Journal of the Scholarship of Teaching and Learning, Vol. 24, No. 4, December 2024, pp.43-60.
doi: 10.14434/josotl.v24i4.35434

The Impact of Norm- and Criterion-Referenced Grading Systems on Students' Course-Related Expectations

Jingxuan Liu
Columbia Business School
jl6297@gsb.columbia.edu

Michelle Wong
Harvard University

Bridgette Hard
Duke University

Abstract: The present study examined how information about different grading systems affects students' course expectations, particularly in ways that may have downstream consequences for learning and other academic outcomes. In an online experiment using a preregistered design, we prompted two samples of current and recent college students (N = 547) with a hypothetical course that adopted either a norm- or criterion-referenced grading system, two common grading policies in higher education. We then examined students' expectations for their own course-related goals, perceptions, and behaviors. We found that, compared to criterion-referenced grading, norm-referenced grading led participants to expect higher performance-goal orientation, lower mastery-goal orientation, lower course self-efficacy, and fewer help-related behaviors. Norm-referenced grading also increased perceptions that the instructor believes intelligence to be nonuniversal and fixed (i.e., not malleable). Some effects of grading system were stronger for students from non-minoritized backgrounds and lacking prior experience with the assigned grading system. Although participants reported experiencing criterion-referenced grading more often in college, both grading policies were commonly experienced by participants. Our findings suggest that norm-referenced grading policies negatively impact student expectations. More broadly, these findings highlight the importance of grading policies in shaping students' course-related expectations.

Keywords: *Norm-referenced grading, Criterion-referenced grading, Competition, Achievement orientation, Grading policies*

When students enroll in a college course, they develop initial impressions based on cues from the instructor and classroom environment. Some of these cues can influence expectations even before a student has stepped foot in a classroom. For example, an instructor's welcome email before class can positively affect students' perceptions and attitudes, generating effects that may persist throughout the semester (Legg & Wilson, 2009). Research suggests that cues ranging from instructor behaviors to course policies can affect students' course perceptions and expectations (LaCosse et al., 2020; Muenks et al., 2020; Murphy & Taylor, 2012). We propose that a course policy that may be particularly important in shaping student expectations is how students will be evaluated and graded. The goal of the present study was to examine how information about different grading policies might affect students' course expectations, particularly in ways that shape downstream consequences for learning and other academic outcomes.

Norm- and Criterion-Referenced Grading Systems

There are many different grading policies that instructors can adopt. *Norm-referenced grading*, also known as "grading on a curve" or "relative grading," assigns grades based on how students' performance compares to the class (*e.g.,* the student earns an "A" if they perform better than 70% of the class, Knight, 2001). Instructors may prefer norm-referenced grading as it guarantees consistency and grants them more control over the grade distribution (Lok et al., 2016). *Criterion-referenced grading*, in contrast, assigns letter grades based on the quality of students' performance according to a predetermined set of criteria and cutoffs (*e.g.,* the student earns an "A" if they achieve an average of at least 90% of available class points). The benefits of criterion-referenced grading include presenting students with clearer goals to work towards (Aviles, 2001).

An instructor's grading policy may not fully disclose their evaluation methods. Regardless of the communicated grading system, they might aim for a specific grading distribution (e.g. 20% "A"s) using adjustments to grade cutoffs or exam difficulty. Nonetheless, communicating a course's use of norm- or criterion-referenced grading may influence different perceptions, behaviors, and outcomes via the different ideologies they convey. Although research in this area remains limited, two prior studies found that students who experienced criterion-referenced grading (vs. norm-referenced grading) showed higher motivation, viewed grades as reflective of their learning (Covington & Omelich, 1984), and demonstrated more improvement over time (Williams et al., 1975).

Instructors often outline their choice of norm- or criterion-referenced grading in the course syllabus, making it one of the first details students learn. Consequently, grading policies may critically shape students' initial expectations for the course. Early expectations can lead to long-term effects on students' behavior and performance through recursive or self-reinforcing processes: students act according to their expectations and receive feedback that reinforces those expectations (Brady et al., 2018; Cohen et al., 2009).

Given the potential importance of initial expectations, the present study investigated how norm- and criterion-referenced grading might affect what students expect from a course. In an online experiment, we simulated the presentation of a grading policy at the beginning of a hypothetical course, manipulating the policy to be norm- or criterion-referenced. Participants were surveyed about their motivation, beliefs about their ability, willingness to assist or seek help, and beliefs about the instructor. This investigation into the unexplored link between grading policies and course expectations aims to inform grading policy design to foster positive course experiences. In the sections that follow, we review relevant literature that informed our hypotheses.

Expected Achievement Orientation

Norm-referenced grading, by limiting high letter grades in a course, fosters a more competitive environment than criterion-referenced grading (Aviles, 2001; Covington & Omelich, 1984; Schinske & Tanner, 2014). This increased competition may influence student motivation, namely achievement goals. In the context of education, achievement goals--students' intentions behind learning activities (Meece et al., 2006)— are defined on two dimensions: mastery and performance. Mastery goals focus on acquiring knowledge for self-improvement, whereas performance goals emphasize being positively perceived for one's competence, especially relative to peers (Ames, 1992; Nicholls et al., 1990).

Research suggests that competitive course environments shape achievement goals (Ames & Ames, 1984; Meece, 1991), with students tending towards performance goals under high perceived competition (Lam et al., 2004; Urdan, 2004). Additionally, feedback that compares students often leads them to attribute success to inherent ability (*i.e.,* "I'm smart") rather than effort (*i.e.,* "I worked

hard"), particularly during difficult tasks (McColskey & Leary, 1985)—a feature of performance-orientation (Meece et al., 2006). Therefore, the competitive impression conveyed by norm-referenced grading might prompt students to pursue performance goals that focus on outshining peers. While mastery and performance goals are not incompatible (Barron & Harackiewicz, 2003), focusing on social comparison may divert student attention from mastery goals, which emphasize self-improvement.

Expected Self-Efficacy

Self-efficacy, defined as one's perceived capability to learn and perform effectively (Wentzel & Miele, 2009), is a strong predictor of academic success (*e.g.,* Hoigaard et al., 2015) and is negatively influenced by competition in the learning environment (Chan & Lam, 2008). Bandura (1997) noted that self-efficacy increases with higher perceived performance. Norm-referenced grading, by explicitly limiting high grades and fostering a competitive atmosphere through comparison-based feedback, may dampen students' self-confidence and self-efficacy in the course.

Expected Willingness to Give and Seek Help

By inducing competition, norm-referenced grading may also diminish students' willingness to help others and seek help for themselves, both crucial for learning and social bonding (Hertz-Lazarowitz, 1989; Karabenick & Newman, 2013). Competitively-focused people view peers as rivals, reducing their likelihood of collaboration (Darnon et al., 2006; Poortvliet & Darnon, 2010) and increasing antagonistic behaviors (Poortvliet et al., 2012). Moreover, norm-referenced grading could attach a cost to helping, as aiding a peer could jeopardize one's own success (Bell et al., 1995; Cook et al., 2013).

Similarly, norm-referenced grading could undermine students' help-seeking by affecting *classroom goal structures*—achievement goals perceived at the classroom level (Karabenick & Newman, 2013). Environments that emphasize learning promote help-seeking, whereas performance-oriented environments, where help-seeking may signal incompetence, discourage it (Bong, 2008; Butler & Neuman, 1995; Karabenick, 2004; Ryan & Pintrich, 1997). By emphasizing competition and performance, norm-referenced grading may increase the perceived risks of seeking help, reinforcing fears of negative judgment from peers or instructors.

Perceptions of the Instructor's Intelligence Beliefs

In addition to impacting students' expectations regarding their own abilities and behaviors, different grading policies may influence students' view of instructors' beliefs about intelligence. Instructors may subscribe to a *universal theory* of intelligence, believing that all individuals can achieve high intelligence, or a *nonuniversal theory,* believing that only some people are capable (Rattan et al., 2012). Instructors can also hold different beliefs regarding the malleability of intelligence (*i.e.*, mindsets). Those with *growth mindsets* believe that intelligence can be developed, whereas those with *fixed mindsets* believe that intelligence is fixed (Dweck, 2008). Perceptions of instructors' intelligence beliefs matter. Instructors perceived as holding a universal theory of intelligence or a growth mindset typically foster better student engagement, lower psychological distress, and narrower achievement gaps based on race and gender (Canning et al., 2020; Fuesting et al., 2019; LaCosse et al., 2020; Muenks et al., 2020; Rattan et al., 2018).

Students infer instructors' beliefs through various situational cues—verbal, non-verbal, syllabi, feedback, and policies (Murphy & Taylor, 2012; Muenks et al., 2020; Restrepo, 2020; Lou & Noels, 2020). Grading policies may serve as similar situational cues that affect students' perceptions of their

instructor's beliefs. By comparing students to one another and limiting success to a select few, a norm-referenced grading system may imply a belief in nonuniversal and fixed intelligence.

The Moderating Effects of Demographics and Student Experiences

Competition in academic settings may disproportionately affect certain student groups. Historically marginalized students, who are already disadvantaged in higher education (Stephens et al., 2014), may be particularly vulnerable to competition pressure. Both experimental and observational studies found that competitive environments make racially minoritized (RM) and first-generation (FG) college students feel less motivated (Sommet & Elliot, 2017), more like "imposters" (Canning et al., 2020), and experience higher levels of anxiety and depression (Posselt & Lipson, 2016) compared to their peers. Notably, such differences tend to narrow in more collaborative, less competitive settings, highlighting how RM and FG statuses may influence the impact of academic competition on student perceptions and learning (Canning et al., 2020). Given the above, RM and FG students might be particularly vulnerable to the competitive pressures exacerbated by norm-referenced grading policies, compared to criterion-referenced policies.

We also explored whether students' prior experiences with grading systems moderated their impact on student expectations. Given the relative rarity of norm-referenced grading, students' previous encounters with it might shape their current expectations. Without specific predictions, the present study sought to understand if experiences with norm-referenced grading heighten or dampen students' sensitivity to the signals of competition conveyed by the grading policy.

The Present Study

In this online experiment, we empirically investigated the effects of norm-referenced versus criterion-referenced grading on student outcomes. Based on prior literature, we hypothesized that participants exposed to norm-referenced grading would report: a higher orientation toward performance goals (H1), lower self-efficacy (H2), reduced willingness to help others (H3), and stronger beliefs that their instructor views intelligence as nonuniversal and fixed (H4). Additionally, we anticipated these participants would demonstrate less willingness to seek help (H5a) and perceive a greater threat in seeking help (H5b). Finally, we explored how racialized minority and first-generation (RM/FG) status, along with prior experiences with grading systems, might moderate these effects.

Method

Study Design

In a pre-registered experiment with 547 current and recent college students, we examined the differential effects of norm- and criterion-referenced grading systems on student expectations, perceptions, and behaviors related to a hypothetical course[1]. Participants were presented with descriptions of this course using one of the two grading policies. Subsequently, we measured their expected achievement goals, self-efficacy, help-related behaviors, and perceptions of the instructor's intelligence beliefs for the hypothetical course. This online experimental design allowed for precise control over the presentation of the grading policies and mitigated the influence of confounding

[1] All methods and analyses were pre-registered. The pre-registration is available here: https://aspredicted.org/blind.php?x=f8jh93

environmental and social factors that exist in real-life classrooms. The study was approved by the University Institutional Review Board, 2018-0046 (EO369).

Participants

Among the 550 participants we recruited, three were excluded due to missing data, lack of college experience, and misunderstanding of the assigned grading condition[2]. In the final sample of 547 participants, 52% identified as female, 42% identified as White[3], and 45% of participants were designated as racial-ethnic minoritized (RM: 23%) and/or first-generation college students (FG: 30%). RM participants were defined as identifying with racial-ethnic minorities (Black, Hispanic/Latino, American Indian or Alaska Native, Native Hawaiian or Pacific Islander, or Multiracial/Other students with at least one identity in the previous categories[4]). FG participants did not have at least one parent who completed a 4-year college degree. In later analyses, we combined participants with status as either RM or FG into a single group: RM/FG. Finally, more than three-quarters (78%) of participants were currently enrolled college students (see Supplemental Materials, Table S1).

Procedure

We recruited participants from two sources to ensure diverse and generalizable results. First, we recruited 330 participants from the university credit pool (SONA) at a selective, private university in the southeastern United States. Next, we recruited 220 participants from Amazon's Mechanical Turk (MTurk) via the CloudResearch platform. These participants were selected based on a minimum 95% approval rating, past college experience, and an age range of 18-30 years.

Participants were randomly assigned to either the criterion-referenced (n = 273) or norm-referenced grading condition (n = 274). They were then presented with descriptions of a hypothetical course employing their assigned grading policy (see Figure 1). After reading the description, participants answered two comprehension checks to confirm their understanding (*e.g.,* "If Sally performs exactly better than 45% of her peers, what grade will she receive?"), and reported their prior experience with such grading policies. They were also asked to write a few sentences describing their expected experience in the course, encouraging reflection on the assigned grading policy and validating their understanding. Participants next self-reported their anticipated experiences in the course, using measures detailed in the next section. Finally, participants reported demographic information including race and parental education status.

[2] 99% of the participants in the criterion-referenced condition and 86% in the norm-referenced condition correctly answered both comprehension check questions correctly. Nevertheless, all but one participant who failed both comprehension checks demonstrated proper understanding of their assigned grading condition, as reflected in their free-response question (judged by the first and third author). One participant who did not demonstrate understanding of the grading system was excluded.

[3] 10% did not report gender and 12% did not report race, likely due to an error in the Qualtrics system that affected the MTurk sample.

[4] Asians are not included in this group as they are generally not considered underrepresented in higher education settings

Norm-referenced condition

Imagine yourself as a college student taking a course that is important to you. The final grade of this course is determined by your standing in the class. Your final course score (an average of your exam and assignment scores across the semester) will be compared to that of your classmates, and your final letter grade will be calculated based on how your performance compares to the whole class. You will receive:

- An "**A**" if your score is in the top 15% of your class
- A "**B**" if your score is higher than 50%, but lower than the top 15% of students in your class
- A "**C**" if your score is higher than 20%, but lower than the top 50% of students in your class
- A "**D**" if your score is higher than 5%, but lower than the top 80% of students in your class
- A "**F**" if you fail to perform better than 5% of your classmates

Letter Grade	Performance Percentile
A range	85th-99th
B range	50th-85th
C range	20th-50th
D range	5th-20th
F	Below 5th

Criterion-referenced condition

Imagine yourself as a college student taking a course that is important to you. Your grade in the course will be calculated based on your own performance in the course, and a letter grade will be assigned based on your final course score (an average of your exam and assignment scores across the semester). You will receive:

- An "**A**" if your score is 90% and above
- A "**B**" if your score is between 80% and 90%
- A "**C**" if your score is between 70% and 80%
- A "**D**" if your score is between 60% and 70%
- A "**F**" if your score is below 60%

Letter Grade	Score
A range	90%-100%
B range	80%-90%
C range	70%-80%
D range	60%-70%
F	Below 60%

Figure 1. Prompts Used for Norm- and Criterion-Referenced Grading Conditions

Measures, A full and detailed list of all measures can be found in the Appendix.

Achievement Goal Orientation

Six items assessed participants' expected performance- and mastery-goal orientation (three items each) on a 5-point Likert scale, adapted from the Achievement Goal Questionnaire-Revised (e.g., "My goal for this course would be to perform well relative to other students*",* 1 = strongly disagree, 7 = strongly agree, Elliot & Murayama, 2008). Reliability was high for performance- and mastery-goal orientation scales (Cronbach's a = 0.85 and 0.83, respectively).

Class Self-Efficacy

Six items with 5-point Likert scales were adopted from the self-efficacy section of the Motivated Strategies for Learning Questionnaire (e.g., "I believe I will receive an excellent grade in this class", 1 = Strongly disagree, 5 = Strongly agree, Pintrich, 1991) and showed strong reliability (a = 0.88).

Help-Related Behaviors

Seven items with 7-point Likert scales were adapted from Fuesting et al., (2019) to measure how willing people are to offer help to their classmates in various situations (e.g., "Provide constructive feedback on a fellow student's project", 1 = extremely unlikely, 7 = extremely likely) (a = 0.95).

Six items measured participants' experience of threat related to help-seeking behaviors on a 5-point Likert scale (e.g., "I would feel like a failure if I need help in this class", 1 = Strongly disagree, 5 = Strongly agree). The measure consisted of items adopted from Karabenick, (2003), Karabenick & Knapp, (1991), and Payakachat et al. (2013) (a = 0.92).

We developed ten novel items that assessed participants' likelihood to engage in different types of common help-seeking behaviors in a course on 7-point Likert scales (e.g., "Attend office hours to ask for help from the instructor", 1 = Strongly unlikely, 7 = Strongly likely) (a = 0.81).

Perceived Instructor Beliefs

One question adapted from (Rattan et al., 2012) measured participants' perception of the instructor's beliefs about universal potential for intelligence (scale of 1 to 20, 1= perceived instructor belief in universal potential for intelligence). A second question measured participants' perception of the instructor's belief in students' ability to succeed in the course on a scale of 1 to 20 (1 = the instructor believes everyone can succeed in the course).

Four items adapted from Dweck (1999) asked participants to rate the instructor's likely beliefs about the malleability of intelligence on a 6-point Likert scale. Higher scores indicate participants' perception of a more "fixed" instructor mindset regarding intelligence (e.g., "The instructor in this class seems to believe that students either 'have it' or they don't", 1 = Disagree a lot, 6 = Agree a lot) (a = 0.91).

Results

Preliminary Analysis

We first examined whether the experimental manipulation exerted consistent effects across the two samples to determine whether they could be combined for subsequent analyses. We conducted a MANOVA with all dependent variables as outcome variables and grading condition, sample (SONA or MTurk), and their interaction as predictors. The analysis revealed a significant overall main effect of sample, $F(9, 469) = 5.91$, $p < 0.001$, $Wilk's\ \Lambda = 0.90$, as well as a significant main effect of grading condition, $F(9, 469) = 28.80$, $p < 0.001$, $Wilk's\ \Lambda = 0.64$. The interaction effect was not significant, $F(9, 469) = 1.46$, $p = 0.16$, $Wilk's\ \Lambda = 0.97$, however, indicating that the effect of grading condition did not differ significantly across the two samples. Thus, the samples were combined for subsequent analyses with sample as a control variable.

Differential Effects of Grading Conditions

Individual linear regression models[5] were used to determine whether student achievement motivation, course self-efficacy, help-related behaviors, perceived help-seeking threat, and perceived instructor mindset differed across the two grading conditions. Condition (0 = criterion, 1 = norm) was entered as the predictor of interest, and age, gender, RM/FG status, and sample (MTurk or SONA) were entered as control variables in the models. The model beta coefficients for condition are presented in Table 1.

[5] This analysis deviated from the pre-registered analysis plan of applying t-tests. We adopted this approach instead of the preregistered t-tests to better control for potentially confounding variables. The regression findings converged with the t-test approach.

Table 1. Results from regression analysis on the effect of condition.

Outcome Measure	Coefficient Estimate	Std Error	Test Statistic
Class self-efficacy (1-5)	-0.38	0.07	-5.57***
Performance-goal orientation (1-5)	0.67	0.08	8.71***
Mastery-goal orientation (1-5)	-0.29	0.08	-3.84***
Help-giving behaviors (1-7)	-1.22	0.13	-9.19***
Help-seeking behaviors (1-7)	-0.21	0.09	-2.33*
Help-seeking threat (1-5)	0.27	0.10	2.79**
Perceived instructor belief in universal potential for intelligence (1 = universal, 20 = nonuniversal)	3.52	0.51	6.91***
Perceived instructor belief in success (1 = universal, 20 = nonuniversal)	5.62	0.53	10.63***
Perceived instructor mindset (1 = growth, 6 = fixed)	1.12	0.12	9.7***

Note. Regression beta estimates for condition (0 = criterion-referenced, 1 = norm-referenced) in predicting each outcome variable. Gender, age, URM status, and data source (MTurk or SONA) were controlled for. The scale for each measure is provided in parentheses. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Consistent with our predictions, participants in the norm-referenced condition reported significantly higher orientation toward performance goals (H1 supported) and lower course self-efficacy (H2 supported) compared to those in the criterion-referenced condition. Although we did not specifically predict an effect of grading condition on mastery-goal orientation based on findings from

a pilot study[6], we found that participants in the norm-referenced condition reported significantly lower mastery-goal orientation in the present sample.

Participants in the norm-referenced condition anticipated lower willingness to offer help to classmates (H3 supported) and to seek help for themselves (H5a supported). They also reported more perceived threat around seeking help (H5b supported) compared to those in the criterion-referenced condition.

Finally, participants in the norm-referenced condition were more likely to perceive the instructor as believing that only some, instead of all, students can achieve a high level of intelligence and succeed in the course. These participants were also more likely to perceive the instructor as having a fixed mindset compared to those in the criterion-referenced condition (H4 supported).

Moderating Effects of RM/FG Status

In light of previous literature, we anticipated that RM/FG participants would be particularly vulnerable to the negative effects of norm-referenced grading. Initial pilot studies, which only assessed RM status, indicated that this demographic factor only affected help-related behaviors under different grading policies. Consequently, this study focused on whether RM/FG status would similarly affect responses to norm- and criterion-referenced grading, especially regarding help-giving, help-seeking, and perceived help-seeking threat.

We employed a linear regression model to assess moderation effects, including grading condition, RM/FG status, and their interaction as predictors. RM/FG status itself did not significantly predict any outcomes or moderate the effect of grading condition on help-seeking threat. However, it significantly moderated the impact of grading condition on help-giving behaviors (Figure 2). Specifically, norm-referenced grading reduced anticipated help-giving among both non-RM/FG (simple slopes: B = -1.60, $p < 0.001$) and RM/FG participants (simple slopes: B = -0.85, $p < 0.001$), with a more pronounced decrease observed in non-RM/FG participants when compared to those under criterion-referenced grading.

RM/FG status similarly moderated the effect of condition on help-seeking behavior. Compared to the criterion-referenced condition, norm-referenced grading decreased help-seeking behaviors for non-RM/FG participants significantly (simple slopes: B = -0.42, $p < 0.001$), but not for RM/FG participants (simple slopes: $B < 0.001$, $p = 1.00$).

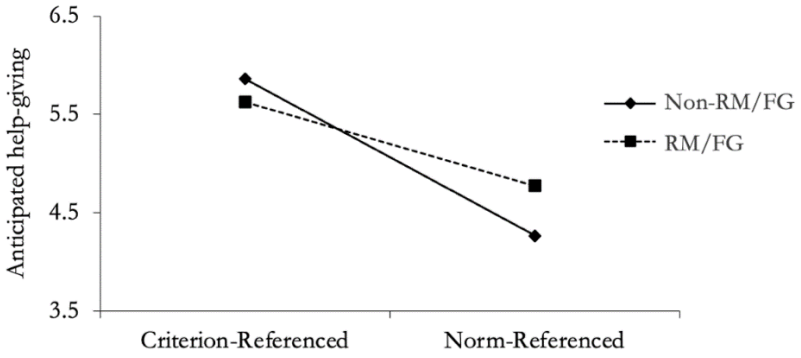

Figure 2. Help-Giving Behaviors (1 = Extremely unlikely, 7 = Extremely likely) as a Function of Grading Condition and RM/FG Status. The negative effect of norm-referenced grading condition on help-giving was more pronounced for non-RM/FG students.

[6] Please see https://osf.io/4vbjw?view_only=684cbd98a0d54f4284a291c29226b0e0 for more details about the pilot study findings

Moderating Effects of Participant Experience

The present study also provided an initial exploration of college students' experience with norm- and criterion-referenced grading. Our findings suggest that norm-referenced grading was commonly experienced by the students we sampled (nearly 40% experienced it at least once), but criterion-referenced grading was more familiar (82.3% of participants indicated experiencing it at least once). As not all participants reported experiencing the grading policy they were assigned to, we performed an exploratory, non-preregistered analysis of whether prior experience moderated grading condition's effect on participant expectations. We conducted a MANOVA test on all outcome measures, with condition, experience with the assigned grading system (have or have not experienced the grading system at least once), and their interaction as predictors. Grading condition had a significant main effect on the collection of outcomes, $F(9, 469) = 29.90$, $p < 0.001$, $Wilk's\ \Lambda = 0.64$, consistent with our previously reported findings, but it also interacted significantly with experience, $F(9, 469) = 3.06$, $p = 0.001$, $Wilk's\ \Lambda = 0.94$.

Given the significant overall interaction effect, we performed follow-up linear regression models for each outcome measure with the same predictors. These analyses revealed a significant interaction between experience and condition in predicting help-giving, performance-goal orientation, and perceived instructor mindset. A close look within each condition revealed that, in the norm-referenced condition, lacking prior experience with that grading system significantly decreased anticipated help-giving ($B = -0.78$, $p < 0.001$, Figure 3) and increased performance-goal orientation ($B = 0.27$, $p = 0.008$) and perceptions of a "fixed" instructor mindset ($B = 0.57$, $p < 0.001$). Whether participants had experience with the assigned grading condition did not affect outcomes in the criterion-referenced condition.

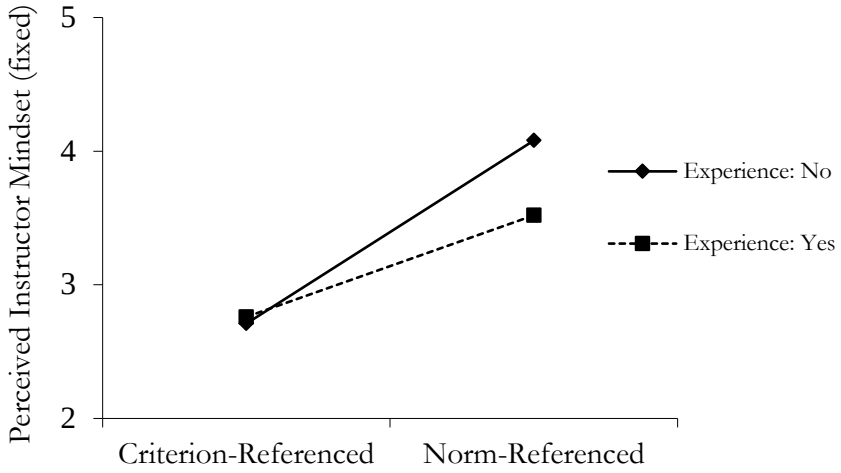

Figure 3. Perceived Instructor Mindset (1 = Growth, 6 = Fixed) as a Function of Grading Condition and Experience. Grading condition had a more dramatic effect when participants lacked experience with their assigned grading system.

Discussion

Despite their widespread use in higher education, the distinct impacts of norm- and criterion-referenced grading systems have been understudied. The current research employed an online

experiment to isolate how these grading policies shape initial student expectations, which are crucial as they often set the trajectory for student engagement and achievement throughout the semester (Legg & Wilson, 2009). Early impressions, formed from limited initial cues, such as grading policies, guide students' thoughts and behaviors. These thoughts and behaviors can then reinforce such impressions, resulting in a recursive process that affects behavior and experience throughout the semester (Cohen et al., 2009). Our findings reveal distinct effects of each grading policy on student expectations, supporting our hypotheses and providing useful insights for educators.

Implications for Course-Related Perceptions and Behaviors

Norm-referenced grading policies, by limiting high grades, signal a competitive environment that orients students toward performance goals rather than mastery goals and diminishes their expected self-efficacy. These findings align with prior research indicating that competitive settings heighten performance goals while reducing perceptions of personal capability (Chan & Lam, 2008; Lam et al., 2004; Urdan, 2004). Furthermore, our findings reveal that norm-referenced grading increased perceived threats to help-seeking and decreased the likelihood of students giving and seeking help, consistent with prior findings of the negative impact of competition on communal behaviors (Karabenick, 2003).

Additionally, the implementation of norm-referenced grading can signal to students that the instructor views intelligence as fixed and success as limited to a select few. This perception potentially contributes to a classroom environment where students feel less engaged and motivated, which can exacerbate racial and gender achievement gaps (Canning et al., 2020; Fuesting et al., 2019; LaCosse et al., 2020; Muenks et al., 2020; Rattan et al., 2018). Such findings underscore the broader implications of grading policies on student attitudes and educational equity.

Moderating Effects of Demographics and Experience

Prior studies have highlighted the susceptibility of students from underrepresented backgrounds—such as those from marginalized racial groups or first-generation college students—to competition and perceived threats (Canning et al., 2020; Rattan et al., 2018; Sommet et al., 2015). Contrary to these findings, our study revealed that grading systems impacted achievement orientation and self-efficacy uniformly across demographic groups. Likewise, perceptions of the instructor's intelligence beliefs were unaffected by students' backgrounds, aligning with research indicating minimal moderating effects of demographic characteristics on classroom cues (Muenks et al., 2020).

However, we did find that marginalized status moderated help-related behaviors in an unexpected way. Non-RM/FG students displayed heightened sensitivity to grading policies, showing decreased willingness to both offer and seek help under norm-referenced grading. In contrast, while RM/FG students also exhibited less willingness to help peers in the norm-referenced condition, this decrease was less pronounced. Remarkably, their willingness to seek help remained consistent across grading systems. This difference could stem from the interdependent cultural values more prevalent among first-generation students and marginalized racial groups, in contrast to the independence favored by their non-RM/FG peers (Stephens et al., 2012).

Participants' prior experience with the grading systems also moderated their expectations. Those lacking personal experience with norm-referenced grading anticipated being more competitive, offering even less help, and perceived a more fixed mindset in their instructor, compared to those with personal experience. This amplified reaction may be attributed to the concept of psychological distance (Liberman et al., 2007). Compared to experienced participants, those inexperienced with norm-referenced grading likely applied schemas that foreground the competitive aspect of such a

grading policy when predicting their experience in the course. Despite this, the negative impact of norm-referenced grading on expectations was universal, affecting even those experienced with it. This observation suggests that the adverse views associated with norm-referenced grading may not only be initial impressions but also reflect ongoing realities experienced by students.

Limitations and Future Directions

Several limitations of the present study suggest directions for future research. First, our analysis was based on participants' initial expectations for a hypothetical course. While initial course impressions shape students' later attitudes, behaviors, and learning outcomes, the impact of grading policy may change throughout a semester due to student performance and other classroom dynamics. Future field studies should examine how norm- and criterion-referenced grading affect students longitudinally in real classroom settings. Second, the grade cutoffs used in our manipulation, designed for clarity (*e.g.,* top 15% gets an A in the norm-referenced condition), may not accurately reflect the variability found in real educational settings, where cutoffs can differ across disciplines and instructors. Investigating the impact of varying cutoffs (*e.g.,* top 10% vs top 30% guarantees an A) could provide deeper insights into how different grading standards influence student behavior. Lastly, our hypothetical scenarios did not specify class size, a factor known to influence social comparison (Thijs et al., 2010). Students, for example, may be more willing to offer help in large classes over small ones under norm-referenced grading. Exploring how class size affects the impact of grading systems, particularly in terms of willingness to help, could further clarify how grading systems operate in different educational contexts.

The generalizability of our findings may also be influenced by the cultural and educational context, as the study was conducted with a U.S.-based college student sample. Grading conventions vary internationally, and cultural dimensions, such as individualism vs collectivism, could significantly influence how grading policies impact student expectations and behaviors. Further research is needed to extend these findings in non-U.S. educational settings.

The present study offers a first step in exploring how different grading policies shape student expectations for a course. Building on these initial insights, future studies might examine whether course difficulty affects the impact of different grading policies. In courses perceived as very challenging, norm-referenced grading might offer advantages that enhance self-efficacy (e.g., the class "curve" may cushion students from low grades). Additionally, examining hybrid grading policies that blend features of norm- and criterion-referenced grading could yield valuable insights (Lok et al., 2016). Finally, the present study was unable to examine the impact of grading policies on actual student achievement. Although comparing learning outcomes in norm- and criterion-referenced courses presents logistical challenges—requiring similar course content across courses, for example—it remains a critical area for future investigation.

Conclusion and Implications for Teaching

The present study examined the effects of norm- and criterion-referenced grading policies on students' course-related expectations. By presenting grading policies as initial situational cues, we demonstrated that such policies substantially shape students' early expectations concerning their motivation, self-efficacy, and cooperative behaviors, as well as their perceptions of instructor beliefs. Our findings notably reveal that norm-referenced grading tends to cultivate more performance-focused, less confident, and less cooperative attitudes among students, while also reinforcing the perception that their instructors have a fixed mindset for intelligence.

The implications of this research extend beyond how students earn grades to suggest that choice of grading policy can shape student expectations in ways that help or hinder their success.

These insights should prompt further research, stimulate reflection, and foster discussions among educators and administrators regarding the selection of grading policies in courses and across departments and institutions. By understanding the broader consequences of these grading policies, stakeholders can make informed decisions aimed at enhancing student engagement and success.

Declaration of Interest Statement

The authors have no conflicts of interest to declare.

Appendix

Supplemental Materials

Please find Supplemental Materials with additional tables, figures, and analyses here:
https://osf.io/puvd7?view_only=684cbd98a0d54f4284a291c29226b0e0

Measures

Course self-efficacy:
Based on the knowledge you have of the course, how much do you agree with the following statements:

1. I believe I will receive an excellent grade in this class
2. I am confident I can learn the basic concepts taught in this course
3. I am confident I can understand the most complex material presented by the instructor in this course
4. I expect to do well in this class
5. I am certain I can master the skills being taught in this class

1 = Strongly disagree, 5 = Strongly agree

Achievement goals:
Based on the knowledge you have of the course, how much do you agree with the following statements:

1. My goal for this course would be to perform well relative to other students
2. I would strive to do well compared to other students in the course
3. My goal for the course would be to perform better than the other students
4. I would strive to understand the content of this course as thoroughly as possible
5. My goal would be to learn as much as possible from this class
6. My goal for the course would be to completely master the material presented in the class

1 = Strongly disagree, 5 = Strongly agree
1-3 Performance goal-orientation, 4-6 Mastery goal-orientation

Help-giving:
Based on the knowledge you have of the course, please rate on how likely you would be to engage in the following behaviors:

1. Offer to explain a course concept to a fellow student who did not initially grasp it
2. Provide constructive feedback on a fellow student's project
3. Share your notes with a classmate who missed last week's class
4. Spend extra time with a classmate before an exam in order to help them
5. Share your favorite strategies for studying for the course with a fellow student
6. Offer to study with a struggling classmate

7. Try to work with other students from this class to complete course assignments and/or study for exams

1 = Extremely unlikely, 7 = Extremely likely

Help-seeking:

Based on the knowledge you have of the course, please rate on how likely you would be to engage in the following behaviors:

1. Attend office hours to ask for help from the instructor
2. Stay after the class to ask the instructor for help
3. Ask questions during class
4. Email the instructor to ask questions
5. Seek help from a teacher's assistant
6. Ask a fellow classmate questions
7. Borrow notes from a fellow classmate
8. Ask to study with a fellow classmate/a group of fellow classmates
9. Try to figure out the material on my own rather than asking for help*
10. Try to guess rather than ask for help*

1 = Extremely unlikely, 7 = Extremely likely

*Reverse scored

Help-seeking threat:

Based on the knowledge you have of the course, how much do you agree with the following statements?

1. I would feel like a failure if I need help in this class
2. I would not want anyone to find out that I need help in this class
3. Getting help in this class would be an admission that I am just not smart enough to do the work on my own
4. People would think less of me if I succeed in this course only because I got help
5. I worry that other students may think that I am incompetent if I ask for help
6. If I ask for help, the instructor may think that I am incompetent

1 = Strongly disagree, 5 = Strongly agree

Perceived instructor belief in the universal potential for intelligence:

Based on this grading system, how much do you think the professor believes that almost all people have the potential to become highly intelligent, or that only some people have the potential to become highly intelligent?

For this question, we define becoming "highly intelligent" as attaining the capacity to acquire knowledge and skills at the highest level possible, for example attaining the knowledge and skills of a Nobel Prize winner

1= almost all people have the potential to become highly intelligent
20 = only some people have the potential to become highly intelligent

Perceived instructor belief in the universal potential for success:

Based on this grading system, how much do you think the professor believes that all students in the course are capable of succeeding in the course?

1= almost all students are capable of succeeding in the course
20 = only some students are capable of succeeding in the course

Perceived instructor intelligence mindset:

Based on your knowledge in this course, how much do you agree with the following statements?

1. The instructor in this class seems to believe that students have a certain amount of intelligence, and they really can't do much to change it
2. The instructor in this class seems to believe that every student can learn new things and significantly grow their intelligence*
3. The instructor in this class seems to believe that some students are smart, while others are not
4. The instructor in this class seems to believe that students either "have it" or they don't

1 = Disagree a lot, 6 = Agree a lot

*Reverse scored

References

Ames, C. (1992). Achievement goals and the classroom motivational climate. *Student Perceptions in the Classroom*, *1*, 327–348.

Ames, C., & Ames, R. (1984). Goal structures and motivation. *The Elementary School Journal*, *85*(1), 39–52. https://doi.org/10.1086/461390

Aviles, C. B. (2001). Grading with norm-referenced or criterion-referenced measurements: To curve or not to curve, that is the question. *Social Work Education*, *20*(5), 603–608. https://doi.org/10.1080/02615470120072869

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman/Times Books/ Henry Holt & Co.

Barron, K. E., & Harackiewicz, J. M. (2003). Revisiting the benefits of performance-approach goals in the college classroom: Exploring the role of goals in advanced college courses. *International Journal of Educational Research*, *39*(4), 357–374. https://doi.org/10.1016/j.ijer.2004.06.004

Bell, J., Grekul, J., Lamba, N., Minas, C., & Harrell, W. A. (1995). The impact of cost on student helping behavior. *The Journal of Social Psychology*, *135*(1), 49–56. https://doi.org/10.1080/00224545.1995.9711401

Bong, M. (2008). Effects of parent-child relationships and classroom goal structures on motivation, help-seeking avoidance, and cheating. *The Journal of Experimental Education*, *76*(2), 191–217. https://doi.org/10.3200/JEXE.76.2.191-217

Brady, S. T., Hard, B. M., & Gross, J. J. (2018). Reappraising test anxiety increases academic performance of first-year college students. *Journal of Educational Psychology*, *110*(3), 395–406. https://doi.org/10.1037/edu0000219

Butler, R., & Neuman, O. (1995). Effects of task and ego achievement goals on help-seeking behaviors and attitudes. *Journal of Education Psychology*, *87*(2), 261–271. https://doi.org/10.1037/0022-0663.87.2.261

Canning, E. A., LaCosse, J., Kroeper, K. M., & Murphy, M. C. (2020). Feeling like an imposter: The effect of perceived classroom competition on the daily psychological experiences of first-generation college students. *Social Psychological and Personality Science*, *11*(5), 647–657. https://doi.org/10.1177/1948550619882032

Chan, J. C. Y., & Lam, S. (2008). Effects of competition on students' self-efficacy in vicarious learning. *British Journal of Educational Psychology*, *78*(1), 95–108. https://doi.org/10.1348/000709907X185509

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive Processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, *324*(5925), 400–403. https://doi.org/10.1126/science.1170769

Cook, K. S., Cheshire, C., Rice, E. R. W., & Nakagawa, S. (2013). Social exchange theory. In J. DeLamater & A. Ward (Eds.), *Handbook of Social Psychology* (pp. 61–88). Springer Netherlands. https://doi.org/10.1007/978-94-007-6772-0_3

Covington, M. V., & Omelich, C. L. (1984). Task-oriented versus competitive learning structures: Motivational and performance consequences. *Journal of Educational Psychology*, *76*(6), 1038–1050. https://doi.org/10.1037/0022-0663.76.6.1038

Darnon, C., Muller, D., Schrager, S. M., Pannuzzo, N., & Butera, F. (2006). Mastery and performance goals predict epistemic and relational conflict regulation. *Journal of Educational Psychology*, *98*(4), 766–776. https://doi.org/10.1037/0022-0663.98.4.766

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

Dweck, C. S. (2008). *Mindset: The new psychology of success*. Ballantine Books.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, *100*(3), 613–628. https://doi.org/10.1037/0022-0663.100.3.613

Fuesting, M. A., Diekman, A. B., Boucher, K. L., Murphy, M. C., Manson, D. L., & Safer, B. (2019). Growing STEM: Perceived faculty mindset as an indicator of communal affordances in STEM. *Journal of Personality and Social Psychology*, *117*(2), 260–281. https://doi.org/10.1037/pspa0000154

Hertz-Lazarowitz, R. (1989). Cooperation and helping in the classroom: A contextual approach. *International Journal of Educational Research*, *13*(1), 113–119. https://doi.org/10.1016/0883-0355(89)90020-7

Hoigaard, R., Kovac, V. B., Overby, N. C., & Haugen, T. (2015). Academic self-efficacy mediates the effects of school psychological climate on academic achievement. *School Psychology Quarterly*, *30*(1), 64. https://doi.apa.org/doiLanding?doi=10.1037%2Fspq0000056

Karabenick, S. A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology*, *28*(1), 37–58. https://doi.org/10.1016/S0361-476X(02)00012-7

Karabenick, S. A. (2004). Perceived achievement goal structure and college student help seeking. *Journal of Educational Psychology*, *96*(3), 568–581.

Karabenick, S. A., & Knapp, J. R. (1991). Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of Educational Psychology*, *83*(2), 221–230. https://doi.org/10.1037/0022-0663.83.2.221

Karabenick, S. A., & Newman, R. S. (2013). *Help seeking in academic settings: Goals, groups, and contexts*. Routledge.

Knight, P. (2001). *A briefing on key concepts: Formative and summative, criterion, and norm-referenced assessment*.

LaCosse, J., Murphy, M. C., Garcia, J. A., & Zirkel, S. (2020). The role of STEM professors' mindset beliefs on students' anticipated psychological experiences and course interest. *Journal of Educational Psychology*, *Advance online publication*. https://doi.org/10.1037/edu0000620

Lam, S.-F., Yim, P.-S., Law, J. S. F., & Cheung, R. W. Y. (2004). The effects of competition on achievement motivation in Chinese classrooms. *British Journal of Educational Psychology*, *74*(2), 281–296. https://doi.org/10.1348/000709904773839888

Legg, A. M., & Wilson, J. H. (2009). E-Mail from professor enhances atudent motivation and attitudes. *Teaching of Psychology*, *36*(3), 205–211. https://doi.org/10.1080/00986280902960034

Liberman, N., Trope, Y., & Stephan, E. (2007). Psychological distance. In *Social psychology: Handbook of basic principles, 2nd ed* (pp. 353–381). The Guilford Press.

Lok, B., McNaught, C., & Young, K. (2016). Criterion-referenced and norm-referenced assessments: Compatibility and complementarity. *Assessment & Evaluation in Higher Education*, *41*(3), 450–465. https://doi.org/10.1080/02602938.2015.1022136

Lou, N. M., & Noels, K. A. (2020). "Does my teacher believe I can improve?": The role of meta-lay theories in ESL learners' mindsets and need satisfaction. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01417

McColskey, W., & Leary, M. R. (1985). Differential effects of norm-referenced and self-referenced feedback on performance expectancies, attributions, and motivation. *Contemporary Educational Psychology*, *10*(3), 275–284. https://doi.org/10.1016/0361-476X(85)90024-4

Meece, J. L. (1991). The classroom context and students' motivational goals. *Advances in Motivation and Achievement*, *7*, 261–285.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology*, *57*(1), 487–503. https://doi.org/10.1146/annurev.psych.56.091103.070258

Muenks, K., Canning, E. A., LaCosse, J., Green, D. J., Zirkel, S., Garcia, J. A., & Murphy, M. C. (2020). Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class. *Journal of Experimental Psychology: General*, *149*(11), 2119–2144. https://doi.org/10.1037/xge0000763

Murphy, M. C., & Taylor, V. J. (2012). The role of situational cues in signaling and maintaining stereotype threat. In *Stereotype threat: Theory, process, and application* (M. Inzlicht&T. Schmader, pp. 17–33). Oxford University Press.

Nicholls, J. G., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing students' theories of success in mathematics: Individual and classroom differences. *Journal for Research in Mathematics Education*, *21*(2), 109–122. https://doi.org/10.2307/749138

Payakachat, N., Gubbins, P. O., Ragland, D., Norman, S. E., Flowers, S. K., Stowe, C. D., DeHart, R. M., Pace, A., & Hastings, J. K. (2013). Academic help-seeking behavior among student pharmacists. *American Journal of Pharmaceutical Education*, *77*(1). https://doi.org/10.5688/ajpe7717

Pintrich, P. R. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. https://eric.ed.gov/?id=ED338122

Poortvliet, P. M., Anseel, F., Janssen, O., Van Yperen, N. W., & Van de Vliert, E. (2012). Perverse effects of other-referenced performance goals in an Information exchange context. *Journal of Business Ethics*, *106*(4), 401–414. https://doi.org/10.1007/s10551-011-1005-8

Poortvliet, P. M., & Darnon, C. (2010). Toward a more social understanding of achievement goals: The interpersonal effects of mastery and performance goals. *Current Directions in Psychological Science*, *19*(5), 324–328. https://doi.org/10.1177/0963721410383246

Posselt, J. R., & Lipson, S. K. (2016). Competition, anxiety, and depression in the college classroom: Variations by student identity and field of study. *Journal of College Student Development*, *57*(8), 973–989. https://doi.org/10.1353/csd.2016.0094

Rattan, A., Savani, K., Komarraju, M., Morrison, M. M., Boggs, C., & Ambady, N. (2018). Meta-lay theories of scientific potential drive underrepresented students' sense of belonging to science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *115*(1), 54–75. https://doi.org/10.1037/pspi0000130

Rattan, A., Savani, K., Naidu, N. V. R., & Dweck, C. S. (2012). Can everyone become highly intelligent? Cultural differences in and societal consequences of beliefs about the universal potential for intelligence. *Journal of Personality & Social Psychology*, *103*(5), 787–803. https://doi.org/10.1037/a0029263

Restrepo, L. (2020). *The effect of instructor mindset on student motivation and self-efficacy* [Thesis, Augusta University]. https://augusta.openrepository.com/handle/10675.2/623439

Ryan, A. M., & Pintrich, P. R. (1997). "Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology*, *89*(2), 329–341. https://doi.org/10.1037/0022-0663.89.2.329

Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE—Life Sciences Education*, *13*(2), 159–166. https://doi.org/10.1187/cbe.cbe-14-03-0054

Sommet, N., & Elliot, A. J. (2017). Achievement goals, reasons for goal pursuit, and achievement goal complexes as predictors of beneficial outcomes: Is the influence of goals reducible to reasons? *Journal of Educational Psychology*, *109*(8), 1141–1162. https://doi.org/10.1037/edu0000199

Sommet, N., Quiamzade, A., Jury, M., & Mugny, G. (2015). The student-institution fit at university: Interactive effects of academic competition and social class on achievement goals. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00769

Stephens, N. M., Hamedani, M. G., & Destin, M. (2014). Closing the social-class achievement gap: A difference-education intervention improves first-generation students' academic performance and all students' college transition. *Psychological Science*, *25*(4), 943–953. https://doi.org/10.1177/0956797613518349

Stephens, N. M., Townsend, S. S. M., Markus, H. R., & Phillips, L. T. (2012). A cultural mismatch: Independent cultural norms produce greater increases in cortisol and more negative emotions among first-generation college students. *Journal of Experimental Social Psychology*, *48*(6), 1389–1393. https://doi.org/10.1016/j.jesp.2012.07.008

Thijs, J., Verkuyten, M., & Helmond, P. (2010). A Further Examination of the Big-Fish–Little-Pond Effect: Perceived Position in Class, Class Size, and Gender Comparisons. *Sociology of Education*, *83*(4), 333–345. https://doi.org/10.1177/0038040710383521

Urdan, T. (2004). Using multiple methods to assess students' perceptions of classroom goal structures. *European Psychologist*, *9*(4), 222–231. https://doi.org/10.1027/1016-9040.9.4.222

Wentzel, K. R., & Miele, D. B. (2009). *Handbook of motivation at school*. Routledge.

Williams, R. G., Pollack, M. J., & Ferguson, N. A. (1975). Differential effects of two grading systems on student performance. *Journal of Educational Psychology*, *67*(2), 253–258. https://doi.org/10.1037/h0077022