International Journal of Educational Research 137 (2026) 102921

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Are teachers objective and impartial in evaluating business school students?

Petr Houdek *, Simon Senčar, Štěpán Bahník

Faculty of Business Administration, Prague University of Economics and Business, Czech Republic

ARTICLE INFO

Keywords:
Business schools
Grading
Grade enhancement
Student assessment
Ethical behavior

ABSTRACT

This paper investigates the prevalence, mechanisms, and drivers of grade enhancement at a large Central European business school, examining how often teachers raise marks, why they do so, and whose grades benefit. A sequential mixed-methods approach was applied: in-depth interviews with academic staff revealed motives and practices (Study 1); next, an anonymous faculty survey assessed self-reported behaviours (Study 2); finally, an archival analysis of 873,899 course outcomes spanning ten academic years searched for statistical abnormalities indicative of grade enhancement, such as the bunching of scores just above pass or letter-grade thresholds (Study 3). We found that approximately 70 % of instructors acknowledged at least occasional grade enhancement. Scores cluster immediately above key thresholds; typically, one or two points are presumably added instead of penalizing borderline cases, supporting the notion of "helpful rounding". The dominant motives are empathy, perceptions of measurement imprecision, and recognition of student effort, whereas institutional pressure via student evaluations of teaching plays a smaller role. Contrary to common claims in the literature, female instructors are not more lenient than male colleagues; the practice has not intensified over the decade, and a small but significant bias favors female over male students. The study offers one of the largest multi-method examinations of grade enhancement, introduces "helpful rounding" as a distinct micro-mechanism, and extends a literature dominated by Anglo-American contexts to a post-transition European setting.

1. Introduction

The average grades of university and business school students have been steadily improving, and concerns over grade inflation have been raised, worrying that the trend is caused by lowered evaluation standards (Peng Lin, 2021; Vítečková & Houdek, 2025), the underfunding of universities at a time of the massification of higher education (Jephcote et al., 2021), and by universities artificially boosting their students' competitiveness in the job market (Funke et al., 2011). Grade inflation may also be a consequence of the introduction of student evaluation of teaching (SET) measures (Eiszler, 2002; Ewing, 2012; Johnson, 2013; Wang & Williamson, 2022) and the broader commercialization of education that has led to the empowerment of students (Sherman, 2020). At the same time, the amount of time allocated by students for studying dropped by a third (Babcock & Marks, 2011), and a similar decrease in learning effort had been recorded (Kuh & Hu, 1999).

* Corresponding author at: Faculty of Business Administration, Prague University of Economics and Business, náměstí Winstona Churchilla 4, Prague 130 67, Czech Republic.
E-mail address: petr.houdek@gmail.com (P. Houdek).

https://doi.org/10.1016/j.ijer.2025.102921
Received 5 August 2025; Received in revised form 18 December 2025; Accepted 18 December 2025
Available online 2 January 2026

Grade enhancement, i.e., practices leading to the upward shifting of grades without a corresponding increase in learning or performance, poses a severe problem for the education system, primarily due to the loss of the informative value of grades and the wiping away of the difference between levels of knowledge, skills, or diligence of students (Nata et al., 2014; O'Halloran & Gordon, 2014).

The issue with such conclusions is their aggregate nature and the inability to filter out the influence of a process of continuous improvement of higher education (both in educational content and technology), the possibility that students getting more intelligent, and a shift towards learning that is focused on reasoning and problem-solving instead of memorizing factual information (Darling-Hammond et al., 2020). On the other hand, others argue that, especially in business schools, there is an increased focus on narrowly defined degrees targeted for specific job requirements instead of enhancing students' critical thinking abilities. Students are then perceived only as purchasers of services (Beatty, 2004; Natale & Doran, 2012).

Moreover, there is a near absence of research on personal differences in teachers' reasoning and individual decision-making leading to enhanced grades in specific academic environments. The research does not detail the various preferences and mental models of teachers who assign the grades (Brooks, 2012; Tóth & Csapó, 2022). As Jephcote et al. (2021; p. 567) wrote, there is "a need to investigate further the local institutional contextual factors that inform grade distribution."

In this article, we seek to identify, describe, and explain the grade-enhancing practices of teachers at a large Central European business school. We adhere to the concept of grade enhancement, rather than grade inflation, because we focus directly on enhancing grades and do not examine the time-varying factors underlying inflation. We investigated whether grade enhancement occurs through the artificial addition of grade points. Various motivations for grade enhancement, its influencing factors, and justification strategies were investigated using qualitative interviews with the academic staff (Study 1), teachers' surveys (Study 2), and an analysis of a large data set of students' grades (Study 3).

We employ a sequential mixed-methods design to move from discovery to confirmation and to reduce any mono-method bias through triangulation. Study 1 (qualitative interviews) elicits instructors' motives, practices, and the language they use around grade decisions, generating initial mechanisms that directly inform the construction of survey items and response options in Study 2 (anonymous survey). Study 2 estimates the prevalence of specific practices and self-reported drivers across a broader faculty sample, examines associations among them (e.g., discretion, perceived inaccuracy, empathy), and identifies patterns suitable for behavioral testing. Finally, Study 3 (archival analysis) tests the predicted behavioral signature (i.e., decreased frequency of scores just below grade thresholds), thereby validating self-reports against objective distributions, assessing stability over time, and probing subgroup differences relevant to equity. Together, the studies provide complementary evidence: Study 1 identifies mechanisms and themes; Study 2 establishes prevalence and the correlational structure of those mechanisms; and Study 3 offers behavioral corroboration.

2. Grade enhancement at universities

Teachers often help students to obtain better grades than they deserve. They are motivated either internally, i.e., by altruism or the wish to correct pre-existing inequalities or an unfair evaluation system (see more in the section Values and characteristics of teachers enabling grade enhancement); or externally, by supervisor pressure, financial incentives, or learning targets set by educational authorities (see more in the section Institutional factors behind grades enhancement). Sometimes, entire universities support morally questionable grade enhancement (Farley, 1995) and ultimately contribute to grade inflation (Bachan, 2017; Wilson, 1999). Multiple mechanisms facilitate this: using overly lenient grading standards, turning a blind eye to students' cheating, permitting the retaking of failed tests, shrinking the course syllabi, accepting late assignments, giving easy exams, or changing the process of calculation of final grades after bad outcomes (Hassan et al., 2020; Zabaleta, 2007). However, rising grades may also be a result of the increased learning effort, knowledge, or intelligence of students (Trahan et al., 2014).

2.1. Institutional factors behind grades enhancement

Institutionalized grade enhancement or inflation have several causes. First, starting in the US and now appearing in other countries, students are perceived as customers, fueling the demand for higher education, and universities feel obliged to keep them happy. Students' perceptions are also changing; they expect to be provided with academic success, not just the opportunities to achieve it (Singleton-Jackson et al., 2010). Entitlement to academic success is correlated with an overestimation of one's own knowledge and performance (Bertl et al., 2019), which can lead to increased pressure on teachers to grade in line with their students' expectations. The commercialization of higher education, along with rising tuition fees, has strengthened the power of the student body (Sherman, 2020).

Second, universities strive to enhance the competitiveness of their students in the job market to attract more applicants. As a case in point, Loyola Law School in Los Angeles decided in 2010 to artificially raise its students' GPAs by 0.333, meaning that all A- grades automatically changed to A, and this change was applied to all grades awarded from 2004 onward (Funke et al., 2011).

Third, grade inflation may also be caused by the higher prevalence of SET measures. The reliance on SET has been found to affect grading (Borch et al., 2020), particularly in cases where student evaluation is highly emphasized, and teachers' financial remuneration or career advancement may be at stake (O'Halloran & Gordon, 2014). Instructors respond by grading more leniently, virtually "bribing" students to earn better evaluations (Clayson et al., 2006; Ewing, 2012; Zabaleta, 2007). Kogan et al. (2022) argue that the primary driver of the correlation between students' grades and their teaching evaluations is grade satisfaction, meaning that students who receive higher grades tend to evaluate their instructor more positively. A meta-analysis conducted by Wright and Jenkins-Guarnieri (2012) estimated that 10 % of the variance in SET measures could be attributed solely to student achievement.

Fourth, conflict avoidance among teachers may be another factor that affects grades (Iris Franz, 2010). Students often feel entitled

to receive a good grade based on their own perception of the effort they put into the course. Dissatisfied students often express their dissent by complaining to classmates, disparaging the teacher, or confronting the teacher (Frisby et al., 2015). When they choose confrontation, they may try to pressure the teacher into changing an evaluation or granting preferential treatment. More agreeable teachers prefer to resolve such conflicts through compromise or back down completely (Jensen-Campbell & Graziano, 2001). Therefore, agreeable teachers, who are predominantly female (see the next section and Costa et al., 2001), may choose to acquiesce to a better grade, even if undeserved.

2.2. Values and characteristics of teachers enabling grade enhancement

Teachers often justify grade-enhancing practices by appealing to the greater good, altruistic values (e.g., increasing students' chances in the job market, helping students win scholarships), or denying responsibility for and the possibility of truly fair evaluation (Diamond & Persson, 2016; Hassan et al., 2020).

Altruism can bolster unethical behavior (Babula et al., 2020), as teachers may enhance student grades by adopting a selfless vision (Karylowski, 1984). To illustrate altruistic but unethical behavior, Babula et al. (2020) allowed participants to decide the punishment for university students who committed plagiarism, examining the effect of the perceived wealth of the offenders. Surprisingly, up to 70 % of participants voted to abolish the punishment for the poor plagiarist, while simultaneously insisting on the punishment for the rich plagiarist. The research suggests that empathy and altruistic thinking may lead to questionable decisions. Similarly, teachers may desire to help students who struggle to meet certain standards or a grade threshold, motivated by compassion. Gino and Pierce (2010) stress that the higher the level of empathy, the higher the propensity to act unethically in the name of altruism. Similar findings indicate that, despite an aversion to lying, a large group of people lie for the benefit of others (Erat & Gneezy, 2012).

Assessment subjectivity in standards-based educational systems may also be a factor influencing teachers' grading practices. Generally, teachers admit that their grading practices are driven by a sense of what is fair to students (Tierney et al., 2011) and are not strictly based on students' academic achievement. Green et al. (2007) examined ethics in classroom assessment and found that a significant number of grading practices perceived by teachers as ethical are considered unethical by professional standards. Teachers at all levels of the education system allow non-achievement-related factors, such as student activities, effort, or class participation, to influence the grading of students (Chen & Bonner, 2017). Since these factors are hard to objectively and precisely assess, room for grade enhancement emerges.

2.2.1. Seniority effects on grading

Adjunct instructors tend to be more lenient in their grading practices. Kirk and Spector (2009) analyzed the grading of adjunct instructors versus full-time faculty who taught the same courses and found that the awarded grades by adjunct instructors are significantly higher. This leniency might be caused by adjuncts scoring worse in student evaluations than the full-time faculty and seeking to boost student feedback by assigning better grades. They feel less obligation to maintain academic integrity and thus may overlook student cheating, leading to better test scores (Sonner, 2000). Student feedback becomes less important for tenured faculty members, as they no longer feel insecure about their job and do not desire high student evaluations to advance in their careers (Griffith & Sovero, 2021). Accordingly, older teachers tend to grade more strictly, and this is also supported by a finding that as teachers gain experience, their grading standards tend to rise (Gershenson, 2020). Generally, the young age of instructors is associated with awarding higher grades.

2.2.2. Gender effects on grading

The effects that a student's or teacher's gender may have on grading are a topic that has not been heavily studied. In general, women tend to demonstrate a greater aversion to conflict than men, who, in turn, are more likely to enjoy arguments and disagreements (Wolak, 2020). Women are also more agreeable, warm, and open to feelings than men, this difference between genders being the greatest across Europe and North America (Costa et al., 2001). Further evidence is provided by Yang and Girgus (2019), who confirm that women prioritize maintaining positive social relationships and are willing to go out of their way to please others. Therefore, female instructors may prefer avoiding disputes over grades with students (especially males) and hence could be more lenient with their grading than male instructors. Consistently, male students receive more praise from female teachers than from male teachers (Malik et al., 2018). Women in insecure teaching roles award significantly higher grades than their tenured female colleagues (Griffith & Sovero, 2021), while among male instructors, no such trend was identified. This could be partly attributed to gender bias in student reviews: students tend to give female instructors lower ratings. Female teachers who award higher grades receive better feedback from students and thus improve their chances for a tenured position (Griffith & Sovero, 2021). The fact that male professors earn better student feedback if they demonstrate competence, but female professors must demonstrate both competence and warmth to obtain the same ratings, corroborates the negative bias in student reviews towards female instructors (Joye & Wilson, 2015).

The student's gender also plays a role in grading. Several studies suggest that a grading bias against male students exists. Berg et al. (2020) found that Swedish male students receive lower final grades in junior high school mathematics despite outperforming female students three months later in high school introductory testing. Protivínský and Münich (2018) also document that teachers' grading is biased in favor of girls due to their perceived superiority in non-cognitive skills, such as in-class behavior or homework completion, which influences teachers' grading decisions but not test scores. Male university students exert less effort and show weaker motivation toward academic performance compared to their female counterparts (Marrs & Sigler, 2012), which may lead to poorer academic outcomes.

2.2.3. Rounding tendency

Grade enhancement can also be a simple heuristic of teachers for round numbers, which manifests itself in the awarding of better, "rounded" scores. Humans have a strong preference for round numbers (Jain et al., 2020; Lynn et al., 2013), customers are much more likely to donate money at the checkout if the donation request rounds up the total they ought to give (Kelting et al., 2019), assets are being sold and bought excessively just at the time their price is round (Fraser-Mackenzie et al., 2015), and marathon runners' finishing times cluster around round numbers, e.g., 3:00, 3:30 or 4 h, but also at 10 or 15-minute marks (Allen et al., 2017). A preference for round price endings of 0 and 5 has been documented in markets for gold, stocks, foreign currencies, real estate, bond futures, and oil futures (Sonnemans, 2006). Round numbers are easier to remember, process, and compute than non-round numbers (Estelami, 1999). Better processing fluency and comprehension of round numbers (Kettle & Häubl, 2010) allow people to function in daily lives with lower mental effort.

2.3. Overview of research question and hypotheses

We investigate grade enhancement through a fairness-repair lens within standards-based grading (see Section 2.1 Institutional Factors Behind Grades Enhancement). Within standards-based grading, grades should represent demonstrated achievement, rather than a combination of achievement, effort, and conduct (Brookhart et al., 2016). When numeric totals near salient cut-points seem to slightly misstate judged mastery, instructors may apply micro-repairs (e.g., rounding up) as small organizational-justice corrections (Colquitt et al., 2013) that reconcile rigid procedures with outcome fairness (see Section 2.2 Values and Characteristics of Teachers Enabling Grade Enhancement). These actions may be motivated by empathic, prosocial concern, especially in concrete borderline cases with high perceived stakes; and since these may differ depending on the gender of students and teachers, we expect gender effects of grade enhancement (see Section 2.2.2 Gender Effects on Grading). Because the motivation to improve a point evaluation is higher if it leads to grade improvement, such micro-repairs should occur mostly when points are just below grade cutoffs (see Section 2.2.3 Rounding Tendency).

This theoretical framework leads us to a research question: How and why do instructors enhance grades?

Prior work suggests that both institutional pressures and instructor values may shape grading practices; however, the underlying processes remain underdescribed (e.g., rounding near thresholds, weighing effort). *This motivates an exploratory qualitative Study 1.*

H1: Instructors will report grade enhancement primarily to correct perceived measurement imprecision and to recognize effort, with institutional SET pressure being secondary. *Explored in Study 1; measured in Study 2.*

H2: There will be a gender difference in instructor leniency. *Assessed in Studies 2 and 3.*

H3: There will be gender asymmetry in grade enhancements among students. *Assessed in Study 3.*

H4 (Helpful rounding): Scores just below key thresholds (e.g., 60, 75, 90) will be less common. *Tested in Study 3.*

3. Study 1 – interviews with academic staff (Exploratory)

As a preliminary, exploratory step, we conducted in-depth interviews to identify mechanisms and language that shaped subsequent measurement in Study 2.

Prague University of Economics and Business is, with approximately 14,000 enrolled students, the largest business-focused university in the Czech Republic. It consists of six faculties, which have considerable autonomy but offer similar education in economics, business, IT, and related fields. At the university, courses are categorized into various types, including mandatory and elective language courses, field-of-study electives, field-of-study mandatory courses, and many more.

Grading at the university is universal across all six faculties and programs. The grading system has five grades: 1 (100 - 90 %; A) excellent, 2 (89 - 75 %; B) very good, 3 (74 - 60 %; C) good, 4+ (59 - 50 %; D) failed with the possibility of repeating the exam, and 4 (<60 %; E) failed. Some courses do not require a grade; instead, only an evaluation passed or failed (based on the 60 % threshold) is given. Students are evaluated based on either: a) the number of grade points earned during the semester and for the final exam, 100 points being maximum and equal to 100 %, or b) only the evaluation of the examiner on the final exam, in which case there may be no need to award an exact number of points, but instead just a grade.

To gain initial insight into how teachers perceive the grading process at the university, we used an in-depth interview method (Kvale, 1996).

3.1. Methods

3.1.1. Sample

Interviews were conducted with six academic employees of the university. The respondents were selected to represent the diversity of the academic staff, ranging from Ph.D. students to full professors. Both full-time employees and one part-time instructor were included in the study. Interviewees were teaching both mandatory and elective courses.

3.1.2. Procedure

The interviews consisted of 14 topics (open questions) and were conducted either personally or through a video call. The interviews began with factual personal questions to understand the interviewees' positions at the university. The subsequent questions were divided into four segments based on prior review. First, we examined the existence and potential external and internal motivations for grade enhancement, including pressure from supervisors or university management, financial incentives, and the values of the teacher.

Second, we asked about grade enhancement and the process of student assessment. Third, the values of teachers and the potential ethical dilemmas associated with grade enhancement were discussed. Finally, we focused on the influence that the student evaluation of teaching and student pressure may have on the assessment of the teachers.

The discussion was led by the second author of this study. Answers were recorded for detailed analysis; these recordings were subsequently deleted to ensure the anonymity of the interviewees. Detailed interview transcripts are available upon request. As the interviews were conducted in Czech, the transcripts are also in the Czech language. The transcripts in English in this article are selected extracts translated by the second author. Two researchers independently open-coded all transcripts, then iteratively built a codebook anchored in four a priori topic areas (motivations; assessment process; values/ethics; student/SET influence) (Braun & Clarke, 2021; Burnard, 1991). The preliminary results were consulted with other teachers, but no new topics or insights emerged; theoretical saturation was thus achieved.

3.1.3. Instruments (interviews)

The semi-structured guide covered four domains: (1) motivations and contexts for grade enhancement, (2) assessment processes and thresholding practices, (3) values and ethical considerations, and (4) perceived SET/student pressures. Prompts invited concrete examples (e.g., decisions at the threshold points) and probed rationales (accuracy vs. empathy). The guide underwent expert review by two senior faculty members to ensure face and content validity and was pilot-tested with one academic, who was not included in the final sample.

3.2. Results

We conducted a thematic analysis following Braun and Clarke's (2021) phases: familiarization, initial coding, theme development, review, definition, and reporting. We interpret grade enhancement through a fairness-repair lens within standards-based grading: instructors perceive small discrepancies between numeric scores and holistic judgments of performance and enact micro-repairs to reduce perceived unfairness.

3.2.1. Causes of grade enhancement

All interviewees agreed that they try to maintain the principles of objectivity in evaluating students. Interviewees agreed that e-tests, which are usually automatically evaluated, are more objective in assessing student knowledge, compared to the traditional oral examinations. The accuracy of oral examinations was deemed questionable due to the influence of superficial factors, such as the feelings of sympathy, drive, or charisma (or lack thereof) of some students, and the general subjectivity of the oral examination process. The interviewees argued that the nature of student evaluation (students can earn a certain share of points during the semester for activities, homework, projects, presentations, papers, and interim tests) inherently carries subjectivity. However, all interviewees revealed that they believe the final scores do not precisely correspond to the students' knowledge, while one interviewee went further to argue that small differences in point scores between students can be attributed to in-semester evaluation subjectivity. An interviewee who is also a course supervisor said:

> "Of course, subjectivity influences the evaluation. I try to apply a nuanced and holistic approach and take into account the work and [non-evaluated] activity of students during the whole semester, and I think this is the right approach."

When a student shows effort, is active during the semester, and might fall just short of some grade threshold, the interviewees are more likely to help them. One of the interviewees admitted to adding points in a case where a student explained that they needed to achieve a certain GPA to participate in an exchange program.

While student satisfaction could potentially influence bonuses for teachers, the interviewees stated that they do not feel pressured or have a financial incentive to minimize the number of students who fail a course or ensure that a certain grade point average is achieved. An adjunct instructor, who teaches an elective course, stated that grading is superfluous; they believe that in optional courses knowledge sharing is the priority, not strict grading.

3.2.2. Grade-enhancing practices

When interviewees want to help students, they often offer the possibility of retaking an exam (even after the scheduled exam period has ended). All interviewees said that it is not customary for them to give points for free; extra effort or demonstration of knowledge is required. The adjunct instructor confessed to reusing tests between course groups and using the same tests for years, allowing students to prepare for specific questions. The full professor said that the course curriculum has been reduced in recent years to minimize the number of students who fail the course. Moreover, they added:

> "If someone gets 58 or 59 points [60 points being a passing grade], we try to find a point or two, for example, we look at the electronic test and search for room to add a couple of points,"

indicating that even courses concluded with an e-test are not immune to grade enhancement.

Most interviewees admitted that they sometimes add a point or two if a student misses a grade threshold, although they usually demand extra work. In the process of deciding whether to enhance a grade or not, they rely on non-achievement-related factors, such as attendance or perceived effort. Moreover, interviewees affirmed that this practice is widespread at the university.

Interviewees who participate in oral examinations stated that they do not strictly adhere to the rule of evaluating student performance in the oral exam and then summing all points (points earned during the semester and points from the final examination).

Instead, they calculate the final grade points by considering the points awarded during the semester and allocating points from the final oral exam to achieve the desired grade.

3.2.3. Teachers' values

Across the interviewees, helping students and enhancing grades is not viewed as morally problematic, and it does not produce any negative emotions. One of the interviewees said:

> "I do not have the heart not to give a second chance to a student in distress".

The majority of interviewees stated that their primary goal is to motivate and give knowledge to the students, and that strict grading is not necessarily the best approach. Motivating students who show effort was stressed by one interviewee in this way:

> "If some moral dilemmas or feelings of regret [caused by enhancing grades] emerge, I am always able to rationalize what I am doing."

3.2.4. Influence of students

Our interviewees reported no personal experience of the positive correlation between lenient grading and good student evaluation of teaching, but stated that they can imagine this correlation exists. Nevertheless, full-time faculty interviewees also emphasized that student evaluations of teaching are taken seriously and are discussed regularly by university management. Corrective measures are implemented in cases of extreme student dissatisfaction, such as when the concerned teacher is unable to continue teaching the course. The faculty management usually confronts its subordinates in cases of negative feedback. The adjunct instructor stated that SET was not important to them; instead, they "valued personal feedback more." Interviewees consider student satisfaction to have a relatively small effect on any financial bonuses or remuneration they may receive.

A majority of interviewees stated that they and other teachers increasingly face unsubstantiated demands from students for retaking exams, the possibility of resubmitting term papers after the deadline, and grade enhancement. The frequency of such demands and the way they are handled differ among the interviewees. Some tend to comply with them, while others reject all requests. The youngest interviewee thinks that age and gender play a role in students' approach toward teachers. They added that:

> "Younger teachers, especially women, can become more often the target of aggressive and unfounded demands for grade enhancement."

3.3. Discussion

Low nuances and the perceived inaccuracy of the grading system seem to be the main reasons for teachers enhancing grades, which is mostly viewed as a morally right thing to do, a finding in line with Green et al. (2007). Grade enhancement might be further exacerbated by the method of awarding points for the final examination. When teachers calculate with points earned during the semester, they can adjust their final assessment of the student in a way that the sum of the points is just high enough to pass a grade threshold. This "helpful rounding" may provide a new insight into the topic of imprecise grading, which has been concerned mostly with the consideration of non-achievement factors in grading (Bowers, 2011; Chen & Bonner, 2017).

In cases where a course is concluded with a written test, teachers lack the opportunity for similar calculations, which may lead to more objective grading and less grade enhancement (although rounding may still occur in these cases as well). Adjunct instructors and younger women teaching elective courses may be most susceptible to lenient grading practices or grade enhancement, findings in line with existing literature (Griffith & Sovero, 2021; Kirk & Spector, 2009; Sonner, 2000). These teachers may not always have the authority to tame audacious students, leading to a greater pressure from those students.

In sum, the interviewed teachers do not regret grade enhancement, nor do they suffer from negative emotions, because they are easily able to rationalize or justify it, contradicting Hassan et al. (2020), who claim that grade-enhancing practices stir up internal conflict. Whether teachers wish to motivate students, appreciate their effort, or correct the inherently limited evaluations, this reasoning allows them to continue with grade-enhancing practices guilt-free.

4. Study 2 – survey among academic staff

Study 2 presents the large-sample empirical test of the themes identified in Study 1, using an anonymous faculty survey. We followed up with a representative questionnaire survey to confirm whether, why, and how grade enhancement occurs at the faculty level. The e-survey informed by the findings was sent through the Center for Science and Research of the Faculty of Business Administration (one of the six faculties of the university). Anonymous responses were collected for 10 days beginning on the 7th of April 2022 (with a reminder after four days).

4.1. Methods

4.1.1. Sample

The questionnaire was sent to all 178 academic employees of the faculty (full-time and part-time), and 68 responses were received, a response rate of 38.2 %. No respondents were excluded. The demographic characteristics and employment status of the sample are

shown in Table 1.

4.1.2. Materials

The questionnaire consisted of 9 questions concerning evaluation and grading strategies (see Table 2) and 4 questions related to demographic characteristics and employment status (shown in Table 1). It assessed: (a) grading procedure, (b) frequency of enhancement, (c) reasons (multi-select list reflecting Study 1 themes), (d) perceived pressures, and (e) student requests. Items were content-validated through expert review (by two faculty members not on the author team) and pre-testing with three academics to verify clarity and interpretation. Item order of the evaluation and grading questions was randomized to reduce carryover and demand effects. To mitigate response bias, the survey was anonymous, emphasized that there were no right or wrong answers, and avoided evaluative wording; however, we acknowledge that social desirability cannot be fully eliminated

4.2. Results

4.2.1. Grading

When asked whether they take into account other factors not explicitly stated in the course syllabus in the final evaluation of students, respondents answered on average around the midpoint of a 4-point scale, with 41 % of respondents agreeing that they take into account other factors (see Table 2 for all the discussed results). Most respondents indicated that they add the points from the final examination to the points awarded during the semester, and the sum of these points determines the final grade. Yet, around a third (33.8 %) of the teachers answered that they do not decide on final grades by making a simple sum of points that students obtained during the semester and at the final exam, some explicitly answering that they enhance grades close to thresholds (20.6 %) and some that they award points based on a grade for which they decided based on the knowledge demonstrated in the final examination (11.8 %). Overall, the respondents tended to believe that the students' scores reflect their actual knowledge and skills.

4.2.2. Grade enhancement

Only 30.9 % of respondents answered that they never grade more leniently or otherwise help students to get a better grade or pass the course. However, only 4.4 % answered that they help students routinely. Correspondingly, the respondents indicated that only a small minority of students ($Mdn = 2.8$ %, $IQR = 4$ %) contact them with requests for a more lenient evaluation, score enhancement, or a grade improvement. The most frequent reasons for grade enhancement were the humaneness of such behavior, the subjectivity of assessment, and viewing sharing knowledge as a priority over grading and testing.

Most respondents (52.9 %) indicated that they accede only to reasonable requests or only to requests from students who are missing, for example, one point (29.4 %). Only a minority of respondents (10.3 %) refuse all such requests. Reasons for approving students' requests were similar to reasons for grade enhancement, with subjectivity of assessment and viewing sharing knowledge as a priority, instead of grading and testing as the most common answers. The respondents mostly reported not feeling any pressure to evaluate students more leniently from the university administration, faculty, or supervisors.

4.2.3. Relationships between responses

Table 3 shows correlations between variables measured in the questionnaire. The results show that teachers who consider non-achievement-related factors when grading and teachers employing enhancement grading strategies tend to help students more. The teachers who report being asked less frequently by students to enhance their grades also feel less pressured to evaluate students more leniently. Academic degree, age, or gender did not correlate significantly with any of the other measures.

4.3. Discussion

Teachers who do not adhere to the required method of making a sum of assessment points from the semester tend to enhance grades. Cross and Frary (1999) and Chen and Bonner (2017) found that teachers allow non-achievement-related factors to influence their grading. We confirmed these findings and further demonstrated that these teachers are less objective and tend to give higher grades, compared to teachers who adhere to the established grading rules. Our research raises an important issue of the teachers who do not believe in the accuracy of the assessment system and seek to correct grades that, in their opinion, do not correspond to the students' knowledge. Moreover, most of the teachers also do not consider testing and grading a priority in their job. Academic enabling (McMillan, 2005) is thus widespread and may have significant consequences.

Following Hassan et al. (2020), we found that teachers who enhance grades primarily justify this by denying the possibility of an accurate and fair evaluation. A significant proportion of teachers who enhance grades are influenced by feelings of empathy towards

Table 1
Characteristics of the participants sample in Study 2.

Gender		Age		Highest academic degree attained		Position	
Male	63.2 %	< 30	2.9 %	Master	25.0 %	Adjunct instructors	10.3 %
Female	36.8 %	30–40	41.2 %	Doctoral	47.1 %	Full-time faculty	80.9 %
		41–50	25.0 %	Assoc./Full Professorship	27.9 %	PhD students	8.8 %
		> 50	30.9 %				

Table 2
Questions and answers in Study 2. Note that the questions and answers are not necessarily in the order displayed to the participants (they were randomized). The numbers in parentheses after answers denote coding used for the answers in a correlational analysis, with "NA" meaning that the answer was not used. Participants could have indicated multiple reasons in questions asking for reasons for helping and reasons for acceptance of students' requests. These questions were given only to those who indicated in a preceding answer that they help the students and that they accept at least some of the requests. The proportions of answers are computed only from these respondents for these two questions.

Variable name	Question	Answers	Proportion / M (SD)
	Do you take into account factors not explicitly stated in the course syllabus (e.g. demonstrated effort, activity, visits to consultation sessions) in the final evaluation of the student?		
Not considering other factors		Definitely yes (1) - Definitely no (4)	2.63 (1.02)
	How do you decide the final grade? In case you apply different criteria in different courses, please indicate your most typical approach.		
Enhancement grading strategy		I add the points from the final examination with the points awarded during the semester, the sum of the points makes the final grade (1)	57.4 %
		I calculate with the points earned during the semester, in case the total score is close to a grade threshold, I will give a better grade (for example, if the score is 59, I adjust my evaluation and enter 60 points to give the student a better grade) (3)	20.6 %
		Assuming the number of points from the semester is sufficient, I decide on a grade based on the knowledge demonstrated in the final examination and then enter the corresponding number of points (2)	11.8 %
		Regardless of the number of points earned during the semester, I decide on a grade based on the knowledge demonstrated in the final examination and then record the corresponding number of points (2)	1.5 %
		The final grade is determined automatically by completing an e-test or other form of automated assessment (1)	1.5 %
		Other (NA)	7.4 %
	By your estimation, to what extent, on average, do the final scores in the courses correspond to the actual knowledge and skills of VŠE students?		
Knowledge correspondence		Students' scores fully reflect their actual knowledge and skills (1) - Students' scores do not reflect their actual knowledge and skills (10)	3.62 (1.79)
	Do you grade more leniently or otherwise help students (e.g. by adding a point or several points) to get a better grade or pass the course?		
Helping students		No, never (1)	30.9 %
		Yes, but only very rarely (2)	35.3 %
		Yes, I offer students an extra activity for which they can earn extra points (3)	22.1 %
		Yes, sometimes (3)	7.4 %
		Yes, I do that routinely (4)	4.4 %
	Why do you grade more leniently or otherwise help students?		
Reasons for helping		I consider such behaviour to be human, it's a shame to lose credits or miss a good grade because of one or a few points	34.0 %
		I want to help students achieve their goals (e.g. go on Erasmus, get a scholarship, etc.)	14.9 %
		I want to avoid being confronted by a student	14.9 %
		I want the students to like me and my course	0.0 %
		Aware of the subjectivity and inaccuracy of the student assessment, I do not consider adding points to be problematic	36.2 %
		My priority is sharing knowledge, not grading and testing	55.3 %
		Other	12.8 %
	Out of 100 students you teach, approximately how many would approach you requesting easier grading, extra points, or an improved grade?		
Proportion of requesting students		0–100	4.90 (8.54)
	How do you deal with student requests for grade improvement?		
Acceptance of requests		I refuse any requests to improve grades or enhance scores (1)	10.3 %
		I only accede to reasonable requests (e.g., an error or ambiguity in the test assignment) (2)	52.9 %
		I usually accede to students who are missing, for example, one point (3)	29.4 %
		I accede to most requests (4)	2.9 %
		I accede to all requests (5)	0.0 %
		Students never approach me with requests to improve their grades (NA)	4.4 %
	Why do you accede to students' requests?		
Reasons for acceptance		I consider such behaviour to be human, it's a shame to lose credits or miss a good grade because of one or a few points	19.0 %
		I want to help students achieve their goals (e.g. go on Erasmus, get a scholarship, etc.)	17.2 %
		I want to avoid being confronted by a student	10.3 %
		I want the students to like me and my course	0.0 %
		Aware of the subjectivity and inaccuracy of the student assessment, I do not consider adding points to be problematic	43.1 %
		My priority is sharing knowledge, not grading and testing	34.5 %
	Do you feel pressure to evaluate students more leniently from the university administration, faculty, or supervisor?		
Felt pressure		Definitely yes (1) - Definitely no (4)	3.35 (0.88)

Table 3
Spearman correlations between variables measured in Study 2. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95 % confidence interval for each correlation. Provided age categories "< 30″, "30 - 40″, "41 - 50″, "> 50", were coded as 1–4. The academic degree was coded as 1–3 for master's degree, doctoral degree, and associate or full professor, respectively. * indicates $p < .05$. ** indicates $p < .01$.

Variable	M	SD	1	2	3	4	5	6	7	8	9
1. Not considering other factors	2.63	1.02									
2. Enhancement grading strategy	1.59	0.84	−0.11 [−0.35, 0.14]								
3. Knowledge noncorrespondence	3.62	1.79	−0.18 [−0.41, 0.06]	.00 [−0.25, 0.25]							
4. Helping students	1.85	0.74	−0.37** [−0.56, −0.14]	.29* [.04, 0.50]	.22 [−0.02, 0.43]						
5. Proportion requests	4.90	8.54	−0.09 [−0.32, 0.15]	−0.15 [−0.39, 0.10]	.06 [−0.18, 0.30]	−0.09 [−0.32, 0.15]					
6. Response requests	2.26	0.69	−0.08 [−0.32, 0.17]	.14 [−0.12, 0.38]	−0.12 [−0.35, 0.13]	.23 [−0.01, 0.45]	.04 [−0.21, 0.28]				
7. Pressure	3.35	0.88	.03 [−0.21, 0.27]	.08 [−0.17, 0.32]	−0.08 [−0.32, 0.16]	−0.08 [−0.31, 0.16]	−0.33** [−0.53, −0.10]	.12 [−0.13, 0.35]			
8. Degree	2.03	0.73	−0.07 [−0.30, 0.18]	.08 [−0.18, 0.32]	−0.00 [−0.24, 0.24]	−0.05 [−0.28, 0.19]	.01 [−0.23, 0.24]	−0.17 [−0.40, 0.08]	.10 [−0.14, 0.33]		
9. Age	2.84	0.91	.05 [−0.19, 0.28]	.01 [−0.24, 0.26]	−0.12 [−0.35, 0.12]	−0.21 [−0.43, 0.03]	−0.20 [−0.42, 0.04]	−0.08 [−0.32, 0.17]	.15 [−0.09, 0.37]	.19 [−0.05, 0.41]	
10. Woman	0.37	0.49	−0.08 [−0.32, 0.16]	−0.08 [−0.32, 0.17]	.01 [−0.23, 0.25]	−0.18 [−0.40, 0.06]	−0.11 [−0.34, 0.14]	.16 [−0.09, 0.39]	−0.10 [−0.33, 0.14]	−0.07 [−0.31, 0.17]	.17 [−0.07, 0.39]

students. On the other hand, we do not find evidence indicating that characteristics of teachers such as age, gender, or academic degree make them more prone to enhance grades, as was indicated in our Study 1 and was the case in earlier studies (e.g., Gershenson, 2020; Griffith & Sovero, 2021; Kirk & Spector, 2009; Sonner, 2000).

Approximately 5 % of students request grade enhancement from their teachers. Although women may seem to be more acquiescent, no statistically significant difference was found between genders in terms of how the requests are handled. We did not corroborate the findings that feminine traits identified by Costa et al. (2001) and Yang and Girgus (2019) lead to lenient grading or excessive acceptance of unsubstantiated student demands. Additionally, age does not appear to be a predictor of teachers acceding to student requests. Our results are in opposition to Gershenson (2020) who found that grading standards rose with teachers' ages. Since older teachers also receive fewer requests for grade enhancements, students are likely aware of their stricter norms; however, we are unable to answer whether the strictness of the older teachers is due to their higher standards, less pressure from the university administration, or their greater job security and tenure.

>40 % of respondents feel pressured by their supervisor to enhance students' grades, but they didn't report that they are enhancing grades more than their colleagues who felt no pressure. The felt pressure might also have been a consequence of more frequent requests for grade enhancement by students, which we found to be associated with the felt pressure.

Since the results of Study 2 are self-reported data and may be biased due to the demand effect and social desirability (Holden & Passey, 2009), we follow up with Study 3, which aims to analyze the actual grade points and grades awarded.

5. Study 3 – students' scores and final grades

Study 3 constitutes our primary behavioral test, utilizing data on grading over a whole decade to examine threshold bunching and subgroup patterns. We test whether the hypothesized helpful rounding mechanism leaves a detectable signature in grade records and whether it varies by time, course type, or gender of students and teachers.

5.1. Data

For the purpose of the final study, a sizeable amount of data on grading was obtained from the university. A total of 2006,928 entries were available, each containing information about a unique final evaluation of a student in a course. The available data begin with the winter semester of the academic year 2008/2009 and end with the summer semester 2020/2021. The data should represent nearly all final evaluations of university courses for students during this period. The data set contained this information: time period, course code, level of study, study form, course category, semester number of the student, total number of points awarded, final grade, order of the exam, gender of the student, and gender of the instructor.

For further analyses, we excluded 549,416 observations from academic years 2008/2009, 2009/2010, 2010/2011, because a significant number of entries missed the information about the number of points. This reflects the evolution of the university's grading system. Second, 395,328 observations that lacked information about the number of points obtained were excluded (as mentioned earlier, in some courses it is usual not to award points). Third, 67,275 observations with final evaluations that were not represented by a regular grade had to be excluded as well. This is because a course can be concluded with "pass/fail", "excused", and other non-graded outcomes irrelevant for our analysis concerning both grades and point scores. Fourth, we excluded 65,284 observations that were not from the first exam-try of each student because the number of points on the second try may not be independent of the number of points on the first try. Fifth, we excluded 5310 observations from courses in categories with fewer than 1000 observations. These are, for example, courses that belonged to minority specializations or study programs that operated only temporarily, or were elective courses available only for a limited number of study programs. Sixth, we excluded 42,082 observations from courses with fewer than 100

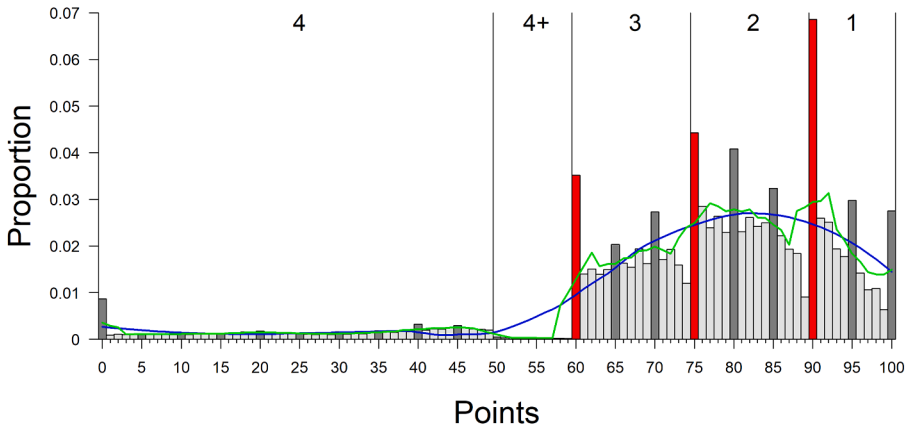

Fig. 1. Relative frequencies of given points in the data set. Thresholds for the three passing grades are shown in red. Other "round" numbers of points (divisible by 5) are shown in dark gray. A fitted local regression line is shown in blue, and a central moving average with a five-point window is shown in green.

observations. Seventh, we excluded 8332 observations where the given grade did not correspond to the assigned number of points, which may be, for example, a result of additional requirements for a passing grade. We checked whether there were any courses for which the number of points would be assigned just to correspond to the grade (if the points were always equal to 60, 75, 90, 100, or lower than 60), but there were none, so we did not exclude any data for this reason. The resulting data set contained 873,899 final evaluations.

5.2. Results

The relative frequencies of points are shown in Fig. 1. It is clear from the graph that the proportion of points just above grade thresholds is much larger than expected by chance. Furthermore, the proportions of points divisible by 5 are also larger than the proportions of points not divisible by 5, showing that teachers tend to round the points they assign students. The most frequent number of points is 90, corresponding to the points required for the grade "1" (A). Note that there are virtually no observations between 50 and 59 points because these points result in the grade "4+", which means that the student can retake an exam. Given that we selected only data from the first tries and the data contain only final evaluations, these observations are not included in the analyzed data.

The round numbers of points can be given directly without deliberately improving students' grades, simply due to the tendency to assign rounded numbers of points. A similar argument cannot be made when two non-round numbers of points are compared. That is, by comparing the proportion of points equal to 74 and 76 or 89 and 91, we can examine the tendency to improve the number of points if the points are just below the threshold for a better grade. We can expect that teachers do not have a similar motivation to decrease points when they are just above the threshold; therefore, the frequencies of points just below the threshold should be lower than those of points just above the threshold. As shown in Table 4, the difference between the proportion of points just above and just below round numbers of points was especially pronounced for the thresholds between grades. Moreover, for the other five examined rounded numbers, the difference between the proportions of points just above and just below the round number of points was significant in four cases, twice in each direction. The results suggest that teachers increase the number of points to improve a grade when students' points are close to the threshold for obtaining a better grade.

Next, we examined whether the studied difference in points just below and just above the threshold differs between male and female teachers and students. First, we conducted a mixed-effects linear regression analysis, including only data for points one point below and one point above the rounded point values between 60 and 95, as shown in Table 3, and excluding 150,497 observations without information on the teacher's gender. We created a new binary variable indicating whether the value was one point above a round value, which served as the dependent variable in the model. The centered gender of the teacher and the centered binary variable indicating whether the obtained points were near a threshold between grades served as predictors alongside their interaction. Random intercepts for the course and random slopes for the threshold variable, gender, and their interaction were included in the model to account for some dependency in the data. The dataset did not contain information about the teacher's identity, so we could not include this information in the model.

The gender of the teacher was not significantly associated with the proportion of studied values above the threshold, $t(266.7) = 1.43$, $p = .154$, $b = 0.006$, 95 % CI [−0.002, 0.014]. Replicating the effect shown in Table 4, the points near a grade threshold were

Table 4
The difference between the proportions of points one point above and one point below the rounded numbers of points. The highlighted rows show thresholds for a better grade. The p-value is obtained from a proportion test.

Points	Below	Above	Difference	*p*
65	1.49%	1.64%	0.14%	< .001
70	1.62%	1.71%	0.10%	< .001
75	1.20%	2.85%	1.65%	< .001
80	2.29%	2.31%	0.02%	.419
85	2.50%	2.22%	-0.28%	< .001
90	0.91%	2.60%	1.69%	< .001
95	1.77%	1.42%	-0.35%	< .001

more likely to be above a round number, $t(821.0) = 36.40$, $p < .001$, $b = 0.242$, 95 % CI [0.229, 0.255]. The threshold effect did not significantly differ between male and female teachers, $t(309.3) = 1.48$, $p = .140$, $b = 0.016$, 95 % CI [−0.005, 0.037].

We conducted a similar analysis with the student's gender rather than the teacher's as the predictor. The model did not include random slopes for the student's gender, which resulted in a singular fit of the model, suggesting overfitting. In comparison to male students, female students were more likely to have points above a rounded number than below, $t(206{,}104.7) = 2.10$, $p = .035$, $b = 0.004$, 95 % CI [0.000, 0.008], even though the effect was small. The effect of the grade threshold was again significant, $t(1063.3) = 41.67$, $p < .001$, $b = 0.249$, 95 % CI [0.238, 0.261], but was qualified by an interaction with student gender, $t(222{,}279.7) = 2.78$, $p = .005$, $b = 0.013$, 95 % CI [0.004, 0.022] (see Table 5). Students of both genders were more likely to have points above a rounded number when it was a threshold for a better grade, but the effect was slightly stronger for female students, $t(988.7) = 39.02$, $p < .001$, $b = 0.254$, 95 % CI [0.242, 0.267], than for male students, $t(939.3) = 34.95$, $p < .001$, $b = 0.237$, 95 % CI [0.224, 0.250].

We also examined the interaction between the genders of teachers and students. A linear mixed-effect regression model did not show a significant triple interaction between grade threshold, teacher's gender, and student's gender, $t(161{,}727.4) = 1.36$, $p = .173$, $b = 0.014$, 95 % CI [−0.006, 0.035].

Next, we examined the difference in the threshold effect between mandatory and elective courses using linear mixed-effects regression. The two-course types did not significantly differ in the proportion of points above a round number, $t(3564.6) = -0.19$, $p = .847$, $b = -0.001$, 95 % CI [−0.009, 0.007], and the threshold effect was not significantly different for the two-course types, $t(6664.4) = -0.14$, $p = .893$, $b = -0.001$, 95 % CI [−0.020, 0.017].

Finally, we employed the same approach to investigate the development of the threshold effect over time. The proportion of points above a round value did not significantly change in time, $t(595.7) = -0.33$, $p = .738$, $b = -0.002$, 95 % CI [−0.012, 0.009], and the threshold effect also did not change in time, $t(639.6) = 0.11$, $p = .916$, $b = 0.001$, 95 % CI [−0.023, 0.026].

5.3. Discussion

The relatively low frequency of points below grade thresholds shown in the study strongly suggests the presence of grade enhancement. This finding corroborates real data on the benevolent attitudes that teachers, on average, express towards students found in both our previous Studies 1 and 2. Assuming that the difference between the frequency of points just above and just below grade threshold values is fully explained by grade enhancement, the rate of grade enhancement would occur in around 3.3 % of given grades. This does not take into account the most important threshold between passing and failing the course, where grade enhancement may be even more prominent than in the case of the studied thresholds, which relate only to improving the grade and do not affect whether the student passes the course. Moreover, it is possible to imagine, as shown in Fig. 1, that teachers can improve grade points by more than one point. This would again increase the estimate of the rate of grade enhancement. Taking into account these additional factors, we can make a rough estimate of the rate of grade enhancement in our data to be about 5–10 % (the considerable size suggests that grade enhancement is more widespread than the results of Study 2 indicated). Using data from a 10-year period, we

Table 5
Differences in proportions of points one point above and below round numbers of points by teachers' and students' gender. The highlighted rows show thresholds for a better grade.

Points	Male teacher	Female teacher	Male student	Female student
65	0.14%	0.19%	0.15%	0.14%
70	0.05%	0.11%	0.08%	0.11%
75	1.54%	1.81%	1.62%	1.68%
80	0.03%	-0.04%	-0.01%	0.04%
85	-0.35%	-0.29%	-0.31%	-0.26%
90	1.39%	1.82%	1.47%	1.86%
95	-0.37%	-0.34%	-0.29%	-0.40%

also investigated whether grade enhancement has become more prevalent in recent years; however, we did not find evidence to support this trend.

We are unable to reject the possibility that other factors, such as arbitrary reference points (Kahneman, 1992) or higher student effort (Allen et al., 2017), may have an effect on the identified grade enhancement: further research would be needed to quantify these effects. However, students can barely know in advance how many points they will score on an exam, so it seems unlikely that they would be more motivated just below the threshold for a better grade.

Even though female teachers have been shown to be more agreeable and conflict-avoidant in previous studies (Costa et al., 2001; Yang & Girgus, 2019), in our sample they were not more prone to enhance grades. However, we found evidence of a bias against male students (Berg et al., 2020; Protivínský & Münich, 2018). Our results confirm that grading slightly favors female students over male students, as female students appear to be more likely to avoid falling just below a grade threshold with their points. Our data do not allow us to uncover the reasons behind this preference. Previous research has identified better course preparation, diligence, and desirable classroom behavior among female students as possible factors; however, our results suggest that their greater willingness and ability to negotiate grades may also play a role. Although previous research suggests that teacher-student gender interactions can either intensely help or harm relationships between the sexes in schools and organizations (e.g., Derks et al., 2016; Duffy et al., 2001; Moss-Racusin et al., 2012), we found no evidence that teacher-student gender interactions played a role in improving grades.

Even though students get significantly more points in elective courses than in mandatory courses,[1] the extent of grade enhancement (rounding up scores) did not differ. Therefore, it may seem that the better outcomes of students in elective courses are due to their greater interest and motivation rather than more frequent grade enhancements by adding undeserved points; however, we cannot rule out the possibility of a simpler curriculum or easier exams in elective courses.

The preference for round numbers has already been well-documented (Estelami, 1999; Lynn et al., 2013), but we were able to identify this heuristic in the assessment process and highlight its implications for grade enhancement.

6. Conclusion

The key takeaway from our research is that teachers view grade enhancement positively as a measure to correct the perceived inaccuracy of the grading system, demonstrate empathy for students, and consider non-achievement factors in assessment. These findings sit somewhat uncomfortably with accounts that frame grade enhancement primarily as the result of external incentives and institutional pressure, as well as the broader commercialization of higher education (Natale & Doran, 2012; Singleton-Jackson et al., 2010). Our results suggest that grade enhancement may persist even in settings where instructors report limited direct institutional pressure, because it is experienced as a legitimate fairness repair in response to perceived measurement imprecision and high-stakes borderline cases (Brookhart et al., 2016; Colquitt et al., 2013).

This paper also develops our understanding of how grade enhancement occurs. Prior work documents lenient standards and the mixing of achievement and non-achievement criteria, but often leaves the micro-mechanism underspecified (Brooks, 2012; Brookhart et al., 2016; McMillan, 2005). Our evidence points to a specific, behaviorally traceable mechanism, namely helpful rounding, where discretion is concentrated at salient cut-off points, resulting in systematic bunching just above the thresholds. This mechanism is consistent with well-documented round-number preferences and clustering at salient numeric points in other domains and helps reconcile two observations in the literature: grades can drift upward without a wholesale lowering of standards, and instructors can sincerely view their actions as both humane and defensible.

Our evidence also complicates common expectations about who enhances grades. Some prior work suggests that instructor characteristics and employment conditions (e.g., gender, job insecurity, adjunct status) may predict leniency through conflict avoidance or incentive sensitivity (Griffith & Sovero, 2021; Kirk & Spector, 2009; Sonner, 2000); nevertheless, we do not find robust evidence that female instructors are systematically more lenient, challenging an implied narrative that gender-linked traits translate straightforwardly into grading outcomes. At the same time, we detect a small but consistent female-student advantage near thresholds, which aligns with evidence of grading bias favoring girls reported in other educational contexts (Berg et al., 2020; Protivínský & Münich, 2018).

The student's partial success in the exam may have been driven by luck, and the subsequent grade enhancement then further exacerbates the injustice instead of rectifying it. Similarly, it could be argued that grades are meant to indicate knowledge and ability in a given field, not just effort. Thus, by grade enhancement, teachers may be giving employers a distorted picture of a student's abilities. Universities can preserve both informational value and humane discretion by adopting light-touch, process-oriented safeguards or nudges (Houdek, 2024; Say et al., 2026). A practical way to channel fairness-repair into consistent practice is to pre-commit, double-check, and offer a concrete regrading path. Before grading begins, a school can record rules for the assessment (e.g., how borderline cases will be handled and what evidence will justify reevaluation). Afterward, there could be a borderline double-check: the instructor re-evaluates the student's work in light of the rubric and, if the boundary is crossed, documents a rationale. Finally, once results are posted, students ought to have access to a transparent regrading path that requires them to reference specific rubric criteria, ensuring that post-hoc corrections (if warranted) remain tied to demonstrable evidence rather than negotiation.

Another avenue for improvement could be an instructor development workshop. Workshops can address when fairness-repair is appropriate versus distortive, showing how measurement error and rubric imprecision should be handled prospectively (through

[1] The median number of points is on average >5 points higher in predominantly elective than in predominantly mandatory courses, $t(1006) = 11.70$, $p < .001$, $d = 0.78$, 95% CI [0.64, 0.91], $M_{\text{elective}} = 85.35$, $M_{\text{mandatory}} = 79.38$.

rubric calibration and clear rounding policies) rather than retrospectively via ad hoc point additions.

The strength of our findings lies in the variety of research methods applied (qualitative approach, survey, and analysis of real grades), while the main results from all studies point in the same direction. While interviews with teachers offered an insight into the issue of student assessment and grading, the survey served to explore teachers' views. Since nearly 900 thousand observations derived from 10 years of student evaluation records were analyzed in Study 3, our results are highly reliable.

Compliance with ethical standards

All the procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethical approval was waived by the Research Ethics Committee at the Center of Science and Research at the Faculty of Business Administration at the Prague University of Economics and Business.

Informed consent

Informed consent was obtained from all individual participants included in the Study 1 and 2.

Funding

We acknowledge funding from the Excellent Teams Project of the Faculty of Business Administration, Prague University of Economics and Business (IP310031), and TAČR project (TQ01000052).

Data availability

The data are available from the authors upon reasonable request, but restrictions apply to the availability of these data as the grades data was provided by the university for the purposes of this research only. The data from the questionnaire survey can be provided freely.

CRediT authorship contribution statement

Petr Houdek: Writing – original draft, Funding acquisition, Conceptualization. **Simon Senčar:** Writing – original draft, Methodology, Data curation, Conceptualization. **Štěpán Bahník:** Writing – original draft, Methodology, Formal analysis.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Nikola Frollová and Karel Pernica for their help in obtaining the grade points data. We would also like to thank Marek Hudík, Nicolas Say, Klára Vítečková, and Marek Vranka for their helpful feedback on the survey. We would like to thank Douglas Shields Dix for his excellent proofreading of the manuscript. We are also grateful to three reviewers for their helpful comments on improving the manuscript. Part of this article was originally Simon Senčar's thesis.

References

Allen, E. J., Dechow, P. M., Pope, D. G., & Wu, G. (2017). Reference-dependent references: Evidence from marathon runners. *Management Science, 63*(6), 1657–1672.
Babcock, P., & Marks, M. (2011). The falling time cost of college: Evidence from half a century of time use data. *The Review of Economics and Statistics, 93*(2), 468–478.
Babula, M., Tookey, M., Muschert, G., & Neal, M. (2020). Altruism as a pathway to unethical action. *International Journal of Ethics and Systems, 36*(2), 235–247.
Bachan, R. (2017). Grade inflation in UK higher education. *Studies in Higher Education, 42*(8), 1580–1600.
Beatty, J. E. (2004). *Grades as money and the role of the market metaphor in management education, 3* pp. 187–196). Academy of Management Learning & Education.
Berg, P., Palmgren, O., & Tyrefors, B. (2020). Gender grading bias in junior high school mathematics. *Applied Economics Letters, 27*(11), 915–919.
Bertl, B., Andrzejewski, D., Hyland, L., Shrivastava, A., Russell, D., & Pietschnig, J. (2019). My grade, my right: Linking academic entitlement to academic performance. Social Psychology of Education : An International Journal, 22(4), 775–793.
Borch, I., Sandvoll, R., & Risør, T. (2020). Discrepancies in purposes of student course evaluations: What does it mean to be "satisfied"? *Educational Assessment, Evaluation and Accountability, 32*(1), 83–102.
Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141–159.
Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage.
Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803–848.
Brooks, V. (2012). Marking as judgment. *Research Papers in Education, 27*(1), 63–80.
Burnard, P. (1991). A method of analysing interview transcripts in qualitative research. *Nurse Education Today, 11*(6), 461–466.
Chen, P. P., & Bonner, S. M. (2017). Teachers' Beliefs about grading practices and a constructivist approach to teaching. *Educational Assessment, 22*(1), 18–34.

Clayson, D. E., Frost, T. F., & Sheffet, M. J. (2006). *Grades and the student evaluation of instruction: A test of the reciprocity effect, 5* pp. 52–65). Academy of Management Learning & Education.
Colquitt, J. A., Greenberg, J., & Zapata-Phelan, C. P. (2013). What is organizational justice? A historical overview. *Handbook of organizational justice* (pp. 3–56). Psychology Press.
Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*(2), 322–331. https://doi.org/10.1037/0022-3514.81.2.322
Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12*(1), 53–72.
Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science, 24*(2), 97–140.
Derks, B., Van Laar, C., & Ellemers, N. (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *Special Issue: Gender and Leadership, 27*(3), 456–469.
Diamond, R., & Persson, P. (2016). *The long-term consequences of teacher discretion in grading of high-stakes tests (No. w22207)*. National Bureau of Economic Research.
Duffy, J., Kelly, W., & Walsh, M. (2001). Classroom interactions: Gender of teacher, gender of student, and classroom subject. *Sex Roles, 45*(9/10), 579–593.
Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education, 43*(4), 483–501.
Erat, S., & Gneezy, U. (2012). White Lies. Management Science, 58(4), 723–733.
Estelami, H. (1999). The computational effect of price endings in multi-dimensional price advertising. *Journal of Product & Brand Management, 8*(3), 244–256. https://doi.org/10.1108/10610429910272547
Ewing, A. M. (2012). Estimating the impact of relative expected grade on student evaluations of teachers. *Economics of Education Review, 31*(1), 141–154.
Farley, B.L. (1995). "A" Is for Average: The Grading Crisis in Today's Colleges. ERIC Clearinghouse. https://files.eric.ed.gov/fulltext/ED384381.pdf#page=46.
Iris Franz, W.-J. (2010). Grade inflation under the threat of students' nuisance: Theory and evidence. *Economics of Education Review, 29*(3), 411–422.
Fraser-Mackenzie, P., Sung, M., & Johnson, J. E. V. (2015). The prospect of a perfect ending: Loss aversion and the round-number bias. *Organizational Behavior and Human Decision Processes, 131*, 67–80.
Frisby, B. N., Goodboy, A. K., & Buckner, M. M. (2015). Students' Instructional dissent and relationships with faculty members' Burnout, commitment, satisfaction, and efficacy. *Communication Education, 64*(1), 65–82.
Funke, R. D., Potthast, A., Boxall, S. F., Diaz-Sprague, R., Myers, G. A., & Funke, M. B. (2011). Retroactive grade inflation. *Association for Practical and Professional Ethics*. http://ethics.iit.edu/eelibrary/node/8547.
Gershenson, S. (2020). Great expectations: The impact of rigorous grading practices on student achievement. *Thomas B. Fordham Institute*, 48. https://fordhaminstitute.org/national/research/great-expectations-impact-rigorous-grading-practicesstudent-achievement.
Gino, F., & Pierce, L. (2010). Robin Hood under the Hood: Wealth-based discrimination in illicit customer help. *Organization Science, 21*(6), 1176–1194.
Green, S. K., Johnson, R. L., Kim, D.-H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education, 23*(7), 999–1011.
Griffith, A. L., & Sovero, V. (2021). Under pressure: How faculty gender and contract uncertainty impact students' grades. *Economics of Education Review, 83*, Article 102126.
Hassan, A., Mohamed, F., & Nelson, J. (2020). Cognitive dissonance in higher education: Inflating grades and not feeling guilty. *Journal of Higher Education Theory and Practice, 20*(16), 74–89.
Holden, R. R., & Passey, J. (2009). Social desirability. *Handbook of individual differences in social behavior* (pp. 441–454). The Guilford Press.
Houdek, P. (2024). Nudging in organizations: How to avoid behavioral interventions being just a façade. *Journal of Business Research, 182*, Article 114781.
Jain, G., Gaeth, G. J., Nayakankuppam, D., & Levin, I. P. (2020). Revisiting attribute framing: The impact of number roundedness on framing. *Organizational Behavior and Human Decision Processes, 161*, 109–119.
Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality, 69*(2), 323–362.
Jephcote, C., Medland, E., & Lygo-Baker, S. (2021). Grade inflation versus grade improvement: Are our students getting more intelligent? *Assessment & Evaluation in Higher Education, 46*(4), 547–571.
Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education, 28*(1), 91–105.
Joye, S., & Wilson, J. H. (2015). Professor age and gender affect student perceptions and grades. *Journal of the Scholarship of Teaching and Learning, 15*(4), 126–138.
Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes, 51*(2), 296–312.
Focus of Attention and altruism. In Karylowski, J., Staub, E., Bar-Tal, D., Karylowski, J., & Reykowski, J. (Eds.), *Development and maintenance of prosocial behavior: International perspectives on positive morality*, (pp. 139–154). (1984) (pp. 139–154). Springer US.
Kelting, K., Robinson, S., & Lutz, R. J. (2019). Would you like to round up and donate the difference? Roundup requests reduce the perceived pain of donating. *Journal of Consumer Psychology, 29*(1), 70–78.
Kettle, K., & Häubl, G. (2010). Numeric fluency and preference. *Advances in consumer research* (p. 37).
Kirk, F., & Spector, C. A. (2009). A comparison of the achievement of students taught by full-time versus adjunct faculty in business courses. *Academy of Educational Leadership Journal, 13*(2), 73–81.
Kogan, V., Genetin, B., Chen, J., & Kalish, A. (2022). Students' Grade satisfaction influences evaluations of teaching: Evidence from individual-level data and an experimental intervention. *EdWorkingPaper*.
Kuh, G. D., & Hu, S. (1999). Unraveling the complexity of the increase in college grades from the Mid-1980s to the Mid-1990s. *Educational Evaluation and Policy Analysis, 21*(3), 297–320.
Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.
Lynn, M., Flynn, S. M., & Helion, C. (2013). Do consumers prefer round prices? Evidence from pay-what-you-want decisions and self-pumped gasoline purchases. *Journal of Economic Psychology, 36*, 96–102.
Malik, R., Javed, N., & Dilshad, M. (2018). Inquiry into gendered teacher-student classroom interactions. *Pakistan Journal of Education, 35*(2). https://www.proquest.com/docview/2362278552/abstract/E5DA587F65DC429APQ/6.
Marrs, H., & Sigler, E. A. (2012). Male academic performance in college: The possible role of study strategies. *Psychology of Men & Masculinity, 13*(2), 227–241.
McMillan, J. H. (2005). Understanding and improving teachers' Classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34–43.
Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings Of The National Academy Of Sciences, 109*(41), 16474–16479.
Nata, G., Pereira, M. J., & Neves, T. (2014). Unfairness in access to higher education: A 11 year comparison of grade inflation by private and public secondary schools in Portugal. *Higher Education, 68*(6), 851–874.
Natale, S. M., & Doran, C. (2012). Marketization of education: An ethical dilemma. *Journal of Business Ethics, 105*(2), 187–196.
O'Halloran, K. C., & Gordon, M. E. (2014). A synergistic approach to turning the tide of grade inflation. *Higher Education, 68*(6), 1005–1023.
Peng Lin, Y. (2021). The distorted grade distribution: Inflating grades post the 1980s. *Journal of Higher Education Theory and Practice, 21*(5), 115–125.
Protivínský, T., & Münich, D. (2018). Gender bias in teachers' grading: What is in the grade. *Studies in Educational Evaluation, 59*, 141–149.
Say, N., Houdek, P., & Vrbová, L. (2026). Improving Forecasting Abilities of Business Students: The Effectiveness of Nudges on Performance Heterogeneity. *Mimeo.*.
Sherman, E. (2020). College tuition is rising at twice the inflation rate—While students learn At home. *Forbes*. https://www.forbes.com/sites/zengernews/2020/08/31/college-tuition-is-rising-at-twice-the-inflation-rate-while-students-learn-at-home/.
Singleton-Jackson, J. A., Jackson, D. L., & Reinhardt, J. (2010). *Students as consumers of knowledge: Are they buying what we're selling?, 35* pp. 343–358). Innovative Higher Education.
Sonnemans, J. (2006). Price clustering and natural resistance points in the Dutch stock market: A natural experiment. *European Economic Review, 50*(8), 1937–1950.
Sonner, B. S. (2000). A is for "adjunct": Examining grade inflation in higher education. *Journal of Education for Business, 76*(1), 5–8.

Tóth, E., & Csapó, B. (2022). Teachers' beliefs about assessment and accountability. *Educational Assessment, Evaluation and Accountability, 34*(4), 459–481.
Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' Interpretations of principles for standards-based grading. *The Educational Forum, 75*(3), 210–227.
Trahan, L., Stuebing, K. K., Hiscock, M. K., & Fletcher, J. M. (2014). The Flynn Effect: A meta-analysis. *Psychological Bulletin, 140*(5), 1332–1360.
Vítečková, K., & Houdek, P. (2025). The rise of business education, the ESG revolution, and the limited impact on students' Values. *Interchange, 56*, 143–157.
Wang, G., & Williamson, A. (2022). Course evaluation scores: Valid measures for teaching effectiveness or rewards for lenient grading? *Teaching in Higher Education, 27*(3), 297–318.
Wilson, B. P. (1999). The phenomenon of grade inflation in higher education. *National Forum, 79*(4), 38–41.
Wolak, J. (2020). Conflict avoidance and gender gaps in political engagement. *Political Behavior*.
Wright, S. L., & Jenkins-Guarnieri, M. A. (2012). Student evaluations of teaching: Combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education, 37*(6), 683–699.
Yang, K., & Girgus, J. S. (2019). Are women more likely than men Are to care excessively about maintaining positive social relationships? A meta-analytic review of the gender difference in sociotropy. *Sex Roles, 81*(3–4), 157–172.
Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education, 12*(1), 55–76.