policy change was a requirement that standards be elevated, such that an A grade required 93% of the points rather than 90%, and a requirement that any instructor awarding more than 50% As provide written justification for the pattern, significantly reducing A grades with no negative impact on student ratings of instruction; in fact, students reported a higher level of intellectual stimulation in the affected classes.

One difference between our program policy change and that described by Ridley et al. (2003) is that ours was an internal decision, not one imposed by administrators outside of the department. In that case, the policy only remained in place for two years because faculty objected to it, and then grading standards reverted to their original inflated state. Our policy, in contrast, has continued for a decade with the outcomes sustained. There has been no faculty uprising against the program; faculty agree it is a good practice. Another difference is that our policy does not dictate a specific grading system, whereas theirs has features of a norm-referenced system.

As a tool for maintaining academic rigor, systematic monitoring of grade distributions seems effective. I speculate that the reason it works is that it keeps the issue of academic rigor in the minds of instructors without dictating the grading strategy. A call to rigor on the grounds that it is necessary to prepare students to do the important work of serving children and families in society speaks to the idealism and professional ambitions of faculty in a family science program such as ours.

### **Limitations and Future Research**

The present study was conducted only in one program at one university campus, and therefore, the results may not be generalizable to other types of programs on other campuses. This is a state school serving mostly local students. It is ANNAPISI and Hispanic-serving. The department offers only bachelor's degrees, and the faculty bear a very heavy teaching load without assigned time for scholarly activity. All of these imply that faculty are deeply invested in high-quality teaching and have much of their academic identity tied to that work. It is possible that a consciousness-raising intervention, such as the one described here, would only work in such a setting.

It is also true that the instruments used in this research are not sophisticated instruments with demonstrated reliability and validity. This is applied research conducted in a real-life setting that is not tightly controlled. The measures are limited by their simplicity. The design includes the normal confounds associated with the passage of time in an academic department.

The current analysis ruled out potential alternative explanations for the change in grades. The students had similar academic profiles before and after the policy change. The courses and curricular requirements were virtually identical. We cannot rule out the possibility that grade distributions changed as a result of turnover in our tenure-track faculty body that coincided with the policy change. However, the change from 2011 to 2018 was also observed in the part-time faculty, where their average class GPA changed from 3.4 in 2011 to 3.0 in 2018 ( $F=6.7$ ,  $df=1$ ,  $p=.01$ ). Part-time faculty was fairly stable during these years, as was the curriculum, so the change in grades can only reasonably be attributed to the policy on reporting grade distributions.

As limited by the quasi-experimental design as these data may be, they highlight the fact that very little empirical work on grading practices has been published in higher education, much less in family science programs. There has been quite a lot of research attention paid to the productivity and prestige of family science programs (e.g., Kamp Dush, 2014; Reifman et al., 2019) as well as academic work more generally (e.g., Altbach, 2006; Kenny, 2017). However, there is no published research about how family science faculty conceptualize grading, how they assign grades, how their students perceive grades, or how grading relates to program quality or reputation, creating a gap in the scholarship of teaching and learning in family science. Research addressing these topics exists, especially regarding elementary and secondary school teachers (see review by Brookhart, 1994), but there is little research at the college level and never in family science programs.

### **Implications**

Grade inflation can cause future problems for students. Students whose grades have been inflated have a better academic record but worse subsequent performance in later coursework (Friedl

et al., 2012). When knowledge and skill are not commensurate with grades, then subsequent learning can be hindered because students are misplaced in classes that are beyond their capability. In the worst-case scenario, students may find themselves in jobs that they are unqualified and unprepared for, causing dissatisfaction for students and potentially damaging the reputation of academic disciplines and institutions (Jaschik, 2018).

Regularly monitoring grade distributions can produce enough awareness and identify potentially problematic grading practices such that grade inflation can be curbed. As a result, college professors may be able to reverse the grade compression seen in recent decades and do so without increasing fail rates or damaging students' perceptions of our classes. Faculty can do this without becoming unreasonable and capricious graders and without infringing on the academic freedom of professors.

It remains to be seen if reversing the compression of passing grades in an undergraduate program will damage students' chances of getting into graduate programs or getting jobs. This may affect some groups of students more than others, as some have received more grade leniency than others have. But those students have been done a grave disservice. They have been given a credential without the education that it represents. To correct grade leniency in an education program is to improve the equity and the justice of that program so that all students have access to the same high-quality education.

To realize this vision of a just and equitable higher education, academic programs have to verify that excellent grades in their coursework truly reflect academic excellence in students. Programs that aspire to excellence will improve the quality of their offerings rather than simply inflate their grades. College grades may become a more useful predictor of who will succeed in those experiences, increasing the reputation and prestige of the academic programs that produce qualified graduates. This outcome benefits all students in the program.

There has been discussion within family science (Hamon & Smith, 2014; Gavazzi et al., 2014) about the professional integrity of the discipline. Maintaining academic standards to continue to make a scholarly contribution, attract students, and build trust with employers is essential to maintaining the integrity of our discipline.

### References

- Adams, J.B. (2005). What makes the grade? Faculty and student perceptions. *Teaching of Psychology*, 32(1), 21-24. [https://doi.org/10.1207/s15328023top3201\\_5](https://doi.org/10.1207/s15328023top3201_5)
- Altbach, P. G. (2006). The dilemmas of ranking. *International Higher Education*, 42, 2-3. <https://doi.org/10.6017/ihe.2006.42.7878>
- Astin, A. W. (1998). The changing American college student: Thirty-year trends. 1966-1996. *The Review of Higher Education*, 21(2), 115-135.
- Babcock, P. & Marks, M. (2010). The falling time cost of college: Evidence from a half a century of time use data (NBER Working Paper No. 15954). *National Bureau of Economic Research*. [https://www.nber.org/system/files/working\\_papers/w15954/w15954.pdf](https://www.nber.org/system/files/working_papers/w15954/w15954.pdf)
- Bakker, A. B., Sanz Vergel, A. I., & Kuntze, J. (2015). Student engagement and performance: A weekly diary study on the role of openness. *Motivation and Emotion*, 39(1), 49-62. <https://doi.org/10.1007/s11031-014-9422-5>
- Blaisure, K. R. (2003). Family science faculty members' experiences with teaching from a feminist perspective. *Family Relations*, 52(1), 22-32. <https://doi.org/10.1111/j.1741-3729.2003.00022.x>
- Bejar, I. I. and Blew, E. O. (1981). Grade inflation and the validity of the Scholastic Aptitude Test. *American Educational Research*, 18(2), 143-156. <https://doi.org/10.3102/00028312018002143>
- Block, J. H. (Ed.) (1971). *Mastery learning: Theory and practice*. Holt, Rinehart & Winston.

<https://doi.org/10.26536/ZVHK6341>

*Family Science Review*, Volume 27, Issue 3, 2022

*This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)*

- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301. [http://doi.org/10.1207/s15324818ame0704\\_2](http://doi.org/10.1207/s15324818ame0704_2)
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T. & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803-848. <https://doi.org/10.3102/0034654316672069>
- Burke, L., Amselem, M.C. & Hall, J. (2016). Big debt, little study: What taxpayers should know about college students' time use. *The Heritage Foundation*. <https://www.heritage.org/education/report/big-debt-little-study-what-taxpayers-should-know-about-college-students-time-use>
- Chamberlin, K., Yasue, M., & Chiang, I. C. A. (2018). The impact of grades on student motivation. *Active Learning in Higher Education*, 2018, 1-16. <http://doi.org/10.1177/1469787418819728>
- Dubey, P. & Geanakoplos, J. (2010). Grading exams: 100, 99, 98....or A, B, C? *Games and Economic Behavior*, 69(1), 72-94. <https://doi.org/10.2139/ssrn.1421986>
- Dweck, C. S. (2007). *Mindset: The New Psychology of Success*. Ballantine Books.
- Ebel, R. L. (1974). Shall we get rid of grades? *National Council on Measurement in Education: Measurement in Education*, 5(4), 1-5.
- England, B. J., Brigati, J. R., Schussler, E. E., & Chen, M. M. (2019). Student anxiety and perception of difficulty impact performance and persistence in introductory biology courses. *CBE-Life Sciences Education*, 18(2). <https://doi.org/10.1187/cbe.17-12-0284>
- Ewell, P. & Cumming, T. (2017). History and conceptual basis of assessment in higher education. In T. Cumming & D. Miller (Eds.), *Enhancing assessment in higher education: Putting psychometrics to work* (pp. 1-30). Stylus.
- Finkelstein, I. E. (1913). *The marking system in theory and practice*. Warwick & York.
- Friedl, J., Pittenger, D. J., & Sherman, M. (2012). Grading standards and student performance in community college and university courses. *College Student Journal*, 46(3), 526-532.
- Gavazzi, S. M., Wilson, S. M., Ganong, L., & Avonkovic, A. (2014). Furthering the conversation on the discipline of Family Science: Comments on the articles by Hamon & Smith and Hans. *Family Relations*, 63(3), 333-342. <http://doi.org/10.1111/fare.12076>
- Glaser, R. & Klaus, D. J. (1962). Proficiency measurement: Assessing human performance. In R.M. Gagne (Ed.), *Psychological principles in system development* (pp. 419-474). Holt, Rinehart & Winston.
- Hamon, R. R. & Smith, S. R. (2014). The discipline of Family Science and the continuing need for innovation. *Family Relations*, 63(3), 309-322. <https://doi.org/10.1111/fare.12068>
- Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016) Interest matters: The importance of promoting interest in education. *Policy Insights in Behavior and Brain Science*, 3(2), 220-227. <https://doi.org/10.1177/2372732216655542>
- Jaschik, S. (2018, October 9). *Falling confidence in higher ed*. Inside higher ed. <https://www.insidehighered.com/news/2018/10/09/gallup-survey-finds-falling-confidence-higher-education>
- Johnson, V. E. (2003). *Grade Inflation: A Crisis in College Education*. New York: Springer.
- Kamp Dush, C. (2014). *The HDFS report 2014: Claire Kamp Dush's ranking of HDFS programs in North America*. Adventures in Human Development and Family Studies. <http://u.osu.edu/adventuresinhdfs/the-hdfs-report-2014>.
- Kenny, J. (2017). Academic work and performativity. *Higher Education*, 74(5), 897-913. <https://doi.org/10.1007/s10734-016-0084-y>

- Kinzie, J. (2019). Taking stock of initiatives to improve learning quality in American higher education through assessment. *Higher Education Policy*, 32(4), 577-595. <https://doi.org/10.1057/s41307-019-00148-y>
- Kostal, J. W., Kuncel, N. R., & Sackett, P. R. (2016). Grade inflation marches on: Grade increases from the 1990s to 2000s. *Educational Measurement: Issues and Practice*, 35(1), 11-20. <https://doi.org/10.1111/emip.12077>
- Letiecq, B. L. (2019). Surfacing family privilege and supremacy in family science: Toward justice for all. *Journal of Family Theory and Review*, 11(3), 398-411. <http://doi.org/10.1111/jftr.12338>
- Lipnevich, A.A., Guskey, T.R., Murano, D.M. & Smith, J.K. (2020) What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice*, 27(5), 480-500. <https://doi.org/10.1080/0969594x.2020.1799190>
- Lok, B., McNaught, C. & Young, K. (2016). Criterion-referenced and norm-referenced assessments: Compatibility and complementarity. *Assessment and Evaluation in Higher Education*, 41(3), 450-465. <https://doi.org/10.1080/02602938.2015.1022136>
- Mahoney, K. T., Buboltz, W. C., Jr., Budkner, V. J. E., & Doverspike, D. (2011). Emotional labor in American professors. *Journal of Occupational Health Psychology*, 16(4), 406-423. <https://doi.org/10.1037/a0025099>
- McCabe, J., and Powell, B.(2004). "In my class? No." Professors' accounts of grade inflation. In W. E. Becker and M. L. Andrews, Eds., *The Scholarship of Teaching and Learning in Higher Education: Contributions of Research Universities* (pp.193-220). Indiana University Press.
- Milton, O., Pollio, H. R., & Eison, J. A. (1986). *Making sense of college grades: Why the grading system does not work and what can be done about it*. Jossey-Bass.
- Reifman, A., Niehuis, S., Fitzpatrick, J., Oldham, C. R., Scott, J. P., Fang, D., Yuan, S., Lieway, M., Walisky, D. & Gregersen, L. (2019). Ranking Ph.D. granting human development and family studies departments via research and professional accomplishments. *Family Science Review*, 23(3), 1-21.
- Reifman, A., Fitzpatrick, J., Niehuis, S., Chapman, J. K., Oldham, C. R., Scott, J. P., Fang, D., Yuan, S., Lieway, M., Walisky, D. N., & Gregersen, L. (2019). The journey of ranking Ph.D. granting Human Development and Family Studies departments. *Family Science Review*, 23(3), 91-97.
- Ridley, D. R., Quanty, M. B., & Sciabica, M. (2003). Grading standards and course challenge: An analytical-empirical approach. *Research in Post-Compulsory Education*, 8(3), 345-365. <https://doi.org/10.1080/13596740300200159>
- Rojstaczer, S., & Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading, 1940-2009. *Teachers College Record*, 114(7), 1-23. <https://doi.org/10.1177/016146811211400707>
- Rojstaczer, S. (2016). GradeInflation.com: Grade inflation at American colleges and universities. <https://www.gradeinflation.com/>
- Schinski, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE Life Science Education*, 13(2), 159-166. <https://doi.org/10.1187/cbe.cbe-14-03-0054>
- Smallwood, M. L. (1935). *An historical study of examinations and grading systems in early American universities: A critical study of the original records of Harvard, William and Mary, Yale, Mount Holyoke, and Michigan from their founding to 1900*, v. 24. Harvard University Press.
- Terwilliger, J. S. (1993). Some thoughts on grading systems and grading practices. In S.L. Wise (Ed.), *Teacher Training in Measurement and Assessment Skills* (pp. 63-95). Buros Center for Testing.
- Thorndike, R. L. (1969). Marks and marking systems. In *Encyclopedia of Educational Research* (4<sup>th</sup> ed., pp. 759-766). Macmillan.
- Wofford, T. (2014, August 7). Princeton looks to scrap "grade quotas". *Newsweek*. <https://www.newsweek.com/princeton-looks-scrap-grade-quotas-263363>

Zinn, T.E., Magnotti, J.F., Marchuk, K., Schultz, B.S., Luther, A. & Varfolomeeva, V. (2011). Does effort still count? More on what makes the grade. *Teaching of Psychology*, 38(1), 10-15.  
<https://doi.org/10.1177/0098628310390907>