

# Undergraduate Grading Practices of International and Domestic Faculty: Evidence From Three Large U.S. Public Universities

Trang Pham 

Singapore University of Social Sciences

Stephanie Potochnick

University of North Carolina at Charlotte

*Given the prominence of international instructors in higher education, understanding their grading practices is essential for informing college grading debates. This first large-scale assessment of undergraduate grading practices highlights how different demographic, classroom and departmental factors shape international instructors' grading behaviors. Using a unique dataset of over 2,000 randomly selected instructors from three public universities, we examine (a) whether undergraduate-level grading practices differ between domestic and international instructors, (b) what factors contribute to the differences, and (c) whether the differences vary across key subgroups. We find that international instructors grade lower than domestic instructors—about 35% of a standard deviation lower on average. Part of this gap is explained by the concentration of international instructors in particular departments. International instructor grading practices differ across regions of origin, prior U.S. higher education experience, gender, and race. Our results provide insights into U.S. college grading debates and supporting the international instructor workforce.*

**Keywords:** *grading behaviors, grading practices, international faculty grading behaviors, higher education grading practices*

Debates about grade inflation and grade equity in U.S. colleges have intensified the focus on instructor grading practices, highlighting the significant impact that grades have on higher education institutions, students, and instructors. Recent controversies surrounding the firing of college instructors over grading practices (e.g., Quinn, 2023; Saul, 2022) demonstrate the challenges that academic institutions face as they struggle to address both rampant grade inflation (Denning et al., 2022; Rojstaczer & Healy, 2012) and discriminatory and unfair grading practices (E. S. Park et al., 2021; J. J. Park et al., 2020). For students, the impact of grades extends well beyond the classroom. Grades determine their eligibility for scholarships, educational opportunities (e.g., clubs/sports, internships), graduation, and can strongly shape future career prospects. For instructors, the grades that they assign can impact their job performance evaluations and opportunities for career advancement (Ginther & Kahn, 2006; Isely & Singh, 2005).

Given the broad importance of college grades, research has begun to examine how instructor characteristics and work-related pressures influence grading practices. To date,

research finds that instructor grading practices often differ based on the gender, race, and position rank (e.g., adjunct, pre-tenure) of instructors (Jewell & McPherson, 2012; Ran & Xu, 2019). These differences are partly due to differential work pressures (e.g., job security) and classroom contexts (e.g., class size) (Franz, 2010; Griffith & Sovero, 2021). But they also reflect differences in teaching philosophy, student-instructor race and gender match, and department-specific grading norms (Harbatkin, 2021; Jewell et al., 2013; Joshi et al., 2018; Koedel, 2011).

Though valuable, this research has yet to examine the grading practices of international instructors, a vital and culturally diverse segment of the U.S. higher education workforce. Although estimates of the size of the international instructor workforce differ based on the data source and definition used (nativity, citizenship, or visa status), all evidence suggests that international instructors make up a sizable share of U.S. instructors (Kim et al., 2012). According to the National Center for Education Statistics (NCES, 2021), in 2019, international faculty measured based on temporary visa status (i.e., nonresident alien) made up 6%



(50,628 positions) of all full-time faculty positions at U.S. higher education institutions. This rate is higher in the Harvard COACHE survey. The estimated share of international faculty, as measured by non-U.S. citizenship status, at 4-year research-intensive universities varies across disciplines and ranges from 8.9% in education to as high as 33% in physical sciences (Kim et al., 2012).

In addition to their prominent size, international instructors are culturally diverse, coming from roughly 200 countries and territories. Given that evidence suggests grading expectations and practices often differ across countries (Anglin & Meng, 2000; Cheng et al., 2018), research is needed to assess whether these same cross-country differences persist between domestic and international instructors in the U.S. higher educational system. Foreign-born faculty who obtained their undergraduate degrees outside the United States may have differing grading expectations and practices. That is because they are less likely to be familiar with the U.S. undergraduate higher educational system and, as suggested by prior research (Kim et al., 2011), may have different undergraduate cultural, social, and educational experiences that shape their academic experience. For instance, foreign-born instructors educated abroad are often distinct from their foreign-born, U.S.-educated and U.S.-born peers: They have higher work productivity but lower job satisfaction than both (Kim et al., 2011). Research, however, has yet to examine whether similar differences exist in grading practices.

Given the prominence of international instructors in higher education, understanding their grading practices is essential for informing broader college grading debates. This article initiates a pivotal exploration by assessing how U.S.-based international instructors' undergraduate-level grading practices compare to that of domestic instructors. We follow Kim et al. (2011, 2012), and define international status based on undergraduate degree country location for theoretical and practical reasons. Theoretically, we want to understand the grading practices of international instructors who are learning to adapt to the U.S. higher educational undergraduate grading system. Foreign-born instructors educated outside the undergraduate U.S. educational system, no matter their citizenship or visa status, are the group of instructors most likely to face unique U.S. undergraduate grading adaptation challenges due to their different undergraduate grading experiences and cultural norms. Practically, we focus on the country of undergraduate degree attainment to overcome data limitations. Because no current large-scale dataset (survey or administrative) simultaneously collects information on instructors' international status and grading practices, researchers have not been able to examine domestic and international instructor grading practices. We overcome this challenge by creating a unique, large-scale dataset that combines institutional administrative grading records with data from individual instructor CVs (*curricula vitae*)—from which country of undergraduate

degree attainment is the strongest proxy available to identify international versus domestic status.

Building on different literatures related to U.S. grading practices, this article develops a framework for understanding the different factors shaping instructors' grading practices. We then apply this framework to international instructors by incorporating literature on immigrant adaptation and assimilation to examine how their differing backgrounds, cultural perspectives, and contextual experiences within the U.S. higher educational system may lead to differing grading behaviors. Using a unique large-scale database of U.S. instructors across three major U.S. universities, we assess (a) whether undergraduate-level grading practices differ between domestic and international instructors; (b) what factors, including instructor characteristics, classroom and departmental contexts, and immigrant background, contribute to international-domestic instructor grading differences; and (c) whether international-domestic instructor grading differences differ across key subgroups (e.g., gender, position rank). The results of this study reveal key differences in the grading practices of domestic and international instructors and identify avenues for future research.

### **Current Evidence on Instructor Grading Practices**

Because grades are a standard form of assessment used across educational systems, evidence of the grading practices of college-level instructors is informed by studies at both the K–12 and higher educational system levels. Grading is a complex process that requires an instructor's objective and subjective judgment about a body of students' work and performance to communicate with wider audiences, such as other institutions, students' families, or recruiters (Brookhart et al., 2016; Schneider & Hutt, 2014). Grades are largely presumed to signal students' academic abilities (Pattison et al., 2013). However, grade decisions can also reveal how instructors navigate their teaching environments, interpreting external norms, practices, and pressures related to their individual characteristics, position rank, classroom structure, and departmental and institutional contexts (see e.g., Ehrenberg & Zhang, 2005; Hermanowicz & Woodring, 2019; Jewell & McPherson, 2012; Kokkelenberg et al., 2008).

#### *Instructor Characteristics and Position Rank*

Extant research finds that instructor grading practices differ based on the instructors' gender, race, and position rank. At the college level, research has largely focused on examining grade inflation and how discrimination and job insecurity pressures lead to differential grading practices by gender and position rank. Using two decades of data from one large public university, Jewell and McPherson (2012) found that female instructors grade students significantly higher than

males. This difference exists because female instructors face higher levels of discrimination in student evaluations (Boring, 2017; Radchenko, 2020) and greater overall job insecurity (Ginther, 2001; Ginther & Kahn, 2006), both of which create pressure to grade higher. Instructors who award higher grades tend to have higher student evaluations (Isely & Singh, 2005; Kostal et al., 2016), and females are more likely to be retained if they award higher grades (Griffith & Sovero, 2021).

Research on instructor rank finds a similar work-pressure, job security grading incentive. This research consistently finds that adjunct, nontenured, and pre-tenure instructors award higher grades than tenured instructors (Kezim et al., 2005) and that these higher grades do not completely reflect better student quality (Ehrenberg & Zhang, 2005; Ran & Xu, 2019). The assumed reason for the higher grades is that instructors with less job security are pressured to give higher grades to improve their student evaluations and to reduce time dealing with student complaints, which can hinder work productivity (Franz, 2010).

Though the same heightened discrimination and job insecurity pressures affect non-White instructors (Arnold et al., 2016; Ward & Hall, 2022), there is limited research at the college level on how grading practices differ by instructor race/ethnicity. The only study that we are aware of found that grading practices differed by gender but not by race/ethnicity (White versus non-White instructors) (Jewell & McPherson, 2012). The authors caution, however, that due to data limitations, they could not distinguish between domestic and international non-White instructors, the latter of which may have different grading motivations and incentives.

Instead, research at the K–12 level provides insights into how grading practices of instructors with different racial/ethnic backgrounds may differ. These studies focus on how a teacher's race/ethnicity may impact the quality of the student-teacher relationship and consistently find that students, particularly Black students, perform better on tests, grades, and behavior indicators (e.g., attendance) when classes are taught by an instructor of the same race/ethnicity (Downey & Pribesh, 2004; Harbatkin, 2021; Joshi et al., 2018; Kozlowski, 2015). Though there is debate as to why, two main assumptions are that non-White teachers are less likely to harbor negative racial/ethnic biases against non-White students and that they are more likely to serve as positive role models for them (Downey & Pribesh, 2004; Kozlowski, 2015). In turn, students are more engaged, and teachers are more effective. Given that colleges and universities are increasingly concerned about racial equity in student performance and grading practices (Bowman & Denson, 2022), these results, combined with Jewell and McPherson's (2012) study, highlight the need to examine how grading practices differ by the racial/ethnic background of the instructor, for both domestic and international instructors.

### *Classroom and Departmental Factors*

Other research on instructor grading practices focuses on how structural differences across classrooms/courses, particularly the role of class size, shape instructor grading practices. In grading, instructors consider what scholars often refer to as achievement (e.g., test scores) and nonachievement factors (e.g., class participation, effort, and motivation) (Brookhart, 2013; Brookhart et al., 2016; Cheng et al., 2020; Olsen & Buchanan, 2019).<sup>1</sup> Although some debates remain about whether and how class size affects course grades (Ake-Little et al., 2020), common views among researchers are that nonachievement factors are harder to track in large classes. Large class size adversely affects students' and professors' motivation and attitudes—nonachievement factors—toward the learning process. In smaller classes, instructors have a better understanding of students' efforts and are likely to account for effort when grading (Millea et al., 2018). Consequently, studies show that class grades decrease when class size increases (Diette & Raghav, 2015; Kokkelenberg et al., 2008) and that class grades are lower in introductory/entry-level courses than advanced/upper-level courses due to the former having larger class sizes (Bean, 1985).

Additionally, research indicates that grading norms and practices differ across fields of study and departments. Courses in the humanities, education, and, to a lesser extent, social sciences tend to have higher grades than courses in STEM (science, technology, engineering, and mathematics) and business fields (Butcher et al., 2014; Jewell & McPherson, 2012; Koedel, 2011; Rojstaczer & Healy, 2012). Though the exact reasons for these differences are unknown, potential explanations are that more quantitative disciplines have fewer subjective grading elements, whereas instructors in the humanities and similar disciplines are more likely to give credit for nonachievement factors (Hermanowicz & Woodring, 2019). Additionally, differences in teacher gender and racial/ethnic biases—which are particularly concerning in STEM—and the use of grading protocols—which have discriminatory effects—may lead to across discipline grading differences (E. S. Park et al., 2021; J. J. Park et al., 2020). And departmental size matters. Small departments may try to attract more students by inflating grades, while large departments may be more research-oriented and focus less on teaching effectiveness (Jewell et al., 2013).

### *Teaching Philosophy and Instructional Approaches*

Finally, research indicates that instructors' teaching philosophy and instructional approaches shape grading practices. Instructors differ on how much they weigh achievement and nonachievement factors, their consideration of external factors (e.g., family demands), their use of instructional approaches (e.g., lecturing vs. active learning activities), views on the purpose of assessments (e.g., hold students

accountable vs. “weed out” students), and inclusion of diversity into the classroom and curriculum—all of which have strong implications for student performance and instructor grading practices (E. S. Park et al., 2021).

### **International Instructors and Their Grading Practices**

The lack of research on international instructors’ grading practices is largely due to data limitations. There is no large-scale dataset that simultaneously collects information on an instructor’s international status and grading practices. Administrative datasets, such as university-published grades, provide the class-average grades an instructor assigns but do not include detailed instructor information. Conversely, large-scale surveys of instructors and doctorate degree-holders (e.g., National Study of Postsecondary Faculty, Survey of Earned Doctorates) have information to identify international status (e.g., citizenship status) but not instructor grades. Consequently, no large-scale study has examined how grading practices differ between domestic and international instructors. Appendix A details this data challenge further.

#### *Instructor, Department, and Classroom Factors*

Instead, research on international instructors’ characteristics and work context provides useful insights into how their differential demographic profiles, position ranks, and departmental backgrounds may lead to differing grading practices compared to domestic instructors. Available studies on international instructors highlight several consistent demographic and professional characteristics of international instructors, despite the fact that each study captures different time frames and U.S. higher education workforce segments (see Appendix A for further details). Overall, these studies indicate that international instructors are more likely to be male and Asian, with the top three origin countries being China, Korea, and India (Kim et al., 2011; Open Doors, 2022b). They are more likely to be in a research-intensive university (Marvasti, 2005), a STEM field (Lin et al., 2009), and a tenure-track, pre-tenure position (Espinosa et al., 2019).

The demographic profiles and position ranks of international instructors suggest that they may grade higher than domestic instructors, while their associated departmental and classroom contexts suggest the opposite. Because international instructors are more likely to be in a pre-tenure-track position, they may feel job insecurity pressures to inflate grades. Like female and racial/ethnic minority instructors, foreign-born instructors are discriminated against in student evaluations (Fan et al., 2019; Hamermesh & Parket, 2005), which may worsen this job insecurity, grade inflation incentive. Additionally, as a majority of international instructors are non-White, they may be more effective at teaching non-White students, leading to higher achievement and grades. Alternatively, given that the

majority of international instructors are in STEM fields, they may adhere to departmental norms and practices, and award lower grades than non-STEM instructors (Hermanowicz & Woodring, 2019; Koedel, 2011). The extent to which they do so will likely depend on the classroom context. But there is limited research on the types of classes that international instructors teach.

Consequently, we have no *a priori* expectation of whether international instructors will grade higher or lower than domestic instructors. We simply suspect that their grading practices will differ and that differences in demographic profile, instructor rank, and departmental and classroom contexts will explain some of this difference.

#### *Immigrant Background and Teaching Philosophy/ Instructional Approach*

International and domestic grading differences are likely to extend beyond just these factors. That is because international instructors are immigrants with their own culturally informed teaching philosophy and approaches. Importantly, how instructors allocate recognition for nonachievement factors in grading is not culturally unanimous. For instance, Cheng et al. (2018) found that Chinese secondary teachers equally weigh achievement (e.g., tests) and non-achievement (e.g., learning habits) factors when grading, while Canadian teachers focus mostly on achievement factors. Additionally, countries differ in terms of how much they encourage the freedom of inquiry by students (Macfarlane, 2012). International instructors from countries where academic freedom is regulated or oppressed may struggle to adapt to the U.S. higher educational system’s approach of encouraging student academic freedom and in-class discussion (Altbach, 2001, 2009). As a result, they may be less receptive to students’ ideas or confrontations, or vice versa, and grade students differently.

These and other cultural differences in grading practices, however, may lessen over time as international instructors adapt to the U.S. higher education system and cultural norms. As suggested by immigrant assimilation theories (Brown & Bean, 2006; Massey & Malone, 2003), upon arrival international instructors, like other immigrants, may face adaptation challenges that affect grading practices. For instance, it may take time for them to adjust to the U.S. standard grading system, which also requires a cultural understanding of what standards to apply when assigning grades (Witte, 2011). Additionally, international instructors may adjust their grading behaviors to be more similar to that of domestic instructors as a way to signal becoming an in-group member (Alba & Nee, 2003). Extant research finds that over time, immigrants often behave more like the U.S.-born population (e.g., speak English) (NAEMS, 2015).

However, cultural minorities may also resist assimilation if certain practices in the host culture are not aligned with their



values (Pennesi, 2016; Schultz, 2016). The practice of grade inflation, which some evidence suggests may be more serious in the United States (Rojstaczer & Healy, 2012; Rodriguez-Planas, 2022), is one potential example. International instructors may resist grade inflation pressures and maintain grading differences over time.

## Methods

### *Data*

To address data limitations, we created the first large-scale dataset that combines instructors' international status and assigned grades. Creating this dataset was a time- and labor-intensive process that required constructing and combining two different data sources: (a) publicly available course-grade data from university websites and (b) corresponding instructor characteristic data compiled from CVs. The latter required a manual search through each instructor's CV and website to identify key characteristics (e.g., international status). Because data collection was labor-intensive, we had to be decisive with our priorities. We pooled the universe of undergraduate course-grade data from three universities, then used this to select a random subsample of faculty within each university. Below, we describe this process further (see Appendix C for more details).

For our university choices, we selected three large R1, public, and flagship universities in the Midwest for several reasons. First, we focused on R1 institutions, given that international instructors are more likely to work at research-intensive universities (Marvasti, 2005). Second, we prioritized selecting three comparably similar (e.g., public, flagship, Midwest) institutions to increase internal validity. This minimizes the risk that unobserved institutional differences confound our international-domestic grading comparison. Finally, all three universities selected publish their course-grade data online, which was necessary to be included. These data were often not available (or complete) at smaller or private institutions. Appendix Table C-1 describes the key features of the three universities selected. Overall, each university offers a wide array of undergraduate degree programs (260+), has a large undergraduate population (between ~23,000 and 33,000), is majority White both in terms of undergraduates (between ~55% and 80%) and instructors (between ~60% and 70%), and has a sizable share of nonresident alien instructors (between ~5% and 15%).

From each school, we collected data on all courses offered during all semesters between 2011 to 2017. Data include class-average grades (on a 5-point scale), the full distribution of grades (percentage As, Bs, etc.), class level, student enrollment, department, semester, and instructor's full name. From this list, we eliminated all lab sections of a course (if lab and main course were graded separately) since the instructor of record may not be the primary day-to-day

instructor (e.g., labs led by graduate student teaching assistants). We also eliminated cotaught courses, given that each instructor is likely to have differing characteristics.

Given the size of the grade data (over 166,000 courses taught by 18,000 instructors), it was not feasible to build a corresponding dataset of all instructors across universities. Thus, we decided to create a random sample of instructors that would be representative of the entire population of instructors at each university. To do so, we used a conservative power calculation on the grade dataset and found that a sample size of ~3,000 instructors would be sufficient to detect a grading difference of at least 10%. Using a list of instructors derived from the grade data, we then randomly selected instructors, stratified by university (~1,000 per university), from this list to create our targeted sample.

We stratified by university and not department for two reasons. First, department structure differed across universities, making it difficult to stratify by department. For example, Public Policy was a school in one university, but in the Political Science Department in another. Second, though prior research suggests that international instructors are more likely to be in STEM fields (Kim et al., 2012; Marvasti, 2005), we had no way to confirm this for our focal universities or to know the distribution of international instructors by field and department at the sampling stage. Thus, we used simple random sampling at the university level (adjusted to ensure exceptionally large and small departments were not overrepresented; see Appendix C). This ensures that the final sample of international and domestic instructors that we observe is representative of all instructors at each of the three universities.

Then, using full names from the grade dataset, we conducted a manual search through each instructor's CV and website to collect information on instructors' qualifications, employment history, demographics (e.g., gender), and information to identify international instructor status (e.g., country where bachelor's degree was obtained). Because we omitted graduate student instructors ( $N=766$ ) and instructors who could not be found during the manual search ( $N=211$ ), the final dataset contains 2,023 instructors who taught 16,021 courses over 21 semesters. This sample size is sufficient to capture statistically meaningful international-domestic instructor grading differences.

### *Measures*

*Instructor Grades.* The key outcome of interest is class-average grade. Following prior research (Ake-Little et al., 2020; Butcher et al., 2014), we calculate the mean value of all students' grades in one class, measured on a 4.0-point scale (rounded to the hundredth).

*International Versus Domestic Instructor.* The key independent variable of interest is a binary indicator classifying

instructors as 1 = international or 0 = domestic instructors. As noted, we follow Kim et al. (2011, 2012) and classify instructors as international if they obtained their bachelor's degree outside the United States. We use a three-stage system to identify bachelor's degree country location. In Stage 1, we identified country location based on CVs and online profiles. For example, if an instructor reported that they received their bachelor's degree from China, we coded them as *international*, and China as their origin country. We coded 80% of instructors in stage one.

For the remaining 20%, we had to use additional methods to identify country location. In Stage 2, we classified instructors as international if they had clear research interests in a specific country and bore a common name of that country (e.g., named Paganini, a common Italian name, who published on Italian politics). We coded an additional 14% of instructors using these criteria. In Stage 3, we used instructors' last names or language spoken to classify international status. For instance, if they reported speaking a non-English language and had a corresponding origin last name (e.g., spoke Korean, last name Kim), we coded them as international. If they had a common American-based last name (e.g., Smith) and no international markers (e.g., language), we coded them as domestic. Six percent of the sample were coded using these criteria. We use all instructors (no matter their identification stage) in our analysis. Sensitivity checks excluding those identified in Stages 2 and 3 were qualitatively similar (see Appendix Table B-2).

Because CVs and online profiles rarely identify a person's nativity or citizenship status, our measure also captures a small share of U.S.-born instructors who completed their undergraduate degrees abroad. These instructors are not our primary focus of interest, but there is no systematic way to identify them. They, however, are likely to make-up an extremely small share. Best available estimates from the Institute of International Education indicate that, in 2022, ~40,000 U.S.-born individuals pursued an undergraduate degree abroad—a dramatically smaller number than the over 16.5 million undergraduates (the vast majority of whom are U.S.-born) pursuing a U.S. undergraduate degree (NCES, 2023; Project Atlas, 2022). Moreover, their inclusion in our “international” classification is likely to attenuate (i.e., underestimate) differences between “domestic” and “international” instructors due to induced measurement error.

*Instructor Characteristics and Position Rank.* Because grading practices often differ based on gender, race/ethnicity, and position rank, we create indicators for each. For gender and race/ethnicity, we create proxy indicators since these are not explicitly listed on CVs. Following prior research (Li & Koedel, 2017), we use a combination of visual inspection, last name origin (e.g., Hispanic surname), and other biographic information (e.g., country of degree attainment, employment history, self-reported citizenship status), to

classify individuals by race and ethnicity (White, Black, Asian, Hispanic, and other) and by gender (male and female). The “other” category for race and ethnicity includes (a) individuals we identify as aligning with a different race and ethnicity category based on U.S. Census standards (e.g., Native American, Pacific Islander, two or more races) and (b) individuals for whom we could not identify their race and or ethnicity (9.46% of the sample). For gender, we are not able to identify the gender of 1% of our sample. Because we do not want to overclaim inferences on this small unknown gender group in our statistical estimates, we combined this group with females (the reference category) to limit sample reduction. Results are robust after excluding observations with unknown or unidentifiable data on race, ethnicity and gender.

Though we would prefer to use self-identified indicators and recognize that this proxy approach is problematic—phenotypic characteristics do not identify race or ethnicity, both of which are products of long-standing power inequities, not biology (Laughter, 2018)—we contend that this is still an appropriate approach given data limitations. In the U.S. context, where race is a social construct, this method presents the social understanding of racial and gender groups rather than a representation of self (Leonardo, 2007). To ensure that we captured a broad social understanding of these groups, our coding team represented different self-identified race, ethnic, and gender categories (i.e., non-Hispanic White male, Asian female). Interrater reliabilities for race and gender were 95% and 98%, respectively. We also verified that race and gender shares in our sample matched with shares reported in the 2014 IPEDS (Integrated Postsecondary Education Data System) Research-I universities data sample and the Li and Koedel (2017) study.

To capture position and rank, we classify each instructor's highest reported university job title into three broad categories: (a) assistant professor, (b) tenured professor (includes associate, full, and emeritus/emerita), and (c) non-tenure instructor (include lecturers, adjunct, teaching specialists, visiting professors, postdoctoral fellows, and other unknown ranks).

*Departmental and Classroom Factors.* We create several indicators to capture departmental and classroom differences in grading practices (Bean, 1985; Kokkelenberg et al., 2008). For departmental differences, we capture unique differences across academic disciplines using three broad disciplinary categories—(a) STEM, (b) social sciences, and (c) arts, humanities, and other—based on the 2010 U.S. Department of Education's official Classification of Instructional Program taxonomy. We use this broad categorization because it is consistent with prior studies (e.g., Butcher et al., 2014) and departmental structures differed across universities. However, in sensitivity analysis, we add department-university fixed effects, which capture more specific disciplinary differences

and unobserved cultural differences specific to each department at each university.

For classroom differences, we measure classroom size using total student hours (divided by 100) and its squared term to capture nonlinearity. We measure course-level by classifying course numbers into one of three categories: introductory (1000 level or equivalent), intermediate (2000 & 3000 level or equivalent; reference category), and advanced (4000+ level or equivalent).

*Immigrant-Specific Background.* We use three measures to capture international instructors' diverse backgrounds. Using the country of bachelor's degree attainment, we classify international instructors into three broad regions of origin—Asia, Europe and Canada, and Other Regions—each of which has a sufficient sample size. We combine Canada with Europe (rather than Other Regions) given its historical connection with England. See Table C-1 for a list of the top 10 origin countries. To capture potential differences in U.S. higher educational system familiarity, we use two measures—U.S. graduate degree attainment, and U.S. academia work time—both of which follow similar logic to standard *time-based* acculturation measures used in the immigration literature (National Academies of Sciences, Engineering, and Medicine [NAEMS], 2015) but are specific to academia. U.S. graduate degree attainment (available for 96% of the sample) is coded as 1 if international instructors received a master's or doctorate degree from a U.S. institution, 0 if no U.S. graduate degree. We measure international (and domestic) instructors' work time (available for 97% of the sample) in the U.S. academy (semester units) from the time they received the highest degree (usually PhD), until the current semester. In a few cases, we use date of the first publication if degree date is not available. Overall, with this method, we assume that instructors start working in U.S. academia upon graduation, which induces measurement error but is the only consistent way to measure U.S. academic work-time. Such measurement error will attenuate estimates toward zero.

### Analysis

First, we use descriptive statistics to examine overall differences in the grading practices of international and domestic instructors, and to identify differences in instructor characteristics and departmental and classroom factors. Next, we use regression analysis to examine how these factors—instructor, departmental, and classroom—contribute to observed differences in international-domestic instructor grading. We use the following specification:

$$Y_{ijknt} = \beta_0 + \beta_1 I_i + \beta_2 X'_i + \beta_3 S'_{ijknt} + \beta_4 T'_i + \gamma_n + \theta_t + \mu_{ijknt},$$

where  $Y_{ijknt}$  is the class-average grade given by instructor  $i$  for course-section  $k$  in department discipline  $j$  in university

$n$  in semester  $t$ .  $I_i$  is the key variable of interest and identifies instructors' international status.  $X'_i$  is a vector of general instructor characteristics (race, gender, and position rank).  $S'_{ijknt}$  is a vector of classroom and department controls (student hours, class level, and academic discipline).  $T'_i$  is a vector of instructor  $I'$  immigration-specific background (U.S. worktime and graduate degrees).  $\epsilon_{ijknt}$  is the error term. To account for differences in grading across semesters and universities, we include semester fixed effects,  $\theta_t$ , and university fixed effects,  $\gamma_n$ . We cluster standard errors at the department level to account for multiple courses nested under one department. Results were similar if we clustered at the instructor-semester level.

We estimated four additive models to assess the unique and combined influence of different theoretical blocks of interest. In each model,  $\beta_1$  is the coefficient of interest. It captures the overall grading difference between international and domestic instructors. Model 1 is a minimally adjusted model that controls for semester and university fixed effects only. Models 2 through 4, respectively, add instructor, classroom and departmental, and immigrant-specific controls. Then, using our full model (Model 4), we add more detailed indicators to assess if international-domestic grading patterns differ based on international instructors' diverse origins (i.e., region/country of origin) and U.S. higher educational system familiarity (i.e., U.S. work time and U.S. graduate degrees). Lastly, we add interaction terms one at a time to our full model to assess whether international-domestic grading patterns differed by gender, race/ethnicity, rank, and discipline.

A multilevel model is not applicable in this case due to a violation of the uncorrelation assumption. A multilevel model assumes that the model levels are uncorrelated with its independent variables. For example, if students are distributed randomly within classes within a school, student's ability is not correlated with their class, that is, good students are scattered rather than gathered within one class. However, in our data, instructors are nested within a department discipline within a university, the key covariate (international status) is correlated with departments and fields (see Kim et al., 2012; Marvasti, 2005, and Appendix Table C-2). The uncorrelation assumption would be violated if we apply a multi-level or hierarchical model in our article.

### Results

#### *Overall Grading Practices and Characteristics of International and Domestic Instructors*

Table 1 presents the descriptive statistics of the instructor sample and their courses side by side. We find that international instructors make up 19% of the teaching workforce and teach 17% of the undergraduate course load. Overall, international instructors assign lower grades on average than domestic instructors (3.17 versus 3.30,  $p < .05$ ). The

TABLE 1  
Descriptive Statistics

	Instructor-Level Sample			Course-Level Sample		
	All Faculty	Domestic Instructors	International Instructors	All Courses	Courses Taught by Domestic Instructors	Courses Taught by International Instructors
Outcome variables						
Class-average grade				3.28 (0.46)	3.30 (0.46)	3.17 <sup>a</sup> (0.46)
Instructor characteristics						
Gender						
Male	0.55	0.54	0.57	0.60	0.60	0.56 <sup>a</sup>
Female	0.45	0.46	0.43	0.40	0.40	0.44 <sup>a</sup>
Race						
White (non-Hispanic)	0.72	0.81	0.40 <sup>a</sup>	0.76	0.82	0.45 <sup>a</sup>
Black	0.04	0.04	0.04	0.04	0.04	0.04
Asian	0.11	0.02	0.43 <sup>a</sup>	0.07	0.01	0.39 <sup>a</sup>
Hispanic	0.02	0.01	0.03 <sup>a</sup>	0.02	0.01	0.04 <sup>a</sup>
Other races	0.12	0.12	0.10 <sup>a</sup>	0.11	0.12	0.08 <sup>a</sup>
Position rank						
Assistant (pre-tenure) prof.	0.15	0.13	0.21 <sup>a</sup>	0.10	0.09	0.17 <sup>a</sup>
Tenured prof.	0.42	0.40	0.51 <sup>a</sup>	0.51	0.50	0.56 <sup>a</sup>
Nontenure prof.	0.43	0.47	0.29 <sup>a</sup>	0.39	0.42	0.27 <sup>a</sup>
Classroom and departmental contexts						
Departmental disciplines						
Social Sciences	0.45	0.46	0.40	0.45	0.45	0.43
STEM	0.30	0.26	0.44 <sup>a</sup>	0.31	0.29	0.40 <sup>a</sup>
Arts, humanities, healthcare	0.25	0.27	0.16 <sup>a</sup>	0.24	0.26	0.16 <sup>a</sup>
Course levels						
Introductory level class				0.44	0.44	0.44
Intermediate level class				0.27	0.27	0.29 <sup>a</sup>
Advanced level class				0.29	0.29	0.27 <sup>a</sup>
Course enrollment						
Total student-hours (x100)				0.56 (0.82)	0.56 (0.84)	0.59 (0.73)
Total student-hours (x100) squared				0.99 (5.92)	1.01 (6.29)	0.88 (3.31)
Immigrant-specific background						
Regions of origin						
Asia			0.51			0.46
Europe & Canada			0.36			0.42
Other region			0.13			0.13
Experience with U.S. higher education						
U.S. graduate degree	0.93	0.99	0.69 <sup>a</sup>	0.95	0.99	0.75 <sup>a</sup>
U.S. work time (x100 semesters)	0.40 (0.36)	0.41 (0.38)	0.35 <sup>a</sup> (0.31)	0.49 (0.36)	0.50 (0.36)	0.44 <sup>a</sup> (0.35)
Number of observations	2,023	1,620	384	16,021	13,458	2,563

Note. Mean statistics, standard deviation in parentheses. The female category also includes the small 1% sample for whom we could not identify gender.  
<sup>a</sup>Indicates differences at 5%.



difference is about 35% of a standard deviation, equivalent to an A– versus B+.

We find important demographic, rank, and classroom and departmental differences between domestic and international instructors that may contribute to these grading differences. Though the majority of domestic (54%) and international (57%) instructors are male, racial/ethnic differences are stark. The vast majority of domestic instructors are White (81%), whereas Asians (43%) make up the largest share of international instructors. International instructors are more likely than domestic to be in a tenure-track position, pre- (21% vs. 13%) or post-tenure (51% vs. 40%), and to be in STEM (44% vs. 26%). In terms of classroom context, international instructors are more likely to teach intermediate-level courses and slightly larger classes.

Lastly, international instructors have diverse backgrounds and U.S. higher educational histories. The majority (51%) are from Asia, 36% are from Europe or Canada, and the rest are from other regions—most notably Latin American countries like Mexico, Colombia, and Brazil. Most (69%) obtained their graduate degrees (master’s and/or PhD) in the United States, meaning that they themselves experienced the U.S. educational system as a student. A sizable share (31%), however, had no U.S. degree. As for work experience, international instructors have worked on average 13.3 years (40 semesters) in the U.S. higher educational system. But there is sizable variation: 44% have worked in the U.S. higher educational system for less than 10 years, 36% for 11 to 25 years, and 20% for over 25 years.

#### *Explaining Grading Practice Differences Between International and Domestic Instructors*

Table 2 presents our four additive models. Across all models, we consistently find that international faculty grade lower than domestic faculty—around 0.13 to 0.16 point lower on a 4-point scale (or about 28% to 35% of a standard deviation). The international coefficient is consistently negative and remains largely robust in size with the inclusion of different controls across models. As seen in Model 4, even after adding the full set of controls, we find that international instructors grade, on average, lower than domestic instructors.

Focusing on Model 4 and the controls, we find that our estimates are largely consistent with the current instructor grading literature with a few exceptions. As expected, nontenured professors, and to a lesser extent, assistant professors assign higher grades compared with tenured professors. The coefficient on both is positive (.06 and .04, respectively), and in the case of nontenured professors, remains statistically significant in the full model. Additionally, grades are lower in larger classes, with about a 0.02-point decrease for every 10 students ( $p < .01$ ). This decrease follows a nonlinear slow and increasing rate. Across class levels, grades are highest in

advanced courses ( $b = 0.14$ ;  $p < .01$ ) and suggestively lower in introductory courses ( $b = -.04$ ;  $p > .1$ ). Compared to faculty in STEM, faculty in the arts and humanities ( $b = 0.20$ ;  $p < .01$ ) and, to a lesser extent, social sciences ( $b = 0.07$ ;  $p < .1$ ) grade higher. Unlike prior research, we do not detect grading differences by gender, nor across racial/ethnic groups, except for the other race/ethnicity category, which consists of multiple groups (e.g., Native Americans, mixed races, and race not determined), making it difficult to interpret.

#### *Grading Practice Differences Among Diverse Subgroups of International Instructors*

Adjusting our full model (Model 4), the three panels in Table 3 examine if international instructor grading differs based on their origin region and prior U.S. higher educational system experience. To do so, we expand our international instructor dummy into respective subcategories, keeping domestic instructors as the reference group, and run separate regressions. In Panel 1, we find suggestive evidence that international instructors across all origin regions—Asia, Europe and Canada, and Other Regions—assign lower grades than domestic instructors. The coefficient on each is negative, but not always statistically significant, a possible reflection of smaller sample sizes. Those from Europe and Canada, however, appear to assign grades with the most significant difference from domestic instructors ( $b = -.18$ ,  $p < .01$ ).

As for differences in prior U.S. higher educational system experience, we find mixed results. Panel 2 suggests that grading practices of international instructors differ based on where they obtained their graduate degree(s). Those who obtained their graduate degree from the U.S. grade lower than domestic instructors ( $b = -.16$ ;  $p < .01$ ) and international instructors with a foreign degree ( $b = -.05$ ;  $p > .1$ ). Tests show that the coefficients are not equivalent ( $p = .06$ ). In contrast, Panel 3 indicates that grading practices of international instructors do not differ based on their U.S. academic work time. Classifying international instructors into two categories—above and below the median (14.6 years)—we found that no matter their length of residence, international instructors assigned similarly lower grades. The coefficient on both terms is negative, similar in size, and statistically significant. We found a similar result using a continuous work-time measure (not shown).

#### *Moderating Effects in Domestic and International Instructor Grading Practices*

Lastly, because domestic instructor grading practices differ across key subgroups, we explore whether similar differences exist for international instructors. Using our full model, we ran four separate regressions, adding (one at a time) an interaction term between *international status* and the

TABLE 2

*Estimates of Differences in Grades Assigned by International and Domestic Faculty*

Variables	(1)	(2)	(3)	(4)
	Baseline Model	Instructor Characteristics	Classroom & Department Factors	Full Model, Immigrant Background
International	-0.14*** (0.04)	-0.16*** (0.05)	-0.13*** (0.05)	-0.14*** (0.05)
General instructor characteristics				
Gender (ref. = female)				
Male		-0.03 (0.03)	0.00 (0.03)	0.00 (0.03)
Race (ref. = White)				
Black		-0.00 (0.06)	-0.01 (0.06)	-0.01 (0.06)
Asian		0.06 (0.05)	0.07 (0.05)	0.06 (0.05)
Hispanic		-0.00 (0.07)	-0.01 (0.07)	-0.01 (0.07)
Other races		-0.11** (0.05)	-0.11** (0.04)	-0.11** (0.04)
Position ranks (ref. = taught by tenured prof.)				
Taught by assistant prof.		0.06* (0.03)	0.05* (0.03)	0.04 (0.03)
Taught by nontenure prof.		0.03 (0.03)	0.06** (0.03)	0.06* (0.03)
Classroom and Department contexts				
Class enrollment				
Total student hours (x100)			-0.19*** (0.03)	-0.19*** (0.03)
Total student hours (x100) squared			0.01*** (0.00)	0.01*** (0.00)
Class levels (ref. = intermediate level class)				
Introductory level class			-0.04 (0.02)	-0.04 (0.02)
Advanced level class			0.14*** (0.02)	0.14*** (0.02)
Department disciplines (ref. = STEM)				
Social sciences			0.07* (0.04)	0.07* (0.04)
Arts, humanities, others			0.20*** (0.04)	0.20*** (0.04)
Immigrant-specific backgrounds				
U.S. graduate education				-0.04 (0.06)
U.S. work time (x100 semesters)				-0.01 (0.04)
Constant	3.20*** (0.03)	3.21*** (0.04)	3.14*** (0.04)	3.19*** (0.07)
Observations	16,021	16,021	16,021	16,021
R-squared	.03	.04	.16	.16
Semester fixed effects	X	X	X	X
Institution fixed effects	X	X	X	X

*Note.* Robust standard errors in parentheses, clustered at department levels throughout. The female reference category also includes the small 1% sample for whom we could not identify gender.

\* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ .

TABLE 3  
*Assessment of International Versus Domestic Grading Differences  
 by International Instructor Subgroups (ref. Domestic Instructor)*

Panel 1. International instructor by region of origin	
International instructor from Asia	-0.09 (0.05)
International instructor from Europe & Canada	-0.18*** (0.07)
International instructor from Other Region	-0.09 (0.07)
Panel 2. International instructor by country of graduate degree	
International instructor with U.S. graduate degree	-0.16*** (0.05)
International instructor with foreign graduate degree	-0.05 (0.06)
Panel 3. International instructor by U.S. work time	
International instructor with U.S. work time <median	-0.12*** (0.04)
International instructor with U.S. work time ≥median	-0.17** (0.07)

*Note.* All panels use the full model (Model 4, Table 2), except in Panel 3, a dichotomy indicator replaces the U.S. work time continuous measure. U.S. work time median in Panel 3 is 44 semesters (or 14.6 years). Robust standard errors in parentheses, clustered at department levels throughout. \*\* $p < .05$ , \*\*\* $p < .01$ .

following subgroups: gender (males vs. females), race (White vs. non-White), position rank (post-tenure vs. nontenure and pre-tenure), and discipline (STEM vs. social sciences and arts & humanities). See Appendix Table B-1 for full results. Focusing on the statistically significant results only, Figure 1 graphs the predicted values for each respective sub-group. Among international instructors, we find that, in contrast to domestic instructor patterns, females grade lower than males (0.11 point lower;  $p < .1$ ), and Whites grade lower than non-Whites (0.22 point lower,  $p < .01$ ). Across-discipline grading patterns, however, were similar: Both international and domestic instructor grades were highest in the arts & humanities and lowest in STEM. Differences by position rank were more distinct among international instructors but often not statistically significant.

#### *Sensitivity Analysis*

We conducted a series of sensitivity tests to ensure the robustness of our results. First, we test whether student sorting biases our results (Table 4). This addresses the concern that students with different instructor preferences may choose to enroll or not enroll in international instructor-led classes. Our results would be biased if students' personal preferences were correlated with students' academic ability, captured by class-average grades. For instance, if academically strong students prefer to enroll in classes led by

domestic rather than international instructors, those classes would have higher grades on average due to students' stronger academic performance, not differential instructor grading practices. To test for potential student sorting, we first assess whether students prefer a specific type of instructor. In column 1, we test whether more students enroll in classes taught by international (versus domestic) instructors using our full model but switching the dependent variable to total student-hours. We reject this hypothesis; international and domestic instructors taught courses with similar enrollment. Next, to alleviate concerns about the potential correlation between students' instructor preferences and academic ability, we use class-average grades and narrow our analysis to two situations where student sorting is not possible: among courses that were taught by only international or only domestic instructors (column 2); and among courses with only one class per semester (column 3). In both columns, the coefficients on international status are similar to our main result (Table 2). We conclude that student sorting does not bias our results.

Next, we assess whether international and domestic faculty are sorted into different departments with distinct grading practices by adding department-university fixed effects to our full model (Appendix Table B-2). These fixed effects capture unobserved differences across more specific academic disciplines than our broader measure (e.g., biology, not just STEM), and also capture cultural differences/practices unique to each university-specific department. With these added fixed effects, the coefficient on international is reduced by half but remains statistically significant ( $b = -0.07$ ;  $p < .1$ ). This suggests the potential sorting of international faculty into lower-grading departments. Nonetheless, even with these more extensive controls, we are still not able to fully explain why international instructors grade lower than domestic instructors.

#### **Discussion**

Given the broad importance of college grades and heightened focus on U.S. college grading practices, this study uses a novel dataset to provide the first large-scale assessment of undergraduate grading practices of international instructors—a sizable yet understudied segment of the U.S. higher education workforce. We make no normative claims as to whether grading students higher or lower is a grading policy/practice that should be valued. Instead, we argue that given the importance of international instructors and U.S. higher education concerns about grade inflation, grade equity, and teacher effectiveness, research is needed to understand what factors shape international instructor grading practices—the focus of our article—and the potential implications of these practices. Our results provide insights into the U.S. higher educational grading system broadly and for supporting the international instructor workforce.

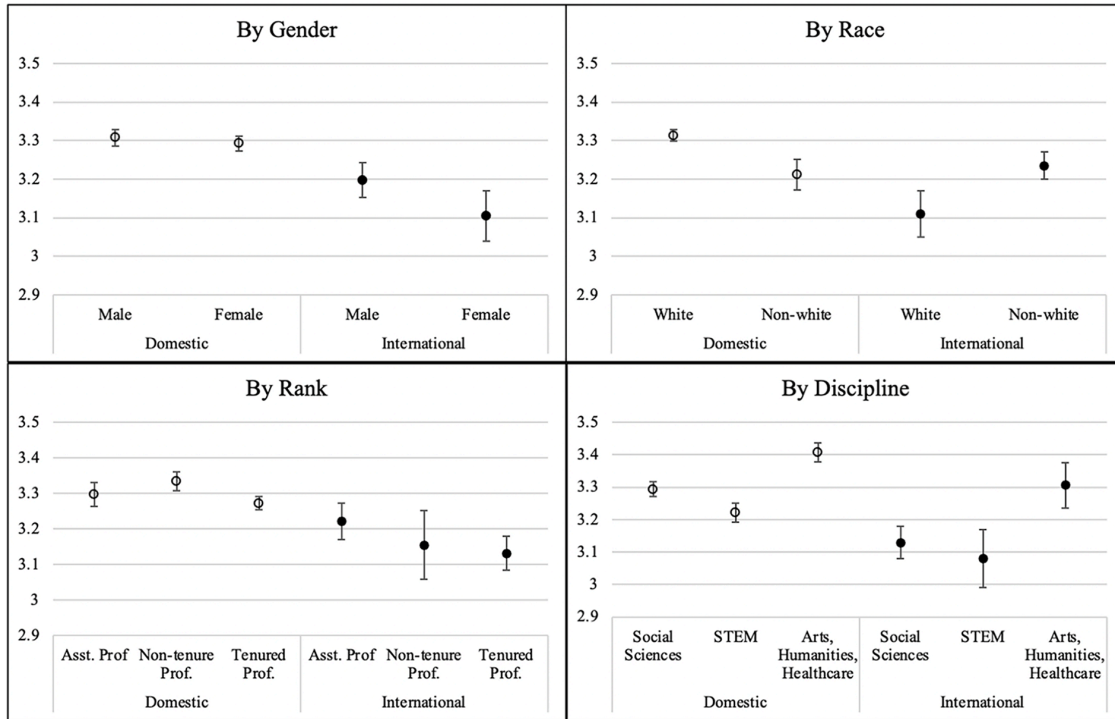


FIGURE 1. Moderating effects in domestic and international instructors' grading practices.

Note. All panels use the full model (Model 4, Table 2) with the addition of an interaction between international status and the corresponding moderator. Confidence intervals are at 95%.  $N=16,021$ . A significant moderating effect is detected when confidence intervals of two groups are not interlaced. Female category also includes the small 1% sample for whom we could not identify gender.

TABLE 4  
Sensitivity Test on Student Sorting

Variables	(1)	(2)	(3)
	Student Self-Selection DV = Total Student-Hours	Courses Taught by Only International or Only Domestic Instructors DV = Class-Average Grades	Only One Course Section per Semester DV = Class-Average Grades
International	0.02 (0.11)	-0.15*** (0.05)	-0.11*** (0.04)
Constant	-0.44 (0.35)	3.20*** (0.07)	3.23*** (0.06)
Observations	532	15,489	11,371
R-squared	.30	.16	.15
Chance of sorting	Yes	No	No

Note. Full model (Model 4, Table 2), except for column 1, where DV is replaced by total student-hours. Robust standard errors in parentheses, clustered at department level.

\*\*\* $p < .01$ .

Focusing on instructors at three U.S. public research universities, we found that international instructors make up almost 20% of undergraduate instructors—a statistic that is consistent with past estimates (Kim et al., 2011)—and that, on average, they grade about 35% of a standard deviation lower than domestic instructors—the equivalent difference

between an A– and a B+. Cumulatively, such a grading difference can have a sizable effect on students' overall grade point average (GPA).

Though international and domestic instructors often differed on key characteristics and classroom and departmental factors known to influence grading practices, these



differences did not explain why international instructors grade lower. Consistent with prior research (Kim et al., 2012; Open Doors, 2022a, 2022b), international instructors in our sample were more likely to be in STEM and were disproportionately male—two factors typically associated with lower grading. But these and other differences, including race/ethnicity, position rank, class size, and class level, did not explain international-domestic grading differences.

Instead, these factors simply contributed to instructor grading practices *overall*. Specifically, all else equal, courses led by tenured professors had lower grades than those led by nontenure or assistant professors. This is consistent with prior study findings that instructors in a more precarious position rank (e.g., pre-tenure) tend to give higher grades due to job security and retention pressures (Kezim et al., 2005). We also found that grades were lower in larger classes, possibly because nonachievement factors are harder to track (Diette & Raghav, 2015; Millea et al., 2018), and that grades are highest in advanced-level courses.

In terms of departmental and disciplinary differences, we found that consistent with prior research (Hermanowicz & Woodring, 2019), faculty in STEM fields were more likely to grade lower than their peers in arts & humanities and social sciences. Our study, however, adds to this literature by showing that both international and domestic faculty seem to adhere to this same broad disciplinary grading pattern. In our interaction results, we found that both domestic and international STEM faculty graded lower than their respective peers in other disciplines. The fact that international faculty were more likely to be in STEM, however, did not fully explain why international faculty grade lower.

The inclusion of department-university fixed effects helped explain a sizable share of the international-domestic instructor grading differences. One potential explanation is that international instructors may sort into subfields within STEM, such as math and chemistry, where the allocation of grades may be even lower (Hermanowicz & Woodring, 2019). The challenge is that these fixed effects are “catch-all” measures that capture both specific disciplinary differences (e.g., subdisciplines within STEM) and unobserved cultural and structural differences (e.g., departmental size, cultural norms, student differences) specific to each academic department. Our study, consequently, reiterates the need for future research to examine how grading practices differ across specific disciplines and departments within a university (Hermanowicz & Woodring, 2019).

Grading practices of international instructors, however, were not homogenous. In our sample, international instructors were highly diverse, spanning broad global regions—Asia (51%), Europe and Canada (36%), and Other Regions (13%)—and with varied U.S. educational system histories. Some had considerable prior U.S. higher education system histories (e.g., 69% obtained a U.S. graduate degree), while others were relatively new (44% had worked in U.S. higher

education for less than 10 years). Overall, we found that international instructors from all regions assign lower grades than domestic instructors, but the difference is the greatest for those from Europe and Canada. One possible explanation, as suggested by cross-national research between Canadian and Chinese instructors, is that Canadian (and European) instructors may be less likely to consider non-achievement factors in grading, which can lead to assigning lower grades (Cheng et al., 2018).

We also found that international instructors’ grading practices differ based on their prior U.S. higher educational system experience, but the results are mixed. On the one hand, we found no evidence that over time international and domestic instructor grades converge, as suggested by immigrant assimilation and adaptation literatures (NAEMS, 2015). Coefficients on U.S. work time were never statistically significant. However, we did find that international instructors who obtained a U.S. (versus foreign) graduate degree graded lower. Presumably, these international instructors should be more familiar with the United States’s higher educational system than their international counterparts who obtained all foreign degrees. A lack of familiarity or cultural understanding of the U.S. higher educational system is not likely to explain why international instructors with a U.S. graduate degree grade lower. Instead, more research is needed.

Lastly, we found that international instructor grading practices differed by gender and race/ethnicity. Among international instructors, males grade *higher*, not *lower*, than females. This contrasts with prior evidence on gender grading dynamics, which finds that female instructors grade higher than males to avoid poor student evaluation and for better job retention (Griffith & Sovero, 2021; Radchenko, 2020). Our result suggests a different gender dynamic is likely occurring among international instructors and highlights the need for future research to examine how gender and international status intersect to influence instructor grading.

Similarly, more research is needed to better understand how race/ethnicity intersects with international status to shape instructor grading. Among international instructors, we found that White (non-Hispanic) instructors graded lower on average than non-White instructors. This racial/ethnic difference likely reflects grading practice differences between White (non-Hispanic) and Asian instructors, each of whom makes up the vast majority of international instructors (43% and 40%, respectively). Consistent with prior research, we found no similar racial/ethnic differences among domestic instructors (Jewell & McPherson, 2012; McPherson et al., 2009). This nonresult is likely because domestic instructors are largely homogenous—81% were non-Hispanic White—making it difficult to assess racial/ethnic patterns. As for international instructors, there is limited insight from current research to understand why White instructors grade lower. It could be that non-White instructors face similar heightened discrimination and job insecurity, grade inflation pressures

as female instructors (Boring, 2017; Radchenko, 2020). Alternatively, non-White international instructors may be more effective at connecting with classroom diversity, leading to higher student achievement (Downey & Pribesh, 2004; Kozlowski, 2015). Our study cannot distinguish between these or other explanations. Instead, we highlight the need for further investigation.

#### *Limitations and Avenues for Future Research*

Though this is the first large-scale study of international instructors' grading practices, our study is exploratory. Future research is needed to address data limitations and to assess whether our findings expand to more nuanced international status definitions and other types of institutions. Our international status indicator—undergraduate degree country location—is better-suited than other indicators (e.g., nativity) to identify the international instructor subgroup of interest as they are most likely to face U.S.-grading adaptation challenges. However, there are several important limitations to this dichotomized measure. As noted, our measure also captures the small share of U.S.-born instructors who completed their undergraduate degree abroad and likely underestimates the international-domestic instructor grading difference. Related, future research should use a more nuanced, multitiered classification of international instructor status that better captures the diverse backgrounds (nativity, citizenship status, length in the U.S. academia) and increasingly global academic training (graduate preparation, work history) of international *and* domestic instructors (e.g., U.S.-born instructors trained abroad). These factors are likely to expose them to different teaching styles and grading practices. To do so, we need better quantitative data on international instructors and more qualitative studies to better unpack the complex, lifetime journeys and academic trajectories of international instructors that shape their overall teaching philosophy and practices.

Additionally, though we are the first to connect international instructor status with grades across a wide variety of courses, we do not have data on the students. Despite the robustness of our results to several sensitivity checks, we cannot completely rule out whether different types of students (e.g., high versus lower performing) may differentially self-select into courses taught by international versus domestic instructors. Nor do we know how the grading difference we observe may impact instructors, particularly if international instructors are evaluated more harshly by students as a result of their lower grading. Given the importance of grading for students' future, instructors' job security, and university reputations, comprehensive data that connect all the pieces—student and instructor characteristics, grades, and instructor evaluations—are needed to further explore these issues. To accurately capture student and instructor characteristics and related experiences, this data must include self-reported indicators of gender, race, and

ethnicity. Without self-reports, studies like ours are forced to rely on proxy indicators, which, though valuable, are prone to measurement error and fail to capture the full spectrum of gender or the social and historical complexities surrounding race and ethnicity (Laughter, 2018; Li and Koedel, 2018).

Lastly, our data and results are not designed to be representative of all U.S. universities. Instead, we provide a foundational starting point for understanding potential differences between domestic and international instructor grading practices. We do so by focusing on a distinct type of university: Three large R1, flagship, public universities in the Midwest, all of which are Predominantly White Institutions. Grading practices of international instructors at Minority Serving Institutions and other types of higher education institutions (e.g., private, liberal arts, community colleges) and geographic locations may differ due to differential work pressures, grading norms, and instructor and student body composition (Butcher et al., 2014; Rojstaczer & Healy, 2012). Future research should examine if our results are generalizable to other types of higher education institutions.

#### **Conclusion**

Overall, we consistently find that international instructors grade lower on average than domestic instructors. Although we are not able to identify the exact structural and/or cultural mechanisms as to why international instructors grade lower, our study serves as a call to action for U.S. higher educational systems to recognize and support its diverse international instructor workforce. Higher educational systems need to ensure that international-domestic grading differences, as well as other possible differences in research collaboration, networking, and service, do not unfairly impact the job security and promotion of international instructors. Finally, much can be learned from international instructors. For instance, they may help curb grade inflation, a pressing problem at many universities (Denning et al., 2022; Rojstaczer & Healy, 2012), or help higher educational systems serve the increasingly diverse student bodies in a more equitable manner. To do so, international instructors must be supported in all endeavors.

#### **Appendix A**

##### *Summary of Prior Studies and Data Sources Used to Examine International Instructors: Data Limitations and Key Characteristics*

Table A-1 highlights the main data sources that studies have used to estimate the share of international instructors and/or examine international instructor experiences. The first two data sources—IPEDS and Open Doors—are aggregated administrative data often used to estimate the size of the international instructor workforce. The other data sources are survey data, all designed for different purposes (e.g., to examine new U.S. PhD recipients) but have information that enables researchers to identify if the participant (a) works in

TABLE A-1

*Overview of Main Data Sources Used to Identify, Estimate and/or Examine U.S. International Instructors*

Database	Organization	Data Overview	Data Years	Types of Instructors / Scholars Included	International Status Indicator(s)	Available Information Related to Teaching
Integrated Postsecondary Education System (IPEDS)	U.S. Department of Education	Collects a variety of characteristics of most U.S. higher education institutions	Annual (since 2004)	Full- & part-time faculty (all ranks and type)	Nonresident alien	Position title only <sup>a</sup>
Open Doors	Institute of International Education	Comprehensive information resource on international students and scholars studying or teaching at U.S. higher education institutions	Annual (since 1999)	International individuals on nonimmigrant visas, engaged in temporary academic activities and not enrolled as a student at a U.S. college or university	Nonresident alien	Primary function is (a) Teaching Only or (b) Research & Teaching
Survey of Earned Doctorate (SED)	National Science Foundation	Provides educational history, funding, and postdoctoral plans for new U.S. PhD recipients	Annual (since 1957)	New doctoral graduates from a U.S. institution; can select subset employed in higher education	U.S. citizenship	(only available in restricted-access data) Position title; primary function is (1) Teaching Only or (2) Research & Teaching
Survey of Doctorate Recipients (SDR)	National Science Foundation & National Institution of Health	Provides data on the characteristics of science, engineering, and health research doctorate degree holders from U.S. academic institutions	Biennial (since 1973)	All doctoral graduates with a degree in science, engineering and health from a U.S. institution up to the current year; can select subset employed in higher education	U.S. citizenship	(only available in restricted-access data) Position title; primary function is (1) Teaching Only, or (2) Research & Teaching
National Study of Postsecondary Faculty (NSOF)	U.S. Department of Education	Survey data on faculty and instructional staff in American colleges and universities	1987–88; 1992–93; 1998–99; 2003–04	Full- & part-time faculty (all ranks and type)	Foreign-born	Primary function is (1) Teaching Only or (2) Research & Teaching
COACHE	Harvard University	Survey of U.S. faculty job satisfaction at participating higher education institutions	Annual (since 2003)	Full-time faculty only (all ranks and type)	U.S. citizenship; foreign-born	Position title & overall satisfaction indicators related to students, teaching, and courses

<sup>a</sup>These data only identify the type of teaching position (e.g., full, associate, or assistant professor, lecturer, adjunct) an individual (or share of individuals) holds but no other detail related to their teaching/instructor duties.

TABLE A-2

*Available Study Evidence on International and Domestic Instructor Characteristics, Publication and Data Year Reported*

	Marvasti (2005)		Kim et al. (2011)		Lin et al. (2009)		Kim et al. (2012)		Espinosa et al. (2019)		Open Doors (2019b)				
	1993		1999		2003		2004		2003-2005		2016		2017		
	Int'l	Domestic	Int'l	Domestic	Int'l	Domestic	Int'l	Domestic	Int'l	Domestic	Int'l	Domestic	Int'l		
% of instructors	12.1%	87.9%	13.3%	86.7%	20.0%	80.0%	22.1%	77.9%	21.8%	78.2%	4.5%	95.5%	—		
Instructor characteristics															
Gender															
Male	—	—	—	—	87.0%	73.5%	71.3%	62.9%	—	—	64.0%	53.0%	—		
Female	—	—	—	—	23.0%	26.5%	28.8%	37.1%	—	—	36.0%	47.0%	—		
Race															
White	—	—	—	—	—	—	46.2%	89.7%	—	—	—	—	—		
Black	—	—	—	—	—	—	7.1%	5.0%	—	—	—	—	—		
Asian	—	—	—	—	—	—	39.2%	2.2%	—	—	—	—	—		
Hispanic	—	—	—	—	—	—	7.3%	1.7%	—	—	—	—	—		
Others	—	—	—	—	—	—	0.2%	1.4%	—	—	—	—	—		
Fields															
STEM	—	—	—	—	84.4%	85.3%	43.5%	31.4%	—	—	—	—	57.9%		
Social sciences	—	—	—	—	15.6%	14.7%	52.6%	58.2%	—	—	—	—	18.0%		
Arts, humanities, healthcare, other	—	—	—	—	—	—	3.9%	10.3%	—	—	—	—	23.6%		
Rank															
Assistant professor	—	—	—	—	70.4%	22.9%	32.2%	25.8%	—	—	55.0%	21.0%	—		
Tenured professor	—	—	—	—	29.6%	77.1%	53.2%	57.8%	—	—	24.0%	43.0%	—		
Nontenure professor	—	—	—	—	—	—	14.7%	16.5%	—	—	21.0%	35.0%	—		
Work at research universities	51.6%	34.5%	65.3%	34.8%	—	—	—	—	—	—	—	—	—		
Regions of origin															
Asia	—	—	—	—	—	—	—	—	—	—	—	—	57.5%		
Europe & Canada	—	—	—	—	—	—	—	—	—	—	—	—	27.6%		
Other Region	—	—	—	—	—	—	—	—	—	—	—	—	13.5%		
Total number of faculty	25,780		17,600		6,938		11,562		14,543		Over 345,000		135,009		
Database used	NSOF				IPEDS & SED				NSOF		COACHE		IPEDS	Open Doors	
International defined as:	Foreign-born				Nonresident aliens				Foreign-born		International faculty		Nonresident aliens		Int'l scholars

higher education, and (b) is international or domestic, using one or more indicators available in the dataset.

Overall, the table highlights several important data limitations. First, none of these databases has information on instructor grading practices. The available teaching-related information is very limited and often only includes position titles. Consequently, these data cannot be used to examine international instructor grading practices. Second, because each data captures different segments of the U.S. higher education workforce (e.g., only new doctoral graduates in Survey of Earned Doctorate [SED], or faculty in STEM and health in Survey of Doctorate Recipients [SDR]), and has different indicators that can be used to identify international status (e.g., nativity, citizenship), estimates of the size and characteristics of the international instructor workforce will differ depending on the data source used.

Table A-2 demonstrates this variation. The table highlights available statistics on the characteristics of international and domestic instructors from several prominent studies that have used one or more of the data sources highlighted in Table A-1. These studies have examined international instructors' career satisfaction and general academic experiences. The table also highlights key instructor characteristics from Open Door, a main data source most studies use to provide a broad overview of international scholars, including instructors.

Overall, the table shows that there is significant variation in the time frame examined, as well as the international status indicator and data source used. Consequently, each of these studies captures different segments of the overall U.S. higher education workforce as well as different international instructor definitions.



Despite this variation, the studies show important consistencies. International faculty make up a sizable share of the U.S. higher education instructors (between 5% and 20% depending on the types of institution and subfaculty group examined). The share of male and Asian instructors

is higher among international faculty than domestic faculty, across all studies. International faculty are more likely to be in a STEM field (58% according to Open Doors) and to be in a tenured position, often pre-tenure/assistant position.

## Appendix B

### Robustness and Sensitivity Analysis

TABLE B-1  
Moderating Effects in Domestic and International Instructors' Grading Practices

Variables	(1)	(2)	(3)	(4)
	Male vs. Female	White vs. Non-White	Tenure vs. Nontenure and Pre-Tenure	STEM vs. Non-STEM
International	-0.20*** (0.07)	-0.20*** (0.06)	-0.14*** (0.05)	-0.14* (0.09)
Main effects				
Male	-0.02 (0.03)			
Non-White		-0.10** (0.04)		
Position ranks (ref. = tenured prof.)				
Assistant prof.			0.02 (0.04)	
Nontenure prof.			0.06* (0.03)	
Department disciplines (ref. =STEM)				
Social sciences				0.07* (0.04)
Arts, humanities, healthcare				0.19*** (0.04)
Interaction effects				
International × Male	0.11* (0.06)			
International × Non-White		0.22*** (0.07)		
International × Assistant Prof.			0.07 (0.07)	
International × Nontenure Prof.			-0.04 (0.09)	
International × Social Sciences				-0.01 (0.09)
International × Arts/Humanities/Others				0.05 (0.09)
Constant	3.19*** (0.07)	3.19*** (0.07)	3.20*** (0.07)	3.19*** (0.07)
Observations	16,021	16,021	16,021	16,021
R-squared	.16	.16	.16	.16
Semester fixed effects	X	X	X	X
Institution fixed effects	X	X	X	X

Note. All models are based on the full model (Model 4, Table 2) with all other control variables and semester and institution fixed effects. Robust standard errors in parentheses, clustered at department level throughout.

\* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ .

TABLE B-2  
Sensitivity Tests With More Restricted Model and Sample

Variables	(1)	(2)
	Department Fixed Effect	Only Tier-1 Data
International status	-0.07* (0.04)	-0.11** (0.05)
Constant	3.36*** (0.05)	3.17*** (0.07)
Observations	16,021	13,199
R-squared	.09	.17

Note. All models are based on full model (Model 4, Table 2) with all other control variables. Column 1 adds a department fixed effect, column 2 restricts the estimation within Tier-1 data only. Robust standard errors in parentheses, clustered at department level throughout.  
\* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ .

## Appendix C

### Data Construction Process

This section describes how the dataset of 2,023 university instructors was constructed for the analysis. The dataset is built from two sources. The first is from the course and grade data from three focal universities spanning all semesters between 2011 and 2017 inclusively (hereafter: grade dataset). The data were publicly available and downloaded from the university registrars' websites. The grade dataset has information about class-average grades, class levels, class enrollment, numbers of each letter grades, departments, semesters, and most importantly, class instructors' full names. The second source is a unique dataset with information on instructors' qualifications and demographics (hereafter: instructor dataset), which were manually collected based on instructors' full names, universities, and departments.

With these two data sources, we address the limitations of widely used survey databases in the existing literature. The grade dataset is based on administrative records of courses taught by faculty at the three universities, which ensures no concerns about non-random sample attrition (e.g., non-survey response). The instructor dataset contains instructors' nationality information, which enables us to attain specific country-level statistics.

The construction of our dataset proceeded in the following steps.

*Step 1: Identify the Sample Size.* First, the initial downloaded grade file yielded a universe of 18,000 instructors who taught over 166,000 classes during 21 semesters. From this we eliminated lab sections and cotaught courses. Given that the instructor dataset requires manual data construction, our first objective was to identify a sample size that would

be large enough to detect a meaningful difference in grading behavior and ensure the feasibility of a manual data search. A preanalysis showed that in order to achieve a *minimum detectable effect size of 10% of a standard deviation of the average-grade distribution*, our target standard error size needed to be 0.025 for the main analysis, per the following:

$$Z = \frac{\sigma * t}{St.Dev(G)} \Rightarrow 0.1 = \frac{\sigma * 1.96}{0.5} \Rightarrow \sigma \approx 0.025.$$

In this equation,  $Z$  is the effect size,  $\sigma$  is the target standard error of the parameter of interest (i.e., international status),  $t$  is the critical  $t$ -value, and  $St.Dev(G)$  is the standard deviation of the class-average grades of the population.

To perform an *ad hoc* power calculation, we collected information from 300 randomly selected instructors as a test sample (100 from each university) and calculated the standard errors of the main effect when we artificially duplicated the test sample repeatedly. Our calculation showed that when the sample increased tenfold to 3,000 faculty observations, the standard error was approximately 0.025. Although this *ad hoc* power calculation is not perfectly accurate because it replicates the same exact 300 observations, it gives a good estimate of the actual standard error as the real sample size grows. Based on this calculation, we collected data for 3,000 instructors to ensure a well-powered model with the ability to statistically detect meaningful grading differences between domestic and international faculty.

*Step 2: Randomly Select 3,000 Instructors.* Next, we utilized a stratified sampling strategy to create a drawing pool of instructors from departments and then randomly selected 3,000 instructors from within. To ensure that exceptionally large or small departments were not overrepresented in the drawing pool, we removed small departments of less than 10 instructors. For large departments of over 100 instructors, we randomly selected 100 to enter the pool. Departments with instructor populations from 11 to 99 enter the pool as they are. Then, we randomly selected 1,000 instructors from the stratified population at each university to make up the instructor dataset of 3,000 instructors. Given that the distribution of international instructors by field and department was unknown to us at the sampling stage, we applied this strategy to ensure that the sampled 3000 instructors were representative of departments in three universities.

*Step 3: Manually Collect Instructors' Data.* With the list of 3,000 instructors in hand, we moved on to the manual data collection. This was the most labor-extensive step. We conducted a manual search to obtain the qualifications and demographic information of each instructor. We collected this information from instructors' curricula vitae and websites. In rare instances when these sources were unavailable, we substituted them with information from other sources,

TABLE C-1  
Features of Three Focal Universities

Key Features	University A	University B	University C
Carnegie classification	R1	R1	R1
Undergraduate student population	>23,000	>33,000	>32,000
% nonresident aliens	<5%	<10%	<10%
% male	≅50%	≅50%	≅50%
% White	≅80%	≅70%	≅55%
Undergraduate degree programs	300+	260+	300+
Instructor population	>3,000	>3,000	>6,000
% nonresident aliens	<5%	<10%	<15%
% White	≅70%	≅70%	≅60%

Note. Authors' consolidation from College Scorecard Data and National Center for Education Statistics (NCES) Data. Statistics are from 2011–2017, rounded.

TABLE C-2  
Breakdown of International Instructors' Countries of Origin  
(Top 10)

Country of Origin	Counts	Percentage
China	67	.17
India	41	.12
Korea	24	.07
Britain	22	.06
Germany	18	.05
Canada	18	.05
Israel	16	.05
Italy	12	.03
Taiwan	9	.02
Turkey	8	.02
Other	117	.30
Total	384	1.00

such as Scopus, LinkedIn, news articles, and university bulletins. This completed the instructor dataset.

*Step 4: Assemble the Final Data.* At this step, we merged the instructor dataset with the grade dataset and removed those whose profiles were unavailable online. For the current article, we also removed graduate instructors from the analysis. Elsewhere (Pham, 2021), we kept this analysis on graduate instructors despite some noises. This completes a dataset of 2,023 instructors who taught 16,021 undergraduate courses in three focal universities during 21 semesters. Features of the three universities are reported in Table C-1. Breakdowns of the top 10 countries of origin of 384 international instructors are reported in Table C-2.

#### Acknowledgments

Our gratitude goes to Cory Koedel for his constant support from the inception of this article. We thank Claire Altman, Lael Keiser,

Emily Crawford, participants in the 2020 APPAM and AEFPP conferences, and all anonymous reviewers for their comments and suggestions; Cade Copeland, Sophie Burke, and Savannah McCauley for their assistance.

#### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Truman School of Government and Public Affairs, University of Missouri-Columbia, supported this research through dissertation funding for Trang Pham. All errors are our own.

#### Open Practices

More details about our data can be accessed at <https://www.openicpsr.org/openicpsr/project/206501/version/V1/view>.

#### ORCID iD

Trang Pham  <https://orcid.org/0000-0001-5929-5577>

#### Note

1. Alternative corresponding terminologies are academic and nonacademic criteria (Allen, 2005), cognitive and noncognitive behaviors (Brookhart et al., 2016; Townsley, 2020), ability and class involvement (Millea et al., 2018), and evidence of achievement versus ability to succeed in the outside works (Olsen & Buchanan, 2019).

#### References

- Ake-Little, E., von der Embse, N., & Dawson, D. (2020). Does class size matter in the university setting? *Educational Researcher*, 49(8), 595–605. <https://doi.org/10.3102/0013189X20933836>
- Alba, R., & Nee, V. (2003). *Remaking the American mainstream*. Harvard University Press.

- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218–223. <https://doi.org/10.3200/TCHS.78.5.218-223>
- Altbach, P. G. (2001). Academic freedom: International realities and challenges. *Higher Education*, 41, 205–219. <https://doi.org/10.1163/9789087903596>
- Altbach, P. G. (2009). Peripheries and centers: Research universities in developing countries. *Asia Pacific Education Review*, 10(1), 15–27. <https://doi.org/10.1007/s12564-009-9000-9>
- Anglin, P. M., & Meng, R. (2000). Evidence on grades and grade inflation at Ontario's universities. *Canadian Public Policy*, 26(3), 361–368. <https://doi.org/10.2307/3552406>
- Arnold, N. W., Crawford, E. R., & Khalifa, M. (2016). Psychological heuristics and faculty of color: Racial battle fatigue and tenure/promotion. *Journal of Higher Education*, 87(6), 890–919. <https://doi.org/10.1080/00221546.2016.11780891>
- Bean, J. P. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome. *American Educational Research Journal*, 22(1), 35–64. <https://doi.org/10.2307/1162986>
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27–41. <https://doi.org/10.1016/j.jpubeco.2016.11.006>
- Bowman, N. A., & Denson, N. (2022). Institutional racial representation and equity gaps in college graduation. *Journal of Higher Education*, 93(3), 399–423. <https://doi.org/10.1080/00221546.2021.1971487>
- Brookhart, S. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69–90. <https://doi.org/10.1080/0969594X.2012.70317>
- Brookhart, S., Guskey, T., Bowers, A., McMillan, J., Smith, J., Smith, L., Stevens, M., & Welsh, M. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>
- Brown, S. K., & Bean, F. D. (2006). Assimilation models, old and new: Explaining a long-term process. *Migration Information Source*, 9. <https://www.migrationpolicy.org/article/assimilation-models-old-and-new-explaining-long-term-process>
- Butcher, K. F., McEwan, P. J., & Weerapana, A. (2014). The effects of an anti-grade inflation policy at Wellesley College. *Journal of Economic Perspectives*, 28(3), 189–204. <https://doi.org/10.1257/jep.28.3.189>
- Cheng, L., DeLuca, C., Braund, H., & Yan, W. (2018). Grading policies in Canada and China: A comparative study. *Comparative and International Education*, 47(1). <https://doi.org/10.5206/cie-eci.v47i1.9324>
- Cheng, L., DeLuca, C., Braund, H., Yan, W., & Rasooli, A. (2020). Teachers' grading decisions and practices across cultures: Exploring the value, consistency, and construction of grades across Canadian and Chinese secondary schools. *Studies in Educational Evaluation*, 67, 100928. <https://doi.org/10.1016/j.stueduc.2020.100928>
- Denning, J. T., Eide, E. R., Mumford, K. J., Patterson, R. W., & Warnick, M. (2022). Why have college completion rates increased? *American Economic Journal: Applied Economics*, 14(3), 1–29. <https://doi.org/10.1257/app.20200525>
- Diette, T. M., & Raghav, M. (2015). Class size matters: Heterogeneous effects of larger classes on college student learning. *Eastern Economic Journal*, 41(2), 273–283. <https://doi.org/10.1057/ej.2014.31>
- Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, 77(4), 267–282. <https://doi.org/10.1177/003804070407700401>
- Ehrenberg, R. G., & Zhang, L. (2005). Do tenured and tenure-track faculty matter? *Journal of Human Resources*, XL(3), 647–659. <https://doi.org/10.3368/jhr.XL.3.647>
- Espinosa, L., Turk, J., Taylor, M., & Chessman, H. (2019). *Race and ethnicity in higher education* (p. 336). American Council on Education. <https://1xfsu31b52d33idlp13twtos-wpengine.netdna-ssl.com/wp-content/uploads/2019/02/Race-and-Ethnicity-in-Higher-Education.pdf>
- Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R., & Johnston, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLoS ONE*, 14(2), e0209749. <https://doi.org/10.1371/journal.pone.0209749>
- Franz, I. W.-J. (2010). Grade inflation under the threat of students' nuisance: Theory and evidence. *Economics of Education Review*, 29(3), 411–422. <https://doi.org/10.1016/j.econedurev.2009.10.013>
- Ginther, D. K. (2001). *Does science discriminate against women? Evidence from academia, 1973-1997* (Working Paper No. 2001-02). Federal Reserve Bank of Atlanta. <https://doi.org/10.2139/ssrn.262438>
- Ginther, D. K., & Kahn, S. (2006). *Does science promote women? Evidence from academia 1973-2001* (Working Paper No. w12691; p. w12691). National Bureau of Economic Research. <https://doi.org/10.3386/w12691>
- Griffith, A. L., & Sovero, V. (2021). Under pressure: How faculty gender and contract uncertainty impact students' grades. *Economics of Education Review*, 83, 102126. <https://doi.org/10.1016/j.econedurev.2021.102126>
- Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24(4), 369–376. <https://doi.org/10.1016/j.econedurev.2004.07.013>
- Harbatkin, E. (2021). Does student-teacher race match affect course grades? *Economics of Education Review*, 81, 102081. <https://doi.org/10.1016/j.econedurev.2021.102081>
- Hermanowicz, J. C., & Woodring, D. W. (2019). The distribution of college grades across fields in the contemporary university. *Innovative Higher Education*, 44(6), 497–510. <https://doi.org/10.1007/s10755-019-09474-w>
- Isely, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations? *Journal of Economic Education*, 36(1), 29–42. <https://doi.org/10.3200/JECE.36.1.29-42>
- Jewell, R. T., & McPherson, M. A. (2012). Instructor-specific grade inflation: Incentives, gender, and ethnicity. *Social Science Quarterly*, 93(1), 95–109. <https://doi.org/10.1111/j.1540-6237.2011.00827.x>
- Jewell, R. T., McPherson, M. A., & Tieslau, M. A. (2013). Whose fault is it? Assigning blame for grade inflation in higher education. *Applied Economics*, 45(9), 1185–1200. <https://doi.org/10.1080/00036846.2011.621884>



- Joshi, E., Doan, S., & Springer, M. G. (2018). Student-teacher race congruence: New evidence and insight from Tennessee. *AERA Open*, 4(4), 2332858418817528. <https://doi.org/10.1177/2332858418817528>
- Kezim, B., Pariseau, S. E., & Quinn, F. (2005). Is grade inflation related to faculty status? *Journal of Education for Business*, 80(6), 358–364. <https://doi.org/10.3200/JOEB.80.6.358-364>
- Kim, D., Twombly, S., & Wolf-Wendel, L. (2012). International faculty in American universities: Experiences of academic life, productivity, and career mobility. *New Directions for Institutional Research*, 2012(155), 27–46.
- Kim, D., Wolf-Wendel, L., & Twombly, S. (2011). International faculty: Experiences of academic life and productivity in U.S. universities. *Journal of Higher Education*, 82(6), 720–747. <https://doi.org/10.1353/jhe.2011.0038>
- Koedel, C. (2011). Grading standards in education departments at universities. *Education Policy Analysis Archives*, 19, 23. <https://doi.org/10.14507/epaa.v19n23.2011>
- Kokkelenberg, E. C., Dillon, M., & Christy, S. M. (2008). The effects of class size on student grades at a public university. *Economics of Education Review*, 27(2), 221–233. <https://doi.org/10.1016/j.econedurev.2006.09.011>
- 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>
- Kozlowski, K. P. (2015). Culture or teacher bias? Racial and ethnic variation in student–teacher effort assessment match/mismatch. *Race and Social Problems*, 7(1), 43–59. <https://doi.org/10.1007/s12552-014-9138-x>
- Laughter, J. (2018). Race in *Educational Researcher*: A technical comment on Li and Koedel (2017). *Educational Researcher*, 47(4), 259–261. <https://doi.org/10.3102/0013189X18769647>
- Leonardo, Z. (2007). The war on schools: NCLB, nation creation and the educational construction of whiteness. *Race Ethnicity and Education*, 10(3), 261–278. <https://doi.org/10.1080/13613320701503249>
- Li, D., & Koedel, C. (2017). Representation and salary gaps by race-ethnicity and gender at selective public universities. *Educational Researcher*, 46(7), 343–354. <https://doi.org/10.3102/0013189X17726535>
- Li, D., & Koedel, C. (2018). A technical comment on Li and Koedel (2017): Author response. *Educational Researcher*, 47(4), 262–263. <https://doi.org/10.3102/0013189X18769304>
- Lin, Z., Pearce, R., & Wang, W. (2009). Imported talents: Demographic characteristics, achievement and job satisfaction of foreign-born full-time faculty in four-year American colleges. *Higher Education*, 57(6), 703–721. <https://doi.org/10.1007/s10734-008-9171-z>
- Macfarlane, B. (2012). Re-framing student academic freedom: A capability perspective. *Higher Education*, 63(6), 719–732. <https://doi.org/10.1007/s10734-011-9473-4>
- Marvasti, A. (2005). U.S. academic institutions and perceived effectiveness of foreign-born faculty. *Journal of Economic Issues*, 39(1), 151–176. <https://doi.org/10.1080/00213624.2005.11506784>
- Massey, D. S., & Malone, N. (2003). Pathways to legal immigration. *Population Research and Policy Review*, 21(6), 473–504. <https://doi.org/10.1023/A:1022914700681>
- McPherson, M. A., Jewell, R. T., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economic Journal*, 35(1), 37–51. <https://doi.org/10.1057/palgrave.eej.9050042>
- Millea, M., Wills, R., Elder, A., & Molina, D. (2018). What matters in college student success? Determinants of college retention and graduation rates. *Education*, 138(4), 309–322. <https://api.semanticscholar.org/CorpusID:149980776>
- National Academies of Sciences, Engineering, and Medicine (NAEMS). (2015). *The integration of immigrants into American society* (p. 21746). National Academies Press. <https://doi.org/10.17226/21746>
- National Center for Education Statistics (NCES). (2021). *Full-time faculty in degree-granting postsecondary institutions, by race/ethnicity, sex, and academic rank: Fall 2019, fall 2020 and fall 2021*. [https://nces.ed.gov/programs/digest/d22/tables/dt22\\_315.20.asp](https://nces.ed.gov/programs/digest/d22/tables/dt22_315.20.asp)
- National Center for Education Statistics (NCES). (2023). *Undergraduate enrollment*. <https://nces.ed.gov/programs/coe/indicator/cha/undergrad-enrollment#:~:text=Between%20fall%202010%20and%20fall,percent%20to%2016.8%20million%20students>
- Olsen, B., & Buchanan, R. (2019). An investigation of teachers encouraged to reform grading practices in secondary schools. *American Educational Research Journal*, 56(5), 2004–2039. <https://doi.org/10.3102/0002831219841349>
- Open Doors. (2022a). *IIE Open Doors / Academic level*. <https://opendoorsdata.org/data/international-students/academic-level/>
- Open Doors. (2022b). *IIE Open Doors / International Scholars*. <https://opendoorsdata.org/data/international-scholars/>
- Park, E. S., Wilton, M., Lo, S. M., Buswell, N., & Sato, B. K. (2021). *STEM faculty instructional approaches to assessment, grading and diversity are linked to racial equity grade gaps* (Working Paper No. 21-7, 18). UCI Education Research Initiative.
- Park, J. J., Kim, Y. K., Salazar, C., & Hayes, S. (2020). Student–faculty interaction and discrimination from faculty in STEM: The link with retention. *Research in Higher Education*, 61(3), 330–356. <https://doi.org/10.1007/s11162-019-09564-w>
- Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, 42(5), 259–265.
- Pennesi, K. (2016). “They can learn to say my name”: Redistributing responsibility for integrating immigrants to Canada. *Anthropologica*, 58(1), 46–59. <https://ir.lib.uwo.ca/anthropub>
- Pham, T. (2021). *Three essays on assimilation and acculturation of international instructors at U.S. research universities* [Doctoral dissertation, University of Missouri–Columbia]. MOspace Repository. <https://doi.org/10.32469/10355/93239>
- Project Atlas. (2022). *U.S. students pursuing full degrees abroad*. Institute of International Education. <https://www.iie.org/research-initiatives/project-atlas/u-s-students-pursuing-full-degrees/>
- Quinn, R. (2023, November 28). Professor says Spelman raised students’ grades, fired him. *Inside Higher Ed*. <https://www.insidehighered.com/news/faculty-issues/academic-freedom/2023/11/28/professor-says-selman-raised-students-grades-fired>
- Radchenko, N. (2020). Student evaluations of teaching: Unidimensionality, subjectivity, and biases. *Education*

- Economics*, 28(6), 549–566. <https://doi.org/10.1080/09645292.2020.1814997>
- Ran, F. X., & Xu, D. (2019). Does contractual form matter? The impact of different types of non-tenure-track faculty on college students' academic outcomes. *Journal of Human Resources*, 54(4), 1081–1120. <https://doi.org/10.3368/jhr.54.4.0117.8505R>
- Rodríguez-Planas, N. (2022). COVID-19, college academic performance, and the flexible grading policy: A longitudinal analysis. *Journal of Public Economics*, 207(3), 104606. <https://doi.org/10.1016/j.jpubeco.2022.104606>
- 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>
- Saul, S. (2022, October 3). At N.Y.U., Students were failing organic chemistry. Who was to blame? *The New York Times*. <https://www.nytimes.com/2022/10/03/us/nyu-organic-chemistry-petition.html>
- Schneider, J., & Hutt, E. (2014). Making the grade: A history of the A–F marking scheme. *Journal of Curriculum Studies*, 46(2), 201–224. <https://doi.org/10.1080/00220272.2013.790480>
- Schultz, L. (2016). Resisting assimilation: Deliberate acculturation by the American English language learner. *Multicultural Perspectives*, 18(2), 103–107. <https://doi.org/10.1080/15210960.2016.1159100>
- Townsley, M. (2020). Grading principles in pandemic-era learning: Recommendations and implications for secondary school leaders. *Journal of School Administration Research and Development*, 5(S1), 8–14. <https://doi.org/10.32674/jsard.v5iS1.2760>
- Ward, L. W. M., & Hall, C. N. (2022). Seeking tenure while Black: Lawsuit composite counterstories of Black professors at historically White institutions. *Journal of Higher Education*, 93(7), 1012–1036. <https://doi.org/10.1080/00221546.2022.2082760>
- Witte, A. E. (2011). Understanding international grading scales: More translation than conversion. *International Journal of Management Education*, 9(3), 49–59. <https://doi.org/10.3794/ijme.93.319>

### Authors

TRANG PHAM is a lecturer at the Singapore University of Social Sciences, Singapore; [phamtrang4652@gmail.com](mailto:phamtrang4652@gmail.com). She studies educational and immigration policy and teaches social policy and social sciences courses.

STEPHANIE POTOCHNICK is an associate professor in Sociology and Public Policy at the University of North Carolina at Charlotte; [spotochn@charlotte.edu](mailto:spotochn@charlotte.edu). Her research examines the social demography of immigration and how programs and policies can promote the education and health of immigrant children and families.