IN DEFENSE OF MANDATORY CURVES

Joshua M. Silverstein[*]

ABSTRACT

This article sets forth the first comprehensive defense of mandatory curves. It begins with a case study of one law school. That institution lacked formal grade normalization policies during the period of the case study. As a result, the school suffered from dramatic grade disparities. This article contains a list and statistical analysis of the most significant disparities. The statistical analysis supports the conclusion that the grade disparities were caused by differences in teacher grading philosophy, and not by student merit or any other factor. Next, this article presents several arguments in favor of mandatory curves. The most crucial is that grade variances that flow from differences in professor grading philosophy are grossly unfair to students. A second important defense of forced curves is that grade disparities distort the process of course selection, inducing students to register for classes based on the grading practices of the professor rather than on substantive concerns, such as topical importance, career relevancy, and skill development. The article then responds to the eight most significant and common objections to mandatory curves. Several of these objections are deeply problematic, such as the contention that curves prevent professors from awarding students the grades they deserve. Others have some merit, such as the

---

* Associate Professor, University of Arkansas at Little Rock, William H. Bowen School of Law. B.A. 1993, Hamilton College. J.D. 1996, New York University School of Law. Completion of this article was made possible by a grant from the University of Arkansas at Little Rock, William H. Bowen School of Law.

I would like to thank the following individuals for their helpful comments on earlier drafts: Adjoa Aiyetoro, Coleen Barger, Terri Beiner, Terrence Cain, John DiPippa, Felecia Epps, Frances Fendler, Lynn Foster, Ken Gallant, Chuck Goldner, Ken Gould, Sarah Hobbes, Nick Kahn-Fogel, Beth Levi, Nancy Levitt, Philip Oliver, Ranko Oliver, Kelly Olson, Rick Peltz, Stephen Sheppard, Tonya Smith, Rob Steinbuch, June Stewart, Tom Sullivan, Phyllis Silverstein, Erik Senko, Kelly Terry, and Carolyn Turturro.

I presented this article to the faculty at the University of Arkansas at Little Rock, William H. Bowen School of Law, three times. My thanks to all those who attended and offered comments. And thanks to the Faculty Development Committee of the Bowen Law School for organizing two of these presentations. I also presented this piece to the faculty at the University of Arkansas Law School in Fayetteville, Arkansas. My thanks to all those who attended and offered comments.

I would also like to thank Joey Price, Amanda LaFever, Hallie Sears, and Joshua West for their invaluable research and editing assistance in preparing this article, and Ryan Lazenby for his outstanding work in performing the statistical analysis in Appendix 2.

Finally, I wish to thank Professor Matthew Silverstein of New York University-Abu Dhabi, both for reading drafts of this article and for engaging in numerous discussions about the issues discussed in this piece. The article is substantially better as a result of his contributions.

argument that mandatory curves encourage excessive competition among students. But the latter set of criticisms ultimately does not undercut the case for curves, principally because mandatory curves are the lesser evil. For example, the competition objection fails because the significant grade disparities that frequently result in the absence of a curve probably cause more competition than mandatory curves do. And even if forced curves do increase competition, the unfairness of grade disparities flowing from differences in professor grading philosophy is the more pressing concern. Finally, the article ends with a discussion of some issues regarding the structure and scope of mandatory curves, including the applicability of curves to smaller classes, seminars, and clinics.

"[A]wards should be 'according to merit'; for all men agree that what is just in distribution must be according to merit."[1]

"We teach that justice matters; let us do our best when it is our turn to hand down the decisions."[2]

I. INTRODUCTION

The purpose of this article is to set forth a comprehensive defense of mandatory curves. A mandatory (or forced) curve is the best method for preventing substantial grade disparities among professors. The grading practices at the University of Arkansas at Little Rock, William H. Bowen School of Law, illustrate the problems that can develop in the absence of such a policy. Until the fall of 2011, that school had no formal guidelines regarding grade normalization.[3] As a result, there were dramatic variations in grading for many years. Professors teaching the same course sometimes differed by an entire grade point (e.g., 3.5 vs. 2.5 on a 4.0 scale).[4] Such variances are unfair to the students because they flow from differences in faculty grading philosophy. Grades determine academic honors, placement prospects, and career paths. It is unjust for the allocation of these benefits to depend on teacher grading philosophy rather than on student merit.

---

1. ARISTOTLE, NICOMACHEAN ETHICS, Book V. § 3(A), ¶ 1 (W.D. Ross trans.), *available at* http://classics.mit.edu/Aristotle/nicomachaen.mb.txt.

2. Jeffrey Evans Stake, *Making the Grade: Some Principles of Comparative Grading*, 52 J. LEGAL EDUC. 583, 618 (2002).

3. *See* UNIVERSITY OF ARKANSAS AT LITTLE ROCK, WILLIAM H. BOWEN SCHOOL OF LAW ACADEMIC RULES (Nov. 9, 2009) (hereinafter "UALR BOWEN LAW SCHOOL ACADEMIC RULES"), at 11–12, *available at* http://ualr.edu/law/files/2011/06/SCAN1891_000.pdf.

4. *See infra* Part III.

Grade disparities also distort the process of course selection by students. Again, the William H. Bowen School of Law ("Bowen Law School" or "Bowen") is illustrative. Students there placed substantial weight on the professor's grading history in deciding whether to register for a course.[5] Indeed, students sometimes referred to the registration process as the "grade selection process" rather than the "course selection process."

While a majority of law schools have adopted a mandatory curve,[6] there is still substantial resistance to such policies. This resistance is driven principally by a series of flawed objections to forced curves. Some of the criticisms are deeply problematic, such as the contentions that mandatory curves prevent professors from awarding students the grades they deserve and that curves infringe on academic freedom. Others have some validity, such as the argument that curves encourage excessive competition among students. But the criticisms with merit do not undercut the case for mandatory curves, primarily because forced curves are the lesser evil. For example, the competition objection fails because the significant grade disparities that frequently result in the absence of a curve probably cause more competition than curves do. And even if mandatory curves induce slightly more competition than grade disparities, the fairness that results from properly-designed curves makes them well worth this cost.

Other critics of mandatory curves do not challenge such policies in full. Rather, they assert that forced curves should not apply to a subset of courses, namely smaller classes, seminars, and clinics. I accept that mandatory curves cannot be applied in precisely the same way to both small and large courses. Additional mechanisms are necessary to ensure that a curve's operation in small classes maximizes fairness. But once those mechanisms are adopted, mandatory curves ought to govern in small classes. In addition, there is no reason to treat seminars and clinics differently from other courses, unless a particular seminar or clinic falls below the size thresholds that warrant the use of size-related mechanisms.

This article demonstrates that mandatory curves are a crucial element of fair grading. All law schools should thus have a forced curve.

Part II of the article sets forth a brief taxonomy of grade normalization. Part III discusses the grade disparities that Bowen Law School experienced in the absence of a forced curve. Part IV presents the primary justifications for mandatory curves. Part V responds to the eight most significant and common objections to forced curves. Part VI addresses several issues regarding the structure and scope of mandatory curves, including the applicability of curves to smaller classes, seminars, and clinics.

---

5. *See infra* note 52 and accompanying text.
6. *See infra* note 206 and accompanying text.

The article also includes three appendices. Appendix 1 contains statistical analysis of the grade disparities at the Bowen Law School listed in Part III. Appendix 2 sets forth the mandatory curve I recommended to the faculty at Bowen. The faculty adopted the curve in the spring of 2011 with only a few, relatively minor changes. The curve took effect in the fall of 2011. Finally, Appendix 3 (available only online) contains a more comprehensive list of the grading disparities at Bowen.

II. A GRADE NORMALIZATION TAXONOMY

In this article, the term "grade normalization" refers to any policy or practice designed to standardize grades among professors, whether formal or informal, mandatory or recommended, rigid or flexible.[7] *Formal* grade normalization principally involves setting course means, medians, or distributions.[8] *Informal* grade normalization includes practices such as circulating grade distribution reports among the faculty and encouraging professors of comparable classes to consult about grades. Grade normalization is *mandatory* if the grading policies must be followed—as with a required mean or required consultation among professors who teach the same course. And normalization is *recommended* if professors have discretion over whether to follow the applicable standards—as with a suggested mean or suggested consultation practices. In this piece, the terms "mandatory curve" and "forced curve" refer to any grade normalization policy requiring a class mean, median, or grade distribution.[9]

---

7. The terms "grade normalization" and "mandatory curve" have disparate uses in the literature and among law schools. *See* Nancy H. Kaufman, *A Survey of Law School Grading Practices*, 44 J. LEGAL EDUC. 415, 417 n.6 (1994) (explaining that the word "curve" may be "[c]onstrued narrowly to refer to the distribution of grades among students rather than a class mean or median," but that a number of law schools responding to a survey about grading practices interpreted the word to include "[c]lass means and medians, as well as distributional schemes"). For examples of differing usages, see Richard A. Epstein, *Grade Normalization*, 44 S. CAL. L. REV. 707, 710 (1970–71) ("Grade normalization is responsive to both criticisms, every bit as much as mandatory curves, because it both sets the median and limits the spread of the curve."); Jay M. Feinman, *Law School Grading*, 65 UMKC L. REV. 647, 652 (1997) ("Grade normalization is the process of transforming individual professors' grades to a common curve. . . . Normalization makes use of an approved distribution of grades; the distribution can be strict or flexible."); Deborah Waire Post, *Power and the Morality of Grading–A Case Study and a Few Critical Thoughts on Grade Normalization*, 65 UMKC L. REV. 777, 778 (1997) ("The subject of this symposium issue is grade normalization, one example of which is the 'strict statistical curve.'").

8. For examples of other types of formal normalization relying upon a statistical methodology, see Paul T. Wangerin, *Calculating Rank-in-Class Numbers: The Impact of Grading Differences Among Law School Teachers*, 51 J. LEGAL EDUC. 98, 101–103, 117 (2001); Nicholas L. Georgakopoulos, *Relative Rank: A Remedy for Subjective Absolute Grades*, 29 CONN. L. REV. 445, 447–50 (1996); VALEN E. JOHNSON, GRADE INFLATION: A CRISIS IN EDUCATION 209-24, 245 (2003).

9. At present, law schools use many different types of grade normalization, including a wide variety of mandatory curves. For comprehensive surveys of law school grading practices, see Andy Mroch, *Law School Grading Curves* 95 (2005), *available at* http://www.aals.org/deansmemos/Attachment05-14.pdf; Robert C. Downs & Nancy Levit, *If it Can't be Lake Woebegone . . . A Nationwide Survey of Law School Grading and Grade Normalization Practices*,

The central purpose of a mandatory curve is the prevention of grade disparities that result from different grading philosophies among the faculty rather than from student merit.[10] The most important type of grade disparity is a variation in mean (or average) GPA across courses. As used in this article, the term "grading philosophy" has two elements—"substantive expectations" and "grading style." Faculty differ in *substantive expectations* when they disagree over what constitutes excellent, satisfactory, and unsatisfactory work, or when they disagree about what constitutes superior and inferior work. For example, Professor X may consider a Contracts final written by a first-year student to be "excellent" only if the student shows a very high level of mastery. On the other hand, Professor Y labels a Contracts test "excellent" if the student demonstrates a more basic understanding of the subject.

Of course, the phrase "substantive expectations" could be used to describe other aspects of assessment, such as whether to administer a closed-book or open-book test, whether to use multiple choice or essay questions, which legal skills deserve emphasis (e.g., extracting rules from legal texts, applying the rules to new circumstances, and conveying the application in writing), or which substantive topics to teach and test on. But that is not how I am employing the term here. In this piece, "substantive expectations" refers strictly to differences in the *degree* of expectation, not the constituent parts. For example, Professors X and Y may both teach and test on the same skills and subjects, and they might even give the same weight to the skills and subjects covered in the exam. But Professor X demands mastery of the various skills and subjects for a test to be considered "excellent," whereas Professor Y only requires basic competency. That is what I mean by different "substantive expectations."

Teachers differ in *grading style* when they disagree over purely formal grading matters—such as whether most "excellent" or "superior" papers should receive As, or only a few. For example, Professor J might believe

---

65 UMKC L. REV. 819 (1997); *id.* at 837–40 (summarizing the types of grade normalization in use); Kaufman, *supra* note 7; *id.* at 417–21 (summarizing the forms of grade normalization in use). *See also* WILLIAM M. SULLIVAN ET AL, THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING, EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW 162–84 (2007) (hereinafter CARNEGIE) (discussing law school assessment generally); ROY STUCKEY AND OTHERS, CLINICAL LEGAL EDUCATION ASSOCIATION, BEST PRACTICES FOR LEGAL EDUCATION 235–63 (2007) (hereinafter BEST PRACTICES) (same); Daniel Keating, *Ten Myths About Law School Grading*, 76 WASH. U. L. Q. 171 (1998) (discussing a variety of grading issues).

10. *See* Post, *supra* note 7, at 786 (explaining that grade normalization "is about 'fair competition'"—it is about preventing grade differences that are "not attributable to differences in students, but to differences in instructors"); *see also* ST. JOHN'S UNIVERSITY SCHOOL OF LAW FACULTY OF LAW STATEMENT ON GRADE NORMALIZATION, *available at* http://www.stjohns.edu/academics/graduate/law/current/handbook/exams/normalization.stj ("Grade normalization . . . is intended to ensure fair and just grading of students based upon their academic performance regardless of the particular course taken, the section to which assigned, the degree of difficulty of the examination, *and the identity of the professor teaching the course*.") (emphasis added).

that all excellent exams deserve an A, while Professor K thinks that A grades should be a rare occurrence, awarded only to truly outstanding papers.

In sum, "grading philosophy" consists of two types of *grading standards*—standards that set the degree of substantive expectation and standards that set the formal labels applicable to each level of achievement.

III. GRADING DISPARITIES AT THE WILLIAM H. BOWEN SCHOOL OF LAW

The Bowen Law School suffered from grade disparities based on professor grading philosophy for many years. This subpart contains a chart listing a selection of the most significant disparities from the fall 2006 semester through the spring 2010 semester.[11] I have included these disparities in order to illustrate the problems that can develop when an institution lacks binding grade normalization procedures. Some past law review articles have offered individual examples of grade disparities, made general statements regarding grade variations at a given school, or presented the disparities in statistical form.[12] For this piece, I felt it critical to present a more comprehensive list in unadjusted form. I want the reader to see, in concrete terms, what dramatic grade variances look like.

In the chart set forth below, the grade disparities are broken down into the following seven categories:

1. First-Year Courses—Disparities in Sections of the Same Course
2. Upper-Level Courses—Disparities in Sections of the Same Course
3. First-Year Courses—Disparities Across Courses
4. Upper-Level Courses—Disparities Across Bar Courses
5. Upper-Level Courses—Disparities Across Electives
6. Upper-Level Courses—Disparities Across Seminars[13]
7. Miscellaneous Additional Disparities

The grade point averages listed for each course are *mean* grade point averages. The number of students in each class is set forth in parentheses. An asterisk (*) denotes a required course. A caret (^) denotes a composite grade point average derived from two sections of the course taught by the same professor. A significantly more comprehensive list of grading disparities at my law school is contained in Appendix 3 (available online).[14]

---

11. There were also significant disparities from the fall 2003 semester through the spring 2006 semester and during the 2010-2011 school year. But I felt that inclusion of these other disparities would unnecessarily lengthen this article without appreciably strengthening my argument.

12. *See infra* note 17.

13. As at many law schools, Bowen uses the term "seminar" to refer to smaller courses in which the final grade is determined principally by a major writing project rather than a final exam.

14. *See* Appendix 3: A Broad Sampling of Grading Disparities at the William H. Bowen School of Law, *available at* http://jurisophia.org/in-defense-of-mandatory-curves-appendix-3-a-b. The list con-

1. First-Year Courses – Variances In Sections of the Same Course

Spring 2010: Legal Writing II*[15] (27) = 3.361
vs.
Legal Writing II*^ (28) = 3.101; Legal Writing II* (29) = 3.034
vs.
Legal Writing II*^ (33) = 2.768
vs.
Legal Writing II*^ (34) = 2.206
Fall 2009: Legal Writing I* (29) = 3.440 vs. Legal Writing I* (32) = 2.859
Fall 2009: Legal Writing I* (29) = 3.440 vs. Legal Writing I*^ (34) = 2.772
Spring 2009: Contracts II* (59) = 3.134 vs. Contracts II* (91) = 2.750
Spring 2009: Legal Writing II* (30) = 3.629 vs. Legal Writing II* (30) = 2.908
Spring 2009: Legal Writing II* (30) = 3.629 vs. Legal Writing II*^ (32) = 2.754
Fall 2008: Torts* (92) = 3.269 (20 As) vs. Torts* (60) = 2.625 (2 As)
Fall 2008: Contracts I* (92) = 3.128 (19 As) vs. Contracts I* (60) = 2.825 (4 As)
Fall 2007: Contracts I (D) = 3.160 vs. Contracts I (N) = 2.839
Fall 2007: Legal Writing I* (30) = 3.442 vs. Legal Writing I*^ (31) = 2.751
Spring 2007: Criminal Law* (84) = 3.176 vs. Criminal Law* (42) = 2.845
Spring 2007: Legal Writing II* (19) = 3.276 vs. Legal Writing II*^ (41) = 2.801
Fall 2006: Legal Writing I*^ (43) = 3.162 vs. Legal Writing I* (20) = 2.684

2. Upper-Level Courses – Variances in Sections of the Same Course

Spring 2010: Legal Profession* (33) = 3.273 vs. Legal Profession* (41) = 2.701
Fall 2009: Const. Law* (57) = 3.482 (20 As) vs. Const. Law* (86) = 2.781 (2 As)
Fall 2007: Evidence* (25) = 3.580 vs. Evidence* (57) = 3.035
Spring 2007: Business Assocs. (40) = 3.219 vs. Business Assocs. (16) = 2.875
Spring 2007: Legal Profession* (19) = 3.342 vs. Legal Profession* (71) = 3.007
Fall 2006: Evidence (16) = 3.422 vs. Evidence (76) = 2.572

3. First-Year Courses – Variances Across Courses

Spring 2010: Contracts II*^ (56) = 3.222 vs. Civ. Pro. II* (94) = 2.774
Fall 2009: Torts*^ (160) = 3.1955 vs. Property I*^ (100) = 2.544
Fall 2008: Torts* (92) = 3.269 vs. Civil Procedure I* (92) = 2.717
Fall 2007: Contracts I* (96) = 3.160 vs. Torts (59) = 2.636
Spring 2007: Criminal Law* (84) = 3.176 vs. Property II* (82) = 2.759

4. Upper-Level Courses – Variances Across Bar Courses

---

tained in Part III has sixty distinct grade disparities among pairs of courses. The list in Appendix 3 has 543 such variances for the same period of time (fall 2006 through spring 2010), though about eighteen of these are not significant.

15. At the Bowen Law School, the legal writing course is actually called "Reasoning, Writing, and Advocacy," or "RWA." *See* http://ualr.edu/law/academics/curriculum/course-descriptions/.

Spring 2010: Secured Transactions (56) = 3.442 vs. Decedents' Estates (63) = 2.968
Fall 2009: Secured Transactions (45) = 3.611 vs. Commercial Paper (40) = 2.944
Spring 2009: Secured Transactions (55) = 3.421 vs. Commercial Paper (52) = 2.860
Spring 2009: Secured Transactions (55) = 3.421 vs. Business Assocs. (66) = 2.742
Fall 2008: Secured Transactions (40) = 3.488 vs. Commercial Paper (33) = 2.727
Fall 2008: Conflicts of Law (20) = 3.463 vs. Evidence* (35) = 2.764
Spring 2008: Business Associations (87) = 3.338 v. Sales (17) = 2.471
Fall 2007: Sales (12) = 3.625 vs. Commercial Paper (66) = 2.845
Fall 2007: Evidence* (25) = 3.580 vs. Business Associations (18) = 2.917
Spring 2007: Secured Transactions (28) = 3.436 vs. Family Law (34) = 2.868
Fall 2006: Conflicts of Law (32) = 3.227 vs. Bus. Assoc. (48) = 2.729

5. Upper-Level Courses – Variances Across Electives

Spring 2010: Law Office Management (15) = 4.000 vs. Admin. Law (17) = 2.941
Spring 2010: Public Health Law (10) = 4.000 vs. Estate Planning (10) = 2.900
Fall 2009: Disability Law (20) = 3.632 vs. Poverty Law (38) = 3.000
Spring 2009: Real Estate Trans. (60) = 3.429 vs. Oil & Gas Law (35) = 2.707
Fall 2008: Advanced Torts Media Law (27) = 3.852 vs. Jurisprudence (28) = 3.259
Spring 2008: Real Estate Trans. (32) = 3.758 vs. Construction Law (11) = 3.023
Fall 2007: Local Government Law (28) = 3.759 vs. Health Law (21) = 2.583
Fall 2007: Administrative Law (49) = 3.597 vs. Juvenile Law (32) = 2.570
Spring 2007: Land Use (23) = 3.727 vs. Intro. To Int'l. Law (39) = 2.974
Spring 2007: Gov. Reg. of Business (39) = 3.667 vs. Fed. Income Tax (28) = 2.889
Fall 2006: Bankruptcy Law (21) = 3.475 vs. Federal Income Tax (36) = 2.951
Fall 2006: Intellectual Property (28) = 3.438 vs. White Collar Crime (65) = 2.906

6. Upper-Level Courses – Variances Across Seminars

Spring 2010: 4th Amend. (18) = 4.000 vs. Family Mediation (16) = 3.367
Fall 2009: Mediation (12) = 3.917 vs. Capital Punishment (16) = 3.266
Spring 2009: Race & Crim. Justice (16) = 3.953 vs. Animal Law (19) = 3.197
Spring 2008: Mediation (12) = 3.932 vs. Bioethics (16) = 3.469
Fall 2006: Capital Punishment (16) = 3.797 vs. Law & Soc. Science (9) = 3.306

7. Miscellaneous Additional Variances

Spring 2010: All 1-L *Night* Courses = 3.075 vs. All 1-L *Day* Courses = 2.890
Fall 2009: All 1-L *Night* Courses = 3.162 vs. All 1-L *Day* Courses = 2.984
Fall 2009: All 2-L/3-L *Night* Courses = 3.303 vs. All 2-L/3-L *Day* Courses = 3.096
Spring 2009: All 1-L *Night* Courses = 3.115 vs. All 1-L *Day* Courses = 2.893
Fall 2008: All 1-L *Day* Courses = 3.078 vs. All 1-L *Night* Courses = 2.847

Disparities such as these (and most of the others listed in Appendix 3) generally cannot be explained by differences in student achievement or ability. Of course, *some* of the grade variances listed above might reflect genuine performance differences, particularly where smaller, elective classes are

involved. But there is only a minute statistical likelihood that any significant number of these courses varied in performance by as much as the grades suggest.[16] Accordingly, the variations most likely flow from differences in faculty grading philosophy. Appendix 1 contains statistical analysis that supports this conclusion.

Crucially, I am *not* suggesting that any member of the faculty at my law school—or any other institution with similar variances[17]—is acting in bad faith. To the contrary, I firmly maintain that all of my colleagues at the Bowen Law School are grading in good faith. However, as the grade disparities show, my colleagues have *dramatically* different conceptions of

---

16. *See* Downs & Levit, *supra* note 9, at 829 & n.32 (explaining that, if there are two classes of students at a law school, each with sixty students, the students were randomly assigned to one class or the other, and one professor assigned grades with a mean of 2.2 and the other with a mean of 2.8, then "something is drastically wrong" because "such a large difference in grading means would be accounted for by chance"—i.e., by chance distribution of the superior students to one class—less than five percent of the time); *see also* CATHOLIC UNIVERSITY OF AMERICA GRADING POLICY, *available at* http://www.law.edu/res/docs/academic/announcements/lawannouncementsbook.pdf ("Where the credentials of entering students have been balanced across sections, as in the first year, or where there is a sufficiently large sample of students in a class, there is no empirical basis to justify widely divergent medians, averages [i.e., means], or distributions of grades among those sections and classes. The law school, therefore prescribes mean/median ranges for its courses."); Wangerin, *supra* note 8, at 112 (concluding that grade disparities at one law school that are comparable to those at Bowen "seem[] conclusively to demonstrate that dramatic differences in the definitions of letter grades exist within a single part of the university and even within different sections of the very same course").

17. *See, e.g.*, Wangerin, *supra* note 8, at 108 (discussing the results of a study of grading at one law school) ("Teachers [at the law school] gave wildly different grades, sometimes even within different sections of the same class. Further, teachers in this single [law] school did this year after year after year, for eight successive years."); *id.* at 110 (providing an example where the mean GPAs in two sections of the same course were 3.2 and 1.91); *id.* at 111–12 (setting forth comprehensive data on the grade disparities in tables 3 and 6); Post, *supra* note 7, at 796 & nn.74, 75 (discussing problematic grade variances at Touro College's Jacob D. Fuchsberg Law Center); *id.* at 802 (same). Downs & Levit, *supra* note 9, at 825 ("Prior to our faculty's adoption of the grade normalization plan at UMKC, grades, at least in the first year and required courses, varied dramatically depending upon which professor taught a particular course. In the first year, for example, we had variations from a high mean of 2.89 (close to a B average) to a low mean of 2.28 (essentially a C+ average) in different sections of the same course in the same semester."); Epstein, *supra* note 7, at 708 (noting that, prior to the implementation of grade normalization at USC, "it was not uncommon in a given subject for one section of the first year class to have a median grade four (or even more) points higher than the median for the other section," despite the fact that "every effort was made at the outset to equalize the strength of the sections"). Unjustified grade disparities have been a problem in legal education for at least 80 years. *See* John L. Grant, *The Single Standard in Grading*, 29 COLUM. L. REV. 920 (1929) (a study of grading practices at Columbia Law School). And they are also an issue in undergraduate education. *See* SHOUPING HU, BEYOND GRADE INFLATION: GRADING PROBLEMS IN HIGHER EDUCATION, 12–13 (ASHE Higher Education Report, Vol. 30, No. 6, 2005) (observing that adjunct and untenured faculty award higher grades than other teachers and that grades in certain departments tend to be much higher or lower than average); *id.* at 38–39 (summarizing studies of grade disparities across college departments); Talia Bar & Asaf Zussman, *Partisan Grading*, AMERICAN ECONOMIC JOURNAL: APPLIED ECONOMICS (forthcoming) (manuscript at 1, on file with the author and *available at* http://pluto.huji.ac.il/~azussman/partisan_grading.pdf) (in this study of grading practices at a particular university, the authors concluded that "student grades are linked to the political orientation of professors: relative to their Democratic colleagues, Republican professors are associated with a less egalitarian distribution of grades and with lower grades awarded to Black students relative to Whites."); *infra* note 67 (discussing studies that demonstrate that different university departments apply varying grading standards).

what constitutes "earning" an A, B, or C grade. That is where the problem lies. A mandatory curve will remedy the situation because it will largely eliminate the impact of my colleagues' divergent grading standards. As Professors Downs and Levit explain: "[G]rade normalization policies should not be thought of as punishment for bad faculty behavior but as a normative structure to help systematize irregularities in what is now an excessively subjective process."[18]

IV. ARGUMENTS IN SUPPORT OF MANDATORY CURVES

There are two principal reasons that law schools should use mandatory curves to eliminate grade disparities. First, grade disparities based on faculty grading philosophy rather than student merit are unfair. Second, such disparities incentivize students to select courses based on the grading practices of the professor, rather than on substantive grounds. Part IV explains these points more fully and sets forth some additional arguments in favor of mandatory curves.

A. Substantial Grade Disparities Are Unfair to Students

Significant grade disparities are unfair to students because they corrupt the process of reward distribution in law school. It is undisputed that grades are critically important.[19] They determine scholastic honors, placement prospects, and career paths.[20] Indeed, a recent study concluded that grades are even more important than the prestige or reputation of the law school one

---

18. Downs & Levit, *supra* note 9, at 822. Of course, one might suggest that there *is* some type of misconduct involved in certain circumstances, a form of negligence perhaps rather than bad faith. *Cf.* Post, *supra* note 7, 800 (observing that debates about grade normalization are not pleasant for faculties because "there is implicit in the entire proceeding an understanding that the behavior of some of the faculty has made" grade normalization necessary").

19. Downs & Levit, *supra* note 9, at 819 ("It is undeniable that grades matter . . . ."); Stake, *supra* note 2, at 585 ("We can argue about the degree to which grades matter, but few doubt that grades do matter."); BEST PRACTICES, *supra* note 9, at 235 ("Grades are important in law school.").

20. Douglas A. Henderson, *Uncivil Procedure: Ranking Law Students Among Their Peers*, 27 U. MICH. J. L. REFORM 399, 405–06 (1994) ("Without question, grades are universally perceived to determine the direction of legal careers no matter the specialty or the setting.") ("Other research demonstrates the important effect of grades later in life, not just in school."); Gerald F. Hess, *Heads and Hearts: The Teaching and Learning Environment in Law School*, 52 J. LEGAL EDUC. 75, 78 (2002) ("Grades and class rank are significant gatekeepers to the reward system during and after law school—law review membership, research or teaching-assistant positions, internships, and jobs."); Barbara Glesner Fines, *Competition and the Curve*, 65 UMKC L. REV. 879, 883 n.19 (1997) ("For example, Dean's list, Latin honors, Law Review membership, probation & dismissal are all often based on grades or rank."); *id.* at 886 ("Financial aid may be based on maintaining or achieving a particular grade average."); Downs & Levit, *supra* note 9, at 819 ("In addition, grades often are important in the determination of which students receive scholarships or other forms of financial aid."); Stake, *supra* note 2, at 584–86 (detailing various ways in which grades matter).

attends.[21] Thus, it is critical that grades reflect academic merit.[22] Allowing faculty to award widely divergent marks based on their individual grading philosophies unfairly substitutes grading approach for merit. Put more generally, grade disparities caused by differences in teacher grading philosophy violate the norms of distributive justice. The purpose of mandatory curves is to ensure that law school rewards are distributed according to student merit.[23]

The unfairness of grade variations is most pronounced in required, multi-sectioned courses—the courses that constitute the bulk of the first-year curriculum at virtually every law school. Students generally do not choose who their instructors will be in such classes. Those assigned to sections with "easier" graders have a crucial advantage over students assigned to courses with "tougher" graders, an advantage that has no relationship to academic merit. Given the importance of first-year grades,[24] providing any group of students with such an advantage is especially unfair. To quote my colleague, Professor Terrence Cain, under the system previously in operation at the Bowen Law School, students received higher grades not because of superior performance, but "based on accident of birth"—i.e., merely because they were assigned to courses with easier graders. Professor Jeffrey Stake makes the same point: "It is plainly unfair for some students to suffer lower grades because of a throw of the dice, and especially so because first year grades are the most important."[25]

---

21. Richard Sander & Jane Yakowitz, *The Secret of My Success: How Status, Prestige and School Performance Shape Legal Careers* 2 (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1640058 ("The consistent theme we find throughout this analysis is that performance in law school—as measured by law school grades—is the most important predictor of career success. It is decisively more important than law school 'eliteness.'").

22. *See* Post, *supra* note 7, at 790 ("Grading is not justified if it produces either false positives or false negatives in judging the quality of a student.") (citing Randall R. Curren, *Coercion and Ethics of Grading and Testing*, 45 EDUC. THEORY 425, 436 (1995)).

23. *See* Fines, *supra* note 20, at 892 (observing that grade normalization policies are "grounded in a distributive justice rationale—given that faculty differ significantly in their grading practices, and given that the rewards and punishments allotted to students may be profoundly affected by the chance of their assignment to particular professors, normalization policies are required to achieve equity"); *see also* Downs & Levit, *supra* note 9, at 825 ("'A grading system cannot be fair and effective . . . unless all faculty abide by the same general standards.'") (quoting one school's response to a survey conducted by the authors).

24. *See* Wangerin, *supra* note 8, at 104 ("First-year grades in law school play a tremendously important role in the lives of individual law students, probably a much more important role than a single year of grades in any other part of the university."); *id.* at 104 n.4 (noting that law review eligibility, moot court eligibility, research assistant positions with law professors, summer association positions with law firms, and scholarship eligibility are all frequently dependent on first-year grades and class rank); Roger C. Cramton, *The Current State of the Law Curriculum*, 32 J. LEGAL EDUC. 321, 328 (1982) ("The incentive and reward mechanism of law school turns almost entirely on first-year performance."); BEST PRACTICES, *supra* note 9, at 235 ("Grades are important in law school, particularly for first year students.").

25. Stake, *supra* note 2, at 588; *accord* Feinman, *supra* note 7, at 652 ("Normalization is used to prevent the inequity that otherwise would result from random section assignment."); Steve H. Nickles, *Examining and Grading in American Law Schools*, 30 ARK L. REV. 411, 457 (1977) ("Whether a student

Matters are only slightly better with respect to electives. Students frequently have little choice over who their instructor will be in bar courses, other doctrinal classes, seminars, and clinics. At smaller law schools, many courses are taught by only one member of the faculty. And even when a class is taught by two (or more), scheduling issues often prevent students from taking courses taught by the professor of their choosing. For example, some classes fill relatively early in the registration process, students in night programs cannot take off work to enroll in day courses, and professors are often unavailable to teach a given class. Therefore, the awarding of grades based on "accident of birth" is a problem across the entire curriculum.

To bring some concreteness to this analysis, consider grade disparities through the lens of class rank. Class rank is critically significant in the legal job market,[26] and schools frequently use rank in determining academic rewards such as law review membership and graduation honors. Indeed, class rank is generally more important to students than absolute GPA.[27] Ranking students is inappropriate, however, when professors grade using different metrics. Professors who give disproportionately low grades injure their own students in the rankings, and those who award disproportionately high grades harm the class ranks of students in other courses.[28] As explained by Professor Deborah Post, "without some standard for grading that makes the grades of the different faculty members comparable, ranking makes no sense. It is like comparing apples to oranges."[29]

---

remains in school, makes the law review, or receives any of the rewards associated with 'good' grades depends to a large degree upon the grades he receives in his first year of law work; and his first year grades depend to some extent, due to varying standards among teachers[,] upon the accident of mandatory course section assignments."); Fines, *supra* note 20, at 892 ("Indeed, it is unfair if a student's chance of being chosen for Law Review or for an interview with the elite corporate firm is based on the chance of being assigned to one particular professor rather than on merit."); Keating, *supra* note 9, at 185.

26. Henderson, *supra* note 20, at 405 ("Empirical research confirms that employers do use class rank to select students: the firms studied consistently used ranking as the key indicator of law school success."); *see also* Fines, *supra* note 20, at 886 (observing that most prestigious employers screen applications using class rank).

27. Keating, *supra* note 9, at 185.

28. *Id.* at 188; *see also id.* at 185 ("Therefore, when one teacher gives more total grade-wealth than other teachers, the effect is the same as if that generous teacher were deducting points from other teachers' students."); Fines, *supra* note 20, at 892 ("Both hard graders and easy graders distort the competitive process."); JOHNSON, *supra* note 8, at 222-23 (noting that when a professor awards all As, he "has tacitly ranked his students higher than all students who elected not to take his class," with the effect being "the elevation of his students' GPAs relative to the GPAs of all students who chose not to take his class").

29. Post, *supra* note 7, at 809; *accord* Stake, *supra* note 2, at 589 ("Variation in the average grade also results in unfairness and inefficiency when grades are combined. The students taking courses in which the teacher awards lower grades are less likely to qualify for honors, more likely to end up with a low class rank, and more likely to flunk out of law school entirely. In short, 'the reliability and validity of GPA as a criterion of academic success are attenuated because the GPA is not comparable for students who take courses with severe grading standards and students who take courses with lenient standards.'") (quoting Stricker et al., *Adjusting College Grade-Point Average Criteria for Variations in Grading Standards: A Comparison of Methods*, 79 J. APPLIED PSYCHOL. 178, 178 (1994)); Wangerin, *supra* note 8, at 113 (explaining that traditional class rank calculations "are corrupted by the different definitions that individual teachers give to the same letter grades); Georgakopoulos, *supra* note 8, at 446 ("The

The impact of disparate grading practices on class rank is not small. Taking courses from professors who use more generous scales can significantly alter a student's class standing.[30] To illustrate, consider the situation of a student ranked just outside the top half of the class at the end of the first year at the Bowen Law School. For the 2009-2010 school year, that would be the student ranked 104th out of 198. This student had a GPA of 3.01.[31] Suppose we elevate a single C+ earned by this student in a three-credit course to a B-. This would raise the student's GPA to 3.06,[32] improving the student's rank to 92nd out of 198.[33] That is a move of twelve places from a single, very small change.

Now, suppose we place this student in two three-credit courses with professors whose class GPAs averaged 0.5 higher than those of the professors teaching the other sections of the same subject, something that happened in the fall of 2008 at my school.[34] Suppose further that, as a result, this student received a B- instead of a C+ and a B instead of a C+ in the two courses. That would raise the student's GPA from 3.01 to 3.135.[35] And that elevates the student's class rank from 104th to 75th out of 198.[36] The student has gone from outside the top half of the class to the top 38%.[37] This could easily be the difference between receiving multiple job offers and receiving none. Indeed, according to Professor Paul Wangerin, "anecdotal evidence suggests that very, very small differences in first-year rank-in-class numbers can have extremely profound consequences for individual students. . . . [A] single rank difference for individual students at the end of the first year of

---

problem is not in the subjectivity of grades but in the aggregation of incomparable ones [into class rank]."); ST. JOHN'S FACULTY OF LAW STATEMENT ON GRADE NORMALIZATION, *supra* note 10 ("Grade normalization . . . is essential if students are to be fairly ranked and law school averages capable of being compared.").

30. *See* Epstein, *supra* note 7, at 707 n.2 ("For example, it is often the case that changes in grade average measured in fractions of a point can alter considerably a student's rank in class."); Georgakopoulos, *supra* note 8, at 454 ("I stress that even a slight difference in grading standards will affect a raw ranking.").

31. The document containing this information is on file with the author.

32. This GPA is reached by using the following calculation: 3.01 x 30 credits in the first year curriculum = 90.3 GPA points. Raising a grade from C+ (worth 2.5) to a B- (worth 3.0) in a three credit course is equivalent to 1.5 additional GPA points (0.5 x 3 = 1.5). 90.3 + 1.5 = 91.8. 91.8/30 = 3.06 GPA.

33. *See supra* note 31.

34. The class GPAs in Torts and Contracts I for one section (the day division) were 3.269 and 3.128, respectively. In the other section (the night division), the GPAs were 2.625 (Torts) and 2.825 (Contracts I). This comes to an average difference of 0.4735. (3.269 – 2.625 = .644; 3.128 – 2.825 = .303; .644 + .303 = .947; .947/2 = .4735.) This data is taken from the chart in Part III, *supra*.

35. Working from the figures in footnote 32, *supra*, raising a grade from C+ (worth 2.5) to a B (worth 3.25) in a three-credit course is equivalent to 2.25 GPA points (.75 x 3 = 2.25). 91.8 + 2.25 = 94.05. 94.05/30 = 3.135.

36. *See supra* note 31.

37. Note that the Bowen Law School has both day and night divisions and that the night division students take fewer than thirty credits. As a result, these examples actually understate the impact when professors in the two first-year sections vary substantially in their grading practices.

law school can mean—literally—everything, in the lives of these students."[38]

Of course, the impact of higher or lower grades is greater when a student is ranked in the middle of the class, at the center of the bell curve, than when the student is an outlier, near the top or bottom of the class. Nonetheless, small changes in class rank—e.g., moving into the top third or top ten percent—can elevate a student from also ran to serious candidate for many types of jobs, including judicial clerkships and associate positions at top-paying law firms.[39]

Professor Paul Wangerin's study of one law school illustrates that my hypotheticals are firmly based in reality. Among the students who started at that school in 1994,

> [t]he rank discrepancies for individual students [at the end of the first year] are astonishing. Several students in the group of 198 were ranked 40 or more places away from the places where they would have been ranked had grades been standardized among the various teachers. . . . [And some] students with the same raw rank would have been ranked almost 50 places apart had statistically standardized grades rather than raw grades been used.[40]

Professor Wangerin studied a total of eight entering classes. As he explains, "[i]n connection with every single entering group studied, I found dramatic rank discrepancies for large numbers of students."[41]

Individual faculty members should not have this type of control on the distribution of rewards at a law school.[42] And this is particularly so in a tightening job market.[43] Professors who award disproportionately high or low grades are imposing a substantial injustice on students.

One possible response to my analysis is that perhaps the "accidents of birth" will balance out as the students progress through school. If the assignment of students to courses taught by high and low graders is essentially

---

38. Wangerin, *supra* note 8, at 104.

39. *See also* Epstein, *supra* note 7, at 708 ("Moreover, in years where all the low (or, to take the other view, high) grading instructors were concentrated in one section, most of the top positions in the class were occupied by students from only one section.").

40. Wangerin, *supra* note 8, at 114.

41. *Id*. For a comparable study in the undergraduate context, see JOHNSON, *supra* note 8, at 209-17; *id*. at 195 ("Effects of these grading disparities on student assessment are severe. For many Duke students, the grading policies used by their instructors were nearly as important to determining their GPA and class rank as was their academic performance.").

42. *See* Fines, *supra* note 20, at 892 ("While we are perfectly willing to allow individual faculty members the freedom and power to teach and assess according to whatever approach they please (assuming minimal competence), we are unwilling to grant individual faculty the ability to control the external distribution of rewards.").

43. *See id*. ("Particularly in those schools where labor market and academic rewards are especially limited, one member of the faculty should not have disproportionate control over the currency that grades represent.").

random, then the disproportionately good and bad grades received by each student might cancel out by the time the student graduates.[44] The problem with this argument is that grades and class rank earned in the first year are the principal method for gaining access to the reward distribution system in law school.[45] As such, any balancing out over time provides small consolation to the students who receive low grades at the start of their legal education. Furthermore, Professor Wangerin's study powerfully demonstrates that any balancing almost certainly does not occur by the end of the first year. Professor Wangerin concluded that grade disparities at the law school that was the subject of his study positively or negatively impacted 166 out of 1350 students at the critical class rank thresholds of (i) first in the class, (ii) the law review membership cutoff, (iii) the scholarship cutoff, and (iv) the good standing cutoff. In other words, at the end of the first year, these 166 students were pushed into or out of one of the four relevant categories because of the different grading philosophies of the faculty.[46] "[G]rading luck did not balance itself out over time and repeated courses. Rather, at least at this school, and at least during the eight studied years, luck-of-the-draw grading differences produced dramatic and perhaps life-changing consequences for a large number of individual students."[47]

In sum, the unfairness of grade disparities among faculty is a compelling justification for the adoption of mandatory curves. Professor Steven Nickles articulates the point superbly:

> Law school policy which permits the standards to vary from teacher to teacher causes its evaluation process to be grossly misleading to the public and arbitrarily discriminatory to its students. These consequences are contrary to the theories of justice and fair play which the law school teaches as being obligatory in all affairs of law and society. If legal education is to meet its cultural and societal responsibilities owed to all of its constituent groups, it must integrate these same precepts into its procedures for student evaluation and eliminate chance, luck and logical scheming by students as bases for academic success.[48]

Other scholars have reached the same conclusion.[49] And such reasoning has motivated the adoption of a mandatory curve at many law schools.[50]

---

44. *See* Wangerin, *supra* note 8, at 99, 104, 113.
45. *See supra* note 24.
46. *See* Wangerin, *supra* note 8, 114–17.
47. *Id*. at 108–09.
48. Nickles, *supra* note 25, at 457–59 (citation omitted).
49. *See, e.g.*, Stake, *supra* note 2, at 588 ("I start with the easiest point: the average grade should be approximately the same for all courses."); Keating, *supra* note 9, at 185 ("Myth No. 9: Mandatory Medians And Other Such Restrictions on a Grader's Autonomy Are Undesirable and Unfair"); Downs & Levit, *supra* note 9, at 857 (explaining that, with certain structural features, grade normalization policies

B. Substantial Grade Disparities Distort the Process of Course Selection by Students

Wide grade disparities detached from student merit are also problematic because they distort the process of course selection by students. It is well-established that students consider the grading philosophy of the instructor in registering for classes.[51] And the greater the disparity in grading philosophy, the greater the weight allocated to this factor by students. Indeed, a recent graduate of my law school designed and administered a website for the express purpose of aiding other students in selecting courses based on the grading practices of the teacher.[52]

"begin to strike a balance between concerns for uniformity and equity"); Post, *supra* note 7, at 793 ("Grading in law schools violates ethical norms and ranking in law schools without any attempt to insure comparability must be some sort of ethical violation."); *see also* Wangerin, *supra* note 8, at 101–03, 117 (arguing for the adoption of a type of statistical normalization that would have roughly the same impact as a mandatory curve); Fines, *supra* note 20, at 892 (observing that both "hard graders and easy graders distort the competitive process" of rewards at a law school, causing unfairness in the distribution of job opportunities and academic honors) (note that Fines is ultimately somewhat skeptical of mandatory curves); Epstein, *supra* note 7 (advocating for a type of formal grade normalization).

50. Downs & Levit, *supra* note 9, at 825 ("In response to our survey question which asked why law schools adopted grade normalization plans, many schools responded that they had noticed grading disparities among professors."); *id.* at 821 (survey was conducted in 1996 and was sent to all 179 of the law schools accredited by the ABA at that time); *id.* at 843 ("We asked schools the reasons why they adopted grade normalization policies. Of the fifty-two schools answering the question, thirty-six (69%) mentioned concerns of fairness, equality among sections, or fears of inequitable grading."); Feinman, *supra* note 7, at 652 (observing that concerns about the fairness of grading disparities based on individual professor grading philosophy "have led some schools to adopt measures to enhance the uniformity of the grading system"). *See, e.g.*, Mroch, *supra* note 9, at 55 (quoting the following language from one school's grade policy: "While complete uniformity in grading is neither necessary nor expected, a significant disparity is not only inequitable to the students, but also undermines a school's systems of student rewards and sanctions"); *see also* UNIVERSITY OF KENTUCKY COLLEGE OF LAW FACULTY RULES AND POLICIES § IX.A., *available at* http://www.law.uky.edu/files/docs/students/Excerpts_from_Policies.pdf ("The average grade rule serves the fundamental interest of fairness to students . . . ."); LOYOLA LAW SCHOOL OF LOS ANGELES SUMMARY OF GRADING POLICIES, *available at* http://reg.lls.edu/gradesandrankings/jd_grades.html (mandatory curve implemented to eliminate grade disparities which "undermine the reliability of student grades and make comparisons difficult for employers and others") ("Normalization reduces arbitrariness in grading and promotes uniform and fixed meaning."); *supra* notes 10 and 29 (quoting the St. John's Faculty of Law Statement on Grade Normalization).

51. *See, e.g.*, Henderson, *supra* note 20, at 423 ("Upper-class [law] students, aware of the differences in professor grading, engage in 'forum shopping' to improve the vagaries in course selection and improve the basis of their grades . . . ."); Keating, *supra* note 9, at 185 (same); Stake, *supra* note 2, at 585 (same); *see also* Hu, *supra* note 17, at 42 ("Grading disparity . . . affects students' choice of courses."); JOHNSON, *supra* note 8, at 172-93 (presenting statistical analysis indicating that professor grading philosophy influences students when they are choosing (1) which instructors to take a particular course from, (2) which courses to take *within* a particular department, and (3) which courses to take *across* departments); GEORGE C. LEEF, DEGRADED CURRENCY: THE PROBLEM OF GRADE INFLATION 18 (American Council of Trustees and Alumni, 2003) ("Finally, grade inflation creates a strong incentive for students to enroll in courses taught by professors who are known to be easy graders.").

52. *See* KEITH PIKE'S HANDY-DANDY LAW SCHOOL GPA CALCULATOR, www.lawgpa.com (setting forth the University of Arkansas at Little Rock, William H. Bowen School of Law grade distributions for Spring 2006 through Summer 2009 "[t]o help [students] estimate what you might get in your

Given the importance of grades, it is perfectly rational for students to grant significant weight to professor grading practices when choosing courses. For example, I strongly suspect that a student usually has a better chance of obtaining a job by taking classes where he earns high grades but learns less, than by taking courses where he earns low grades but learns more.[53] Likewise, it is probably more advantageous to receive high grades in classes a student finds boring than to receive average or low grades in classes the student finds interesting.

It is axiomatic that faculty members do not want students choosing courses based on the grading philosophy of the professor. We would rather they select classes using other criteria, such as (1) substantive importance, (2) student interest, (3) career relevancy, (4) skill development, (5) fit between the instructor's teaching style and the student's learning style, (6) curricular balance, and (7) whether the topic is on the bar exam. By removing grading concerns from the decision-making process, mandatory curves properly incentivize student course registration. And the grading policies of many schools identify the prevention of "grade-seeking course selection" as a basis for their curves.[54]

C. Other Arguments in Favor of Mandatory Curves

Four additional arguments in favor of mandatory curves are worth noting briefly. First, grade disparities have incentive effects on students beyond class selection. For example, because variances cause grades to convey inaccurate information, students might be unduly encouraged or discouraged about their performance. A student whose grades are too high may "gain a false sense of confidence that he can do legal work without as much preparation as others."[55] And a student whose grades are too low may give up on studying or even drop out of school.[56] In addition, professor grading practices can distort students' allocation of their limited studying time,[57] causing them to put less effort into "easy grade" classes and more into "hard grade" classes.

---

class"). This student also sent e-mails over our student listserv detailing the grading practices of faculty who awarded particularly low grades.

53. Of course, the student may not be as prepared for the job. But I believe most students would accept that tradeoff, and rationally so.

54. *See, e.g.*, KENTUCKY FACULTY RULES AND POLICIES, *supra* note 50, at § IX.A. ("The average grade rule serves the fundamental interest of fairness to students and the secondary interest of encouraging students to use sound bases for selecting courses."); LOYOLA LAW SCHOOL GRADING POLICIES, *supra* note 50 (mandatory curve implemented to eliminate grade disparities, which has the effect of minimizing "grade shopping").

55. Stake, *supra* note 2, at 585.

56. *Id.*

57. *Id.*

Second, while there is research indicating that students sometimes dislike mandatory curves,[58] there is also evidence that students tend to prefer a mandatory curve when the alternative is substantial grading variances based on professor grading philosophy.[59] Students resent such grade disparities.[60] And they "generally appreciate the reality that grade normalization is designed to protect them from professors who use an inordinately low scale in assigning grades, and from the professors who give very high grades to students in other courses."[61] This is certainly true of the students at the Bowen Law School, whose elected representatives voted *unanimously* in support of the curve that I proposed to the faculty there.[62]

Third, grading disparities can cause dissension among faculty members.[63] This was certainly a problem at the Bowen Law School prior to the adoption of the curve that I proposed.

Fourth, mandatory curves prevent undue grade inflation.[64] Grade variances put pressure on teachers to raise grades to match the practices of their colleagues, which can result "in spiraling grade inflation."[65] As some of the examples set forth in Part III demonstrate, this clearly was an issue at my institution. For example, I teach Contracts I & II, Secured Transactions, and Jurisprudence. Recently, the other Contracts teacher and the other Secured Transaction teacher dramatically raised their grades. Out of a concern for fairness to my students, I followed suit. Grades in Secured Transactions at my school reached particularly, and unreasonably, high levels. Similarly, while I am the only Jurisprudence professor, I recently raised the grades I awarded in that course to match the grading practices of those who teach

---

58. *See* Lawrence S. Krieger, *Human Nature as a New Guiding Philosophy for Legal Education and the Profession*, 47 WASHBURN L.J. 247, 298 (2008).

59. *See* Downs & Levit, *supra* note 9, at 832 n.36 ("Our experience at UMKC with students regarding grade normalization is that students . . . much prefer a normalization among sections to the specter of having as their professors those with lower grading standards."); Stake, *supra* note 2, at 591 (allowing a teacher to give consistently higher grades raises a perception of unfairness and "generate ill will among students."); *see also* Post, *supra* note 7, at 796 (explaining that students at the author's school engaged in a "massive protest" when two members of the faculty gave unusually low grades); *id.* at 803–04 (discussing the effort by students at the author's law school to implement grade normalization policies to address grade disparities and the effect such disparities "had on their ability to obtain gainful employment").

60. *See* Downs & Levit, *supra* note 9, at 825; Stake, *supra* note 2, at 591; Post, *supra* note 7, at 796.

61. Downs & Levit, *supra* note 9, at 855.

62. Minutes from SBA Meeting Sunday, April 18, 2010, University of Arkansas at Little Rock, William H. Bowen School of Law (on file with the author).

63. *See, e.g.*, Keating, *supra* note 9, at 185 (detailing the situation at one law school). *Cf.* Post, *supra* note 7, at 800 ("Grade normalization places tremendous pressure on faculty, especially when it is forced to debate publicly its own behavior. . . . A faculty is asked as faculty to consider the consequences of their actions on the students they teach and to deliberate on the usefulness of self-regulation. Self-regulation is not a pleasant process, particularly when there is implicit in the entire proceeding an understanding that the behavior of some of the faculty has made this self-regulation necessary.").

64. *See* Leef, *supra* note 51, at 19.

65. Stake, *supra* note 2, at 588.

other electives. Prior to the adoption of a mandatory curve, I was awarding grades in Secured Transactions and Jurisprudence in the 3.4 to 3.5 range.

V. OBJECTIONS TO MANDATORY CURVES

Opponents of mandatory curves offer a number of criticisms. This section addresses the most powerful and most popular critiques.

A. Objection One: "Mandatory Curves Prevent Teachers from Giving Students the Grades They Deserve"

Probably the most common criticism of mandatory curves is that they restrict professors from awarding the grades that students actually earned.[66] This argument is really two distinct criticisms. First, mandatory curves prevent grades from reflecting students' *absolute* performance. Second, curves prevent grades from reflecting students' superior or inferior *relative* performance. Both positions are fatally flawed.

The problem with the *absolute* performance argument is *conceptual*. To establish that mandatory curves stop teachers from awarding the "correct" grade, there must be an independently valid or shared concept of desert that consistently establishes when a student deserves an A, a B, or some other grade. But no such concept exists. Professors have substantial disagreements regarding the standards that should be used to assess student performance. And, more specifically, we have different understandings of what constitutes "A work," "B work," and "C work." These designations are used in dramatically varying ways by (1) different academic fields, (2) different schools within the same field, (3) different professors within the same school, (4) different professors within the same department, and (5) even different professors who teach the same class.[67] For example, in some programs, a C means satisfactory performance.[68] In others, it denotes failing

---

66. *See* Post, *supra* note 7, at 806 ("'There is no such thing as a normal class or a normal distribution of students in a class. . . . Normalization fails to give students the grade they deserve.'" (quoting Response by Professor X to Author's Memorandum to Faculty soliciting comments on grade normalization) (on file with Professor Post). This argument has been a constant refrain at my school.

67. *See* Wangerin, *supra* note 8, at 99 ("For years, statisticians and educational researchers have explored the foregoing grading-differences problem in complex statistical studies of grading data. These studies have universally demonstrated that different departments in the overall university grade differently."); *id.* at 100 n.1 (identifying four studies described in the previous parenthetical); *id.* at 112 (finding that "dramatic differences in the definitions of letter grades" existed within the law school that was the subject of this study, and that the such differences existed "even with different sections of the very same course" at the law school); Fines, *supra* note 20, at 882 ("However, grades have as many meanings as the criteria with which we test. . . . In sum, there are too many variations in teaching and testing to be able to say that grades have a fixed meaning outside the classroom."). For an excellent literature review and a separate study of this issue, see JOHNSON, *supra* note 8, at 197-209.

68. *See* UALR BOWEN LAW SCHOOL ACADEMIC RULES, *supra* note 3, at 12 (providing that a C+ = 2.5, a C = 2.25, a C- = 2.0, and that the GPA necessary to remain in good standing is 2.0).

work.[69] Letter-grade designations clearly lack the objective content necessary to settle such differences.[70]

Of course, none of this means that grading is either arbitrary or wholly subjective. There certainly is a great deal of objectivity in grading. *In clear cases*, there is much agreement among professors about what constitutes "excellent," "satisfactory," and "failing" work. The problem is that there is not *enough* consensus to establish that a mandatory curve bars professors from awarding students the grades they "deserve" in any absolute sense.

My contention here that the concept of "desert" or "merit" is too subjective to support the claim that mandatory curves prevent professors from giving students the grades they "objectively" earned is perfectly consistent with my argument in Part IV.A., above, that grade disparities based on professor grading philosophy rather than student merit are unfair. There is nowhere near enough objectivity in the concept of academic merit to support the assertion that mandatory curves stop professors from awarding students the correct, absolute grades. But there is more than enough objectivity in the concept to justify the conclusion that grades based on professor grading philosophy are unfair. That is because grading philosophy is clearly not an aspect of student merit. To make the point syllogistically: Fairness demands that grades reflect merit. Professor grading philosophy is not a part of merit. Therefore, awarding grades based on teacher grading philosophy is unfair.

The *relative* performance argument suffers from a different problem. According to this position, sometimes the students in a class perform better or worse than the average class. In these cases, a mandatory curve forces the teacher to give grades that are either too high or too low.[71] The weakness with this argument is *evidentiary*: Professors lack the information required in order to justifiably claim that their students "merited" or "deserved" higher or lower grades than students in other classes. Professor Daniel Keating's explanation of this point is illuminating and worth quoting in full:

---

69. *See* TENNESSEE TECH UNIVERSITY GRADUATE CATALOG, *available at* http://www.tntech.edu/gcatalog/grading/ ("A grade of 'C' is *considered a failing grade* in the doctoral program.") (emphasis added).

70. *See* Keating, *supra* note 9, at 178–79 (explaining that it is a "myth" the grades "A" or "B" or "C" have some "absolute meaning . . . wholly apart from where that grade places a student within a particular class"); LOYOLA LAW SCHOOL GRADING POLICIES, *supra* note 50 (explaining that "grades convey no information on a student's absolute performance, such as 'fair,' 'good' or 'excellent'"). One might argue that faculty at a given law school should develop a common understanding of "A work," "B work," and "C work," and that this will lead to consistency in grading. Given the differences among faculty when it comes to assessment, this is not realistic. For a discussion of why that is, see *infra* Part V.E.2.

71. *See* Feinman, *supra* note 7, at 652 ("For many reasons, students in one section of a course may achieve at higher levels than students in another. Sections also may vary in their achievement from year to year. Normalization does not permit the recognition of these differences."); *see also* Stake, *supra* note 2, at 590 ("Another argument for disparate averages is that the students are better in some classes.").

> The most common argument against the imposition of a mandatory mean was that perhaps the current differences in grade means among professors were in fact justified by the differing performances of students in different classes. The simple response to this is maybe so, but how in the world would anyone ever be in a position to know that? There are really only two things that a professor grading a set of exams can determine about those exams with any degree of certainty: the relative rank ordering of those exams and the approximate raw-point distance apart of those exams from one another. As to any other assessment of those exams, we simply lack sufficient information. Are we in a position to assert that our exams are better overall than those being graded by another professor who teaches a different course? Are we in a position to assert that our exams are better overall than those being graded by another professor who teaches the same course? *If we are not in a position to assert that our students' overall performance was better or worse than the overall performance of some other class, then why should we be able to distribute more (or less) grade-wealth than some other professor distributes to his or her class?*[72]

The same point stands even if a professor could be reasonably certain that this year's students performed better than last year's students (an unlikely scenario).[73] Such certainty would still not justify variations in grades among professors:

> The reason is that we would still lack at least two key pieces of information to make an informed judgment about how much total grade-wealth to distribute to our class relative to those taught by other faculty. First, we do not know how well students are performing in other classes. It may be that the entire student body is stronger or weaker, in which case the rise and fall that we are seeing in our classes' performances would similarly be reflected in other professors' classes. Second, even if we know something about the performance of students in classes we are not teaching,

72. Keating, *supra* note 9, at 186 (emphasis added); *accord* LOYOLA LAW SCHOOL GRADING POLICIES, *supra* note 50 (explaining that "[g]rades [only] indicate how well students perform in a class compared to other students in that class").

73. Keating, *supra* note 9, at 186 ("We might believe that we are at least in a position to assess how the performance of this class compares with that of classes from previous years in which we have taught this course. I am personally dubious about the likely accuracy of even this sort of assessment, given memory lapses, differences in the particular performance-measuring device we use from year to year, and even differences in how well we taught the same course."); *see also* Stake, *supra* note 2, at 590–91 ("Substantial year-to-year variation in students is possible, but it is also possible, indeed likely, for a teacher's test or grading to vary from year to year. . . . It is much more likely, however, that his teaching, testing, or grading mood had changed.").

> we do not know how other faculty will choose to assess that performance. To put this point another way, *we lack information not only about the relative performance of our particular class compared to others but also about the way in which that performance is being assessed*.[74]

Of course, we all understand how a faculty member might believe that he can distinguish between the performances of his classes from year to year, and thus that he is in a position to award higher grades in some years. But even if this belief is true, it is ultimately insufficient. That is because it "ignores the reality that when we give our own class more or less total grade-wealth than that given by a colleague, we are making an implicit (if unwitting) statement about the relative effort or performance of our group compared to theirs."[75] However, professors lack the information necessary to make such a judgment.[76]

One normative assumption underlying Keating's position is that "grade-wealth" should be distributed evenly across courses in the absence of evidence supporting an alternative distribution. Basic principles of equality support this assumption. Suppose that ten people are placed on a desert island with fifty items of food of equal nourishment value. Assume further that no evidence is available as to the caloric and health needs of the ten individuals, evidence that would assist in deciding how to allocate the food. The only fact known is that each of the ten people on the island is in need of food. Fairness demands that we divide the fifty items of food equally even if it turns out (as it almost certainly would) that the ten individuals have different caloric and health needs and thus that, given better information, an equal division would not be the fairest. Grade wealth *across classes* is subject to the same calculus. If we have no evidence regarding which classes of students are performing better (or who needs more food)—if we know only that each class of students is performing in some form and needs grades (needs food)—but we still must divide up the grades (the food), then distributing the grades (the food) equally is the morally best approach. This explains why opponents of mandatory curves bear the burden of proof. The

---

74. Keating, *supra* note 9, at 186–87 (emphasis added); *accord* Stake, *supra* note 2, at 590 ("Few would object to giving better students better grades, but again there is a problem of proof.").

75. Keating, *supra* note 9, at 187.

76. *Id.* There is no need to assume that the students in two different classes are substantially equivalent when their incoming GPAs in prior law school work are not the same. And, as discussed in Part VI.A.2., *infra*, that becomes a significant concern in courses with less than thirty students. As a result, the mandatory curve I proposed to my faculty permits professors to give higher or lower grades in smaller courses where the students' law school GPAs—"[t]he only objective and discernible information that we have about the relative differences in the quality of various groups of students," Keating, *supra* note 9, at 187—suggests that such grades are justified. *See infra* Appendix 2; *see also* Epstein, *supra* note 7, at 709 (noting that the prior law school records of the students in a course are the "best evidence" of how the students will perform in the class).

moral presumption is equal division—of food or grades. Those against equal division on the desert island or at a law school must demonstrate that the presumption fails by empirically establishing that need on the island or performance in the classroom is not in fact equal.[77]

In sum, professors do not possess the information necessary to make fair assessments of relative merit across classes. Unless a professor has a firm basis for believing that *his students* performed better than the students in *another class* taught by a *different instructor*, the professor is in no position to assert that his students "deserve" or "earned" better grades than the other students.

The information problem described above is also relevant in four other ways. First, since the burden of proof falls on opponents of forced curves, any weakness in my statistical analysis of the grade disparities in Part III does not significantly undercut the case for mandatory curves. A mere lack of evidence that professor grading philosophy plays a substantial role in assigning marks does not justify unregulated grading practices. Instead, critics of forced curves must demonstrate that professors can consistently proffer reliable evidence that their students performed better or worse than students in other classes. The critics have not met this burden.

Second, some commentators contend that mandatory curves are premised on the unrealistic assumption that all students perform the same in every class.[78] This is not true. Mandatory curves are premised on the fact that we do not *know* whether students are performing better in one class than another, not on the view that students *in fact* never perform better in one class than another.

Third, some schools permit grade normalization policies to be waived in particular circumstances, such as where the professor believes the students' performance in the course was particularly strong or weak.[79] But such a belief does not justify the awarding of higher or lower grades since the professor does not know whether his students performed better or worse than students in courses *taught by other professors*. Given this reasoning,

---

77. Downs and Levit present a somewhat similar argument:

> So, if we can't prove that one randomly selected group of students is, in fact, better than another, and if we can't prove whether the professor was any good or not, what can we do? We can, and do, assume that the student groups are substantially equivalent, and that the professors are all good teachers. The only things left are whether professors are good graders and are all using the same standard of measurement. Since we cannot control the problems inherent in scoring blue books, the focus needs to be on the standard of measurement. Our sense is that, all other things being neutral, statistically equivalent student groups in different sections should have grade means which are reasonably close together.

Downs & Levit, *supra* note 9, at 831.

78. Krieger, *supra* note 58, at 299.

79. *See* Downs & Levit, *supra* note 9, at 845, 847; Mroch, *supra* note 9, at 7 ("Many policies specify that dean approval may override the grading policy.").

there is no mechanism for requesting an exception to the mandatory curve in the grading system I proposed at the Bowen Law School.[80]

Fourth, it is often argued that some law school courses are either more difficult or easier than others and that this justifies lower or higher grades in those courses.[81] Keating's analysis suggests the flaw with the first part of this argument: We have nowhere near enough systematic, reliable evidence about which law school subjects are more difficult or easier than others to justify awarding higher or lower grades in particular classes.[82] In addition, the second part of the argument does not follow from the first; the mere fact that a subject is more challenging does not justify lower grades. Rather, if a course is more difficult, the teacher's expectations should be lower for grading purposes. To illustrate, I teach Jurisprudence, a course that, based on anecdotal evidence, my students and I find to be one of the most demanding at the law school. As best I can tell, my students' level of mastery of the legal philosophy and constitutional theory I teach in that class always falls *well* short of their level of mastery in Secured Transactions, my other upper-level course. Nonetheless, under the prior grading system at Bowen, I gave roughly the same grades in each subject. That stems from the following reasoning: Since Jurisprudence involves more difficult material, I do not expect my students to grasp the topics at the same level I expect of my students in Secured Transactions, which is a traditional doctrinal class. Frankly, holding my students in both courses to the same standard of proficiency strikes me as grossly unfair. And, in the absence of a mandatory curve, it would deter even most philosophy majors from registering for Jurisprudence. The same reasoning applies in reverse to courses that are allegedly easier than average.

Finally, there is another, critical response to the objection that mandatory curves prohibit professors from awarding students the grades they earned: The harm caused by a mandatory curve is the lesser evil. Assume away the conceptual and informational problems articulated above. As a statistical matter, there should be very few classes of students who perform substantially better or worse (for reasons having to do with academic merit) than the average class. Thus, in only a small fraction of cases will a mandatory curve prevent professors from giving students the grades they earned. Yes, it will occur now and then, even in large doctrinal classes with a random sampling of students. But the odds of this happening, even in small

---

80. *See also* Downs & Levit, *supra* note 9, at 847 (noting that waiver protocols based on a teacher's subjective sense that the students in a class performed above or below average threaten to swallow the rule).

81. *See, e.g.*, Fines, *supra* note 20, at 891 ("Moreover, some subjects may be more intrinsically difficult than others, so that student achievement will be lower.").

82. *See also* Nickles, *supra* note 25, at 455 n.149 ("There is also little merit in the contention that a wide divergence in grades is attributable to variances in the difficulty experienced by students in dealing with the subject matter of respective courses. In fact, evidence suggests the divergence in grading bears no relation to the profundity of the materials.").

electives, are *much* lower than the odds of a student ending up in a class with a professor who gives disproportionately high or low grades based on that professor's individual grading philosophy.[83] And thus, the relatively few times a curve might force a teacher to award grades that are either too high or too low constitute a small harm in comparison to the systematic and unfair grading disparities that can exist without a mandatory curve.[84] At the Bowen Law School, instead of a class or two suffering every few years pursuant to a mandatory curve, we were burdened with a system in which large numbers of students were unjustly harmed every semester. And the same is true at other law schools.[85] "Obviously we cannot entirely eliminate inequity in our sorting processes."[86] What we can do, however, is adopt grading systems that minimize unfairness. Mandatory curves best accomplish that end.

B. Objection Two: "Grade Disparities Do Not Justify Mandatory Curves Because Many Disparities Are Between Courses That Are 'Apples' and 'Oranges'"

One of my colleagues at the Bowen Law School offered an objection to mandatory curves that is related to the contention that curves stop professors from awarding students the grades they deserve. Her argument is that many of the courses in the law school curriculum—including one course that she teaches and one that I teach—are so different that comparing grades in the courses is like "comparing apples to oranges." From this, she concluded that the grade disparity data I presented does not justify a mandatory curve. A number of critics at our law school and elsewhere have expressed sympathy for my colleague's argument. But this argument is clearly invalid. Rather than undercutting the relevancy of my grading data, the fact that two courses are "apples and oranges" compellingly establishes the importance of the data and the need for a mandatory curve.

---

83. *See* Stake, *supra* note 2, at 617 (explaining that, for large classes, a teacher's above- or below-average grades are far more likely the result of a defective assessment instrument than of an abnormal class population; the risk of having a group of non-average students is greater in small courses, but still smaller than the risk of a defective assessment); Georgakopoulos, *supra* note 8, at 453 (indicating that a student's risk, in an unrestrained system, of having a professor with undesirable grading tendencies is much greater than the student's risk, under grade normalization, of enrolling in a course with unusually talented students); *see also* Wangerin, *supra* note 8, at 103 (The statistics-based normalization system proposed by the author "will produce distorted grading information in the relatively rare instances where the random assignment of individual students to different groups produces groups of significantly different overall ability. Conversely, the letter-grade process will produce distorted grading information in the relatively common instances where the random assignment of individuals to different groups produces groups of roughly comparable overall ability.").

84. Georgakopoulos, *supra* note 8, at 454 ("We should be ready to accept the remote danger that a class of top students might be hurt in order to eliminate the more likely injustice of some courses systematically giving higher grades than others.").

85. *See, e.g.*, Wangerin, *supra* note 8, at 108–17 (detailing the results of a study of one law school's grading practices).

86. Fines, *supra* note 20, at 895.

First, even if two classes are in fact so different that they cannot be compared, the courses are both graded using the official law school scale. This means that grades in the two classes operate in exactly the same way. For example, they have the same impact on class rank—i.e., high grades raise class rank and low grades reduce class rank—and they appear the same on transcripts. In other words, even if my colleague is grading apples and I am grading oranges (as she contends), all of the grades we award are used in identical ways for purposes of reward distribution at the law school. Second, recall Keating's information argument—professor's lack the evidence necessary to justify giving higher or lower grades to their students than are awarded in other courses taught by different teachers. Now, bring these two points together. Because my colleague contends that she is assessing oranges in her course and I am assessing apples in mine, she not only lacks any *evidentiary* basis to believe that her students are performing better than mine, she is also acknowledging that there is a *conceptual* problem in comparing our students. She is implying that even if we conducted the necessary empirical investigation, the differences between our two classes are so great that any data we uncovered would not enable us to make fruitful comparisons. This means that my colleague is conceding that it is impossible *in principle* to compare our courses in a useful way. If this is true, what possible justification could my colleague have for awarding her students higher grades than mine? What could legitimize giving one set of students higher grades than another set of students on the same scale when we are unable to compare the two groups of students for purposes of applying the scale? In sum, my colleague's "apples and oranges" argument, if it is correct, establishes that no opponent of a mandatory curve could *ever* proffer evidence justifying awarding higher grades in one course than another when the classes are "apples and oranges." And without such evidence, no teacher should be allowed to award higher or lower grades to his or her students.

C. Objection Three: "Mandatory Curves Prevent the Best Teachers from Awarding Grades That Reflect Their Superior Teaching"

Some opponents of mandatory curves contend that such policies prevent grades from reflecting differences in the quality of teaching.[87] Indeed, one of my colleagues at Bowen argued that some of the grade disparities at our school were caused by variations in the quality of teaching rather than by differences in grading philosophy. According to my colleague, the reason Professor A gave higher grades than Professor B was that Professor A's students performed better because A is a superior teacher.

87. *See* Stake, *supra* note 2, at 589 ("I have been asked to address the argument that we should allow teachers to give higher or lower grades because some teachers teach better.").

This criticism is essentially another version of the first objection—that mandatory curves prevent professors from awarding students the grades they earned. As such, it is flawed for the same reason: We lack the evidence that would justify allowing some teachers to award higher marks.

It is extremely difficult to identify better and worse teachers.[88] But even if we could, higher grades would only be justified if the students receiving the better teaching verifiably performed better than other students. To elaborate, the link between teaching quality and the amount learned, on the one hand, and grades, on the other hand, is limited. Grades are clearly about more than what students learned. If a student enters a course knowing the subject and learns nothing, the student will still receive an A if she performs at the top of the class. Consider another example. Suppose that Professors X and Y both teach Contracts. X is the better teacher. As a result, X's students learned more than Y's in the course. But Y's students were much more talented and so they displayed greater mastery of Contracts on their finals than X's students. Who deserves the higher grades? Virtually all professors would say Y's students. Accordingly, unless an instructor can proffer evidence that his students outperformed students in another course, the instructor has no basis to award the students higher marks, even if he could prove that he is a better teacher.[89] And as discussed in Part V.A., above, evidence of superior performance by the students will generally not be available.[90]

Beyond the lack of information, there is a second problem with the "teaching quality" objection to mandatory curves: Superior teaching has little to do with student merit. To illustrate, assume X's students verifiably perform better than Y's students because X is the superior teacher. Assume further that the students in both classes were randomly assigned to each because we are dealing with a first-year, required course. How is it fair to punish Y's students with lower grades simply because they were assigned to a course with a worse instructor? If X is permitted to award higher marks, X's students will receive a windfall because of sheer luck.[91]

---

88. Stake, *supra* note 2, at 590; *see also* Downs & Levit, *supra* note 9, at 847 ("'[I]t is only natural for us to believe that, compared to most of our colleagues, we can inspire a greater effort from our students than those same students would put out in other classes. . . . The problem with this logic, aside from the human tendency to self-delusion, is that we cannot all be right.'") (quoting a response to a survey of law schools on grading practices).

89. *Cf.* Stake, *supra* note 2, at 590 ("Unless a teacher can produce decent evidence of better learning on the part of her students, we should hesitate to accept such a contention as sufficient to warrant higher grading.").

90. As explained in note 76, *supra*, evidence of superior or inferior performance *will* exist in smaller classes filled with students whose GPAs, taken together, are higher or lower than the average GPA at the law school.

91. *See* Stake, *supra* note 2, at 590 (concluding that "it would be unfair to the students to allow poor teaching to result in low grades in addition to weaker training than that of their luckier schoolmates"). Or, suppose Professor Z does a good teaching job one year and a bad teaching job the next. Why should the lucky students in Z's class the first year receive higher grades than the unlucky students in Z's class the next year?

Alternatively, suppose Professor J teaches a very popular and important bar course, such as Secured Transactions, but is only an average teacher. Professor K teaches a relatively unpopular elective, such as Jurisprudence, but is an excellent teacher. How is it fair to impose lower grades on J's students simply because they are interested in the more popular topic or are concerned about the bar exam?

Focusing on the first example, one might reply that since X's students developed greater mastery of Contracts, awarding similar grades to the students from X's and Y's classes will misinform consumers of transcript information. A prospective employer, for example, might be misled into believing students in each class are at comparable levels of proficiency in the subject. But any unfairness flowing from such misinformation pales in comparison to the unfairness that results from awarding grades based on factors having nothing to do with student merit—e.g., luck of the draw in course registration. Our first duty as law professors is to our students, not to future employers.[92]

A related reply is that perhaps X's and Y's students will themselves be confused about their degree of competency in Contracts. This is a legitimate concern, to be sure. However, once again, the unfairness of assessing students on grounds that are beyond their control is the greater harm.[93] Admittedly, that is a moral claim. But I am reasonably certain that the vast majority of students would agree with me.[94]

My analysis in the previous two paragraphs is largely unchanged in the context of Professor J's Secured Transactions course and Professor K's Jurisprudence course. Students generally do not have control over who teaches which subjects. Thus, the students with a legitimate preference for Secured Transactions over Jurisprudence have little choice but to take the class with the inferior teacher. Punishing them with lower grades because Professor J is a less-skilled instructor is unfair.

In sum, (1) we do not know who the superior teachers are; (2) even if we did, that does not tell us which classes of students are actually perform-

---

92. Professor Stake makes an additional, related argument: "Law employers often want to know not how much the student has learned in law school but rather how capable the student is of learning . . . . The student who learned less in the poorer teacher's course may be just as capable of learning as the student who learned more in the better teacher's course, contrary to the implication of the inferior grade." *Id.* Stake ultimately concludes that "allowing the grade averages to vary according to the quality of the teaching might send the wrong signals to employers." *Id.*

93. Of course, grades are partly based on factors beyond students' control—e.g., students do not get to choose their parents. Indeed, those skeptical of free will would say that grades are based *entirely* on factors beyond the control of students. I do not believe that an extended digression into the nature of "control" is warranted. What I will say is that reducing the extent to which grades are determined by factors beyond students' control is a worthwhile aim, especially where the external factors are essentially imposed by the law school.

94. *See supra* note 62 and accompanying text (explaining that the Bowen Student Bar Association voted unanimously in favor of the mandatory curve I proposed at the school).

ing better; and (3) even if we knew which classes were performing at higher levels because of superior teaching, it would be unfair to award higher grades to the students in those classes.

D. Objection Four: "Mandatory Curves Infringe on Academic Freedom"

Another popular attack on mandatory curves is that they limit academic freedom.[95] Professors Downs and Levit persuasively argue against this position, demonstrating that it is one of the weakest objections to forced curves.

First, they explain that academic freedom can be used to both criticize *and justify* mandatory curves. "Academic freedom is a concept that is rooted in the fear of suppression of ideas, and the victimization of those who express them."[96] It serves to protect intellectual freedom in teaching and research.[97] But the concept is not monolithic; rather, academic freedom contains three essential strands involving (1) the personal autonomy of professors, (2) freedom from manipulation by the government, and (3) institutional autonomy.[98] Any claim by instructors that an institutionally established grading system limits their freedom puts the first and third branches of academic freedom at war;[99] it raises a conflict between professor autonomy and institutional autonomy. Since the concept of academic freedom lends support to both sides of the grade normalization issue, it does not provide a strong basis to oppose a mandatory curve.

Second, and more importantly, "even considered on its own merits, the idea that an institutionally imposed grading system violates the personal autonomy strand of academic freedom is problematic."[100] By contending that a mandatory curve limits the freedom of teachers, one is essentially maintaining that "freedom of a particular grade distribution is tantamount to freedom of thought."[101] But this position has little to support it.

Consider first what mandatory curves do *not* do. As Downs and Levit explain, grade normalization places no restriction on a "professor's ability to call the performance categories whatever he or she wants;" the system of

95. Downs & Levit, *supra* note 9, at 848; Post, *supra* note 7, at 807 (explaining that in one law school's deliberation about grade normalization, "[t]he faculty began with an assumption that we should balance 'academic freedom' against the students' demands for relief").

96. Downs & Levit, *supra* note 9, at 849.

97. *Id.*

98. *Id.*

99. *Id.*

100. *Id.* It should be noted that there are compelling understandings of academic freedom under which personal autonomy is not relevant to the concept, or in which personal autonomy is a value merely on instrumental grounds—i.e., as a tool designed to protect other, more central aspects of academic freedom. *See, e.g.*, MATTHEW W. FINKIN & ROBERT C. POST, FOR THE COMMON GOOD: PRINCIPLES OF ACADEMIC FREEDOM (2009).

101. Downs & Levit, *supra* note 9, at 849.

grades is already set at every school by other policies.[102] Nor do mandatory curves limit a professor's freedom to "make relative judgments about the merits of students' performances."[103] In addition, mandatory curves have no impact on instructors' authority over which substantive factors to consider in awarding grades, such as analytic depth, concision, and case citation.[104] Finally, I would add, mandatory curves impose no limits on scholarly inquiry or teaching methodology and content.[105] All a mandatory curve does is create "some relatively weak parameters on a professor's ability to place more than a certain number of individuals in a given performance category."[106] And even that limitation provides substantial discretion to the instructor under normalization policies that use means or medians rather than ranges for each grade level.[107] "Professors may feel that their judgment as to student competencies is being challenged. But is this a matter of academic freedom? How are ideas being suppressed? The answer is—only very indirectly, if at all."[108]

In short, (1) numerous restrictions on grading already exist;[109] (2) the addition of a mandatory curve based on mean GPA is a mild, additional restriction that still provides critical flexibility to professors; and therefore (3) the imposition on academic freedom is, at most, minimal.

Professor Keating puts the point somewhat more bluntly: "What about academic freedom? The response to this argument is that academic freedom does not include making relative determinations, in the absence of necessary information, about how much total grade-wealth that we can distribute to our students compared to that given out by a colleague."[110] Under a mandatory mean system of grade normalization, professors retain the "freedom to determine the only two facts about which we have reliable information concerning our students' performance: their rank-order within the group being assessed and their approximate distance from one another."[111]

---

102. *Id.* at 850.

103. *Id.*

104. *See id.*

105. *See id.* at 849 ("Grading does not involve freedom of religious or political thought; it has nothing to do with limitations on scientific inquiry.").

106. *Id.* at 850.

107. Downs & Levit, *supra* note 9, at 850.

108. *Id.* Indeed, "[a] greater amount of 'academic freedom' was lost when law professors were told they must give grades in the first place." *Id.* at 852.

109. *See also id.* at 851–52 (identifying various other restrictions on grading practices: professors cannot (1) fail 95% of a class for not living "up to a Platonic ideal of a law student," (2) "sell grades to the highest bidder," or (3) "exercise caprice in grading by throwing papers down the stairs").

110. Keating, *supra* note 9, at 187.

111. *Id.* at 187–88; *see also* Post, *supra* note 7, at 808 ("The connection between the two, academic freedom and grading, is not immediately obvious to me.").

Grade normalization is designed to insure the equitable treatment of our students. "Sweeping claims of academic freedom fall when they run up against legitimate concerns that students are not being treated fairly."[112]

E. Objection Five: "Mandatory Curves are Incompatible with Criterion-Referenced Grading"

An increasingly common attack on mandatory curves is that they are not compatible with criterion-referenced grading. This argument was advanced in two influential, recently-written books: *Educating Lawyers*, better known as the "*Carnegie Report*,"[113] and *Best Practices for Legal Education*, published by the Clinical Legal Education Association.[114]

"Criterion-referenced grading" means to "measure student performance against an external objective standard."[115] A good example is awarding As to all students who score 90% or better on an exam. Under such an approach, all students can, in theory, earn As, Bs, or lower, "depending on the degree to which they demonstrate mastery of the criteria."[116] The performance of other students in the class is irrelevant because the students are assessed pursuant to an absolute standard.[117]

"Norm-referenced grading" means to measure the performance of students in relation to the other members of the class.[118] Under this approach, students earn A's because they scored near the top of the class.[119] In other words, norm referencing focuses on comparative assessment.[120]

---

112. Downs & Levit, *supra* note 9, at 851; *see also* Post, *supra* note 7, at 808 ("Why faculty examination preparation and grading practices should escape completely the process by which teaching is evaluated, I have no idea. There is no other aspect of teaching which is as zealously guarded as the grading process.").

113. *See* CARNEGIE, *supra* note 9, at 168–71.

114. *See* BEST PRACTICES, *supra* note 9, at 243–44.

115. Lynn M. Daggett, *All of the Above: Computerized Exam Scoring of Multiple Choice Items Helps to: (A) Show How Exam Items Worked Technically, (B) Maximize Exam Fairness, (C) Justly Assign Letter Grades, and (D) Provide Feedback on Student Learning*, 57 J. LEGAL EDUC. 391, 398 (2007).

116. *Id.* at 398–99; *accord* CARNEGIE, *supra* note 9, at 170.

117. Daggett, *supra* note 115, at 399.

118. BEST PRACTICES, *supra* note 9, at 243 ("Norm-referenced assessments are based on how students perform in relation to other students in a course rather than how well they achieve the educational objectives of the course.").

119. Daggett, *supra* note 115, at 399.

120. Criterion-referencing and norm referencing are actually not as different as this description suggests. Critically, both forms of grading involve the usage of an external standard. When a teacher assesses her students under a comparative grading system, the teacher is determining which students reached superior levels of achievement pursuant to some independent metric—e.g., quality of legal analysis, comprehension of the substantive law, issue spotting, or a combination of these types of skills. As one commentator observes, that is the only way norm referencing can operate: "Student A is better than student B *at what*?" Fines, *supra* note 20, at 891 (emphasis in original); *accord* Post, *supra* note 7, at 788 ("Assessment is normative. Competency—in teaching or in lawyering—cannot be judged without reference to some external standard."); Nickles, *supra* note 25, at 413–14 ("The grade serves to represent a teacher's opinion of a student's absolute or comparative achievement *in attaining some standard* . . . .")

Mandatory curves require the use of some norm-referencing in grading since even if every student in a course demonstrates complete mastery of the subject, the forced curve will typically prevent the instructor from awarding A's to all students.[121] Only the students who perform the best can receive the highest grade.

The *Carnegie Report* and *Best Practices* advocate that individual professors use criterion referencing in their classes.[122] Both works contend that, in each course, assessment should be based on "detailed, explicit criteria that identify the abilities students should be demonstrating . . . and the bases on which the instructor will distinguish among excellent, good, competent, or incompetent performances."[123] And the standards should be provided to students early in the semester to guide their studies.[124] Using criterion referencing will help students understand what is expected of them and why they receive particular grades.[125] In addition, the two books assert that criterion-referenced grading provides students with greater incentives to achieve excellence because, unlike with norm referencing, (1) "the possibility of success is not limited by the performance of their classmates,"[126] and (2) the criteria enable students to identify where they stand.[127] Given the benefits of criterion-referenced assessment, both works recommend that law schools cease grading on a curve.[128]

There are crucial flaws with the analysis proffered by the *Carnegie Report* and *Best Practices*.

1. *Response One: Criterion Referencing Is Substantially Compatible with a Mandatory Curve*

The first problem is with the assumption, implicit in each book, that a mandatory curve is incompatible with criterion referencing. This assumption

---

(emphasis added) (footnote omitted). At most, it is how the standard is used that separates the two forms of grading. For example, if the two systems are used in their purest form, a criterion-referencer would mark Student A's legal analysis in an essay question as "excellent," "good," or "average," whereas a norm-referencer would mark Student A's analysis as "better" or "worse" than B's.

121. Daggett, *supra* note 115, at 399.

122. BEST PRACTICES, *supra* note 9, at 243–45; CARNEGIE REPORT, *supra* note 9, at 168–70; *accord* Krieger, *supra* note 58, at 301–03.

123. BEST PRACTICES, *supra* note 9, at 244 (citing Sophie Sparrow, *Describing the Ball: Improve Teaching by Using Rubrics—Explicit Grading Criteria*, 2004 MICH. ST. L. REV. 1, 6–15).

124. *Id.* at 245; Krieger, *supra* note 58, at 301 ("The faculty member articulates the criteria that she will use to evaluate a student's learning, with enough specificity to guide her students' study and preparation. Such a system demystifies the preparation process for students and allows each to obtain a grade that objectively corresponds to her learning and testing performance.").

125. BEST PRACTICES, *supra* note 9, at 245.

126. Feinman, *supra* note 7, at 650.

127. BEST PRACTICES, *supra* note 9, at 243–44.

128. *See id.* at 244; CARNEGIE, *supra* note 9, at 168–70.

is false. A mandatory curve does not prevent individual professors from using primarily criterion-referenced assessment.

To illustrate, in my first-year courses (Contracts I & II), I use the type of criterion referencing recommended by the *Carnegie Report* and *Best Practices*. Before the mid-point of the first semester, I provide my students with a model answer to an essay question. This model answer is my best estimation (based on top papers I have received and various other factors) of what constitutes outstanding work for a first-year law student. I also give the students a ten-page, single-spaced memorandum with twenty-eight tips on taking exams in my courses. I then spend an hour of class time going over the model answer, the tips memo, and exam preparation (including outlining). As part of that class, I explain to students both the points-based methodology I use to score tests and what A, B, and C exams typically accomplish, using examples from the model answer and excerpts of the tips memo to illustrate. By the end of this process, the students have a basic grasp of the criteria I use in assessing their answers to essay questions. I then follow-up later in the term with two more model answers to practice essay questions (the first of which I also review in class) in order to solidify the students' understanding of my expectations. I also regularly note in class how my exam criteria relate to the legal skills and substantive topics they are learning. As for the short-answer portions of my tests, the questions I ask are always similar to the hypotheticals we cover in virtually every class session. The students are thus more than ready for the non-essay aspects of the final.

In short, going into the fall exam, the students know what their essays need to look like and roughly how many short-answer questions they must answer correctly to reach high levels of achievement. Moreover, because my tests are open book, the students may bring the model answers with them to the final to serve as guides in crafting their own essays.

Early in the spring semester, I spend an hour of class reviewing the fall exam. After this, any student who wishes to go over their first-semester test with me personally—so they can better prepare for the spring test—is free to do so. Roughly half the students schedule such a meeting. Using their answers, the top paper, the model answers I covered in class, and the exam tips memo, I am always able to show the students precisely what they did well, the mistakes they made, and how their grade would have changed had they corrected particular mistakes. I also review another practice essay question late in the spring term to illustrate how the students should approach the mild differences between the fall and spring finals.[129]

---

129. I use a similar process in Secured Transactions, though it is less extensive because the students are in their second or third year. I cover only one model answer and the tips memo rather than three model answers. Note that in an ideal world, I would use formative and summative assessments beyond the final exam—*e.g.*, a mid-term exam—particularly in Contracts I, as recommended by both the *Carnegie Report* and *Best Practices*. *See* BEST PRACTICES, *supra* note 9, at 245–53, 255–60; CARNEGIE

What is the result of this process? Every semester, the raw scores of my Contracts students fall approximately into a bell curve. *Every semester*.[130] Despite my adoption of much of the criterion-referencing approach, only a few students are able to match the level of accomplishment reflected in the model answers.[131] The raw scores of the rest of the exams distribute themselves nicely into a bell curve, falling various degrees short of the model answers.[132] What this means is that criterion-referenced grading, at least as I practice it, is fully compatible with a mandatory curve.

Other law professors have relayed similar experiences. According to Professor Lynn Daggett, "in fifteen years of full-time law school teaching I have had only one large class in which the grade curve requirements prevented me from assigning course grades that reflected levels of mastery as well as class rank."[133] And Professor Deborah Post observes that "I have always found that the distribution of raw scores among students clusters around a midpoint and tapers off at both ends."[134]

Given modern statistics, these results should not be surprising. "[T]he central limit theorem posits that in a large group, performance approximates a bell curve, in which case typical law school grade curve requirements will likely not prevent course grades from reflecting each student's level of mastery as well as each one's standing in comparison to the group."[135] A mandatory curve is thus substantially consistent with criterion referencing, contrary to the position of the *Carnegie Report* and *Best Practices* books.[136]

Of course, in smaller classes, the students' performances are less likely to organize themselves into a bell shape.[137] But that is one of the reasons I recommend forced curves that operate using course means rather than a re-

---

REPORT, *supra* note 9, at 165–67, 171. I hope one day to incorporate additional assessment into my courses.

130. I have taught Contracts I eight times and Contracts II nine times.

131. On average, one or two students do so each fall in Contracts I, and between two and four students do so each spring in Contracts II.

132. Note that this has been the result whether my test was too difficult, too long, both, or just right (by my estimation). When the test is both too long *and* too difficult, however, the top of the bell curve tends to be pushed somewhat to the left (i.e., the back). "Shockingly," I have never written a test that was either too short or too easy.

133. *See, e.g.*, Daggett, *supra* note 115, at 400.

134. Post, *supra* note 7, at 806.

135. Daggett, *supra* note 115, at 399–400; *see also* Post, *supra* note 7, at 805 n.107 (discussing the normal, bell-shaped curve that reflects the distribution of "an enormous range of data").

136. For another source that makes this same mistake, see Feinman, *supra* note 7, at 652 ("Indeed, the use of normalization entails a decision by the law school that criterion-referencing cannot be used in any course (or at least that it must be subordinated to norm-referencing). Accordingly, normalization limits the ability of individual professors to evaluate students on the basis of their performance relative to objective standards.").

137. *See* Daggett, *supra* note 115, at 399 n.21 ("According to the central limit theorem, law school exam scores will approximate a bell curve the larger the group being measured."); Downs & Levit, *supra* note 9, at 846 ("Students in small samples are less likely to be distributed evenly.").

quired distribution of the available grades.[138] Mean systems provide teachers with the flexibility necessary to allocate their marks in accordance with objective achievement levels even in classes that do not fit a normal distribution.[139]

Best Practices contends that a bell curve in student performance reflects a failure of instruction.[140] "[I]f our teaching is effective and successful, all students should learn what we want them to learn and earn high marks on assessment."[141] Certainly, when the teaching and admission practices of law schools are successful, virtually all of our students should learn enough to become competent lawyers. As explained above, however, performance will vary significantly among the students, typically reflecting a bell curve. Therefore, bell-curve, grade distributions are no indictment of law school assessment. And mandatory curves, especially those that use means or medians rather than a required distribution, are, at most, a mild limit on criterion referencing.[142]

---

138. For more on my decision to base my curve proposal on course means, see *infra* Part VI.A.1.

139. This concern extends beyond individual, small classes. Professor Jeffery Stake has argued that law school student bodies might in general not reflect a normal, bell curve distribution in aptitude and work ethic. *See* Stake, *supra* note 2, at 599–600. Under Stake's analysis, the distribution of legal ability at most schools will look like the front half of a bell curve, with many schools also having a thinned out right tail. *Id.* at 599. In other words, students will be congregated towards the bottom of the distribution, while the top of the distribution will frequently have a smaller number of superior performers. Once again, this is not a problem with forced curves that operates using means or medians rather than a strict distribution.

140. BEST PRACTICES, *supra* note 9, at 244.

141. *Id.*

142. One of the reasons the *Carnegie Report* goes wrong in arguing that a mandatory curve is incompatible with criterion referencing is that the authors misunderstand the point of a curve:

> Underlying these two approaches to grading—on the curve and criterion-referenced—are two rarely stated but fundamentally opposed philosophies about the purpose of assessment in professional education. Those who champion grading on the curve assume that legal education largely serves a sorting function. The intent is to identify the best and the brightest in legal theory, analysis and scholarship. . . . On the other hand, the implicit pedagogical philosophy underlying criterion-referenced assessment is that the fundamental purpose of professional education is not sorting but producing as many individuals proficient in legal reasoning and competent practice as possible.

CARNEGIE, *supra* note 9, at 168. This is nonsense. As explained previously, the principal justification for grading on a curve is basic fairness. *See supra* note 10 and accompanying text; *supra* Part IV.A. I am aware of no scholar or school that has expressly or implicitly justified a mandatory curve on "sorting" grounds. Those of us who support formal grade normalization are just as interested as the *Carnegie Report* authors in "producing as many individuals proficient in legal reasoning and competent practice as possible." CARNEGIE, *supra* note 9, at 168. We simply believe, with *ample* justification, that a mandatory curve is necessary to fairly allocate the rewards that flow from becoming a competent legal practitioner. The *Carnegie Report* also contends that "[g]rading on the curve is often coupled with a tacit belief that there is little possibility of doing much more than sorting in the first place, since there is little possibility of raising the performance of all or most students." *Id.* Again, this is nonsense. Grading on a curve says nothing about the possibility of raising student performance. It is principally about allocating law school rewards on the basis of student merit rather than professor grading philosophy.

Perhaps I am being unfair in my interpretation of the *Carnegie Report*. Perhaps the quotations above are only criticisms of grading on a curve by *individual professors* rather than of *institutional* mandatory curve policies. If that is the case, then the *Carnegie Report* authors and I have some common ground. I believe professors should use something akin to the type of criterion-referenced grading that I

What about those instructors (and I have encountered a few) who contend that they are unable to distinguish among the performances of their students because every student satisfies the grading criteria? Frankly, I think what this really says is that the assessment tools these teachers are using are insufficiently challenging. If a law professor cannot distinguish among the performances of the students in a given class for purposes of applying a mandatory curve, that professor needs to reconsider the assessment practices being employed.

Best Practices also criticizes norm referencing because it does not assist students in understanding "the degree to which they achieved the educational objectives of the course."[143] But in practice, mandatory curves do not require that professors use rigid norm-referenced assessment. Moreover, even if pure criterion referencing—criterion-referenced grading unrestricted by grade normalization—does provide students with somewhat greater guidance than the mixture of norm- and criterion-referenced grading that must be used when a mandatory curve is in place, a forced curve is still the lesser evil. Any loss of guidance the students suffer under a mandatory curve is greatly outweighed, morally, by the unfairness caused by grade disparities that result in the absence of grade normalization. And, once again, I believe that students would come close to universally agreeing with me on this point.[144] Finally, it is questionable whether pure criterion referencing would in fact provide more guidance than criterion referencing subject to a mandatory curve. Grading is not as objective and predictable as the authors of the *Best Practices* book suggest. Even with detailed criterion-referencing standards, significant levels of subjectivity are unavoidable in assessing skills and topics as complicated as legal analysis, legal writing, legal research, contracts, property, torts, criminal law, and civil procedure. This subjectivity, combined with the inherent difficulty of most legal topics, will always make it hard for law students to know how far they have progressed and what is necessary going forward to obtain a particular grade.

2. *Criterion Referencing is Neither Practical Nor Ideal at an Institutional Level*

Pure criterion-referenced assessment—again, criterion referencing unrestricted by a mandatory curve—is only fair to students if all professors at a law school use the same grading standards and apply those standards con-

---

practice. But criterion referencing is only fair when it is subject to the parameters of an institutional mandatory curve. And so, norm referencing must continue to play a role in law school assessment.

143. BEST PRACTICES, *supra* note 9, at 243 (further noting that this "can have a negative effect on student motivation and learning"); *see also* Krieger, *supra* note 58, at 302 (concluding that criterion-referenced grading "provides increased transparency").

144. *See supra* note 62 and accompanying text (explaining that the Bowen Student Bar Association voted unanimously in favor of the mandatory curve I proposed at the school).

sistently throughout the institution. In other words, all teachers must adopt and consistently apply the same grading philosophy. This raises two problems for advocates of pure criterion referencing. First, it is very unlikely that the entire faculty at a law school will both agree to the same grading standards *and* apply those standards consistently. As a result, a mandatory curve is necessary to control for the impact of different expectations among professors. Second, even if an institutional grading philosophy and consistent application of that philosophy were possible, they would likely require an unhealthy degree of uniformity in teaching, curricular, and assessment practices at the school.

Beginning with the issue of fairness, one of the claimed benefits of criterion-referenced assessment is that grades produced using that system reflect absolute levels of achievement. But this is only true if all professors at a single institution consistently apply the same grading standards—i.e., they share the same grading philosophy. If professors use varying standards or apply the same standards in dissimilar ways, course marks no longer accurately convey absolute levels of achievement. Instead, grades reflect some combination of student performance *and* the different grading philosophies. For example, assume Professor A gives more high grades in Contracts than Professor B does in the same course because A applies a more generous understanding of what constitutes excellent work (i.e., Professor A has different substantive expectations). Or suppose Professor A believes that a higher percentage of excellent students deserve A's (i.e., Professor A has a different grading style). In reviewing transcripts of students from this law school, an outside observer, operating on the understanding that both A and B use criterion referencing, will conclude that the students in A's course reached higher levels of mastery in Contracts than the students in B's class. But this conclusion is unjustified. In reality, the grade disparities reflect the different grading philosophies of the two Contracts instructors. Thus, when professors use different standards, pure criterion referencing conveys misleading information about student performance, which is unfair to the students and confusing to consumers of transcript information.[145]

The unfairness actually goes further. Most law schools calculate GPA and use it for class rank purposes. The students in Professor A's Contracts course will have a class rank advantage over Professor B's students because the two professors use different grading standards. This, of course, is also

145. My thanks to Professor Matthew Silverstein of New York University in Abu Dhabi for helping me develop this argument. In elaborating on the issue, Professor Silverstein offered the following: "Absolute grading only makes sense when each grade is taken in isolation. When you're looking at a student's transcript—or several students' transcripts—absolute grading obscures far more than it reveals." E-mail from Matthew Silverstein to Author (April 16, 2010, 6:45 PM CST) (on file with the author). *See also* Stake, *supra* note 2, at 587 ("Employers and other readers of law school transcripts do not have the information or the time they would need to figure out what grades mean for individual teachers.").

unfair to B's students. And it once again conveys inaccurate data to those who rely upon transcripts.[146]

Here is another way to think about the issue: *Norm referencing is inevitable*. Students' grades are going to be viewed and used comparatively. Prospective employers compare grades on student transcripts. And grades are used to calculate GPA and class rank, which are critical factors in the distribution of law school rewards.[147] For these comparisons to be valid, either professors must use and apply consistently the same grading standards, or a mandatory curve is needed to assure that grades do not reflect differences in professor grading philosophy.

Implicitly acknowledging this choice, the *Carnegie Report* recommends the development and consistent implementation of institutional grading criteria.[148] Reaching the required level of uniformity in grading philosophy, however, is simply not practical.[149] Consider first what is necessary. Agreement over purely formal grading standards would not work. For example, faculty could not simply set 90% as the cutoff for A grades. We first need to agree on what constitutes 100%. And that requires consensus on substantive, rather than formal, standards. What *is* 100% in Secured Transactions? How is it similar to and different from 100% in Contracts? In Jurisprudence? In a clinical course? Likewise, faculty agreeing that all "excellent" work shall receive an A would be insufficient without consensus on the substantive issue of what constitutes "excellent" work. Put simply, mere agreement over grading *style* is not enough; faculty must *also* reach a common understanding over grading standards that set the degree of *substantive* expectation.[150] And then, of course, we must apply the substantive standards consistently throughout the law school. In short, to establish institutional grading criteria as recommended by the *Carnegie Report*, a law faculty must (1) reach consensus on grading style, (2) reach consensus on substantive expectations, and (3) be able to apply consistently across the school the

146. *See* Post, *supra* note 7, at 793 (concluding that ranking students without insuring the comparability of grades they receive causes the ranking to "convey misinformation to the public and to the various constituencies that care about grades.").

147. *See* Stake, *supra* note 2, at 586 ("[G]rades will often be read comparatively. And even if no one reads an individual teacher's grades comparatively, those grades will in all likelihood be combined into a GPA which will be used to make comparisons between students.").

148. *See* CARNEGIE REPORT, *supra* note 9, at 170–71; *see also* Krieger, *supra* note 58, at 303 ("Most faculties should be able to reach consensus regarding grading practices, perhaps with the adoption of criterion-referenced grading. . . .").

149. *Cf.* CARNEGIE REPORT, *supra* note 9, at 170 ("Perhaps the most serious criticism leveled at criterion-referenced assessment is the difficulty in getting faculty to agree on standards of performance. One professor's A is another's B, or even C.").

150. There is one type of formal agreement that is an exception: A faculty could agree to a set distribution of grades for every class—e.g., the top ten percent of students shall receive an A. But that, of course, is simply a mandatory curve.

grading standards that reflect those two agreements. The barriers to accomplishing all of this are substantial.

Let me begin, however, by granting that law faculty *can* resolve their disagreements over grading *style*—element (1). At schools like mine, there is little question that teachers apply contrasting grading scales; we differ in how we label various levels of student performance. Some instructors believe A's should be reserved for very few students. Others are more generous with high grades. Some teachers think Cs should be used frequently to signal to students that they need to perform better. Others feel that C is generally too low of a grade. I think that a law faculty can probably overcome these types of disagreements. More specifically, I believe faculties are able to reach consensus on whether As, Bs, and Cs should reflect "excellent, good, and satisfactory" work, respectively, or "good, satisfactory, and failing" work. After all, we generally can reach consensus on what level to set mandatory curves—e.g., C+ mean, B mean, or B+ mean. As explained in the previous paragraph, however, agreements over grading style do little good if professors differ over the degree of substantive competency necessary for performance to be considered "excellent," "good," "satisfactory," or "failing." In other words, the real challenge is reaching consensus over substantive expectations—element (2), and translating that agreement into standards that can be consistently applied—element (3).

Numerous factors complicate our reaching consensus on substantive expectations. First, professors teach different types of courses, including doctrinal classes, skills classes, seminars, and clinics. Second, courses that fall into the same category often emphasize different skills. For example, Secured Transactions focuses on understanding and applying a system of numerous, relatively clear rules that fit together in complicated ways. Contracts and Sales Transactions involve a smaller number of more ambiguous rules that fit together in a more straightforward manner. Torts and Constitutional Law emphasize particularly ambiguous rules that, many would argue, frequently do not fit together at all. Third, even in comparable courses, teachers place varying weight on different skills, such as rule application, comprehension of legal texts, advocacy, and legal writing. Fourth, professors emphasize different subjects in their classes. To illustrate, some doctrinal teachers focus almost exclusively on legal reasoning. Others supplement such reasoning with moderate to heavy doses of public policy, strategic planning, and/or theory. Fifth, law professors assess mastery of legal skills and topics in different ways. In doctrinal classes, alone, we vary over whether to administer closed-book or open-book tests, take-home or in-class exams, essay questions or multiple-choice and short-answer questions (or a combination), testing on many subjects or just a few, and assessing many skills or a limited number, *inter alia*. In addition, some of us administer a single final exam. Others augment the summative assessment with class participation, papers, quizzes, or group projects. Sixth, teachers employ var-

ying grading methods, such as point systems or holistic/gestalt grading.[151] Seventh, some instructors assess only the level of competency, while others weigh additional factors, such as effort or improvement.[152] Many of these seven differences are reflected in the studies that have concluded that when professors grade the same essay exams, they vary significantly in their assessments, including over the rank ordering of the answers.[153] The disagreements over rank ordering are especially pertinent because they establish that the professors valued different skills or valued the same skills in disparate ways.[154]

Professors have crucial disagreements over what constitutes excellent, good, satisfactory, and failing work, as illustrated by the grade disparities at the Bowen Law School and other institutions. What the previous paragraph shows is that these disagreements implicate multiple differences in (a) the types of courses we teach, (b) the skills and subjects we emphasize in those courses, and (c) the assessment practices we use. This means that law professors do not just vary *quantitatively* in our degree of substantive expectation; we also vary *qualitatively* in ways that makes it difficult to even compare our differing quantitative standards. Given all of these variations, how realistic is it to believe that we can craft substantive grading standards for A, B, and C work (and the multiple flavors of each grade—e.g., B+ and B-) that can be consistently applied across the entire law school curriculum? Can we construct criteria that will properly equate A- work in Secured Transactions, taught by a professor who focuses almost exclusively on the law and gives a largely short-answer, objective test, with A- work in Constitutional Law, taught by a professor who heavily emphasizes theory and policy, and administers an essay exam? What about with A- performance in a seminar where the reading consists entirely of secondary sources and the final grade is determined by a paper and class discussion? What about with A- performance in a negotiation and transactional skills course with twenty graded exercises? I submit the answer is rather clearly "no."[155] Indeed, I

151. Stake, *supra* note 2, at 586 (observing that in grading, some law professors "add up points" while others "use a gestalt method, holistic and sometimes intuitional, assigning a final letter grade without having assigned numerical scores at any stage of the process").

152. *See* Fines, *supra* note 20, at 882 ("Different instructors produce grades in different ways, taking into account different criteria, including comparison to peers, comparison to evaluative criteria, effort, growth, or behavioral compliance.").

153. *See* Downs & Levit, *supra* note 9, at 826–27 (collecting and summarizing several studies); Henderson, *supra* note 20, at 410–11 (same); *but cf.* CARNEGIE REPORT, *supra* note 9, at 170 (observing that "a long history of assessment research on rank orderings supports the notion that faculty will agree on performance rankings to some reasonable degree").

154. *Cf.* Downs & Levit, *supra* note 9, at 827 (observing that the differences among professors grading the same essays could reflect that the professors use different grading scales, value different abilities in distinct ways, or a combination of each).

155. *See* Keating, *supra* note 9, at 179 (explaining that law school assessment is not like testing multiplication tables in a third-grade math class where all teachers agree on the absolute standard of performance and how to measure it) ("Even if the faculty could agree on which skills were necessary for

think it would be difficult to craft such standards even between two Contracts professors who use the same book, emphasize the same basic skills, cover largely the same topics, and administer substantially similar tests.

The *Carnegie Report* counters that criterion referencing is used in other education programs, particularly medical school.[156] But the authors say little to establish that assessment is substantively comparable in the areas of medicine and law.[157] And their contention that "in other fields, such as medical education, there is no evidence that using criterion-referenced grading instead of grading on the curve harms either student learning or faculty morale"[158] misses the point. The principal problem with pure criterion referencing is the lack of fairness that results from the application of different substantive criteria and grading styles, not the impact on student learning and faculty morale.

Similarly unpersuasive is the *Carnegie Report's* observation that "[p]rofessional test developers now routinely achieve impressive levels of rater agreement on standards of performance in many different contexts, from standards for high school graduation to standards for professional licensure in a wide variety of occupations."[159] The differences between high school graduation standards and law school performance standards are too obvious to merit comment. And the existence of national, standardized licensure tests establishes little since such tests typically measure only basic or minimum competence. In law school, we are measuring much more—namely, multiple degrees of achievement above basic proficiency. Indeed, *Best Practices* expressly admits that criterion-referenced assessment generally operates using a pass/fail scale.[160] That makes it less appropriate in contexts where the focus is on gradations of success rather than merely basic competency.

---

minimum competence, we could probably not agree on how to objectively assess them. And even if we could agree on how to assess them, there would be no way to ensure that different graders would be consistent in applying those assessments."); Downs & Levit, *supra* note 9, at 853 ("The idea that grades reflect a specific standard of performance is troubling on a number of levels. . . . Specifically with respect to grading, the idea of absolute measures of performance is simply not realistic."); *see also* Fines, *supra* note 20, at 882 ("Faculty can develop criteria against which student achievement can be measured. This is by no means an easy task—some may say impossible."); Richard Kamber, *Understanding Grade Inflation*, *in* GRADE INFLATION: ACADEMIC STANDARDS IN HIGHER EDUCATION 45, 63 (Lester H. Hunt ed., 2008) ("[M]ost disciplines are unable to agree upon measurable standards of objective mastery beyond basic comprehension, rote knowledge, and threshold skills. The notion of 'grades as an objective measure of mastery' has limited application in postsecondary education.").

156. *See* CARNEGIE REPORT, *supra* note 9, at 170; *see also id.* at 168 ("Other forms of graduate and professional education are more likely to use what is known as criterion-referenced grading—that is, an absolute standard of performance determines who will receive A's and who will receive other grades.").

157. *Id.* at 168–76.

158. *Id.* at 170.

159. *Id.*

160. BEST PRACTICES, *supra* note 9, at 244 (quoting ALISON BONE, NATIONAL CENTRE FOR LEGAL EDUCATION, ENSURING SUCCESSFUL ASSESSMENT (Roger Burridge & Tracey Varnava eds., 1999)).

Moreover, even the bar exam, the principal licensing tool in the field of law, does not operate using pure criterion referencing. The Multistate Bar Exam (or "MBE"), the multiple choice portion of the test, is scaled using a type of grade normalization.[161] Many states further scale the essay portions of their tests using MBE scores.[162] It is also worth noting that states vary significantly in their minimum passing scores.[163] And then there are the numerous, powerful criticisms of the bar exam.[164] All of this strongly suggests that agreement over general assessment standards on anything beyond minimum competency is highly unlikely in the legal field. If law professionals struggle to reach consensus on basic licensure requirements, how realistic is agreement on multiple levels of performance?[165]

The adoption of school-wide, substantive assessment standards would be significantly more realistic if professors did not vary so greatly in (a) the types of classes we teach, (b) the skills and subjects we stress, and (c) our assessment practices. Importantly, greater uniformity in these areas may be coming. Curriculum reform and the increasing focus on outcome assessment *could* eliminate some of the qualitative differences that make agreement on grading philosophy so unlikely.[166] But any significant reduction in the diversity of our curricular, teaching, and assessment approaches would be pedagogically harmful. Variations across these dimensions reflect the fact that instructors emphasize different aspects of legal practice in their classes. Students thus learn a more diverse set of skills and subjects, and receive more comprehensive and multi-faceted assessment. If, for example, all law professors adopted the same testing protocols in doctrinal classes, much of the latter benefit would be lost. Additionally, the qualitative and quantitative

---

161. Lorenzo A. Trujillo, *The Relationship Between Law School and the Bar Exam: A Look at Assessment and Student Success*, 78 U. COLO. L. REV. 69, 75 (2007) ("The MBE raw score, based on the number of questions answered correctly out of two hundred items, is converted into a scaled score so that every test indicates the same *relative* proficiencies.") (emphasis added).

162. Gary S. Rosin, *Unpacking the Bar: Of Cut Scores and Competence*, 32 J. LEGAL PROF. 67, 68 (2008).

163. *Id.*

164. *See, e.g.*, Denise Riebe, *A Bar Review for Law Schools: Getting Students on Board to Pass Their Bar Exams*, 45 BRANDEIS L.J. 269, 272–82 (2007) (setting forth multiple criticisms of the bar exam); Society of American Law Teachers, *Statement on the Bar Exam*, 52 J. LEGAL EDUC. 446, 446–52 (2002) (same); Trujillo, *supra* note 161, at 77–85 (same).

165. One of my Bowen colleagues has pressed me to admit that few law schools have tried to develop substantive grading standards of the type necessary to implement institutional criterion referencing. *See also* CARNEGIE REPORT, *supra* note 9, at 170 (observing that "actual data on the contention that faculty cannot agree on standards of performance is less well established"). That may be so, but it is entirely possible that one reason the trend in legal education has been in favor of adopting grade normalization is that most law professors recognize the futility of attempting to construct school-wide grading criteria.

166. Aida M. Alaka, *Learning Styles: What Difference do the Differences Make?*, 5 CHARLESTON L. REV. 133, 134 (2011) ("At the national level, the Standards Review Committee of the American Bar Association has been promulgating new standards that focus on learning outcomes assessment measures."); *id.* at 135–37 (briefly summarizing contemporary legal education curriculum reform).

differences in professor expectations prepare students for the similar variation they will face in their legal careers. Our students will practice in front of numerous judges, under the tutelage of many supervisors, and in service to countless clients. These judges, supervisors, and clients will employ a wide variety of standards in assessing the work product produced by our graduates. Law students need training for this diversity of expectations. Of course, on fairness grounds, it is critical that each professor communicate her particular standards to students as clearly as possible. But as long as students receive the proper notice, the diversity in our curricular, teaching, and assessment practices is pedagogically invaluable.[167]

In sum, the establishment of an institutional grading philosophy grounded in criterion referencing, and the consistent application of the standards that reflect that philosophy, are neither realistic goals nor aims worth achieving. Despite the aspirations of the *Carnegie Report* and *Best Practices*, comparative assessment of our students is inevitable at an institutional level. We can conduct these comparisons fairly, using a mandatory curve, or we can perform these comparisons unfairly, allowing differences in substantive expectations and professor grading style to play critical roles in reward distribution. The preferable approach is clear.

F. Objection Six: "Mandatory Curves Induce Excessive Competition Among Students"

Another oft-stated argument against mandatory curves is that they induce excessive competition among students.[168] Such competition damages the learning environment, inhibiting student progress on course objectives. This is the strongest attack against mandatory curves. It raises a critical concern about the possible impacts of forced curves on the core educational mission of law schools. However, the argument ultimately fails as an indictment of grade normalization, principally because unfair grade disparities likely create a more corrosive learning environment than does a mandatory curve. In addition, even if forced curves do increase overall competition, the unfairness of grade disparities flowing from differences in professor grading philosophy is the more pressing concern.

The contention that mandatory curves increase competitiveness focuses on the norm referencing inevitable under any curve. Norm-referenced grading measures students against one another. This leads to more competition

---

167. *Best Practices* argues that it "should not matter . . . whether one teacher or another is conducting the assessment." BEST PRACTICES, *supra* note 9, at 243. This is not only deeply unrealistic, *see supra* notes 151–155 and accompanying text, but, as argued in this paragraph, ultimately undesirable. Variation in law school assessment is a positive, not a negative.

168. *See, e.g.*, CARNEGIE REPORT, *supra* note 9, at 165 ("Equally aversive to students was the competitive atmosphere engendered by the widespread practice in legal education of grading on the curve.").

because the students are contending directly for grades.[169] The competition is exacerbated by the grading limits imposed by the curve itself. Such limits create a "zero-sum game in which students who do well benefit by the performance of those who do poorly."[170]

The competition resulting from a mandatory curve, in turn, compromises the educational objectives of the law school. "Educational psychologists have consistently demonstrated that competitive learning gives the poorest results in educational attainment."[171] It does so by reducing student effort and motivation, and increasing student anxiety.[172] The competitive structure also "can lead to a variety of behaviors designed to gain unfair advantage: ranging from subtle undermining of peers by discouraging active learning or hoarding resources to outright cheating and plagiarism."[173] Mandatory curves are intended to bring fairness to grading. But if they fundamentally disrupt learning, perhaps the benefits are not worth the price.

The most important response to this criticism is that significant grade disparities probably do more to create an overly competitive and disruptive learning environment than mandatory curves. Bowen Law School is illustrative. The students there were consumed with the grade disparities prior to our adoption of a mandatory curve. They regularly complained to the faculty and administration. They focused heavily on professor grading practices in registering for courses (and publicly advised their classmates to do the same). And they brought the issue to the attention of our recent reaccreditation team.

In addition, rivalries between the students in the day and night programs, between sections of various classes, and between other groups of students were a regular feature of the school once the grade disparities became significant. For example, in 2009, on our student listserv, the day and night students had a large-scale "flame war" over grading issues. Students

---

169. *See* Fines, *supra* note 20, at 899 ("One of the aspects of grading that is most detrimental to teaching and learning is the practice of required grade curves. . . . If students are measured against each other on a competitive scale, rather than against an objective norm, competitive learning strategies will predominate."); CARNEGIE, *supra* note 9, at 166 ("Students report . . . that the intensely competitive atmosphere militates against a cooperative learning environment.").

170. Downs & Levit, *supra* note 9, at 855 (describing the argument, though disagreeing with it); *see also* Feinman, *supra* note 7, at 650 ("A ranking system that emphasizes the differences in student achievement, and a system in which students are presumed and permitted to succeed only at the expense of other students, is more likely to engender a competitive rather than a cooperative atmosphere. . . .").

171. Fines, *supra* note 20, at 901 (citing David W. Johnson et al., *Effects of Cooperative, Competitive, and Individualistic Goal Structures on Achievement: A Meta-Analysis*, 89 PSYCHOLOGICAL BULL. 47, 53 (1981) (concluding that sixty-five studies found cooperative learning superior to competitive learning, eight studies found competitive learning superior, and thirty-six studies found no distinction between the two types of learning)).

172. *Id.* at 902; *see also* Krieger, *supra* note 58, at 297–98 ("As a zero-sum system, a mandatory curve also creates anxiety and undermines the security and relatedness needs."). The competitive environment may also have greater negative impacts on female students and students of color. *See* Fines, *supra* note 20, at 902–05.

173. Fines, *supra* note 20, at 907.

from both programs wrote multiple emails that, to put it mildly, were highly insulting to the opposing program.[174] This type of public eruption was a depressingly common occurrence at our school under the old grading system.[175]

One of the most telling examples of the harm caused by the grade disparities at Bowen took place in my Contracts course. In the fall of 2008, the grade disparities between Bowen's two sections of first-year students were particularly egregious. The mean grades in Torts and Contracts in the day program were 3.269 and 3.128, respectively. In the night program, they were 2.625 and 2.825. Needless to say, the students in our night division were furious. I taught in the night program that year. At the start of the spring class session in which I planned to review the fall exam, the first question I received from a student was about the grade disparities. From there, it became clear that the class had no interest in talking about the final itself. They did not want to hear about the degree to which they met my expectations, what top papers did to earn their grades, or how they could have performed better. The students had no interest in any of the objective learning criteria I established for the course. Instead, they were interested in one issue: Why were there such dramatic grade disparities? They blamed the school, the Torts professor in the night division, and me for their grades. The final itself got about seven minutes of substantive attention out of a fifty-five minute session. Contrast this with my experiences as a student at a law school with a recommended curve that effectively operated as a mandatory curve. I never encountered anything at that school which I could attribute to the curve that was remotely as disruptive to the learning environment as what happened in my Contracts class at the Bowen Law School in the spring of 2009.

Given all of these events, I think that the institution of a mandatory curve at my institution will almost certainly *lessen* the tensions on campus and substantially *reduce* competitiveness among the students. I suspect the same is true for other law schools suffering from significant grade disparities. Students resent variations in grades.[176] And with good reason. They realize that their career paths are being determined by professor grading philosophy rather than merit. It should thus come as no surprise that variances in faculty grading practices have disruptive impacts.

As for the hoarding, cheating, plagiarism, and other forms of bad behavior allegedly caused by mandatory curves, my school appears to have

---

174. Copies of every e-mail are on file with the author.

175. A similar e-mail exchange took place in 2006 over variances in the grades in our "Law Skills" course, a required legal skills class taught in small sections by adjuncts. Copies of these e-mails are on file with the author.

176. *See* Downs & Levit, *supra* note 9, at 825; Stake, *supra* note 2, at 591; *see also* Post, *supra* note 7, at 796 & n.75 (discussing "a massive protest" by students over grade disparities at one law school).

suffered from these problems every bit as much as other schools despite our lack of any formal grade normalization policies. I am thus skeptical of the notion that these negative behaviors are substantially attributable to mandatory curves.[177] Indeed, given the numerous other features of law school and the legal marketplace that likely facilitate excessive competition, I suspect that mandatory curves are, at most, merely one cause of such competition among many. The mere assigning of grades leads to competition because grades are the principal currency in the attainment of most law school rewards. Class rank also causes competition because it is fundamentally both a norm-referencing metric and zero-sum in operation.[178] Competition is likely also stimulated by the fact that many professors grade on a curve even in the absence of formal grade normalization policies. Professors Downs and Levit elaborate on this last point, observing that the argument that mandatory curves impose zero-sum incentives on students "assumes unlimited grade wealth in the absence of a grade normalization policy, which certainly does not appear to be the case at most schools."[179] Moreover, as explained in the previous part, forced curves are largely compatible with criterion-referenced assessment, limiting the amount of competition-inducing norm referencing that is necessary under a curve. In sum, the implementation of a mandatory curve should not significantly alter the incentives for competition at schools that do not yet have one in place.

Perhaps forced curves do cause some uptick in overall competitiveness. If so, that is a price well worth paying for the fairness that results from their implementation. As explained above, the harm caused by grading disparities is substantial.[180] I believe that students would generally agree that this damage is significantly greater than any injuries caused by a small rise in competition on campus.[181] I share this moral view. In addition, there are colorable arguments that increased competition is actually beneficial. First, while a preponderance of the research suggests that cooperative learning environments are superior, the evidence is mixed; some studies have concluded that competitive learning is more effective.[182] Second, a competitive learning

177. *See* Downs & Levit, *supra* note 9, at 856 & n.69 (explaining that, in the authors' experience, students have not behaved badly in response to grade normalization systems).

178. *See* Keating, *supra* note 9, at 190 ("The fact is, however, any grading system is a zero-sum game to the extent that higher grades in one class necessarily deflate the currency on which overall class ranks will be based. In other words, when the absolute grade numbers of a student's classmates go up, the student's own relative class rank will go down."). Some commentators have argued we should eliminate public dissemination of class rank in order to reduce competition. *See, e.g.*, Krieger, *supra* note 58, at 300; *see also* Post, *supra* note 7, at 793 ("Ranking in the first year of law school is premature."). While I have some sympathy for this idea, particularly since it seemed to work at the law school I attended, it seems unlikely that many schools are willing to follow this path.

179. *See* Downs & Levit, *supra* note 9, at 855–56.

180. *See supra* Part IV.

181. *See supra* note 62 and accompanying text (explaining that the Bowen Student Bar Association voted unanimously in favor of the mandatory curve I proposed at the school).

182. *See supra* note 171.

environment might better prepare law students for the operation of the marketplace.[183] Even if one finds these two arguments unpersuasive, they are at least plausible. However, there are no colorable arguments in support of grade disparities flowing from differences in professor grading philosophy. Such grade variances are wholly unjustified. This too supports the conclusion that increased competition caused by mandatory curves is the lesser evil.

G. Objection Seven: "Mandatory Curves Induce Excessive Apathy in Law Students"

Some critics contend that mandatory curves are a principal source of law student apathy, particularly in the second and third years of school.[184] There are three problems with this objection.[185] First, even if the claim is true, it is difficult to give it much weight without a comparison to the apathy caused by significant grade disparities. If the latter is a greater cause of student disengagement, then mandatory curves are once again the lesser evil.

Second, there are numerous possible sources of student apathy, including (1) confrontational versions of the Socratic method, (2) "hide-the-ball" teaching practices that are sometimes employed, (3) the focus on doctrinal courses throughout all of law school, (4) the single summative assessment and the related lack of professor feedback, (5) the large size of many law school classes, (6) the disproportionate weight accorded to first-year grades, and (7) part-time jobs during the semester, to name just a few. We do not know how much apathy is attributable specifically to mandatory curves. Without greater clarity on this point, the unfairness of grade disparities is the more pressing concern.

Third, the explanations offered for why curves reduce student motivation are typically unpersuasive. For example, one commentator asserted that mandatory curves make it more difficult for students who performed poorly in the first year to improve their class rank, thus decreasing the incentive to

---

183. *See* Downs & Levitt, *supra* note 9, at 854 ("One gets ahead in the employment arena not by having a fixed quantity of good stuff, but by demonstrating more good stuff than other workers. In short, the world operates on a curve."); Fines, *supra* note 20, at 904 ("One response to this critique may be that we are simply preparing students for the world of law as it is. If students cannot operate in a competition atmosphere, they will not be able to perform well as lawyers because the practice of law is competitive.").

184. *See* CARNEGIE REPORT, *supra* note 9, at 163 ("The danger of using inter-student competition as a primary motivation is that in a situation in which, by deliberate plan, all cannot excel, the school faces an endemic problem of retaining student interest and effort after the first cut, which is decisive, is made at the end of the first year."); *id.* at 165–66 (explaining that some students claim that curves reduce their motivation to study).

185. It is also somewhat ironic that opponents contend that mandatory curves cause both excessive competition and disengagement.

study.[186] But mandatory curves do not reduce the prospects for raising one's class rank as a second or third-year student. If anything, curves have the opposite impact. At least with a mandatory curve, there are limits to the grade wealth that professors can allocate in their classes. Grades thus have substantially the same value throughout law school. This means that slower learners can turn around their performance and pass other students in the second or third year. Without a curve, there is a danger of grade inflation in upper level classes. This decreases the range of grades awarded in later years, making it more difficult to overtake those with higher GPAs. To illustrate, if professors at a given law school award As, Bs, and Cs in the first year, but only As and Bs in the second and third year, the relative value of grades in upper level courses is lower. If Student X received mostly Cs in the first year and Student Y received mostly As, X could not make up this difference in just one year. X would need two years of As to Y's two years of Bs to catch up. Of course, the likelihood of such performance reversals by either Student X or Y is small. But the key point is that a mandatory curve *improves* Xs chances of catching Y (if it requires the same range of grades during each year of law school).[187]

A second explanation of apathy is that mandatory curves weaken performance incentives because students can only improve their grades by passing their classmates—classmates who are frequently perceived as being more talented. This explanation is somewhat more plausible. It founders, however, because students near the bottom of the class *can* improve their performance and obtain higher grades under a mandatory curve. First, recall that forced curves are largely compatible with criterion-referenced grading.[188] As a result, students that come closer to satisfying objective performance standards will receive better grades. Second, the norm referencing inherent in mandatory curves still leaves substantial room for students to improve their grades. For example, students can outperform those close to, but slightly ahead, of them. Indeed, students who performed poorly in the first year merely because they are slower learners have reasonable prospects of outperforming students who achieved much higher grades early on. Additionally, when students who struggled at the start of their legal education objectively improve in later years, this may induce their professors to give slightly lower grades to those near the top so the professors can award higher grades to the improving students, as is permitted under more flexible curves, like those structured around mean GPA.

Another commentator, Professor Larry Krieger, offers a more compelling explanation for how mandatory curves might induce apathy. Professor

186. Henderson, *supra* note 20, at 415.

187. *See* Georgakopoulos, *supra* note 8, at 446 & n.3 (making this essential point in a slightly different context).

188. *See supra* Part V.E.1.

Krieger observes that forced curves cause students to perceive their grades as emanating from an external source of control rather than from their own efforts.[189] "This experience . . . displaces the intrinsic motivations to learn and become capable with compelled competition for relative worth."[190] And since performance generally improves as one moves from motivations that are "external and controlled to internal and chosen," decreased effort is a likely result of imposing mandatory curves.[191]

Krieger's primary target is curves structured around rigid ranges for each specific grade, rather than those built around means or medians.[192] But since his argument could be used against the type of grade normalization I support, a response is in order.

Assume that Professor Krieger is correct that mandatory curves cause students to believe that their grades are emanating from an external source. Grade disparities have the *exact same effect*. Recall the example of my Contracts class in the fall of 2008.[193] My students placed blame for the grade variances not on themselves and their efforts, but on the school and me. Moreover, the students were *correct* in blaming us. This critically distinguishes student views of grade disparities from their views of forced curves. Let me elaborate.

Student perception that a mandatory curve based on mean GPA is a critical source of the grades they receive only loosely tracks reality. Recall that mandatory curves are substantially compatible with criterion referencing. Thus, grades awarded under a curve generally reflect the professor's assessment of student progress on course objectives. And law teachers can explain this to students, hopefully reducing the mistaken perception that the curve plays a major role in determining grades. On the other hand, when a law school suffers from significant grade variances, students are justified in blaming an external source for their marks. Why? Because if grade disparities exist, differences in teacher grading philosophy account for a substantial portion of the grades students receive. This means that a crucial piece of grades *is* determined by an external source entirely beyond the students' control—the professors. When that is the case, there is nothing teachers can tell students that will eliminate confusion about their marks because the students are not actually confused; their perceptions are largely accurate. In sum, if Professor Krieger is correct, both mandatory curves and grade disparities cause students to believe their marks are emanating from an external source, reducing the incentive to study. But only in the case of grade vari-

189. Krieger, *supra* note 58, at 298.

190. *Id.*

191. *See id.*; *see also* Fines, *supra* note 20, at 900 (noting that "extrinsic orientations tend to undermine learning" by, among other things, reducing student effort).

192. *See* Krieger, *supra* note 58, at 297.

193. *See supra* notes 174–175 and accompanying text.

ances are the students correct, eliminating the option of educating them to reduce misperceptions. Thus, unjustified grade disparities, once again, are worse than mandatory curves.

H. Objection Eight: "Recommended Curves and Informal Grade Normalization Practices Are Superior to Mandatory Curves"

The final objection to mandatory curves acknowledges that some form of grade normalization is necessary, but concludes that other types are superior to forced curves.

One option is to make curves recommended rather than required.[194] The principal justification for such an approach is greater flexibility. If a professor believes the students in a particular class performed in an especially superior or inferior manner, the professor can reward the students accordingly. However, as discussed in the section addressing the objection that curves bar professors from giving students the grades they earned, it is not enough for a teacher to believe his students performed better or worse than *his* standard class. Because grading is inevitably comparative, the professor must have evidence that his students outperformed students in *other courses* taught by *different instructors*. Such evidence will rarely, if ever, be available. A second problem with recommended curves is that they are less effective than mandatory approaches.[195] Schools that adopt such a policy often continue to suffer from unjustifiable grade disparities.[196]

Another option is the usage of informal normalization practices. The most common such practices are (1) consultation among professors regarding grades, and (2) the provision of historical grade distributions.[197] For example, some schools have written policies recommending that certain professors, typically those teaching the same section of a course, consult with each other about grades prior to submitting their final marks to the administration.[198] Other schools circulate the grade distribution for every course after the grades become official.[199] My law school previously used the latter

---

194. *See* Downs & Levit, *supra* note 9, at 839 (noting that some law schools make parts of their curve policies recommended rather than mandatory); *see also* Krieger, *supra* note 58, at 303 (proposing that "a faculty member with consistent or substantial variations from a normal distribution of grades, or from a realistic grade range, could be required to support those variations and receive approval from the administration or a designated committee").

195. *See* Post, *supra* note 7, at 778 (concluding that voluntary grade normalization "does nothing to curb the excesses of faculty").

196. *See id.* at 798, 812 (explaining that the author's law school adopted a recommended curve and that each semester afterwards, there were what the author considered "significant deviations from the norm").

197. Downs & Levit, *supra* note 9, at 847.

198. *Id.* at 847–48.

199. *Id.* at 848; Kaufman, *supra* note 7, at 418 (noting that many schools "distribute information about recent grading practices to the faculty").

approach. Roughly two weeks after grades were submitted, the administration provided the faculty with a report listing the precise grades given in every class that semester. A third approach is regular, faculty-wide discussions of grading practices.[200] Each of these mechanisms is inferior to a mandatory curve.

The problem with the consulting approach (the first option) is that, at best, it will only resolve disparities among sections of the same course. While that type of variance is probably the most unfair, other grade disparities are also deeply problematic. And one-on-one consulting is simply not practical across an entire faculty. The third option—full faculty discussions—makes more sense if the goal is to eliminate grade disparities generally, not just those among sections of the same class. But if such a dialogue actually results in consensus, there is no reason to avoid taking the next step and instituting a mandatory curve that reflects the agreement. The only plausible basis not to do so is the added flexibility found in an informal system that allows teachers to deviate from the norm when they believe a particular class is deserving. But as explained above with respect to recommended curves, individual professors will virtually never have sufficient information to justify such a deviation since they will have little evidence indicating that their students performed better than students taught by other professors in different classes.

As for the second option, circulating grade reports was an abject failure at my law school. Virtually no one read the reports carefully. Perhaps the grade reports have some influence at other institutions.[201] But if so, once again, there is no reason not to implement a mandatory curve that is consistent with the general trend in grade distributions.[202]

Recommended curves and some informal normalization practices are premised on a need for greater flexibility in grading so that professors can award grades consistent with variations in performance from year to year. As I explained, this justification is unpersuasive. Nonetheless, the curve I recommended at my school does provide *some* of the desired flexibility. My proposal sets a mandatory mean range of 3.2 to 3.4, with a target of 3.3. The primary purposes underlying such a range are (1) to provide flexibility in setting grade distributions so they reflect differences in student achievement within a given class, and (2) to prevent the curve from appearing unduly restrictive to students and faculty. To the extent the range also allows teach-

---

200. Downs & Levit, *supra* note 9, at 848. For commentators supporting this third approach, see Krieger, *supra* note 58, at 301, and Feinman, *supra* note 7, at 652.

201. *See* Downs & Levit, *supra* note 9, at 848 ("The posting or circulating of grading information may allow some informal grade normalization to occur among professors.").

202. Another concern with circulating grade distribution reports is that it probably takes time for this practice to bring about the necessary degree of uniformity. Until then, the grading practices at a school suffering from grade disparities will continue to unfairly impact the careers of the students.

ers to moderately adjust their grades from year to year based on their subjective assessment of student performance, this small measure of discretion is substantially consistent with the goal of eliminating unfair grading disparities.[203]

I. A Final Note on the Objections

No grading system is perfectly fair.[204] And this obviously includes mandatory curves. In Part V, I have contended that mandatory curves are the lesser evil. To the extent they have flaws, the harms caused are substantially smaller than the injustices that result from the alternatives.

The trend over the last thirty years has been in favor of adopting grade normalization mechanisms.[205] Moreover, a majority of schools now use a mandatory curve for grade normalization in at least some courses.[206] Many scholars support this approach.[207] And even those who are more skeptical frequently acknowledge that such curves are the lesser evil when the alternative is significant grade disparities.[208]

In sum, all law schools should adopt mandatory curves.[209]

---

203. For discussions of some potential concerns with a mandatory curve based on a mean GPA range, see *infra* notes 220–226 and accompanying text.

204. Fines, *supra* note 20, at 895 ("Obviously we cannot entirely eliminate inequity in our sorting processes.").

205. *See* Downs & Levit, *supra* note 9, at 836 ("Apart from the self-selection effect of those responding, the cumulative data indicate a dramatic trend toward adoption of grade normalization policies: 9% of schools in 1976, 66% in 1993, and 84% in 1996."); Kaufman, *supra* note 7, at 417–18 (in this 1993 survey of 175 accredited law schools, of the 119 responding, 79 (66.4%) stated that some type of curve was in place for at least some courses); *id*. at 423 ("The popularity of grading curves . . . appears to be increasing.").

206. Mroch, *supra* note 9, at 2–3 (in this 2005 AALS survey of grading practices at 188 ABA-accredited law schools, 145 schools responded to the survey; of the respondents, 115 have a formal grading policy; and of those 115, 81 reported that the policy is mandatory in at least some courses); *id.* at 6 ("For those schools with policies, the majority of policies were mandatory. . . ."); Fines, *supra* note 20, at 888 (noting that most schools require faculty to grade according to a curve).

207. *See, e.g.*, Downs & Levit, *supra* note 9, at 821 ("[W]e conclude that there is a need for standardization of grades, particularly in first year, sectioned courses and upper level courses with multiple sections."); Stake, *supra* note 2, at 583, 588–92 (defending the equalization of means as a grading principle); Keating, *supra* note 9, at 185–88 (defending mandatory means and similar forms of grade normalization); Post, *supra* note 7; Georgakopoulos, *supra* note 8, at 446 (outlining the need for the standardization of grades); Epstein, *supra* note 7, at 708 (same).

208. *See, e.g.*, E-mail from Lawrence S. Krieger, Clinical Professor, Fla. State Univ. Coll. of Law, to Author (Aug. 12, 2010, 2:06 PM CST) (on file with the author) ("If there are disparities that the administration is unable or unwilling to address through direct intervention, then it is probably better to have a system like mandatory curve that students perceive to be limiting and unnecessarily inflexible, than one that students perceive to be arbitrary and dependent largely on allocation or selection of one's professors."); Krieger, *supra* note 58, at 297–99, 302 (setting forth a variety of concerns with mandatory curves).

209. Given the grade disparities that exist in other parts of the academy—see *supra* notes 17 and 67—I lean towards the view that all undergraduate and most graduate institutions should implement mandatory curves as well.

VI. THE STRUCTURE AND SCOPE OF MANDATORY CURVES

This part addresses some critical issues regarding the structure and scope of mandatory curves.[210]

A. Structural Features of Mandatory Curves

1. *Structuring Curves Around Mean GPA*

Mandatory curves based on "mean GPA" are the best type of curve. To calculate the mean GPA for a course, one simply adds the grade point values of every final grade given, and then divides that sum by the number of students in the class.[211]

Curves using mean GPA require that instructors assign grades within a reasonably tight range, but still provide flexibility in the precise distribution. Under the system I proposed at my school, professors are restricted in that the total value of the grades they award must fall between 3.2 and 3.4. Thus, no teacher can give significantly higher or lower overall grades than any other instructor. But professors can meet the GPA requirement in a variety of ways—for example, (1) by distributing the grades in a bell curve with a significant grouping around the grades 3.1 through 3.5, (2) by spreading the grades more equally across all of the available marks, or (3) by assigning grades in a "U-shape" via a large number of low and high marks.[212]

The benefits of using mean GPA are best illustrated by considering the important deficiencies in the two, primary alternative types of curves. The first alternative is "median GPA." The median GPA of a course is the grade at which half the students have a higher mark and half the students have a lower mark. Median GPA systems do not sufficiently restrict the distribution of grades because they do not prevent a professor from giving significantly higher or lower overall grades than other teachers. For example, assume two instructors are teaching forty-student classes. Assume further that the mandatory curve requires a *median* of 3.2 to 3.4, rather than a mean of 3.2 to 3.4. Under such a regime, the first professor could give nineteen 4.0s and twenty-one 3.4s, while the second professor could give twenty-one 3.2s and

---

210. There are, of course, structural and scope issues beyond those discussed in this part, such as whether to apply the same curve standards to both first-year and upper-level courses. *See* Downs & Levit, *supra* note 9, at 839.

211. For example, assume that the forty students in a class received the following grades: 3 A+, 5 A, 6 A-, 11 B+, 9 B, 6 B-. Assume further that these grades have the following values: A+ = 4.3, A = 4.0, A- = 3.7, B+ = 3.3, B = 3.0, B- = 2.7. To calculate the mean, one first determines the sum of the grade point values of every final grade given: (A+: 3 x 4.3 = 12.9) + (A: 5 x 4.0 = 20) + (A-: 6 x 3.7 = 22.2) + (B+: 11 x 3.3 = 36.3) + (B: 9 x 3.0 = 27) + (B-: 6 x 2.7= 16.2) = 134.6. Then, one divides the sum by the total number of students: 134.6/40 = 3.365 mean GPA.

212. *See* Daggett, *supra* note 115, at 400.

nineteen 2.7s.[213] The difference in mean GPAs between these two classes is more than three times what is permitted under my curve.[214] Of course, such dramatic differences are unlikely. But the critical point still holds: Median-based systems are much more susceptible to manipulation than mean-based systems.[215] Thus, median systems are significantly less likely to remedy grade disparities of the type that exist at the Bowen Law School.[216]

The second alternative is the "grade range" approach, under which ranges are set for each available mark—e.g., 4% to 6% of the students must receive an A+, 8% to 12% an A, 14% to 18% an A-, 23% to 27% a B+, 28% to 32% a B, and 12 to 16% a B-. Like mean-based systems (and unlike median-based systems), grade range systems *do* sufficiently control professor discretion with respect to the overall value of grades given. However, grade-range systems are generally less flexible than mean-based and median-based systems. In other words, while median-based curves permit too much discretion, grade-range curves do not provide enough. Most grade-range systems are based on a normal distribution. But the students at a given school might not be normally distributed in ability or work ethic.[217] Moreover, smaller classes are much less likely to consist of students who reflect a normal distribution.[218] Mean-based systems do not cause any problems when student performance in a given class does not reflect a bell curve.[219]

---

213. In each of these classes, the median falls within the range because when there are an even number of marks, "the median is the midpoint between the two middle scores." *Id.* at 402 n.31.

214. Since my proposed curve sets a mandatory range of 3.2 to 3.4, the maximum permissible difference between two required courses and between two electives with more than thirty students is 0.2. But the difference between the two classes in my median-system example is 0.7225. For the high class, the mean GPA is 3.685: (4.0 x 19 = 76) + (3.4 x 21 = 71.4) = 147.4; 147.4/40 = 3.685. For the low class, the mean GPA is 2.9625: (3.2 x 21 = 67.2) + (2.7 x 19 = 51.3) = 118.5; 118.5/40 = 2.9625; 3.685 – 2.9625 = 0.7225. Note that a mandatory median of 3.3 would restrict professors more than a median of 3.2 to 3.4. But it would still fall well short of a mandatory mean of 3.2 to 3.4.

215. *See* Stake, *supra* note 155, at 588 n.11 ("[M]edians can be easier for a teacher to manipulate in undesirable ways. One thing a teacher can do is shift the grades of some students above the median to higher grades without shifting the median. . . . Another undesirable thing a teacher might do in response to a forced median is just change the grades of a few students near the median, rather than shifting the whole grading scale to meet the mandate. It is harder to shift the mean by changing the grades of only a few students because the amount the mean changes is the total shift divided by the number of students."); JOHNSON, *supra* note 8, at 241 (explaining that this type of manipulation actually happened at one law school); *see also* Daggett, *supra* note 115, at 402 ("The mean is the best measure of central tendency unless the distribution is skewed (meaning scores tended to be clustered disproportionately at the high or low end of the range); in that case, the median is the best measure.").

216. *See* Keating, *supra* note 9, at 185 (explaining that mandatory median systems give professors too much discretion and that the mandatory median at the author's law school failed to prevent problematic "variations in the distribution of grade-wealth").

217. Stake, *supra* note 2, at 599–600 (explaining two reasons why students at many schools may not reflect a normal distribution).

218. Henderson, *supra* note 20, at 421 (explaining that in samples with fewer than thirty to forty-five students, "statisticians would conclude that the sample would be insufficient to produce a normal distribution"); Downs & Levit, *supra* note 9, at 846 ("Students in small samples are less likely to be distributed evenly."); *see also* Stake, *supra* note 2, at 601 ("For a class of thirty students, however, forcing the scores onto a curve will be troublesome because some of the intervals will contain very few students."); Leslie M. Rose, *Norm-Referenced Grading in the Age of Carnegie: Why Criteria-Referenced*

Of course, the flexibility in setting the grade distribution that is permitted by a mandatory mean is also a potential weakness. For example, two professors teaching the same course might consistently give two different types of distributions.[220] Grade-range systems are less susceptible to this problem.[221] Nonetheless, I believe that few professors will consistently adopt extreme grade distribution practices that might significantly compromise the goals of a mandatory curve. And the danger can be limited by including language in the forced curve policy recommending that professors typically award grades via a normal distribution.[222]

Another concern with my proposed curve is the inclusion of a mean range—under my proposal, 3.2 to 3.4. As I explained previously,[223] the primary purposes underlying such a range are (1) flexibility in setting grade distributions to reflect variations in student accomplishment within a given class,[224] and (2) preventing the curve from appearing unduly restrictive to students and faculty. But what if one professor consistently aims for the top of the range, while another regularly aims for the bottom?[225] That is actually a weakness for all but the most rigid normalization systems. Even the bulk of grade-range systems permit enough flexibility such that two professors teaching the same course could regularly provide their students with different levels of grade wealth. Under my proposal, I address this issue by setting

---

Grading Is More Consistent with Current Trends in Legal Education and How Legal Writing Can Lead the Way, 17 LEGAL WRITING: J. LEGAL WRITING INST. 123, 154 (2011) ("In addition, minimal differences between the students may be exaggerated under certain norm-referenced systems, particularly when a certain percentage of each grade is mandated.").

219. Despite these issues, it should be noted that grade range systems are actually the most common; mean-based curves are the second most popular. *See* Mroch, *supra* note 9, at 3–4 (in this 2005 AALS study of grading policies, 145 ABA-accredited schools out of 188 reported their data to AALS; 115 had formal grading policies; of those 115, 72 schools used the grade range approach, 52 used the mean GPA approach, 28 used the median GPA approach, and 8 used some other alternative); Downs & Levit, *supra* note 9, at 838 (reporting similar results for a survey conducted in 1996).

There are grade-range systems with substantial flexibility. *See, e.g.*, Mroch, *supra* note 9, at 13 (setting forth an example), but such systems do not sufficiently cabin the discretion of graders. *See* Stake, *supra* note 2, at 601 (noting that greater flexibility in each grade range can lead to major variations in mean grades). There is also some indication in the literature that grade-range systems have statistical problems unless standard deviation calculations are used. *See, e.g.*, *id.* (explaining this issue).

220. Fines, *supra* note 20, at 893; *see also* Henderson, *supra* note 20, at 419 ("A few law schools do realize the essential unfairness of mandating means for the law school curriculum, but not mandating standard deviations.").

221. *See also* Stake, *supra* note 2, at 592–99 (arguing that standard deviations should be equalized across all courses).

222. *See* E-mail from Lawrence S. Krieger, Clinical Professor, Fla. State Univ. Coll. of Law, to Author (Aug. 23, 2010, 9:06 AM CST) (on file with the author). However, I included no such recommendation in my proposal. *See infra* Appendix 2.

223. *See supra* note 203 and accompanying text.

224. For example, in a doctrinal course, the range makes it easier for professors to assign grades that reflect genuine differences in raw scores among the students.

225. *See* Henderson, *supra* note 20, at 420.

3.3 as an aspirational target within the 3.2 to 3.4 range.[226] Thus, consistently awarding grades at the low or high end of the range is not compatible with a good faith application of the system. My hope and belief is that professors who consciously attempt to defeat the goals of a grade normalization scheme will be rare.

A mandatory curve based on mean GPA provides the best balance. It restricts professors sufficiently to avoid unjustified grade disparities and provides the flexibility necessary in classes where students are not normally distributed.[227]

2. *Altered Standards for Smaller Courses*

Under the curve I proposed at my school, all required courses are strictly governed by the 3.2 to 3.4 GPA range. The same is true for electives with thirty or more students. Electives with nine to twenty-nine students are governed by the range, subject to the following proviso: If the mean GPA of all students enrolled in the course—based on prior course work at the law school—is lower than 3.2 or higher than 3.4, the permissible grade range for the course expands to encompass the students' incoming mean GPA average, plus 0.1 GPA points if the incoming mean GPA average is greater than 3.4, and minus 0.1 GPA points if the incoming mean GPA average is lower than 3.2. For example, if the combined GPA of the students enrolled in a course is 3.5, then the GPA range for the course is 3.2 to 3.6. If the combined GPA of the students enrolled in a course is 3.1, then the GPA range for the course is 3.0 to 3.4. Electives with eight or fewer students are governed by the same rules as those with nine to twenty-nine students, except the curve switches from mandatory to recommended.[228]

The reason behind the altered standard for smaller courses is statistical. A mandatory curve must be adjusted when applied to classes in which the pupils are not representative of the overall student population. Under the Central Limit Theorem, the danger of this happening becomes statistically

---

226. *See* Downs & Levit, *supra* note 9, at 831 ("Indeed, those schools which have a range, but no target within the range, may permit significant differences in grading standards and results which are not accounted for by chance."); *see also id.* at 838–39 (discussing law schools that use a combination of a required range and an aspirational target).

227. *See* JOHNSON, *supra* note 8, at 243 ("In my opinion, carefully designed constraints on mean course grades provide the most comprehensive solution to the problems associated with disparities in grading practices."). There is also an important administrative argument in support of mean GPA systems. As discussed in Part VI.A.2., *infra*, the required GPA range shifts up or down in small classes when the incoming GPA of the students is higher than 3.4 or lower than 3.2. It is easier to determine how much the curve should be changed when the measure of student quality that is the basis for such a change (mean GPA) and the standard that sets the parameters of the mandatory curve (mean GPA) are the same.

228. *See infra* Appendix 2.

significant when the course enrollment falls under roughly thirty students.[229] In other words, in classes with fewer than thirty members, there is a statistically significant likelihood that the registered students will be above or below average when compared to the student body as a whole.[230] Recognizing this problem, many schools consider the number of students enrolled in determining whether and how the curve applies to the course.[231]

There are two principal ways to address the concern about representative samples. The first is to exempt courses with under thirty students from the mandatory curve.[232] However, this creates significant fairness problems. Most law schools offer so many courses with an enrollment under thirty that excluding such courses would severely weaken the curve's corrective effect on the problem of grade disparities.[233]

For example, at the Bowen Law School, in the fall of 2010, we offered forty-three elective courses open to second and third-year students—twenty-one in the day program and twenty-two in the evening program. Of these

---

229. *See, e.g.*, Downs & Levit, *supra* note 9, at 835 ("The Central Limit Theorem suggests that as a 'rule of thumb' a sample of thirty is reasonably sufficient to reduce the probability that differences among samples are accounted for by chance."); E-mail from Allen Kamp, Professor of Law, The John Marshall Law School, to Author (Feb. 3, 2010, 7:37 PM CST) (on file with the author) ("[I]t is a rule of thumb in statistics that you can't use a population of 30 or less . . . because the 30 would not be representative of the general student population."); Stake, *supra* note 2, at 591 n.19 ("Thirty to thirty-five observations is a typical rule of thumb used in statistics to distinguish between large and small samples. But the number of degrees of freedom and the nature of the sample are important too."); *see also* Daggett, *supra* note 115, at 399 n.21 ("According to the central limit theorem, law school exam scores will approximate a bell curve the larger the group being measured."); Henderson, *supra* note 20, at 421 ("[S]tatisticians typically assume that when the number of items sampled is greater than thirty to forty-five, the samples from most populations would include a wide enough range of samples to approximate a normal curve. Below this number, statisticians would conclude that the sample would be insufficient to produce a normal distribution."). Interestingly, of the schools that exempt small courses from their curve or apply a different standard to such classes, the mean class size used to distinguish a "small" course is only *twenty* students. Downs & Levit, *supra* note 9, at 840.

230. *See also* Daggett, *supra* note 115, at 400 (observing that smaller classes are more likely to be populated by a disproportionate number of top students); Stake, *supra* note 2, at 591 ("In the second and third years better students may gravitate to some courses, and lesser students to others. If that happens, forcing teachers to a single mean could easily increase miscommunication rather than decreasing it."). Professor Stake offers the following example that helps to illustrate the point. At many law schools, the course on federal jurisdiction gets disproportionately talented students. Strictly applying a mandatory curve to that type of class is unfair since the students are likely to perform better than those in other courses. Such application will deter average and weaker students from taking the class. These students will rationally conclude that they are probably going to finish near the bottom of the course, which means they will receive particularly low grades since the more talented students will earn the bulk of the high and average grades. *See* Stake, *supra* note 2, at 591; *see also* Downs & Levit, *supra* note 9, at 845 (offering additional examples).

231. Downs & Levit, *supra* note 9, at 840 ("Forty-six schools explicitly considered the size of the course as either one or the only determinant of whether the policy applied; forty of these schools specified a particular number. . . . There seemed to be general recognition, expressed in various different ways, that small enrollment or self-selected groups may not reflect the school as a whole.").

232. *See id.* at 846 (noting that some schools exempt courses with fewer than thirty, forty, or fifty students).

233. For additional discussion of whether small courses should be exempt from mandatory curves, see *infra* Part VI.B.1.

classes only six in the day division and four in the evening division had more than thirty students. If courses with less than thirty enrolled were exempted from the curve, then the curve would have virtually no impact on the unjustifiable grade disparities in upper-level courses at my law school; such an exemption would eviscerate the corrective effect of the policy. Moreover, it might actually discourage students from taking bar classes since those electives tend to exceed thirty members. Students would focus on the smaller classes excluded from the curve, since grades in those classes would likely be higher than in courses subject to the policy.[234] For the same reason, exempting courses with fewer than thirty members also disadvantages students who are interested in the most popular topics.

The second method for dealing with the problem of unrepresentative course registrants is to use the mean GPA of the students enrolled in the course—based on their prior work at the law school—to shift or expand the mandatory GPA range.[235] As explained above, under my proposal, if the students' incoming GPA is higher than 3.4—i.e., if the enrolled students are above average performers—the instructor is permitted to give higher grades reflecting the capacities of the students. Similarly, if the students' incoming GPA is lower than 3.2—i.e., if the enrolled students are below average performers—this approach provides similar flexibility. Note that because my proposal advocates *expanding* the mandatory range rather than simply *shifting* it, there is no mandate that professors give grades outside the 3.2 to 3.4 range. Above average students do not automatically get the benefit of their prior superior performance. And below-average students are not locked into lower grades based on their previous academic history.[236] The GPA range expansion simply gives professors greater flexibility in assessing the output of the students.

---

234. *See* Keating, *supra* note 9, at 190 ("Everyone knows that the absence of forced grade standardization in small classes would invariably lead to lots of grade-wealth-per-student being distributed in these small classes, along with all of the attendant collective action problems that occur in an unregulated grading environment.").

235. *See* Downs & Levit, *supra* note 9, at 846 ("One solution to the problem of probable non-random distribution of students is to set the grade normalization target mean or range of acceptable means at a level which reflects the actual ability of the students in the course. This can be done by using the mean of student grade averages of the class members compiled for all prior law school course work.").

236. Professors Downs & Levit noted that some students at UMKC School of Law were strongly opposed to a system that uses incoming GPA to shift the permissible mean up or down because "such a system would allow the 'rich to get richer'" while "[t]he weaker students would continue to pull themselves down by their prior performances." Downs & Levit, *supra* note 9, at 846 n.56; *see also* Georgakopoulos, *supra* note 8, at 454 (contending that a mean-shifting system harms slow learners who perform poorly early in their legal education). While Professor Epstein has offered a powerful rejoinder to this concern, his argument centers on student course selection. *See* Epstein, *supra* note 7, at 709–10. When issues of fairness and student motivation are considered, the expanded-range approach is superior to the shifted-range approach.

Note further that this approach does not allow an expansion of the GPA range based on the subjective assessment of the professor, for all the reasons discussed previously about the limited value of an individual professor's personal belief that his students performed better or worse than average.[237] Only objective, reliable data that the students are not representative can justify expanding the mean range, and the GPA of the course enrollees is the only such data we have.[238]

Using the incoming mean GPA to shift or expand the permissible grade range is preferable to excluding small courses from the curve because the former avoids the unfairness that results from a complete exemption while addressing the danger that courses with less than thirty students are not representative of the full student body.[239] This second approach has been defended by multiple scholars[240] and has been adopted by many law schools,[241] including at least one school that uses the approach of expanded rather than shifted ranges.[242]

One objection to shifting or expanding the grade range based on prior course work is that upper-level courses may not test precisely the same skills and abilities as first-year classes. If that is so, then "[r]elying on past grades to predict performance (and to control grade allocations in future classes) is unfair to those students whose greater abilities are not tested or revealed (or developed) until later in the education process."[243] While I doubt that more than a few upper-level courses are sufficiently different from the first-year curriculum to justify this concern, I cannot discount the worry entirely. But even assuming that the objection has merit, the mean-expanding system I recommend is clearly the lesser evil. There is little question that more overall unfairness will result if a mandatory curve is applied

---

237. *See supra* Part V.A.

238. *See supra* note 76.

239. Keating, *supra* note 9 at 190 ("[A]pplying the forced standardization (particularly when adjusted to account for the past performance of those in each group) is still superior to a system of unregulated grading for small classes.").

One small concern with the way I designed the mechanism is that if the combined GPA of the students in a course is 3.39, the GPA range for the course is 3.2 to 3.4. But if the combined GPA is 3.41, the GPA range for the course is 3.2 to 3.51. Of course, this is not ideal. I felt, however, that the type of curve structure necessary to avoid a small unfairness like this would have unduly complicated my proposal, reducing the prospects of its passage. Nonetheless, I do support a curve structure that would resolve this unfairness issue.

240. *See, e.g.*, Epstein, *supra* note 7, at 709–10; Stake, *supra* note 2, at 589 ("The best way out of this dilemma is to set the average grade for the regular students according to their GPAs in previous course work.").

241. *See, e.g.*, Kaufman, *supra* note 7, at 420 (detailing the policy at one school that uses "a fixed class mean in larger classes but sets a population-specific mean for smaller classes" based on the "cumulative grade point average to date of the students in the class."); FLORIDA STATE UNIVERSITY COLLEGE OF LAW ACADEMIC RULES, POLICIES, AND PROCEDURES § 4.3, *available at* http://www.law.fsu.edu/current_students/rules/grading.pdf (same); LOYOLA LAW SCHOOL GRADING POLICIES, *supra* note 50 (same).

242. *See* Downs & Levit, *supra* note 9, at 846 (explaining the approach of that school).

243. Fines, *supra* note 20, at 895.

in an unadjusted fashion to small courses than if my expansion system is used. And there is essentially no doubt that exempting classes under thirty students entirely will be worse than using the mean-expansion approach.

One other option for dealing with the statistical problems that arise in applying mandatory curves to small classes deserves brief attention. Some schools expand the mean range both up and down in all small classes.[244] Since smaller groups are statistically more variable, expanding the mean range in this way makes some sense.[245] The weakness of such an approach is that it is not sufficiently targeted. All small courses are covered, including those where there is little basis to believe the enrollees are not representative of the broader student body. The approach I recommend is tailored to course-specific student populations.

It should be noted that, under the Central Limit Theorem, a sample of thirty is only considered appropriately representative when the sample is selected randomly. Registration for elective classes is not random.[246] Indeed, registration for *required* courses is often not random.[247] There is thus a danger that a class with more than thirty students will be non-representative. Applying the mean-shifting or mean-expansion system to *all* classes, regardless of size, would resolve this concern. At least one school has adopted that approach.[248] Ultimately, my belief is that the danger of a course with more than thirty students being unrepresentative is not sufficiently high to necessitate applying the mean-expansion system to all electives, at least at law schools like the one where I teach. Registration operates in a sufficiently random fashion to make the Central Limit Theorem applicable for practical purposes. Only in rare cases—such as the course on federal jurisdiction at many law schools—will there be a genuine danger that a class with more than thirty students is populated by an unrepresentative group.[249]

With that said, in an ideal world, I would have recommended that my faculty apply the mean-expansion system to *all non-first-year courses*, re-

---

244. *E.g.*, UNIVERSITY OF HOUSTON LAW CENTER STUDENT HANDBOOK 18 (2006), *available at* http://www.law.uh.edu/student/Handbook.pdf (establishing a mandatory mean of 2.9 to 3.1 in large courses and 2.8 to 3.2 in small classes).

245. *See* Epstein, *supra* note 7, at 709 (explaining that the author's school adopted the expanded range approach because of the statistical variability concern).

246. *See* Downs & Levit, *supra* note 9, at 845 ("Indeed, where students self-select, either their course or section, the usual assumption of randomness is inappropriate.").

247. For example, at the law school I attended, Constitutional Law was a required second-year course. But there were four sections, each taught by a different professor. Thus, students had some flexibility in choosing their Constitutional Law instructor. Even first-year courses are sometimes not random. At the Bowen Law School, we have two first-year sections—a day section and a night section. The students are not randomly placed into one section or the other. Because admitted students tend to prefer the day section, the students in that section frequently have stronger admission statistics than those in the night section.

248. *See* Mroch, *supra* note 9, at 65; *see also* Keating, *supra* note 9, at 187 (explaining that the author proposed this approach at his law school).

249. Federal Jurisdiction has never had more than thirty enrollees at Bowen.

gardless of size and regardless of whether the class was required or an elective. Applying the mechanism to every upper-level course with eight or more students safeguards against the mandatory curve harming *any* non-representative class populations. I did not make this recommendation at my institution on political grounds. I feared it would cause too much opposition. But I believe it is the optimum approach.

A note regarding courses with less than nine students is in order. While the incoming-mean-GPA system works with courses between nine and thirty students, I have some concerns about its efficacy in the smallest courses. Small groups behave less predictably,[250] creating a greater potential for substantial variations in student performance each year.[251] Thus, additional flexibility is warranted.[252] However, over time, "any given teacher's mean in such classes ought to fluctuate around the school average and not be biased in one direction or the other."[253] Some grade normalization is thus justified in very small courses, and that is why my proposal made the curve recommended in classes with eight or fewer students,[254] rather than not applying it at all.

I chose the precise number—eight or fewer—for several reasons. To begin with, it is clear that a curve cannot be applied effectively to tiny courses—say, those with one, two, or three students—without having a number of pernicious effects. For example, if the course has just one student, that student's grade will essentially be set by the curve before the course begins. With two or three students, the grades of each member will be too heavily contingent upon the performance of others in the class. It is difficult to determine at precisely what point these types of effects are no longer a concern. My intuition is that it is around five or six. Eight thus provides some cushion.[255]

---

250. Stake, *supra* note 2, at 592.

251. *Id.* at 590 n.16 ("Substantial year-to-year variation is likely only in very small classes, or when admissions standards have changed.").

252. *Id.* at 592 ("[A] school's grading constraints might properly allow greater variation in the average grade in small classes.").

253. *Id.* at 592.

254. *See* Mroch, *supra* note 9, at 21 (describing a school that changes the curve from mandatory to recommended once enrollment falls below twelve students; note, however, that the recommendation is very strong—"there should still be a reasonable distribution of grades unless there are very unusual circumstances (e.g., only top students are enrolled)"); *see also id.* at 67 (describing a school that distinguishes between courses with more than twenty students, thirteen to twenty students, nine to twelve students, and eight or fewer students); HOUSTON LAW CENTER STUDENT HANDBOOK, *supra* note 244, at 18 (setting different *mandatory* standards for (1) first-year courses and upper-level courses with more than twenty students, (2) "small classes" with eleven to twenty students, and (3) seminars, and a *recommended* standard for (4) "very small classes" with ten or fewer students).

255. *Cf.* E-mail from Professor Cyn Yamashiro, Loyola Law School (Los Angeles), to Annette R. Appell and Melissa Swain (Feb. 20, 2010, 6:58 PM CST) (on file with the author) ("I believe that for a credible curve using a workable standard deviation, you need at least 8 students.").

On the other hand, I am firmly opposed to a number that is any higher than eight. Grade disparities involving Bowen's smaller classes have been deeply problematic. For example, seminar courses (i.e., paper classes) at my institution are restricted to sixteen or fewer students, and most have between twelve and sixteen enrolled. Professors in these classes regularly award much higher grades than in other classes. Moreover, differences among the seminars have been dramatic.[256] Our law school is not alone in suffering from these types of problems with seminars.[257] Setting the line at eight addresses this and similar situations.

B. The Scope of Mandatory Curves

For a mandatory curve to accomplish its purpose, the curve must apply to virtually all courses. Limiting the scope of the curve "undermine[s] the equity goal"[258] because grade disparities will likely develop (1) among exempted courses, and (2) between covered and exempted classes.[259] These disparities will also compromise the aim of motivating students to register for courses on substantive grounds. Students will have an incentive to select classes exempted from the policy in the hope of receiving higher grades.[260] Indeed, according to a student at one law school whom I have spoken with, third-year students at this person's institution regularly press second-year students not to register for a course if adding members to the course will subject it to the institution's grade normalization policies.

Nonetheless, arguments are frequently made that certain types of classes should be exempt from mandatory curves. This subpart responds to such

---

256. *See supra* Part III.

257. *See, e.g.*, Keating, *supra* note 9, at 190–91 (Committee report recommending adoption of a mandatory curve noted in part: "The Committee unanimously believes that seminars should be treated the same as regular courses for grading purposes. The tremendous discrepancy between grades in seminars and regular courses and among seminars is perceived as a grossly inequitable part of our current system, and we agree.").

The reader may have noticed that the mandatory portion of my curve applies to *required* courses with fewer than eight students. Such courses are rather rare in legal education. But there is one class at Bowen that fits into this category—an upper-level, skills course with seven to eight students per section, taught by adjuncts, known as "Law Skills II." The grade disparities in that class have been particularly egregious. As a result, I concluded that any harm caused by applying the curve to Law Skills II would be outweighed by the benefit of consistent grades across sections of the course.

258. Fines, *supra* note 20, at 893.

259. *See* Keating, *supra* note 9, at 190 ("Everyone knows that the absence of forced grade standardization in small classes would invariably lead to lots of grade-wealth-per-student being distributed in these small classes, along with all of the attendant collective action problems that occur in an unregulated grading environment.").

260. Fines, *supra* note 20, at 900 ("Competitive students may no longer be shopping for the 'easy graders' but they are quite likely to shop for the courses to which the normalization policies do not apply or to arrange their schedules with a percentage of pass/fail credits so as [to] maximize their investment in and chances for beating the curve in their graded classes.").

contentions with respect to smaller classes and classes that do not follow the "traditional" model of Socratic lecturing and a final exam.

1. *Mandatory Curves Should Apply to Smaller Courses*

Some believe that courses with a low enrollment should be excluded from grade normalization entirely. Law Schools have taken varying positions on this issue.[261] Some schools exempt courses that fall below a certain enrollment level.[262] Others apply a mandatory curve or other normalization policies to all classes regardless of size.[263] However, at many institutions in the latter category, the curve operates under different parameters in smaller courses. For example, some shift or expand the grade range based on the incoming GPAs of the students,[264] as I recommend. At other schools, the permissible GPA range is wider in all small courses.[265] And at still others, the curve is automatically shifted up in smaller sections.[266]

There are two arguments typically offered by supporters of the view that low enrollment courses ought to be entirely exempt from mandatory curves or other grade normalization. The first focuses on the problem of representative samples.[267] As explained above,[268] once class size falls below thirty pupils, there is a statistically significant danger that the enrolled stu-

---

261. *See* Downs & Levit, *supra* note 9, at 836, 840 (In this 1996 survey, 97 of 116 schools responding had some type of grade normalization policy, and of the 97, 46 "explicitly considered the size of the course as either one or the only determinant of whether the policy applied.").

262. Downs & Levit, *supra* note 9, at 840; Mroch, *supra* note 9, at 7; *see, e.g.*, *Grading & Ranking*, UNIVERSITY OF WASHINGTON SCHOOL OF LAW, *available at* http://www.law.washington.edu/Students/Academics/Grading.aspx#litContent (exempting courses with fifteen or fewer students).

263. Downs & Levit, *supra* note 9, at 840; Mroch, *supra* note 9, at 7; *see, e.g.*, *Academic Policies*, UNIVERSITY OF VIRGINIA SCHOOL OF LAW, § I.H., *available at* http://www.law.virginia.edu/pdf/academics/policies_march2011.pdf (all courses, regardless of size, are governed by the mandatory curve); PEPPERDINE UNIVERSITY SCHOOL OF LAW, ACADEMIC POLICY 14, *available at* http://law.pepperdine.edu/academics/content/academicpolicystatement-post-8.1.09.pdf (same); *Academic Regulations*, UNIVERSITY OF KANSAS SCHOOL OF LAW, *available at* http://www.law.ku.edu/academics/regulations.shtml (same).

264. *See supra* notes 235–236 and accompanying text; *supra* notes 241–242 and accompanying text.

265. *See, e.g.*, *Law School Rules*, UNIVERSITY OF WISCONSIN LAW SCHOOL, § 2.07, *available at* http://www.law.wisc.edu/current/rules/chap2.html (In required courses and electives with more than thirty students, the mean GPA must fall between 2.85 and 3.1; in electives with thirty or fewer students, the mean GPA must fall between 2.7 and 3.3); HOUSTON LAW CENTER STUDENT HANDBOOK, *supra* note 244, at 18 (establishing a mandatory mean of 2.9 to 3.1 in large courses and 2.8 to 3.2 in small classes).

266. *See, e.g.*, PENN STATE UNIVERSITY DICKINSON SCHOOL OF LAW GRADING NORMS (on file with the author) (mandatory mean of 2.9 to 3.1 on 4.0 scale in required courses, 2.9 to 3.2 in large electives, and 3.0 to 3.6 in seminars and small classes); *Grading Guidelines*, UNIVERSITY OF PITTSBURGH SCHOOL OF LAW, *available at* http://www.law.pitt.edu/students/policies/grading (*recommended* median of B in large classes and B+ in small classes); Mroch, *supra* note 9, at 9, 13, 84 (listing the grading policies of three schools that follow this approach).

267. *See also* Fines, *supra* note 20, at 893 ("The first rationale exempts small classes from a curve because these classes do not provide a statistically significant sample—there is less likely to be a random distribution of ability in these classes, so we cannot expect a curve to result.").

268. *See supra* notes 229–230 and accompanying text.

dents will not be representative of the student body as a whole. This argument has little force when applied to curves like the one I proposed at my school: In courses with under thirty students, the mean, incoming GPA of the enrolled students may be used to expand the permissible grade range. And in electives with fewer than nine students, the curve shifts from mandatory to recommended. These mechanisms substantially resolve the concern that a small class might have an unrepresentative selection of students.

The second argument is that students perform better in smaller classes, and thus professors should be free to award higher grades in those courses.[269] In regard to the first part of this argument, there is much academic literature supporting the conclusion that lower student/teacher ratios enhance student performance.[270] However, the second step of this argument does not follow from the first. The fact that smaller classes promote greater achievement does not justify awarding higher grades in such classes. Students perform better in small classes primarily because of the structure of the class,[271] not for reasons relating to merit, such as aptitude or work ethic. Thus, awarding higher grades in small classes is unfair.

To elaborate, a lower student-teacher ratio facilitates class discussion and increases student-teacher interaction.[272] It also creates more opportunities for assessment and feedback. Put more simply, smaller classes often force students to work harder; thus, they learn more and perform better.[273] That is a significant benefit, to be sure. But it has little to do with student merit. The students work harder and perform better because of the structure of the class, *not* because they are better students than those in larger courses.[274]

---

269. Fines, *supra* note 20, at 894 (describing this argument) ("This rationale accepts the legitimacy of higher grades when instructional effectiveness is higher.").

270. *See, e.g.*, DAVID W. GRISSMER ET AL., IMPROVING STUDENT ACHIEVEMENT: WHAT STATE NAEP TEST SCORES TELL US at xxv–xxx (2000) (concluding that smaller class sizes correlate with improved performance by students); ELIZABETH WORD ET AL., THE STATE OF TENNESSEE'S STUDENT/TEACHER ACHIEVEMENT RATIO (STAR) PROJECT: FINAL SUMMARY REPORT 1985-1990 (1990) (finding a significant causal relationship between reducing class size and improved student performance); *id.* at 17 (concluding that "[s]tudents in small classes have higher performance than regular and regular/aide classes in all locations and at every grade level"); *see also* Note, *Federal Funding for Newcomer School: A Bipartisan Immigrant Education Initiative*, 120 HARV. L. REV. 799, 818 n.111 (2007) (collecting authorities). *But see* Fines, *supra* note 20, at 894 n.74 (noting that the research on this subject "is complex and somewhat contradictory").

271. *See generally* National Center for Research on Early Childhood Education, *Promoting Children's School Readiness: Rethinking the Levers For Change*, NCRECE IN FOCUS, Vol 1., Issue 1, at 1 (Feb. 2010), *available at* http://ncrece.org/wordpress/wp-content/uploads/2010/03/NCRECEInFocus_v1n1.pdf (referring to smaller class size as a "structural feature" that impacts learning).

272. *See* Riebe, *supra* note 164, at 333.

273. *But see* Keating, *supra* note 9, at 190 ("Some faculty will insist that students work harder in smaller classes, but I fail to see how an individual professor is in a reliable position to assess this.").

274. Put in slightly different terms, if students perform better in smaller classes for reasons that are separate from merit, the instructor ought to hold these students to a higher standard than those in larger courses. The pupils in the small class should have to accomplish more to warrant the same grades. Ap-

Consider the following example which helps to illustrate the unfairness in allowing better grades in small courses. Suppose Professor X chooses to teach a large, doctrinal, first-year course in sections that fall below the school's size threshold for the applicability of grade normalization, while Professor Y teaches the same course without sectioning. Should the students in the sectioned class receive higher grades?[275] By luck of the draw, they were assigned to a class in which they will probably work harder and possibly perform better. Do they deserve better marks because, by chance, they ended up in X's class rather than Y's? Do *Y*'s students deserve *worse* grades because, simply by a roll of the dice, they ended up in the non-sectioned class?

Here is another example. My law school has both a day program and a night program. If we were to mark the cutoff for application of a curve at any point above ten students, we would penalize the day students vis-à-vis the night students because night courses generally have lower enrollments. How would the day students at Bowen judge the fairness of such a system?

In essence, applying different standards in smaller classes creates unfairness when students do not have equal access to such classes.[276] Without similar access, a greater portion of grades is determined by "factors irrelevant to merit: . . . flexibility in scheduling, . . . savvy in understanding the enrollment process, . . . [and] luck in the course lottery."[277]

Let me offer one, final example that illustrates the injustice in treating smaller courses differently on any ground other than statistical necessity when students lack comparable access to such classes. Suppose Professors Q and R both teach the same large, first-year course with random enrollments. Professor Q publishes the minimum number of articles necessary to remain in good standing at the institution. Because of the time Q saves by writing less, Q administers and grades five tests per semester, in addition to the final. Q also uses the Socratic method extensively, calling on ten or more students each class session. Professor R, who publishes more, has less time to devote to grading, and thus gives only a final exam. R also uses the Socratic method sparingly, preferring to lecture. Because of the professors' different approaches to teaching and assessment, Q's students work substantially harder than R's. If students deserve higher grades in small classes because

---

plying a mandatory curve to smaller courses effectively forces this result. My thanks to Professor Matthew Silverstein of New York University in Abu Dhabi for helping me develop this point. *See also* Keating, *supra* note 9, at 189–90 ("Yet it would seem odd to foster a system in which students' overall class standing should be in part a function of how many small classes they take.").

275. *See* Fines, *supra* note 20, at 894 (offering a similar example).

276. Fines, *supra* note 20, at 894 (noting that the equity rationale for mandatory curves is undercut if certain smaller classes are exempted from the curve, but access to these classes is not equally distributed); *see also* Stake, *supra* note 2, at 592 (contending that if small classes are exempt from a curve, "[c]are should be taken . . . to make sure . . . that all students take the same number of hours in any group of courses.").

277. Fines, *supra* note 20, at 894.

they work harder and thus likely perform better, then Q's students deserve higher grades than R's students for the same reason. But that is grossly unfair. R's students are being punished simply because they were randomly assigned to a professor with different priorities. Do Q's students deserve compensation for their additional work? Maybe so. But they are being paid with greater knowledge. That provides them with a critical advantage over R's students. Q's students are not entitled to further reward in the form of higher grades. Note also that the force of this example is largely unchanged if an upper-level elective is substituted for the first-year course. As I explained in Part IV.A.,[278] students frequently have little control over who their instructor will be in second and third-year classes.

Critically, the unfairness of using alternative grading policies in smaller courses is not eliminated by procedures that guarantee equal access to such classes. To illustrate, suppose Student A is worried about the bar exam. Student A thus takes all of the bar courses at her law school, courses which tend to have large enrollments. Student B is not worried about the bar exam, and focuses on seminars and other small classes. Should A receive lower overall grades than B because A is apprehensive about the bar? No. B may have had to work harder in his courses, and thus might have learned somewhat more than A did in her classes. But A should not be punished for emphasizing bar courses in her education—courses that many schools strongly encourage their students to take.[279]

Likewise, suppose that Student C is interested in the most popular subjects, subjects which are taught in classes with large enrollments. Student D has more idiosyncratic tastes and thus focuses on courses with smaller numbers of students. Should C receive lower overall grades than D because C would like to learn the most popular subjects? Again, the answer is "no." Why punish a student with lower grades because that student happens to be interested in the most popular topics?

Admittedly, one might counter that some students take large classes so they can avoid work by getting "lost in the crowd." And if grades are the same in large and small courses, students will have less incentive to register for the latter because they typically must expend greater effort in small classes. To the extent these are concerns, though, we should directly encourage students to enroll in small courses—perhaps by mandating that students take a certain number of such classes—rather than by giving higher grades in smaller courses. The latter strategy penalizes students with legitimate rea-

278. *See supra* notes 25–26 and accompanying text.

279. Similar reasoning defeats the argument that "pure electives" should be excluded in general because the enrollees tend to be more interested in the subject matter and thus work harder. *See* Stake, *supra* note 2, at 589 (setting forth this argument). Why should student A be punished for taking courses she is less interested in because of a desire to better prepare for the bar exam, or for her likely field of practice?

sons for enrolling in large classes, such as the desire to fully prepare for the bar exam or interest in the most popular topics.

One might also reply that because Students B and D learned more in some of their courses than Students A and C did in theirs, awarding similar grades to these individuals will present inaccurate information to readers of transcripts. For example, a potential employer might mistakenly conclude that these students reached comparable levels of mastery in their courses. But this argument fails for the same reasons it did in the context of the "superior teaching" objection to mandatory curves: Any unfairness that is caused by misinformation of this type is small in comparison to the unfairness in awarding grades based on factors having little to do with student merit—including factors beyond the control of students, such as the popularity of certain subjects.[280] Once again, our first duty as law professors is to our students, not future employers and other consumers of transcript information.[281]

What about the student who chooses to register for small electives because he *wants* to work harder? Maybe this describes students A and C. Shouldn't they receive some benefit for their choices? Isn't the decision to enroll in smaller courses reflective of an aspect of merit—a strong work ethic? Yes. But A and C *are* receiving compensation for their choices in the form of greater knowledge. As with Professor Q's students from the preceding example, they do not deserve additional reward in the form of better grades. Note also that Students B and D are receiving a different benefit—B is learning more about bar exam topics, and D is learning more about the most popular subjects. In other words, A, B, C, and D are receiving rewards commensurate with their registration choices, independent of grades. This is

280. *See supra* note 91 and accompanying text.

281. The reader may have noticed that sometimes I consider misinformation on transcripts to be grossly unfair, and other times I treat it as only a small injustice. There is no inconsistency here. A single principle is driving my superficially different conclusions: Grades should be determined, to the maximum extent practicable, by student merit. On the one hand, when misinformation is the result of factors that have nothing to do with merit, I believe the misinformation is deeply unfair. For example, if students Y and Z are randomly assigned to professors with different grading philosophies, and there is no mandatory curve in place, then transcripts will inaccurately convey that Y and Z differed in achievement, when the difference was really in the grading standards of the professors. On the other hand, when misinformation is the unintended byproduct of a grade normalization policy designed to preserve the role of student merit in the awarding of grades, the unfairness caused by the misinformation is of lesser concern. For example, if Y is assigned to a section with an inferior teacher, leading Y to perform worse than Z, a mandatory curve that requires the professors to award Y and Z the same grade will inaccurately convey that Y and Z were equivalent in achievement, disguising the fact that Z performed better. But in this case, the unfairness of such miscommunication is outweighed by the unfairness that would result if Y received a lower grade than Z because of a roll of the dice in the registration process. In other words, when students perform worse for reasons *beyond their control* (e.g., registration in required courses) or for reasons *that reflect legitimate academic choices* (e.g., enrolling in bar courses or interest in the most popular topics), allowing transcripts to reflect the lesser performance will divorce transcript information from merit to a significantly greater degree than when a mandatory curve requires that similar grades be awarded to differing levels of performance.

all that is needed to ensure equity. Adding grades to the calculus is not only unnecessary, it is decidedly unfair.

Five other arguments against exempting small courses from mandatory curves are worth mentioning briefly. First, excluding such classes can eviscerate the corrective effect of grade normalization. As explained above,[282] at a school like Bowen, if courses with fewer than thirty students are not governed by the curve, the curve will do virtually nothing to address the problem of grade disparities in upper-level classes. We simply have too few electives with more than thirty members.

Second, as also discussed above,[283] it is more difficult for students to overcome poor performance in the first year if a law school applies its curve to a smaller proportion of upper-level classes. The grade inflation in second and third-year courses exempted from the curve makes it harder for students to improve their class rank later in their legal educations.[284]

Third, there is a crucial line-drawing problem. Members of the faculty at the Bowen Law School have proposed that critical learning thresholds are crossed when class size falls below four, eight, sixteen, twenty-five, sixty, eighty, and ninety-five students. If we exempt "small" classes based on "greater learning" rather than based on statistical sampling concerns, where should the cutoff be?[285]

Fourth, there are administrative problems with exempting smaller courses. For example, suppose a student takes several exempt courses during his second year, enabling him to raise his GPA. If this student takes courses during his third year in which the class curve is based on the incoming GPAs of the students, this might bias the curve upwards in those classes.[286]

Fifth, accepting for the sake of argument that it is justifiable to treat smaller courses differently because students perform better in them, excluding these classes from the curve is the wrong remedy. The proper response is to use a higher mean range in these classes—e.g., 3.2 to 3.5 or 3.3 to 3.5—rather than the 3.2 to 3.4 range that applies in larger courses.[287] Such an approach allows grades to reflect superior achievement in small courses

---

282. *See supra* notes 233–234 and accompanying text.

283. *See supra* notes 186–187 and accompanying text.

284. Georgakopoulos, *supra* note 8, at 446 & n.3.

285. *Cf. generally* Mroch, *supra* note 9 (identifying schools that have drawn the line at 55, 50, 45, 40, 35, 30, 25, 20, 15, 12, 10). *But see* Fines, *supra* note 20, at 894 n.74 ("The research on class size and its impact on learning is complex and somewhat contradictory. However, the most recent, well documented and designed research, indicates that in classes smaller than thirty students, a number of positive changes in both teacher and student attitudes and behaviors occur.").

286. *See* Stake, *supra* note 2, at 592 (also explaining that this last administrative problem can be remedied by not including grades from previously completed exempt courses in determining the class GPA for purposes of setting the curve).

287. Of course, if the students' incoming GPAs fall outside this range, the range should be shifted further, as justified by the statistical concerns discussed in Part VI.A.2., *supra*.

while preventing (1) grade disparities among small classes, and (2) dramatic grade variations between smaller classes and larger classes. If students perform better in smaller courses, the *most* that superior performance entitles them to is a higher curve. It does not justify unregulated grading.

In conclusion, smaller classes ought to be governed by mandatory curves.

2. *Mandatory Curves Should Apply to Seminars, Clinics, and Skills Courses*

Some assert that seminars, clinics, and skills courses should be exempt from grade normalization. As with the size issue, law schools are divided here.[288] Under the curve I proposed at my school, "non-traditional" courses are subject to the same rules.[289] Thus, in my system, seminars, clinics and skills classes are only treated differently if they fall below the pertinent size thresholds. That is because all other bases for providing special treatment to such courses—or an outright exemption—are unpersuasive.

Perhaps the most common justification offered for excluding non-traditional classes is that the students' close working relationship with the instructor improves student performance.[290] For example, professors provide more feedback in these courses and students often redo assignments based on that feedback. This contention is identical to the second argument offered for excluding *small* classes from mandatory curves: Because students perform better, they should receive higher grades.[291] And it is unpersuasive for

---

288. *See* Downs & Levit, *supra* note 9, at 836, 840 (in this 1996 survey, 76 of 116 schools responding had formal grade normalization policies, and of the 76, "twenty-seven . . . treated paper courses, seminars, skills courses, or research and writing courses by a different standard or exempted them from the policy altogether"; some implicitly exempted such courses under their "size" policies); *see also* Kaufman, *supra* note 7, at 416 (this 1993 survey of law school grading practices found that 91% "of the respondents grade seminars on the same basis as other courses"). *See, e.g.*, Mroch, *supra* note 9, at 41 (according to the policy of one school, "[t]he fact that a course involves a paper, a project, skills training, or a heavy workload does not justify a deviation from the 3.3 median"); *compare* UNIVERSITY OF KENTUCKY COLLEGE OF LAW FACULTY RULES AND POLICIES, *supra* note 50, at § IX.A. (curve applies to "every course and seminar"), *with Grading Policy*, THE UNIVERSITY OF IOWA COLLEGE OF LAW, *available at* http://www.law.uiowa.edu/documents/Grading-PolicyandRanking.pdf (mandatory curve is not applicable to "seminars and other classes in which the student's grade is based primarily on the student's performance on graded skills-oriented tasks (including writing) other than a final examination"). Note that rather than exempting "non-traditional" classes, some schools apply the types of mechanisms used with smaller classes, such as higher or expanded mean ranges. *See* Fines, *supra* note 20, at 894.

289. *See infra* Appendix 2.

290. *See* Downs & Levit, *supra* note 9, at 845 ("There may also be classes in which the grade is determined by a paper, which involves a degree of interactive work between professor and student and perhaps successive refinements, or by the demonstration in a skills course of certain competencies."); E-mail from Professor Annette R. Appell, Washington University in St. Louis School of Law, to Melissa Swain (Feb. 20, 2010, 6:59 AM CST) (on file with the author) (offering the same point in the clinic context).

291. *See supra* notes 269–270 and accompanying text.

all the same reasons.[292] Most importantly, granting that students do superior work in seminars, clinics, and skills courses (an assumption that can be challenged in many cases), the better performance is a byproduct of the structure of these classes, not merit.[293]

This justification for exemption is not strengthened by the fact that clinical faculty, as explained by one of my clinic colleagues, have an obligation to ensure that their clients receive "A work."[294] The quality of work ultimately performed *for the client* is not the same as the quality of work performed *by the student* receiving the grade. Indeed, as this same colleague admitted, the teachers in our clinic often "fill the gaps" left by the students. A student's grade should be based on the student's performance, *not* the work product of the teacher.

I also dispute the assumption that students always do superior work in non-traditional courses. Every seminar teacher at Bowen Law School who has shared their opinion with me on the subject believes that our students actually perform worse in these classes than they do in doctrinal subjects. Yet the grades in our seminars are almost uniformly higher than in doctrinal courses. There is plainly no justification for this result.

A second reason frequently pressed for exempting seminars, clinics, and skills classes is that grading in these courses is too subjective for curving. But it is difficult to see how grading in seminars, clinics, and skills classes is any more subjective than the "gestalt" grading system many professors use for essay exams in doctrinal courses. Indeed, given the subjective nature of "point" grading systems, assessment in seminars, clinics, and skills courses is arguably no more subjective than essay grading in *any* doctrinal class. Professors in non-traditional courses can assess the quality of rough drafts, student responsiveness to constructive criticism, class participation, client interaction, the quality of final drafts, and other aspects of class performance. Students almost certainly vary across these dimensions. That provides sufficient basis for distinguishing the students and implementing a mandatory curve.[295]

In addition, students generally perform a larger number of assignments in seminars, clinics, and skills courses than they do in doctrinal classes. Teachers awarding grades in the former thus have substantially more infor-

---

292. *See supra* notes 271–287 and accompanying text.

293. *See supra* notes 271–281 and accompanying text.

294. *See also* E-mail from Professor Annette R. Appell, *supra* note 290 ("Still, I am nervous about putting clinics on a rigid curve because my feeling is that if I am doing my job, everyone would be doing high level work within a much smaller grade range . . . .").

295. *See* E-mail from Professor Cyn Yamashiro, *supra* note 255 ("I am of the opinion that [a] student's performance can be credibly evaluated and should be. Moreover, while what we do as clinical faculty is much different, that doesn't mean that there aren't meaningful ways of assigning empirical values to [a] student's performance."). My seminar-teaching colleagues at the Bowen Law School have generally expressed the same point with respect to those courses.

mation to work with. This makes assessment *more* reliable in non-traditional courses.[296] Law professor Vanessa Merton, the director of the immigration clinic at Pace Law School, explains this point well:

> Isn't it obvious that the lowest-credit, least demanding clinical program imaginable still provides the professor with probably 1000 times as many data points for evaluation as an average classroom course? . . . During a semester, clinical students probably submit five or ten or fifteen times as much written work for feedback and review by clinical professors as is submitted in classroom courses . . . . Most clinical professors spend at least two hours per week interacting with each student, either individually or in pairs or very small groups. . . . This is not even to consider the vastly broader range of assessment topics that clinicians can use for evaluation.[297]

The more information we have available for assessing the abilities and work product of students, the *easier* it should be to grade them according to a mandatory curve. Greater levels of data provide additional bases for distinguishing the performances of our pupils, enabling faculty to distribute marks in accordance with grade normalization requirements, even when using criterion referencing.

A third (and related) argument for exempting non-traditional courses is that the work product is too varied to make comparisons fruitful.[298] As a result, professors cannot accurately distinguish between better and worse student output. For example, seminar students often write on dramatically different topics. Additionally, as one of my clinic colleagues pointed out, clinic students sometimes have clients with quite different needs. One student may have a contested trial, while another never goes to court. One student may have to research a novel legal issue, while another performs intensive factual discovery. Despite these variations in output, the argument from "variety" is unpersuasive on multiple grounds.

First, I fail to see how it is more difficult to compare performances in legal seminars and clinics than in courses covering English, Political Science, Philosophy, and the countless other subjects where teachers have little

---

296. *See* Lynn M. Daggett, *Teaching Torts by Integrating Ethical, Skills, Policy and Real-World Issues, and Using Varied Pedagogical Techniques: Reflections on Using the Henderson, Pearson and Siliciano Casebook*, 25 SEATTLE U. L. REV. 63, 74–75 (2001) ("Multiple exams in varied formats maximize the reliability and validity of grades.").

297. E-mail from Professor Vanessa H. Merton, Pace University Law School, to Melissa Swain (Feb. 21, 2010, 12:58 AM CST) (on file with the author).

298. *See, e.g.*, Fines, *supra* note 20, at 893 ("Even if there is a random distribution of ability, many seminar classes allow such a degree of flexibility in the products by which students are measured (e.g. seminar papers) that we do not trust our ability to adequately differentiate student ability in these settings.").

trouble making distinctions. Students in these fields often complete assignments that are quite different in topic and structure.

Second, my colleagues at the Bowen Law School who teach seminars and clinics do not seem to have any trouble distinguishing among the performances of their students. Several have expressly admitted to me that this is not difficult, and their actual grade distributions support the point. Indeed, the fact that so many clinical faculty have defended awarding letter grades rather than pass/fail marks[299] concedes that they can differentiate between gradations of student performance. The real problem at my school (and many others) is that seminar and clinical professors make all of the distinctions at higher levels, rather than awarding marks across a broader range.

Third, as explained in response to the "subjectivity" argument, students often perform more assignments in non-traditional courses. The greater number of data points available to teachers in these classes compensates for the variety in work product.

Fourth, the argument from "variety" proves too much. If it were correct, then I could assign papers in Contracts instead of administering a final exam and claim exemption from the curve because my students all wrote on such different topics that I could not tell which ones were better or worse. This makes no sense.

The "subjectivity" and "variety" arguments for excluding non-traditional courses from mandatory curves are really arguments for eliminating grades entirely in these classes. If grading in a particular field is too subjective, or if the work product is too varied to permit valid comparisons, what basis could a teacher possibly have for giving the students grades at all, other than perhaps "pass/fail?" Grading within a course is inevitably comparative. If a teacher awards Student X an A and Student Y a B, the teachers is necessarily stating that X performed better than Y. But giving X a higher mark than Y is only justifiable when the professor can validly distinguish among the performances of the students. The arguments from subjectivity and variety deny that such distinctions are possible. Therefore, these objections to applying mandatory curves in seminars, clinics, and skills courses are actually objections to the usage of standard grading practices in non-traditional classes. Under the proposal I made to the Bowen Law School, pass/fail courses are (obviously) exempted from the mandatory curve. If seminar, clinic, and skills course professors want an exemption from grade normalization, they should request a shift to pass/fail grading.[300]

---

299. *See, e.g.*, E-mail from Professor Vanessa H. Merton, *supra* note 297; E-mail from Professor Cyn Yamashiro, *supra* note 255 (explaining that clinics should be graded rather than pass/fail because student performance "can be credibly evaluated" and there are "meaningful ways of assigning empirical values" to student performance).

300. My thanks to Professor Matthew Silverstein of New York University in Abu Dhabi for helping me develop this argument. For a related perspective on this issue, see JOHNSON, *supra* note 8, at 236

Finally, I want to relay some experiences I had in private practice that support the proposition that clinic and seminar work can be assessed on a curve. During my time as a senior associate at a mid-size law firm, I regularly worked with junior associates—attorneys zero to three years out of law school. Since I was more senior, the management of the firm frequently asked me to review these associates. I spent countless hours working with these younger lawyers on every dimension of litigation. I suspect that I reviewed substantially more of their work than most clinic faculty review of their students since my working relationships with these junior associates stretched beyond one year. Critically, when it came time to perform the reviews, I had no difficulty distinguishing between the A-type junior associate, the B-type junior associate, and even the C-type junior associate. If I had needed to grade the junior associates on a curve, I could easily have done so, even though virtually every one I worked with had capacities comparable to the best students I have seen as a professor. Indeed, there were times I *was* asked to grade them on curve—primarily when we had to make staffing decisions with respect to a new case. And my perspective is not unique. All of the partners that I worked closely with graded associates (including me) on a curve to a degree.

In sum, seminars, clinics, and skills courses should generally be governed by the same grade normalization policies as other types of courses.

VII. CONCLUSION

According to professors Robert Downs and Nancy Levit, "[g]iven the overwhelming importance of grades in determining professional success and influencing personal esteem, . . . it is essential that law schools design grading systems that are scrupulously fair."[301] I agree with this sentiment. Unfortunately, no grading system is perfect. Every approach will have benefits and costs. Thus, I do not believe that what I propose in this article is ideal or free of problems. That is impossible for *any* grading system.[302] In the field of law, as elsewhere, academic assessment will always be a mixture of science and art. Therefore, the standard "scrupulously fair" is best thought of as an aspiration rather than a mandate. But it is an aspiration worth pursuing.[303]

(explaining that a professor's inability to distinguish the quality of work among students in a class does not justify the awarding of higher grades to the students, it only justifies awarding every student the same grade; further arguing that, in such a class, since all students "performed at an *average* level" all of them should be "given *average* grades").

301. Downs & Levit, *supra* note 9, at 820.

302. Numerous commentators have expressed this point. *See, e.g.*, Stake, *supra* note 2, at 618 ("Assessment of student performance is never perfect. Choosing the types of instruments for measuring performance will always involve tradeoffs.").

303. Stake, *supra* note 2, at 587 ("I assume that grading is never perfect. . . . Subjectivity in grading will never be eliminated. . . . We should take due care to prevent the errors we can eliminate and minimize the cost of the errors we cannot eliminate."); Downs & Levit, *supra* note 9, at 857 (given that law

And there are powerful arguments that a mandatory curve satisfies the greatest set of educational values.

APPENDIX 1
STATISTICAL ANALYSIS OF CRITICAL GRADE DISPARITIES AT THE WILLIAM H. BOWEN SCHOOL OF LAW USING THE WELCH T-TEST

Introductory Notes.

a. This appendix contains statistical analysis supporting my conclusion that the grade disparities listed in Part III generally flow from professor grading philosophy rather than from substantive differences in student performance. Note that some of the disparities listed in Part III were not analyzed on grounds of administrative convenience, and two were eliminated as bad data for purposes of the type of statistical analysis conducted.

b. In order to establish the statistical relationship between the mean GPAs of the classes in each class pair, a variation of the standard two tailed t-test known as a Welch t-test was performed.[304] The Welch variation of the t-test was chosen because it has the ability to compare sample sizes with unequal variations. The data used to perform the t-test included both the mean GPA of each course and the grade distribution for each course (i.e., the number of As, Bs, Cs, Ds, and Fs in the class).

c. The t-test was used to determine the probability that there was a significant statistical factor driving the mean GPAs of both classes in each class pair. Like all t-tests, the Welch t-test is designed to disprove a null hypothesis. In this case, the null hypothesis is "there is no statistically significant difference between the means of the classes in question." To test the null hypothesis with respect to a pair of classes, the mean GPA and grade distribution for both classes were used as the inputs in the t-test. The t-test returns a p value. If the p value is less than .05, then the null hypothesis has been disproven, and there is a 95% likelihood that the variation in the means of the two classes was not random or due to sampling error.

d. If the t-test shows that the variation in the means of two classes was probably not random or due to sampling error, then there is some factor the two classes do not have in common that is driving the variation, such as professor grading philosophy, course material, student interest in the subject, or difference in student abilities that might results from selection bias in the course registration process. An example of the latter would be students that are above average in ability disproportionately registering for an elective course, such as Federal Jurisdiction.

---

schools teach "the idealism of the way the world should work[,] . . . [w]e should strive to make the [grading] system as just as we possibly can").

304. *See* SARAH BOSLAUGH & PAUL ANDREW WATTERS, STATISTICS IN A NUTSHELL 151–68 (Mary Treseler, ed. 2008) (containing a thorough description of t-tests).

e. The p value of each class comparison is included in this appendix. As the results illustrate, in every class comparison, the null hypothesis was disproven. There are only two factors each class pair do not have in common that should play any significant role in explaining variation in the means—professor grading philosophy and student abilities. While there is certainly some bias in the registration process at the William H. Bowen School of Law, the sample size of individual classes is generally large enough such that differences in student abilities in the courses in each class comparison should be relatively rare, especially in required courses, and most especially in the class pairs where the students are largely the same in both courses. Accordingly, the bulk of the variation in the means in each class pair is probably caused largely by differences in professor grading philosophy.

f. Due to the large number of samples and their widely disparate nature, the standard deviations and degrees of freedom that are customarily included with t-test results are omitted. In addition, the actual t-test calculations for each class comparison and the grade distributions for each course are not included. The t-test was performed using a spreadsheet that is on file with the author.

g. All data in this appendix is derived from grade reports produced by the Associate Dean's office at the University of Arkansas at Little Rock, William H. Bowen School of Law.

h. The grade point averages listed for each course are *mean* grade point averages. The number of students in each class is set forth in parentheses. An asterisk (*) denotes a required course. A caret (^) denotes a composite grade point average derived from two sections of the course taught by the same professor.

1. First-Year Courses—Variances In Sections of the Same Course

Spring 2010

Legal Writing II* (27) = 3.361 vs. Legal Writing II*^ (34) = 2.206—P value = .00000576
Legal Writing II*^ (28) = 3.101 vs. Legal Writing II*^ (34) = 2.206—P value = .000056
Legal Writing II* (29) = 3.034 vs. Legal Writing II*^ (34) = 2.206—P value = .0001
Legal Writing II*^ (33) = 2.768 vs. Legal Writing II*^ (34) = 2.206—P value = .008

Fall 2009

Legal Writing I* (29) = 3.440 vs. Legal Writing I* (32) = 2.859—P value = .005
Legal Writing I* (29) = 3.440 vs. Legal Writing I*^ (34) = 2.772—P value = .001

Spring 2009

Contracts II* (59) = 3.134 vs. Contracts II* (91) = 2.750—P value = .0000139
Legal Writing II* (30) = 3.629 vs. Legal Writing II* (30) = 2.908—P value = .0000121
Legal Writing II* (30) = 3.629 vs. Legal Writing II*^ (32) = 2.754—P value = .000000675

Fall 2008

Torts* (92) = 3.269 (20 As) vs. Torts* (60) = 2.625 (2 As) —P value = .0000000475
Contracts I* (92) = 3.128 (19 As) vs. Contracts I* (60) = 2.825 (4 As)—P value = .002

Fall 2007

Contracts I (D) = 3.160 vs. Contracts I (N) = 2.839—P value = .0058
Legal Writing I* (30) = 3.442 vs. Legal Writing I*^ (31) = 2.751—P value = .0002

Spring 2007

Criminal Law* (84) = 3.176 vs. Criminal Law* (42) = 2.845—P value = .028
Legal Writing II* (19) = 3.276 vs. Legal Writing II*^ (41) = 2.801—P value = .0108

2. Upper-Level Courses—Variances in Sections of the Same Course

Spring 2010

Legal Profession* (33) = 3.273 vs. Legal Profession* (41) = 2.701—P value = .0001

Fall 2009

Const. Law* (57) = 3.482 vs. Const. Law* (86) = 2.781—P value = .000000704

Fall 2007

Evidence* (25) = 3.580 vs. Evidence* (57) = 3.035—P value = .00000069

Spring 2007

Legal Profession* (19) = 3.342 vs. Legal Profession* (71) = 3.007—P value = .005

Fall 2006

Evidence (16) = 3.422 vs. Evidence (76) = 2.572—P value = .00000000329

3. First-Year Courses—Variances Across Courses

Spring 2010

Contracts II*^ (56) = 3.222 vs. Civil Procedure II* (94) = 2.774—P value = .00000031

Fall 2009

Torts*^ (160) = 3.1955 vs. Property I*^ (100) = 2.544—P value = .0000000000031

Fall 2008

Torts* (92) = 3.269 (20 As) vs. Civil Procedure I* (92) = 2.717—P value = .000000121

Fall 2007

Contracts I* (96) = 3.160 vs. Torts (59) = 2.636—P value = .0000267

Spring 2007

Criminal Law* (84) = 3.176 vs. Property II* (82) = 2.759—P value = .0000733

4. Upper-Level Courses—Variances Across Bar Courses

Spring 2010

Secured Transactions (56) = 3.442 vs. Decedents' Estates (63) = 2.968-P value = .0002

Fall 2009

Secured Trans. (45) = 3.611 vs. Commercial Paper (40) = 2.944—P value = .000000383

Spring 2009

Secured Trans. (55) = 3.421 vs. Commercial Paper (52) = 2.860—P value = .000000512
Secured Trans. (55) = 3.421 vs. Business Associations (66) = 2.742—P value = .000000257

Fall 2008

Secured Trans. (40) = 3.488 vs. Commercial Paper (33) = 2.727—P value = .000000334
Conflicts of Law (20) = 3.463 vs. Evidence* (35) = 2.764—P value = .000000257

Spring 2008

Business Associations (87) = 3.338 v. Sales (17) = 2.471—P value = .0047

Fall 2007

Sales (12) = 3.625 vs. Commercial Paper (66) = 2.845—P value = .0000324
Evidence* (25) = 3.580 vs. Business Associations (18) = 2.917—P value = .0024

Spring 2007

Secured Transactions (28) = 3.436 vs. Family Law (34) = 2.868—P value = .0002

Fall 2006

Conflicts of Law (32) = 3.227 vs. Business Associations (48) = 2.729—P value = .0404

5. Upper-Level Courses—Variances Across Electives

Spring 2010

Law Office Management (15) = 4.000 vs. Administrative Law (17) = 2.941—P value = .0002
Public Health Law (10) = 4.000 vs. Civil Liberties (8) = 2.719—P value = .00238

Fall 2009

Disability Law (20) = 3.632 vs. Poverty Law (38) = 3.000—P value = .00000468

Spring 2009

Real Estate Transactions (60) = 3.429 vs. Oil & Gas Law (35) = 2.707—P value = .0000469

Fall 2008

Adv. Torts: Media Law (27) = 3.852 vs. Jurisprudence (28) = 3.259—P value = .00000411

Spring 2008

Real Estate Transactions. (32) = 3.758 vs. Construction Law (11) = 3.023—P value = .0039

Fall 2007

Local Government Law (28) = 3.759 vs. Health Law (21) = 2.583—P value = .000031
Administrative Law (49) = 3.597 vs. Juvenile Law (32) = 2.570—P value = .000000000343

Spring 2007

Land Use (23) = 3.727 vs. Introduction to International Law (39) = 2.974—P value = .0009
Gov. Reg. of Business (39) = 3.667 vs. Fed. Income Tax (28) = 2.889—P value = .000039

Fall 2006

Bankruptcy Law (21) = 3.475 vs. Federal Income Tax (36) = 2.951—P value = .006
Intellectual Property (28) = 3.438 vs. White Collar Crime (65) = 2.906—P value = .0008

6. Upper-Level Courses—Variances Across Seminars

Spring 2010

Fourth Amendment (18) = 4.000 vs. Family Mediation (16) = 3.367 —P value = .0000714

Fall 2009

Mediation (12) = 3.917 vs. Capital Punishment (16) = 3.266—P value = .0009

Spring 2009

Race & Criminal Justice (16) = 3.953 vs. Animal Law (19) = 3.197—P value = .0001

Spring 2008

Mediation (12) = 3.932 vs. Bioethics (16) = 3.469—P value = .0001

Fall 2006

Capital Punishment (16) = 3.797 vs. Law & Social Science (9) = 3.306—P value = .0331

APPENDIX 2
THE MANDATORY CURVE PRESENTED TO THE FACULTY OF THE WILLIAM H. BOWEN SCHOOL OF LAW

This appendix sets forth the critical features of the mandatory curve I presented to the faculty of the law school where I teach.[305] The faculty adopted my proposal in the spring of 2011 (effective Fall, 2011), with only a few, relatively minor changes. The appendix explains the changes as well.

A. Summary of Grading Scale and Basic Curve Features

4.0 = A	
3.9	
3.8	
3.7 = A-	
3.6	
3.5	
3.4	Top of Mandatory Mean Range
3.3 = B+	Target Mean
3.2	Bottom of Mandatory Mean Range
3.1	
3.0 = B	
2.9	
2.8	
2.7 = B-	Good Standing GPA
2.6	
2.5	
2.4	
2.3 = C+	
2.2	
2.1	
2.0 = C	Lowest *Credit* Grade
1.9 = F	
1.8 = F	
1.7 = F	
1.6 = F	
1.5 = F	Lowest *Possible* Grade

305. The information in this appendix is taken from two memos that I wrote, one to the full faculty and one to the law school curriculum committee, of which I am a member. A few, minor features of my proposal are not included because they are highly specific to our law school.

B. Detailed Description of Grading Scale and Curve Features

Grades to Be Awarded. Professors will not award letter grades. Rather, professors will award numerical grades ranging from 4.0 to 1.5. In other words, the following grades shall be used: 4.0, 3.9, 3.8, 3.7, 3.6, 3.5, 3.4, 3.3, 3.2, 3.1, 3.0, 2.9, 2.8, 2.7, 2.6, 2.5, 2.4, 2.3, 2.2, 2.1, 2.0, 1.9, 1.8, 1.7, 1.6, 1.5. Letter grades are listed only so that faculty, students, and employers will know the approximate letter value of each numerical grade.[306]

Mandatory Mean in Required Courses. For all required courses, the *mean* GPA of final grades for the course *must* fall between 3.2 and 3.4, with a target GPA of 3.3—i.e., the target mean is a B+.[307]

> Note: The grades of students (i) who receive an incomplete, (ii) who do not take the final exam or otherwise do not finish the course requirements, or (iii) whose exams or other course work were not graded for any reason (e.g., academic dishonesty) are *not* included in calculating the mean GPA for a course.

---

306. Washington University in St. Louis uses a system similar to this. *See Explanation* of *Grading System*, WASHINGTON UNIVERSITY IN ST. LOUIS SCHOOL OF LAW, *available at* http://law.wustl.edu/Registrar/pages.aspx?id=2236 (permitting every grade from 70 to 100).

Originally, I recommended a traditional letter grading system. However, for a variety of reasons, I amended the proposal to incorporate number grades.

My proposal increased the number of grade levels to 26 from 10 under the previously existing grading system. *See* UALR BOWEN LAW SCHOOL ACADEMIC RULES, *supra* note 3, at 12. This is inconsistent with the trend among law schools, which has been towards *fewer* grade levels. Downs & Levit, *supra* note 9, at 842. Professors Downs and Levit suspect the reason for the trend is that

> [w]e [professors] may have a greater comfort level in deciding between an A and a B than we do with smaller increments. Our more generalized concerns about teaching ability and technique, test crafting and scoring, and so on, may influence our willingness to make fine distinctions among students where we fear no real difference exists.

Id.

However, a grading system with a large number of available marks is superior to a system where there are fewer grades. As explained by Professor Stake, under a "coarse system"—i.e., one with few levels—errors are large. Stake, *supra* note 2, at 608. For example, under the previous system at the Bowen Law School, if a professor misgraded an exam, causing the student to move up or down a grade level, the grade was off by 0.25 or 0.5. A finer grading scale—i.e., one with many grade levels—"increases the *number* of errors but reduces the *size* of the errors." *Id.* (emphasis added). Under a system in which every grade value to the tenth of a point may be awarded, if a professor misgrades an exam, the grade will be off by only 0.1 (except in very rare cases). Stake explains that "[b]oth the theory of declining marginal utility of income and the actual and common sales of insurance suggest that the harm from putting large losses on a few persons is worse than the harm from placing small losses on many persons." *Id.* at 609. He also observes that GPA and class rank calculations support the finer system: "It seems less likely that errors in grading will average out when the errors are large and infrequent than when they are small and common." *Id.* at 609 n.55. Stake thus concludes that a finer grading system is superior. *Id.* at 609; *accord* Epstein, *supra* note 7, at 707; Keating, *supra* note 9, at 181–82 (1998); William K.S. Wang, *The Injustice of Reducing the Number of Levels in a Grading System*, 57 J. LEGAL EDUC. 423–26 (2007).

307. B+ is typically valued at 3.3 or 3.33 on a 4.0 scale.

> Note: Mean GPA shall be calculated based upon the final grade *for the course*, including mid-terms, class participation, and any other graded assignments; it is *not* calculated using just the final exam, final paper, or other final project, unless that is the only manner in which the students are evaluated.

Mandatory Mean in Elective Courses with Thirty or More Students. For all elective courses with thirty or more students, the mean GPA for the course *must* fall between 3.2 and 3.4, with a target GPA of 3.3.

Mandatory Mean in Elective Courses with Nine to Twenty-Nine Students. For all elective courses with nine to twenty-nine students, the mean GPA for the course *must* fall between 3.2 and 3.4, with a target GPA of 3.3, *subject to the following proviso*. If the mean GPA of all students enrolled in the course—based on prior course work at the law school—is lower than 3.2 or higher than 3.4, the permissible grade range for the course shall expand to encompass the students' incoming mean GPA average, plus 0.1 GPA points if the incoming mean GPA average is greater than 3.4, and minus 0.1 GPA points if the incoming mean GPA average is lower than 3.2. For example, if the combined GPA of the students enrolled in a course is 3.5, then the GPA range for the course is 3.2 to 3.6. If the combined GPA of the students enrolled in a course is 3.1, then the GPA range for the course is 3.0 to 3.4.

Recommended Mean in Elective Courses with One to Eight Students. For all elective courses with one to eight students, the mean GPA for the course *should* fall between 3.2 and 3.4, with a target GPA of 3.3, *subject to the same proviso* applicable in courses with nine to twenty-nine students.

> Note: The rules applicable to elective courses with nine to twenty-nine students also apply to elective courses with one to eight students. However, for classes in the latter category, the parameters set by the rules are merely *recommended* rather than mandatory.

Multiple Sections of the Same Course. When a professor teaches multiple sections of the same course and administers the same assessment tool (or tools) in each section—e.g., the same final exam—the professor *must* apply the mandatory curve across both sections as if the two sections together constitute a single course.[308]

---

308. While my proposal was pending before the law school's curriculum committee, we made one change to this language, adding the following: "There are five courses excepted *in part* from this policy. If the professor is teaching both the day and night sections of Property I, Property II, Civil Procedure I, Civil Procedure II, or Legal Profession, the professor need not treat the day and night sections as if they are a single course." The reason for the change was that the sections of these courses are not all filled with students at the same stage of their legal education. For example, Property I is a first-year course in our day program and a second-year course in our night program.

Minimum Required GPA to Remain in Good Standing. To remain in good standing, a student must maintain a GPA of at least 2.70—i.e., the student must maintain a B- average.[309]

Grades Below 2.7. All grades below 2.7 reflect performance that falls below minimum competency. However, students shall receive credit for the course if they earn a 2.6 to 2.0. Grades from 1.9 to 1.5 constitute "failing grades." The student will not receive credit for the course. If the course is required, the student must retake the class.

Pass/Fail Courses. Courses taught on a pass/fail basis are exempt from the mandatory and recommended means.

Enforcement Mechanism. If a professor turns in non-complying grades, the grades are returned to the professor. The professor must then re-curve the class. If the professor chooses not to comply, the Associate Dean for Academic Affairs shall act in his or her discretion to bring the grades into compliance, such as by setting the curve based on the professor's raw scores.

> Note: Turning in non-complying grades constitutes failure to turn in grades for purposes of the Late Grade Policy.

Transcript and Website Notation.[310] A short description of the law school grading system and curve shall be conspicuously placed at the bottom of each page of all official transcripts. The full grading system will be added to the academic rules and thus will be available on the website. However, to assist employers and others, the grading system shall be placed in a prominent location on the website. A link to that location will be included in the transcript notation.

> *Transcript Language*: The law school uses number grades ranging from 4.0 to 1.5. Every grade to the tenth of a point within this range may be awarded (i.e., 4.0, 3.9, 3.8, etc.). The equivalent letter values are as follows: A= 4.0, A- = 3.7, B+ = 3.3, B = 3.0, B- = 2.7, C+ = 2.3, C = 2.0, and F = 1.9 to 1.5. The law school also utilizes a mandatory curve in all courses with more than 8 students. Under the curve, the mean (or average) grade must fall between 3.2 and 3.4 with a target mean of 3.3. For further information on the law school grading scale and mandatory curve, please see www.ualr.edu/.

* * *

309. B- is typically valued at 2.7 or 2.67 on a 4.0 scale.

310. I added this provision while my proposal was pending before the law school's curriculum committee. I had always intended to include something like this, but I did not work on the specifics until well after writing my initial proposal.

A slightly modified version of this proposal was adopted by the Bowen Law School faculty (18 to 3) in April, 2011. The proposal was approved with three changes. Two of the modifications were very minor. First, clinic courses with more than eight students were shifted from the mandatory curve to the recommended curve that applies to electives with eight or fewer pupils. At present, there are no clinic courses at Bowen with more than eight students graded on a non-pass/fail basis. So this change had no substantive effect under our existing curriculum. Second, *required* courses with eight or fewer students were shifted from the mandatory curve to the recommended curve that applies to electives with eight or fewer pupils. Bowen currently has only one course that fits into this category—a required skills course.

The third change was more significant. The scale was shifted downward—the permissible grade range was moved from 3.2 to 3.4 with a target mean of 3.3 down to 2.9 to 3.1 with a target mean of 3.0. As part of that change, the good standing GPA was lowered from 2.7 to 2.3, the lowest credit grade was dropped from 2.0 to 1.7, and the lowest possible grade was reduced from 1.5 to 1.2. While this change was substantively important, I labeled the change "relatively minor" in Part I and at the start of this appendix because, for purposes of *this* article, it *was* minor. Unlike the first two alterations, this change did not implicate the mandatory curve *qua* curve. It concerned the *level* of grades we award as an institution, not *disparities* in grades among our faculty or other policies significantly implicated by grade normalization.[311] In a subsequent article, I will be presenting the case for setting B-/2.7 as the good standing GPA and for setting the mean grade at roughly the B+/3.3 level.

311. *See* Downs & Levit, *supra* note 9, at 854 ("Suggesting that grades fall within certain ranges [(grade normalization)] is an entirely different matter than suggesting that higher grades be given [(grade inflation)]."). Put differently, grade normalization is about *equalizing* the pot of grade wealth that each teacher may allocate to his or her students; grade inflation and deflation are concerned with the *size* of the pot of grade wealth each instructor is given.