Real or fake: Is student learning related to grades?

Autumn N. Wild* and Joseph M. Kush†

Department of Graduate Psychology, James Madison University
Center for Assessment & Research Studies, James Madison University

June 21, 2024

Abstract

While graduation rates have increased since the 1990s, modern students are spending less time on academic tasks, ultimately entering the workforce without the skills necessary to succeed. One explanation of this phenomenon is that of grade inflation, in which students' course-assigned grades are higher than what their performance warrants. Although recent literature has attempted to quantify the prevalence of grade inflation, many have relied on cohort-level data or other indirect measures (e.g., SAT). Utilizing pre/post scores on a standardized assessment of content knowledge, the current study examined the relationship between GPA and student learning. Results indicated that while improvement in knowledge was significantly related to grades, this relationship was not of practical significance. Implications for additional research are discussed in-depth.

Keywords: grade inflation; student learning

Institutions of postsecondary education in the United States have seen graduation rates continually increasing since the 1990s (Denning, Eide, Mumford, Patterson, and Warnick, 2022b; Irwin et al., 2023), providing an optimistic view of the state of education. Upon graduation, college students are expected to have training in various skills including reading, writing, and critical thinking. Yet these skills have been found lacking in graduates, leading to young professionals entering the workforce without the qualifications necessary to succeed (U.S. Department of Education, 2007). More recent research has found that as compared to prior cohorts, modern college students spend less time on academic tasks (Babcock and Marks, 2011) yet receive higher grades (Kostal, Kuncel, and Sackett, 2016; Rojstaczer and Healy, 2012). The idea that graduation rates are increasing while students are simultaneously spending less time on task and lacking the necessary qualifications for their profession raises serious concerns regarding the state of higher education.

*Correspondence concerning this article should be addressed to Autumn Wild, James Madison University, 298 Port Republic Road, Harrisonburg, VA 22807. Email: wildan@jmu.edu

†Portions of this manuscript were presented at the International Conference on Assessing Quality in Higher Education, Berlin, Germany, June 2024

One common explanation of this phenomenon is *grade inflation*, whereby students receive grades higher than what their actual performance warrants. While researchers have hypothesized why grade inflation may be occurring (Denning, Eide, Mumford, Patterson, and Warnick, 2022a; 2022b; Kostal et al., 2016), many have noted the challenges in attempting to disentangle warranted versus inflated grades (Jephcote, Medland, and Lygo-Baker, 2021; Kostal et al., 2016). Though difficult to implement, one of the most straightforward solutions to answering this question is to administer a standardized, external assessment of student knowledge that corresponds to increases in performance or learning.

The current study directly investigates the issue of grade inflation at a particular university by examining the relationship between learning and grades. Utilizing pre/post scores from a university-wide standardized assessment, changes in content knowledge were compared with assigned grades. To the best of our knowledge, this paper is the first to investigate grade inflation by leveraging individual-level pre/post scores on a direct measure of student learning to compare with course grades.

1 What is Grade Inflation?

Denning et al. (2022a; 2022a) argued that based on expected trends in various student and institution characteristics (e.g., less time spent studying, fewer available instructional resources) college graduation rates should be *decreasing*. However, the opposite has been observed. Denning et al. propose that this shift is due to a relaxation of, "standards for degree receipt" (Denning et al., 2022b, p. 10) - otherwise referred to as grade inflation. Schutz, Drake, Lessner, and Hughes (2015) define grade inflation as, "... an increase in grades awarded over time unwarranted by corresponding student achievement" (p. 181). Key to this definition is the lack of related student achievement; students may be receiving 'inflated' grades that are not a result of increased learning or knowledge content.

With grade inflation comes fear that the integrity of higher education has been compromised (Kostal et al., 2016). When grade inflation is present, grades begin to show a ceiling effect with a larger number of students receiving grades of A, making it increasingly difficult to distinguish amongst those whose grades are valid versus inflated (Kamber and Biggs, 2004). When an institution's capability of identifying whether a student has or has not learned is affected, the institution is left unable to justify whether a student is prepared to be successful post-graduation. As a result, students may slip through the cracks, believing they are prepared to enter the workforce when in reality they are not (Chowdhury, 2018).

2 Why Might Grade Inflation Happen?

A growing body of literature has focused on identifying potential causes of grade inflation. One commonly offered theme is the presence of course/professor evaluations (Chowdhury, 2018; O'Halloran and Gordon, 2014). Student evaluations of teaching (SETs) have implications for instructors (e.g., tenure, promotion) (Benton and Ryalls, 2016; Clayson, 2022; Park and Cho, 2023), as they provide an assessment of the quality of instruction. However, while SETs may be useful in gauging feedback from students regarding their educational experience, researchers have noted many drawbacks with their use in practice. Recent literature has noted that SETs may be reflective of opinions beyond the instruction the student has received (Clayson, 2022; Curby, McKnight, Alexander, and Erchov, 2019; Stroebe, 2020). For example, students may receive a poor grade in the course or on a particular assignment and as a result evaluate the professor more harshly. To avoid this negative consequence, professors may feel inclined to boost student grades despite poor performance in the course, thus

introducing grade inflation (Stroebe, 2020).

Other causes of grade inflation involve what Schutz et al. (2015) described as sympathy for the personal circumstance of the student. Finefter-Rosenbluh and Levinson (2015) detail the 'ethic of care' involved in an instructor-student relationship, suggesting that teachers may inflate grades to help alleviate the psychological stress of students who wish to perform well. Beyond these internal pressures, students also face societal and parental pressures to get good grades (Chowdhury, 2018). Instructors may grade more leniently as a sympathetic response to such pressures.

Schutz et al. (2015) also list maintaining financial aid as a reason for why grade inflation might occur. Students are often required to maintain certain grades in order to continue receiving financial support (e.g., Pell Grants; Schudde and Scott-Clayton, 2016); thus, instructors may boost grades to keep students in academic good standing (Tucker and Courts, 2010). Additionally, inflating grades can be a mechanism to avoid issues with students and parents. O'Halloran and Gordon (2014) note that conflict avoidance incentivizes grade inflation, as it can make the instructor's (and the student's) life easier.

Regardless of the reason behind why grade inflation occurs, the issue remains that once inflated, grades are no longer a valid indicator of knowledge. The question then becomes, how do we know if grade inflation is happening?

3 How Do We Detect Grade Inflation?

Using a variety of different methods, previous studies have made attempts to discern the degree to which grade inflation may be occurring at institutions of higher education. One common approach is to demonstrate relations (or lack thereof) between grades and achievement. For example, as part of their analyses, Baird, Carter, and Roos (2019) assessed the correlation between academic preparedness variables (e.g., high-school GPA, SAT score) and college academic performance among a sample of students from nonprofit and for-profit institutions. Further, multivariate analyses were utilized to examine the predictive utility of college preparedness variables in both college GPA and degree attainment. Baird et al. (2019) found that academic preparedness variables had predictive utility for collegiate success in public institutions, but little predictive power for for-profit institutions, leading to the conclusion that grade inflation was likely occurring in their sample of for-profit institutions.

A similar approach was taken by Paskausky and Simonelli (2014) who explored the relationship between written final exams and faculty assigned grades in a sample of undergraduate students from a private nursing program. Results indicated a weak to moderate correlation between course grades and exam scores ($r = 0.357$). The authors went on to further propose a new "clinical grade discrepancy measure", defined as the difference between faculty assigned grades and scores on a licensure-style exam. Findings showed that the vast majority of students (98%) received grades higher than their licensure-style exam score, leading to the conclusion that grade inflation was present.

Although not necessarily a method of detecting grade inflation, it is common for colleges and universities to examine graduation rates across time in tandem with GPA. Seeing increases in both measures across time, universities often conclude (incorrectly) that the curriculum is successful in fostering student learning. Without a direct assessment of student achievement or learning, it is impossible to determine the degree to which grade inflation may be occurring. Perhaps the closest to achieving this, Denning et al. (2022a) describe their "ideal test" for grade inflation. Drawing on course grades and exam scores for cohorts from 2001 through 2012, the authors found that as time went on, final exam scores remained relatively unchanged, while grades increased.

Reconsidering the definition of grade inflation provided by Schutz et al. (2015, p. 181: "an increase in grades awarded over time unwarranted by corresponding student achievement"), it is clear Denning et al. (2022a) were able to exactly demonstrate increases in grades unwarranted by student achievement. Yet there remain issues with this conceptualization; namely, that an inflation in grades can only be seen across time. By comparing distinct cohorts, researchers are unable to determine the degree to which any single grade or cohort may be inflated. For example, perhaps grades for the 2001 cohort were already inflated. Oppositely, it is possible the grades observed for the 2009 cohort were not inflated, but were in fact reflective of students' achievement that semester. Without some measure of student learning administered to the same group of students across multiple timepoints, arguments of grade inflation fall short.

Incongruence in methodologies has led to ambiguity in determining the extent to which grade inflation is occurring. Further, without pre/post measures of knowledge administered to the same group of students as a benchmark, previous researchers have been unable to fully determine if grade inflation is present for any given student. The need for answers regarding grade inflation has been further emphasized due to the COVID-19 pandemic, as course instruction and grading practices were necessarily adjusted to suit the sudden shift to an online education format (Park and Cho, 2023). While the studies discussed previously have all explored grade inflation, none make the direct tie to a measure of student learning, administered to the same group of students across multiple time points, and a measure of academic achievement (e.g., GPA). The current work aimed to make this connection.

4 The Present Study

The current study aimed to investigate the issue of grade inflation by exploring the relationship between GPA and pre/post scores on a standardized measure of student content knowledge. We hypothesized that grade inflation would be present, as evidenced by a weak relationship between changes in standardized test scores and GPA (i.e., course-assigned grades would not be related to increases in knowledge). This approach expands upon the limitations of prior research which typically examined differences in scores from different measures of student competency (e.g., SAT and GPA), or from the same measure of learning administered to different cohorts of students (e.g., changes in test scores between seniors in 2023 and seniors in 2024). To truly ascertain whether grades are connected to learning, it is necessary to administer the same measure to the same group of students at multiple timepoints (before and after receiving curriculum). Though this approach may be relatively simple in nature, most institutions of higher education likely do not have access to such widespread standardized assessment data. Thus, the present study is the first to seek evidence of grade inflation by measuring the relationship between student learning and academic success.

5 Method

5.1 Data

Data for the current study were obtained from a Mid-Atlantic R2 university. Students at this university are assessed on general education outcomes at two time points in their undergraduate career. First, incoming first-year students complete assessments prior to their first fall semester to ascertain knowledge before experiencing general education programming. Students are then re-assessed once they have completed 45-70 credit hours (typically in the spring semester of their sophomore year). Test

scores before and after experiencing curriculum can then be matched, allowing for investigations into questions of growth or improvement (i.e., as compared to baseline, do students show an improvement in some learning area) rather than competency alone. The ability to demonstrate growth over time is fundamental to understanding whether learning has occurred – what we theorize underlies grades.

5.2 NW9 – A Measure of Student Learning

Student learning was assessed via the Natural World Test, Version 9 (NW9). The NW9 is a 66-item multiple-choice exam designed to assess quantitative and scientific reasoning. The NW9 is theoretically mapped to four requirements of the institution's general education curriculum, including: 1) quantitative reasoning (mathematics and statistics courses), 2) physical principles (physics courses), 3) natural systems (science courses), and 4) lab experiences. A previous content alignment successfully mapped every item on the test to a given objective, with the test demonstrating sufficient reliability (α = 0.80; Sundre, Thelk, and Wigtil, 2008). Based on these strong properties, scores on the NW9 have been used by faculty in the mathematics and science departments to inform curriculum and improve instruction.

5.3 Procedures & Sample

Each year, the NW9 is administered to a random sample of first-year students in the fall, then to another random sample of second-year students in the spring. NW9 scores and demographic data were obtained from fall 2008 to spring 2020. For each student, an NW9 change score was calculated as the difference between pre and post test scores (e.g., a student with a fall score of 50% and spring score of 65% would have a change score of +15%). Only students with NW9 scores observed in both the fall and spring were retained. To avoid relying on students' overall GPA (which may be comprised of course grades across a wide range of curricula), an NW9-specific GPA was calculated using grades from courses only pertaining to scientific and quantitative reasoning. Specifically, letter grades were converted to numeric values, summed, and divided by the number of courses the student had taken pertaining to the NW9. For example, a student who completed two relevant classes and received an A- (3.7) and a B+ (3.3) would receive a weighted average GPA of 3.5, while a student who completed three relevant classes and received an A- (3.7), B+ (3.3), and B (3.0) would receive a weighted average GPA of 3.33. Students who had not taken any courses related to the NW9 at the spring assessment time point were not included in the study (as any changes in scores across time would not be related to relevant curricula). Fall and spring data were then merged with data from the university's Office of Institutional Research to tie in additional demographic information including sex (female, male) and race (Indigenous, Pacific, Hispanic, Black, Unspecified, Asian, Multi-Race, and White). A total of N = 8,420 students were included in the present study. Descriptive statistics for the analytic sample are shown in Table 1. We examined the extent to which GPA was associated with NW9 change scores using the following linear regression model:

$$GPA_i = \beta_0 + \beta_1 sex_i + \beta_2 race_i + \beta_3 pre_i + \delta change_i, \quad (1)$$

in which GPA_i is the NW9-specific GPA for student i, sex_i is their sex (0=female, 1=male), $race_i$ is a set of dichotomously coded racial identity variables (e.g., Asian = 0/1, Black = 0/1, etc.) with White set as the reference group, pre_i is the pre-test NW9 score from the fall semester, and $change_i$ represents the difference in NW9 scores across time (calculated as $NW9spring$ – $NW9fall$), with δ representing the coefficient of interest. Importantly, by including fall NW9 scores, the model is able

Table 1
Descriptive Statistics of Sample

		Female		Male		Total	
		N	%	*N*	%	*N*	%
Race							
	Indigenous	3	0.05	6	0.10	9	0.15
	Pacific	11	0.18	5	0.08	16	0.26
	Hispanic	44	0.71	28	0.45	72	1.17
	Black	129	2.09	55	0.89	184	2.99
	Unspecified	144	2.34	79	1.28	223	3.62
	Asian	186	3.02	110	1.79	296	4.81
	Multi-Race	229	3.72	130	2.11	359	5.83
	White	3237	52.55	1764	28.64	5001	81.19
		M	*SD*	*M*	*SD*	*M*	*SD*
Fall NW9		68.43	10.44	72.42	10.91	69.84	10.78
NW9 Δ		4.77	9.49	4.16	10.48	4.56	9.86
GPA		2.17	0.99	2.04	0.99	2.12	0.99

to account for individuals who scored high at the initial time point and thus had little to no room for change from fall to spring. All analyses were conducted in R version 4.3.1 (R Core Team, 2024).

6 Results

Prior to conducting analyses, individuals with missing data were removed, resulting in an analytic sample of $N = 6{,}160$ students. Other inspections, including homoscedasticity and residual diagnostics, revealed no issues. Results from the final regression model are provided in Table 2. The overall model was significant, $F(9, 6150) = 6.98$, $p < .001$. Two covariates were significant, including sex ($\beta = -0.160$, $p < .001$), and fall NW9 score ($\beta = 0.007$, $p < .001$). That is, males had significantly lower GPAs than females, while those who scored higher on the NW9 pretest had significantly higher GPAs. Notably, there were no racial differences regarding GPA. Finally, the focal variable NW9 change score was found to be significantly related to GPA ($\beta = 0.004$, $p < .01$), indicating that those who showed greater improvement had significantly higher GPAs (i.e., student learning is related to GPA).

To understand the practical significance of these findings, measures of effect size were further calculated. First, the linear regression model had an $R^2 = 0.01$, indicating that approximately 1% of the variance in GPA was accounted for in the model. That is, 99% of the variation in GPA is due to something other than improvement in NW9 scores (and the other covariates). Likewise, NW9 change score had a Cohen's $d = 0.04$, indicating no meaningful relationship with GPA (regardless of statistical significance).

7 Discussion

7.1 Summary of Key Findings

As mentioned previously, universities are often unable to administer wide-scale standardized assessments in a pre/post design, allowing for questions into student learning or growth across time. As such, previous researchers investigating the presence of grade inflation have instead focused on measures of competency across different cohorts (see Denning et al., 2022a; 2022b; Paskausky and Simonelli, 2014). The current study is one of the first to implement a true longitudinal measure of student learning in

exploring whether grade inflation is present, utilizing scores from the same assessment given to the same group of students before and after receiving curriculum.

Results from our study demonstrated that student learning was statistically significantly related to GPA, providing preliminary evidence that grade inflation may not be occurring at the Mid-Atlantic university in question. These findings have important implications for those in pursuit of investigating grade inflation, which are discussed below.

7.2 Implications for Research and Future Practice

7.2.1 *Low Effect Size*

Though the statistical model containing fall NW9 score, NW9 change score, sex, and racial identity was statistically significant, the model failed to be practically meaningful. The resulting R^2 for the model was 0.01, indicating that the model accounted for approximately 1% of the variance in GPA. This in turn limits the credibility of the inferences one can make from our results. Though statistically it appears grade inflation is not occurring, the extreme lack of practical effect suggests that student learning (i.e., NW9 change score) is not meaningfully related to a student's GPA. Given the context of the present study, the implications for this are extensive. We cannot definitively conclude that grade inflation is or is not occurring, given the conflicting statistical vs. practical significance. As a result, the question of whether grades are valid at the Mid-Atlantic university utilized in the present study remains unanswered. While effect sizes are clearly out of the control of the researcher, future studies demonstrating large effect sizes using a similar pre/post design with a single assessment would be able to more fully determine whether or not grade inflation is occurring. The lack of practical significance in the current study suggests, if anything, that grade inflation is likely occurring. While the current study used scores on a highly valid and reliable measure of quantitative and scientific reasoning, future studies should consider exploring different assessments in conjunction with relevant course-assigned grades.

7.2.2 *Competency vs. Growth – What Matters?*

Throughout this paper, we argue that learning is presumed to underlie grades, and that in order to measure learning, changes in knowledge across time are required. However, others have argued that competency alone demonstrates ability and subsequent grades (see Lipnevich, Guskey, Murano,

Table 2
Multiple Regression Results

Variable	Est.	S.E.
Intercept	1.681***	0.092
Indigenous	0.067	0.135
Pacific	0.089	0.144
Hispanic	−0.005	0.064
Black	−0.102	0.065
Unspecified	−0.037	0.067
Asian	0.057	0.051
Sex	−0.160***	0.027
Fall NW9 Score	0.007***	0.001
NW9 Change Score	0.004***	0.001

Note: *$p < .05$. **$p < .01$. ***$p < .001$.
'White' used as reference group.

and Smith, 2020's discussion of 'product', 'process', and 'progress' criteria). There very well may be instances in which students demonstrate a high level of content knowledge prior to receiving any instruction. In such cases, students may show no change across time, leading to the incorrect conclusion that their grades are inflated. In fact, such students may have perfectly valid grades, reflecting a mastery of the content knowledge required for the course. The present study was able to account for this in part by including a baseline measure of competency. Results indicated this pre-test score was significantly, positively related to GPA, suggesting that, on average, those who scored higher at baseline also had higher GPAs (an argument against grade inflation). However, this finding was of small practical significance ($d = 0.04$, an argument for grade inflation). Thus, regardless of viewpoint (growth vs. competency) the model failed to carry any substantive meaning.

7.2.3 *A Unique Opportunity*

Before investigations of grade inflation can take place, an external, standardized assessment must be administered to a single cohort of students before and after experiencing programming. To the best of our knowledge, this is the first paper to implement such a methodology in investigating grade inflation. Universities with such assessment systems in place can explore if assigned grades are valid by comparing with changes in standardized test scores across time. The Mid-Atlantic university utilized in this study was in the unique position of being able to examine student growth with respect to general education learning outcomes. Although the results of the present study were somewhat ambiguous and ultimately inconclusive, the current study helps to establish a new standard for investigations of grade inflation. This work should serve to encourage institutions of higher education to build such assessment systems.

References

Babcock, P., & Marks, M. (2011). The falling time cost of college: Evidence from half a century of time use data. *Review of Economics and Statistics*, *93*(2), 468–478. https://doi.org/10.1162/REST_a_00093.

Baird, A. F., Carter, J. S., & Roos, J. M. (2019). Seeking evidence of grade inflation at for-profit colleges and universities. *Sociological Focus*, *52*(4), 343–358. https://doi.org/10.1080/00380237.2019.1668321.

Benton, S. L., & Ryalls, K. R. (2016). Challenging misconceptions about student ratings of instruction. idea paper# 58. *IDEA Center, Inc.* https://www.ideaedu.org/idea_papers/challenging-misconceptions-about-student-ratings-of-instruction/.

Chowdhury, F. (2018). Grade inflation: Causes, consequences and cure. *Journal of Education and Learning*, *7*(6), 86–92. https://doi.org/10.5539/jel.v7n6p86.

Clayson, D. (2022). The student evaluation of teaching and likability: What the evaluations actually measure. *Assessment & Evaluation in Higher Education*, *47*(2), 313–326. https://doi.org/10.1080/02602938.2021.1909702.

Curby, T., McKnight, P., Alexander, L., & Erchov, S. (2019). Sources of variance in end-of-course student evaluations. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2019.1607249.

Denning, J. T., Eide, E. R., Mumford, K. J., Patterson, R. W., & Warnick, M. (2022a). Lower bars, higher college gpas. *Education Next*, *22*(1), 56–62. https://www.educationnext.org/lower-bars-higher-college-gpas-how-grade-inflation-boosting-college-graduation-rates/.

Denning, J. T., Eide, E. R., Mumford, K. J., Patterson, R. W., & Warnick, M. (2022b). Why have college completion rates increased? *American Economic Journal: Applied Economics*, *14*(3), 1–29. https://doi.org/10.1257/app.20200525.

Finefter-Rosenbluh, I., & Levinson, M. (2015). What is wrong with grade inflation (if anything)? *Philosophical Inquiry in Education*, *23*(1), 3–21. https://doi.org/10.7202/1070362ar.

Irwin, V., Wang, K., Tezil, T., Zhang, J., Filbey, A., Jung, J., ... Parker, S. (2023). Report on the condition of education 2023. *National Center for Education Statistics*. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2023144.

Jephcote, C., Medland, E., & Lygo-Baker, S. (2021). Grade inflation versus grade improvement: Are our students getting more intelligent? *Assessment & Evaluation in Higher Education*, *46*(4), 547–571. https://doi.org/10.1080/02602938.2020.1795617.

Kamber, R., & Biggs, M. (2004). Grade inflation: Metaphor and reality. *Journal of Education*, *184*(1), 31–37. https://doi.org/10.1177/002205740418400103.

Kostal, J. W., Kuncel, N. R., & Sackett, P. R. (2016). Grade inflation marches on: Grade increases from the 1990s to 2000s. *Educational Measurement: Issues and Practice*, *35*(1), 11–20. https://doi.org/10.1111/emip.12077.

Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020). What do grades mean? variation in grading criteria in american college and university courses. *Assessment in education: Principles, policy & practice*, *27*(5), 480–500. https://doi.org/10.1080/0969594X.2020.1799190.

O'Halloran, K. C., & Gordon, M. E. (2014). A synergistic approach to turning the tide of grade inflation. *Higher Education*, *68*, 1005–1023. https://doi.org/10.1007/s10734-014-9758-5.

Park, B., & Cho, J. (2023). How does grade inflation affect student evaluation of teaching? *Assessment & Evaluation in Higher Education*, *48*(5), 723–735. https://doi.org/10.1080/02602938.2022.2126429.

Paskausky, A. L., & Simonelli, M. C. (2014). Measuring grade inflation: A clinical grade discrepancy score. *Nurse Education in Practice*, *14*(4), 374–379. https://doi.org/10.1016/j.nepr.2014.01.011.

R Core Team. (2024). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rojstaczer, S., & Healy, C. (2012). Where a is ordinary: The evolution of american college and university grading, 1940-2009. *Teachers College Record*, *114*(7), 1–23. https://doi.org/10.1177/016146811211400707.

Schudde, L., & Scott-Clayton, J. (2016). Pell grants as performance-based scholarships? an examination of satisfactory academic progress requirements in the nation's largest need-based aid program. *Research in Higher Education*, *57*, 943–967. https://doi.org/10.1007/s11162-016-9413-3.

Schutz, K. R., Drake, B. M., Lessner, J., & Hughes, G. F. (2015). A comparison of community college full-time and adjunct faculties' perceptions of factors associated with grade inflation. *The Journal of Continuing Higher Education*, *63*(3), 180–192. https://doi.org/10.1080/07377363.2015.1085951.

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, *42*(4), 276–294. https://doi.org/10.1080/01973533.2020.1756817.

Sundre, D., Thelk, A., & Wigtil, C. (2008). The natural world test, version 9: A measure of quantitative and scientific reasoning, test manual. *Unpublished manuscript, James Madison University, Harrisonburg, Virginia*. https://www.madisonassessment.com/uploads/NW-9%20Manual%202008.pdf.

Tucker, J., & Courts, B. (2010). Grade inflation in the college classroom. *Foresight*, *12*(1), 45–53. https://doi.org/10.1108/14636681011020155.

U.S. Department of Education. (2007). *A test of leadership: Charting the future of us higher education.* https://doi.org/10.1086/512957.