Journal of Writing Assessment

Watson, Emily. (2023). Achieving High Goals: The Impact of Contract Grading on High School Students' Academic Performance, Avoidance Orientation, and Social Comparison. *Journal of Writing Assessment*, 16(1). **DOI**: 10.5070/W4jwa.249

Achieving High Goals: The Impact of Contract Grading on High School Students' Academic Performance, Avoidance Orientation, and Social Comparison

Emily Watson, Lancaster University, UK, emilywatsonphd@gmail.com

Abstract: This article examines American high school students' (N=439) self-worth protection behaviors, maladaptive coping mechanisms, and academic performance under a contract grading system, which has been understudied in contemporary secondary classrooms. The quantitative analysis revealed that under the contract grading system, 97% (n=421) earned a passing grade (i.e., A, B, or C) on the assessment and 90% (n=390) fulfilled the contract by reaching mastery (A) or proficiency (B). Compared to the previous year, students with prior experience were 19% more likely to earn an A and 16% more likely to earn a B under the grading contract despite increased workload demands. The qualitative analysis of 40 semi-structured interviews revealed that performance improved as a result of the contract's clarity of purpose, which limited task avoidance and facilitated task-oriented effort toward a desirable goal. Students enrolled in regular courses experienced the most significant grade improvement due to clear expectations that helped them place their effort on the right tasks. The findings of this study lead to a call to action for teachers to implement contract grading in high school classrooms to clarify work expectations, improve task-oriented effort, and help students set and achieve high goals.

Keywords: writing assessment, academic performance, secondary education, self-worth protection behaviors, contract grading

Journal of Writing Assessment is a peer-reviewed open access journal. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

OPEN ACCESS

Problem Statement

With their far-reaching consequences on students' lives and self-concept, grades (e.g., marks on assignments, standardized test scores, cumulative GPAs, report cards, and transcripts) amplify the pressures and stresses of school, particularly for secondary students with eyes on the future. More than twenty years ago, before college admissions went into hyperdrive, de Anda et al. (2000) found that the greatest stressors for high school students were grades, career expectations, and future life plans. High school students in college preparatory environments, like those in this study, reported more stress when their focus was earning high grades (Ainslie et al., 1996). For most students, whose developing brains are particularly vulnerable, high stress can impact neural maturation, increase anxiety and depression, and lead to disengaged coping that lowers academic performance (Arsenio & Loria, 2014; Eiland & Romeo, 2013).

Achievement pressure is pervasive at Good Shepherd High School (GSHS, anonymized), a private, religious school on the West Coast of the United States, where I taught in the English Department from 2013 to 2021. To prepare students for the demands of college and beyond, GSHS dedicates at least five weeks of instruction, or 11 class periods (825 instructional minutes), each school year to teaching each step of the academic research process to all grade levels (9-12) and then assesses their performance using a points-based analytic rubric. With each grade level, the task's difficulty and workload demands increase, making it likely that academic performance will remain relatively stable (see Table 1). Unfortunately, department data revealed that students who fail in 9th or 10th grade continue to experience failure: in 2018, 26% of the approximately 300 first-year 9th graders and 20% of regular 12th graders also earned a low or failing grade (e.g., D or F). Since submitting an assignment for grading involves the vulnerable act of opening up oneself for critique (McArthur, 2018), students, in acts of self-protection, engage in maladaptive coping strategies, such as self-handicapping and the intentional withdrawal of effort to avoid showing low ability, that ultimately cause further harm (Covington, 1984; Thompson & Parker, 2007).

This study examines the intimate relationship between self-worth protection behaviors and academic performance. Self-worth protection stems from a situation's perceived evaluative threat, or the "forecast of failure which is likely to impact negatively on the self, threatening self-estimates of ability and challenging the already fragile estimates of global self-worth" (Thompson & Parker, 2007, p. 131). Students employ self-protective mechanisms, such as self-handicapping, to avoid showing low ability and "protect the already uncertain self-images against further assail" (Thompson & Parker, 2007, p. 131). Perceiving that one's abilities are not well-matched with the task, or that poor performance will harm self-image, can increase evaluative threat and maladaptive coping strategies, like social comparison, avoidance orientation, or procrastination, to protect self-worth (Thompson & Parker, 2007) and avoid harm (Lazarus & Folkman, 1984). Students who are high in the trait of self-handicapping experience greater anxiety (Thompson, 2004; Thompson & Richardson, 2001), higher rates of helpless thinking and behavior (De Castella et al., 2013), greater negative affect (Thompson & Richardson, 2001), and lower overall academic performance (Martin et al., 2001).

Thompson and Dinnel (2007) found that self-worth protection was best understood as "an outcome of choking under pressure, fueled by evaluative threat" (p. 509), which GSHS teachers and counselors observed prior to this study (personal correspondence). Unfortunately, rather than work to reduce fear of failure and evaluative threat, teachers often use it as a motivator, relying on its presence by frightening children into work (Jackson, 2010). In fact, GSHS teachers often

Table 1
Workload Expectations

	9th Grade	10th Grade	11th Grade	12th Grade
Word Count	1,200-1,500	1,500-1,800	1,800-2,100	2,100-2,400
# of Sources Requirement	4	6	6	8

Note. This chart shows the requirements for the paper, which increase with each grade level. Each student in the study completed the largest paper they had written in high school.

used what Putwain et al. (2016) call fear appeals, which can elicit either a challenge or threat appraisal. How students perceive the appeal can vary according to the message's severity (Putwain & Remedios, 2014; Putwain & Roberts, 2012). Depending upon the perceived evaluative threat, students high in self-worth protection perform differently (Thompson, 1999): in situations with limited risk to self-worth, a student will generally perform well, yet high-risk situations can lead to greater evaluative threat and anxiety, which can impair performance and lead to underachievement (Thompson & Dinnel, 2007). Hancock (2001) found that classrooms with higher evaluative threat led all students, including those *without* a propensity toward test anxiety, to earn lower grades. He concluded that "all students are more motivated to learn in classrooms deemed less evaluative" (Hancock, 2001, p. 289).

To mitigate evaluation pressure, emphasize learning over performance, and help all students reach their highest potential, the GSHS English Department implemented an alternative assessment approach called contract grading, a "social agreement with the entire class about final course grades will be determined" (Inoue, 2019, p. 130). Without eliminating grades entirely, and thus allowing teachers to fulfill institution requirements to quantify learning with a single score, contract grading attempts to deemphasize grades and "make them less present and exert less pressure, by ironically paying attention to how grades are constructed" (Inoue, 2019, p. 142). Litterio (2018) found that college students enjoyed clear expectations and autonomy; however, she concluded that "students are still fixated on traditional grades" but instructors should "recognize and return to contract grading as a powerful system not only to break through barriers of traditional assessment but also to reinforce the collaborative and individual practices involved with writing" (p. 9).

Although the earliest known work on contract grading advocated for its use in high schools (McLaughlin, 1961), most empirical work with contract grading has been relegated to the 1970s and focused on the development and implementation of contracts (Amsden, 1970; Bowers & Howard, 1975; James, 1977; Kokus & Mussoff, 1975). Contract grading, then, is underutilized and understudied in U.S. secondary schools. As I wrote in the first part of this study (Ward, 2021b), which focused on students' perceptions of stress under the contract system:

> The unorthodox yet democratic contract grading system is uncommon in schools and, in my experience, unfamiliar to most secondary teachers. As evidenced by recent research (Lindemann & Harbke, 2011; Litterio, 2016; Litterio, 2018), Inoue's (2019) recent book,

and articles in *Inside Higher Ed* (Warner, 2016, 2017), contract grading may be used more in college classrooms where instructors often have more autonomy to implement alternative approaches, but empirical research is needed on contract grading's impact on adolescents at college-preparatory high schools.

Cowan (2020) observed that "sadly, no research has been published that directly compares the *anxiety* students feel using a conventional system versus using a grading contract" (p. 6). She also wrote that "most scholarship on grading contracts in composition focuses on individual case studies...with the occasional study of students in those classes," thus the field "could use more comparative, large-scale studies of grading contracts" that reveal "how much grading contracts impact students academically or emotionally compared to other grading schemes" (p. 8). This study, with 439 participants, fills the gap Cowan described. While contract grading is typically used across entire courses, this study examines the use of contract grading during a single unit and thus serves as a soft entry point for teachers and departments accustomed to traditional grading practices.

The pilot study (Ward, 2021a) revealed that among 12th graders with a history of low or failing grades on the assessment, those who received the contracts showed a statistically significant decrease in their perception of workload demands while also earning significantly higher grades compared to the traditional grading group. The present study took place in Spring 2020 when all grade (9-12) and course levels (i.e., regular, AP/honors, and those accommodated for learning disabilities) adopted the grading contract, instead of an analytical rubric, for a high-stakes writing assessment. The findings of this research are split over two articles, each with a distinct focus: the first article (Ward, 2021b) examined the primary focus of the study—students' psycho-emotional well-being—and found that perceptions of stress and evaluative threat significantly decreased under the grading contract system. This article focuses on academic performance and self-worth protection behaviors under improved psycho-emotional conditions. The following questions guided this study:

1.How does contract grading impact students' academic performance as measured by their final grade?
2.How does contract grading affect students' self-worth protection behaviors, particularly the coping mechanisms of task avoidance and social comparison?
3.Which factors (e.g., prior experience, class level, etc.) impact academic performance?

Contract Grading

> Not everyone is highly talented, but most are capable of attaining an adequate level of achievement. (Covington & Beery, 1976, p. 143)

> Conventional grading rests on two principles that are patently false: that professions in our field have common standards for grading, and that the 'quality' of a multidimensional product can be fairly or accurately represented with a conventional one-dimensional grade. In the absence of genuinely common standards or a valid way to represent quality, every grade masks the play of hidden biases inherent in readers and a host of other a priori power differentials. (Danielewicz & Elbow, 2009, p. 249)

Are most students able to achieve proficiency? Inspired by Danielewicz and Elbow (2009) and Covington and Beery (1976), this question preceded the creation of the contracts. Contract grading rests on the equitable principle that most students are capable enough to achieve proficiency and helps teachers "plan for students' success, not how to document their fall" (Wormeli, 2006, p. 25). Before this study, the English Department used an analytic rubric that awarded points to each aspect of the paper (e.g., 15 points for the thesis, 25 points for supporting evidence, etc.). While recognizing that uniform standards are admirable and rubrics can be quick and efficient for teachers to use, Kohn (2006) argues that they are an inauthentic "tool to promote standardization, to turn teachers into grading machines or at least to allow them to pretend that what they are doing is exact and objective" (p. 12). Deducting points encouraged a punitive posture, in which teachers looked for mistakes and then subtracted points without careful consideration of the number and its impact. In this way, given the rate of low and failing grades, this grading practice appeared to motivate students away from learning and future effort and produced practices that did not accurately reveal what students knew and could do.

This work builds upon research that critiques the bias and inequity that seeps into traditional writing assessment (Breland, 1983; Brimi, 2011; Rachal, 1984). Starch and Elliott's (1912) landmark study asked high school English teachers (*N*=142) to evaluate two papers according to their school's standards. The results varied widely, with the first paper earning scores between 64-98 and the second from 50-97—one paper received 30 different scores. They concluded that "the promotion or retardation of a pupil depends to a considerable extent upon the subjective estimate of his teacher" (Starch & Elliott, 1912, p. 454). More recently, Brimi (2011) trained teachers to use a 100-point rubric and then replicated Starch and Elliott's study with almost identical results: a 46-point variation in scores. As Guskey (2013) observed, "even if one accepts the idea that there are truly 100 discernible levels of student writing performance, it's clear that even well-trained teachers cannot distinguish among those different levels with much accuracy or consistency" (p. 70). To Danielewicz and Elbow (2009), teachers are "experts about writing, but individuals, nevertheless, who cannot pretend to be wholly impersonal or fair" (p. 247).

This study adds to a growing body of recent work on equitable assessment and students' psycho-emotional well-being (Feldman, 2019; Inoue, 2019; McArthur, 2018) by examining the impact of contract grading on high school students during high-stakes writing assessment. While many changes have taken place to adapt educational practices for the 21st-century beliefs that all students can meet demanding standards and deserve the opportunity to succeed (Feldman, 2019), McArthur (2018) observed that "assessment has proven a sticky practice, reluctant to change, immune to innovation" (p. 6). Recently, Feldman (2019) tackled important equity issues for secondary teachers, such as the mathematical unfairness of 100-point grading scales, the use of zeros for missing and/or perceived low-quality work, and the myth of the motivational F. To move toward greater equity, he proposes using an evenly distributed four-point grading scale, awarding no less than 50 percent in the grade book, and removing subjective "soft skills" behaviors (e.g., listening) from the grading process, and allowing students to revise and resubmit. Inoue and Elbow, who both influenced the present study, appear motivated by two pillars Feldman outlines—bias-resistance and intrinsic motivation—yet, in an apparent difference of worldview, we differ most on the final pillar: accuracy. Feldman appears to hold onto hope that accuracy—that is, a fair and objective evaluation—is possible if the teacher designs assessment practices as he suggests. In the end, he argues that the grade will signal what the student knows and can do.

Conversely, for Inoue (2019), a written product represents learning, but it is not learning itself. All learning comes from labor, which is neglected in traditional systems; thus, in his labor-based contract system, he outlines the expectations (most simply, more work for higher grades) and then removes himself from the grading process as much as possible, allowing students to take ownership over their learning and grade. The hybrid contract system used in this study and proposed by Danielewicz and Elbow (2009) does not require more work for higher grades but higher quality work, which they admit rests on the teachers' subjective evaluation of the work. While Feldman (2019) argues that behaviors and soft skills limit the accuracy of a grade, the learning practices identified on the B contract are written as objectively as possible (that is, yielding a yes or no response) and are crucial to the learning process, particularly for the developing academic writers in this study. In the end, both the grade on the learning practices (e.g., taking notes on sources) and the final product (e.g., cohesive paragraphs that utilize information from notes) reveal whether the student has achieved mastery or proficiency or whether they are still developing competence.

In January 2019, the GSHS English Department not only changed our assessment tool but our philosophy, moving from one focused on evaluation (and its counterparts of numbers and percentage points) to learning and skill acquisition. To minimize bias, we utilized the hybrid contract grading system put forth by Danielewicz and Elbow (2009), in which students were offered two contracts: A and B. Contract grading has been critiqued as a lowering of academic rigor; however, by asking students to contract for only A or B, the department sought to buoy each student to "achieve the best that is possible for them" (Nicholls, 1979, p. 1071). The goal was not equality of outcome but the highest possible fulfillment of potential. In this way, the B contract served the minimum threshold for each course and for all students. For some teachers, this would be a C contract, yet the English Department was influenced by Johnston and O'Neill (1973), who found that criteria control performance to a high degree. Even as the criteria changed, students adjusted their behavior to meet the minimum threshold, no matter the challenge. Based on their findings, they urged teachers to set high standards for the minimum pass criteria, outline them as precisely as possible at the beginning of the course, and not to successively lower the criteria.

Teachers in this study (N=13) worked collaboratively to create both the A and B grading contracts for each English course (N=13) (see Figure 1). While college instructors often have considerable freedom in their course design and assessment approach, this same autonomy is not extended to many secondary teachers who need to be in alignment with other teachers of the same course. Admittedly, the process of creating each contract required significant time investment and backward curriculum design, which "may improve instruction" by first identifying the desired results and then "what will constitute acceptable evidence of learning" (Mazur, 2018, p. 2). In this way, the contracts served as holistic tools for learning and metacognition as students self-monitored and self-assessed while drafting and revising. For example, compared to the analytic rubric, which awarded a set number of points for each writing feature (e.g., 25 points for analysis), each contract had 10 items that yielded a yes-or-no response (e.g., analysis is consistently present and explained within each paragraph). Building upon the B contract, the A contract did not demand *more* work but a higher quality of work. Each action on the A contract was infused with the subjective qualities of exceptional writing, such as 'compelling' analysis and 'well-crafted' thesis.

Figure 1

Grading Contract

Contract for an A

The following contract is intended to detail the major elements of a senior-level research paper.

High (98%) **Medium** (95%) **Low** (92%)

You will earn an A if you meet all of the following:

1. Format the paper and Works Cited Page in **MLA format** with no errors.
2. Write a minimum of 100 paraphrased **notes** and use those only to write your paper. Cite each one with an **MLA parenthetical citation**.
3. Find **eight-ten sources** from the library databases. No more than ten sources.
4. Write a well-crafted **thesis statement and corresponding topic sentences that** includes an argument regarding a world problem that has not yet been solved.
5. **Analysis** is consistently present, compelling, and well explained within each paragraph.
6. Dissect **cause** and **effect** related to the issue as well as a well-constructed and supported **solution** (both failed and proposed). Includes a well-written introduction, body paragraphs, and conclusion.
7. Cite at least **two different sources** in each body paragraph without repeating citations showing seamless and balanced use of sources.
8. Meet **minimum word** count of 2100 without exceeding 2400 words (-1% per 1% over or under).
9. Submit a printed essay, and submit to TurnItIn **on time** and receive a **clean report**.
10. Writing must be edited and pruned for **sophisticated use of grammar, mechanics, diction, and sentence structure**.

Contract for a B

The following contract is intended to detail the major elements of a senior-level research paper.

High (88%) **Medium** (85%) **Low** (82%)

You are guaranteed a B if you meet all of the following:

1. Format the paper and Works Cited Page in **MLA format** with no major errors
2. Write a minimum of 100 paraphrased **notes** to support your argument and use only those notes to write your paper. Cite each one with an **MLA parenthetical citation**.
3. Find and use **eight sources** from the library databases. No more than ten sources.
4. Write a **thesis statement and corresponding topic sentences** that include an argument regarding a world problem that has not yet been solved.
5. **The analysis** is consistently present and reasonably explained within each paragraph.
6. Write **cause**, **effect**, and **solution** (both failed and proposed) as well as an introduction, body paragraphs (that follow proper academic structure), and conclusion.
7. Cite at least **two different sources** in each body paragraph without repeating citations. Balanced use of sources.
8. Meet **minimum word** count of 2100 without exceeding 2400 words (-1% per 1% over or under).
9. Submit a printed essay, and submit to TurnItIn.com **on time** and receive a **clean report**.
10. Writing must be edited and pruned for **proper grammar and mechanics**.

Note. These contracts were created for regular 12th grade English courses.

Additionally, while the administration granted permission to use this alternative grading approach, teachers were still required to submit one number grade, between 0 and 100, to our electronic grade book, which is always accessible to parents and students. In this way, contract grading fulfilled institutional requirements by using the letters (e.g., A and B) and numbers (e.g., 95 and 85) of the traditional grading system; however, on the first day of instruction, students received the grading contracts for A and B and then signed for their desired goal by the end of the first week. Previously, teachers used an analytic grading rubric, which was available to students course learning management platform but not regularly referenced by the teacher. Consequently, contract grading focused both students' attention and teachers' feedback on the learning goals while encouraging metacognition on task completion.

To encourage accuracy, each teacher had a not-for-student-use list of qualities of C and D papers. 'Off the contract' papers that met an adequate standard of achievement could earn the passing grade of C (i.e., 72, 75, 78). Incomplete papers warranting a non-passing grade could be revised and resubmitted for a passing grade. Since D and F are non-passing grades that require course remediation, the revise-and-resubmit policy encouraged all students to submit a passing paper, yet revised papers received a penalty: at most, they could earn a C. Consequently, revised papers were still 'off the contract' and not included in the contract fulfillment rates in the coming sections.

To intentionally highlight the subjective nature of assessment, the lead teacher of the pilot study (Ward, 2021a) taught the instructors to grade holistically using a checkmark system (see Table 2) for each contract item to arrive at a number for the grade book (see Table 3). The goal was to shift mindsets and mitigate teachers exercising their authority over the final grade and 'throwing students off the contract' due to perceived low quality. If a teacher deemed a contract item as completed poorly, they could award a check-minus (thus, reducing the B grade from 85 to 82), but the teacher's opinion of quality did not merit an off-the-contract grade (below B) insofar as the item was, in fact, complete. Similarly, the teacher could increase the grade by three points (from 85 to 88) for mastering a contract item. In this way, the contract vastly limited the extent of a teacher's subjective influence as students fulfilled the contract as long as they completed all learning tasks. While Danielewicz and Elbow (2009) argue that the culture of assessment "obscures unfairness in how institutional power and authority determine success and failure" (p. 248) as students come to accept the outcome of their just desserts in our so-called meritocracy, contract grading "fight[s] a large, societal, and culturally enshrined system that looks fair when it is not" (p. 249). The hybrid contract limits the teacher's judgment for most (but not all) grades while offering more transparency about how grades are constructed. While no grading system can completely remove bias, contract grading works to dismantle some of the injustices built into highly subjective traditional grading practices.

Methodology

This study used a mixed methods approach with an explanatory sequential design, generating data in three distinct phases with parallel construction, measures, and instruments with the same cohort of participants. After analyzing the quantitative data from 439 matched-pair pre-post survey responses, I conducted 40 semi-structured interviews with students of all course types (e.g., accommodated, regular, and honors) and courses (e.g., English 1, Honors English 1, etc.). The primary goal of the interviews was to generate a detailed understanding of students' experiences in

Table 2

Checkmark System to Assess Papers in the Contract System

Mark	Description
✓+	Exceeds expectations for contract item
✓	Meet expectations for contract item
✓-	Below expectations for contract item
X	Missing

Note. Teachers were encouraged to use a simple checkmark system for each of the ten items to assess each paper holistically with the grading contract.

Table 3

Percentage Grades from Holistic Assessment

Assessment	A Contract	B Contract
High	98+	88
Medium	95	85
Low	92	82

Note. Each contract was associated with three different numerical grades, each corresponding with the teacher's final assessment of the paper as 'high,' 'medium,' or 'low.' Off-the-contract papers with missing elements earned scores lower than 80.

context and investigate what role, if any, the grading contract had on significant findings related to self-worth protection behaviors. The interviews asked questions about the grading contract (e.g., can you tell me about your experience with and perception of the grading contract?), the perceived difference between contract and traditional grading (e.g., to you, what is the difference between the old rubric and the new contract?), and self-worth protection behaviors (e.g., in the past, have you played it safe with your research paper goals and effort?), with ample follow-up questions to understand students' experiences in context. In this way, while the quantitative strand described "what is" through the distribution of variables across multiple dimensions, the qualitative strand understood and explained "why" by uncovering participants' meanings behind each phenomenon (Creswell, 2015; Merriam, 2009). In the final stage of data collection, I collected final grades. At the time of the interview, not all students had received their final grades; however, to ensure the record reflected their final experience with the grading contract, all were invited to a follow-up interview.

Participants

Participants (N=439) were 9-12th graders at a private, religious high school on the West Coast of the United States, where 96% of students attend college after graduation. The majority (n=284) were 10-12th graders with prior experience completing the annual five-week research paper unit, while 155 were first-year GSHS students completing the project for the first time. All students in the study received a grading contract for the first time.

The gender identity of participants was 51% female and 48% male, while one identified as a transgender male, one as non-binary, and two as other. The study sample was more diverse than the school, with 50% (*n*=226) of participants identifying as European American; 16.1% (*n*=72) as Asian or Pacific Islander; 15.9% (*n*=71) as mixed-race; 9.9% (*n*=44) as Hispanic or Latinx; 2.7% (*n*=12) as Middle Eastern; 2% (*n*=9) as African American; 1.3% (*n*=6) as Native American; and 1.3% (*n*=6) as other.

Interview participants volunteered on the post-survey and then all volunteers whose parents signed a consent form were invited to an interview. Interviews took place until the English course, course type, gender, and ethnicity were as balanced as possible to represent a diverse cross-section. While the pilot study focused on an identified high-need group of regular 12th graders with a history of earning low or failing grades on the paper, this study included all course types: 45% (*n*=18) were enrolled in regular English, 37.5% (*n*=15) were in honors or AP English, and 17.5% (*n*=7) had a diagnosed learning need. Additionally, two interviewees identified themselves as English Language Learners studying abroad.

Ethics

This research was conducted with full institutional consent and formal ethical clearance from the Department of Educational Research's Research Ethics Officer. My current courses, as well as my students in my course during the pilot study (Ward, 2021a), were excluded to minimize my positionality as a researcher studying students at my institution. Parents could opt their children out for any reason, and all interviews were scheduled after parents signed consent forms.

Instrument

Students with prior experience completed adapted versions of the Self-Worth Protection Scale (SWPS) (Thompson & Dinnel, 2007), Perceptions of Academic Stress Scale (PASS) (Bedewy & Gabriel, 2015), and Primary and Secondary Appraisal Scale (PASA) (Gaab, 2009); however, this article only presents findings on academic performance and SWPS. When applicable, "research paper" replaced a generic term for schoolwork. To standardize the surveys, all statements utilized a seven-point Likert scale.

Data Analysis

Descriptive statistics tests were used to understand the impact of different variables on students' academic performance. In addition to t-tests that compare the matched pairs, individual t-tests were also run for demographic variables, including gender, course type, grade level, sports season status, and health during the unit. Using the Benjamini-Hochberg procedure, p-values were adjusted to control for the familywise error rate.

Interview questions followed the analysis of the quantitative strand. I transcribed all interviews in full using Trint and coded them in NVivo 12 by first using the broad-brush, or 'bucket' coding, technique and then taking stock of the "diversity of opinions in each code, the volume of data and the relative importance participants assign to them while simultaneously coding to more discrete subcodes" (Jackson & Bazeley, 2019, p. 69). In this way, I took a deductive stance to generate initial codes and subcodes, which then led to the generation of broad themes, which are presented over two articles: one focused on perceptions of stress (Ward, 2021b) and this one, focused on academic performance and self-worth protection behaviors. Each dataset was

analyzed separately. The findings of the qualitative data are used to elucidate how and why the quantitative findings occurred.

Findings

Quantitative Findings on Academic Performance

Of all participants (*N*=439), 90% (*n*=390) fulfilled the grading contract to earn either an A or a B on the final assessment and 97% (*n*=421) earned a passing grade (i.e., A, B, and C), including 97.5% of Prior-Experiencers (*n*=269) and 97% of First-Timers (*n*=152) (see Table 4). More specifically, 84% (*n*=130) of First-Timers fulfilled the grading contract, including 84% (*n*=111) of 9th graders and 88% (*n*=14) of students who transferred from another high school for 10th, 11th, or 12th grade (see Table 5). Of Prior Experiencers, 94% (*n*=260) fulfilled the contract. The previous year, only 75% (*n*=208) of the same students earned a B or higher with the traditional grading rubric, which had the same requirements. Since the difficulty and demands of the task increase with each grade level, academic performance is expected to remain relatively stable, yet Prior Experiencers had a 25% increase in the numbers of A and B earned under the contract.

Table 4

Descriptive Statistics of Academic Performance of Students with Prior Experience

Variable	Total *N*	*N* Fulfilled Contract	% Fulfilled Contract	Earned C	Not Passing (>69%)	Not Passing %
All students	278	260	94%	15	7	2.5%
Accommodated English	23	19	83%	2	2	9%
LD in Regular English	10	9	90%	1	0	0%
LD in Honors English	2	2	100%	0	0	0%
Regular English	147	132	90%	12	3	2%
Honors English	112	109	97%	2	1	<1%
10th Grade	56	50	89%	5	1	2%
11th Grade	121	114	94%	6	1	<1%
12th Grade	100	95	95%	4	1	1%
Males	124	112	90%	9	4	3%
Females	154	146	95%	6	2	>1%
In-Season Athletes	126	114	90%	10	2	1%
Non-Athletes	156	146	94%	5	5	3%
Ill During Unit	96	87	91%	6	3	3%
Healthy During Unit	186	173	93%	9	4	2%

Note. This table reveals the rate of contract fulfillment for each demographic variable among students with prior experience with the assessment.

Table 5

Descriptive Statistics for First-Year Students' Academic Performance

Variable	Total *N*	*N* Fulfilled Contract	% Fulfilled Contract	Earned C, Passing but Below Contract	Not Passing (>69%)	Not Passing %
All FT Students	154	130	84%	20	4	2.6%
9th Grade	140	117	84%	19	4	3%
Transfers (Grades 10-12)	16	14	88%	1	1	6%
Accommodated English	12	9	75%	3	0	0%
LD in Regular English	4	2	50%	1	1	25%
Regular English	94	82	87%	7	5	5%
Honors English	48	39	81%	9	0	0%
Males	84	67	80%	13	4	5%
Females	71	63	89%	7	1	1%
In-Season Athletes	92	78	85%	12	2	2%
Non-Athletes	63	52	83%	8	3	5%
Ill During Unit	46	38	82%	7	1	2%
Healthy During Unit	105	92	88%	9	4	4%

Note. This table reveals the rate of contract fulfillment for each demographic variable among students completing the assessment for the first time.

The next section presents the qualitative data, which revealed a chain of events that resulted in strong academic performance: a clear path to a desirable outcome → increased control and task manageability → decreased maladaptive coping and self-handicapping → increased confidence → increased academic performance.

Qualitative Findings on Academic Performance

The qualitative data revealed the conditions that made the project significantly less stressful and threatening (Ward, 2021b) and facilitated high academic achievement. The clarity of the contract promoted a sense of control over the project's demands that reduced evaluative threat because success was within reach, as Malcolm (FT honors 10th) explained:

> I knew exactly what was expected of me to get a certain grade, so that made it very straightforward to what I needed to do.

Since effort (what you had to do) produced an outcome (a certain grade), the grading contract clarified the performance-reward relationship for the students in this study, as others have observed with college students (Danielewicz & Elbow, 2009; Polczynski & Shirland, 1977). The students described prior writing experiences in which the path to success was opaque while

the grading contract provided a 'clear,' 'straightforward,' and 'direct path,' a phrase used by 12 interviewees (30%), to a desirable outcome that encouraged task-oriented effort. By spending less time deciphering the expectations, or re-doing work he did incorrectly, Eric (PE regular 11th) described focusing more on the "quality of the content and the quality of the little parts" this year compared to previous years. Similarly, Ella (PE 11th honors) observed:

> It was just so clear about what you're supposed to do, so then you could focus your energy on using good diction and improving your argument.

The contract allowed students to 'focus [their] energy' on quality and high-order thinking, rather than concerns of what they were 'supposed to do.'

Qualitative Findings on Setting and Achieving High Goals

Most (76%, *n*=334) students contracted for an A, including 85% (*n*=213) of Prior Experiencers. The previous year, however, only 43% of Prior Experiencers earned an A on the final paper. Covington (1992) observed students set unattainable goals as a self-handicapping strategy, but only if the goal includes a degree of irrationality, making their inevitable failure to achieve their goal the result of an exceedingly difficult goal, not their ability status. However, for each interviewee who contracted for an A, the choice was strategic, following a logical syllogism: if they contracted for A and fell short, they would earn a B, but if they contracted for a B and fell short, they would earn a C; thus, it was most logical to at least strive for an A, even if it was likely out of reach, as Tristan (PE regular 11th) revealed:

> Truth be told, as I'm sitting here now, I think it might have been a better idea to go ahead and get the A contract and then just let stuff slip because when you contract out for the A, your starting bar's up here instead of down here and you've got that little bit of give if you don't do well.

With clear expectations offered to them and a perceived manageable workload that appeared less than in previous years, they might as well try for the highest grade. After the assessment, Micah (FT accommodated 9th) observed the benefit of contracting for an A and regretted his decision to contract for a B:

> The grading contract made me feel confident enough in my own abilities that I could do the A one, and if I did choose the A one, then I might have gotten a better grade. Now, I know what it's like, and I can try to go for the A contract next time.

Choosing the A meant working for the destination of A with the contract as their map. As William (PE AP 12th) observed, "I'm promising myself to get an A, and I will get an A. That kind of promise is very binding morally and in the brain, too, and freeing—I got to be more passionate about writing."

Vanessa (PE regular 12th) had a history of B grades on the assessment, which she credited to her busy sports competition schedule; however, she contracted for and then earned her first A on the assessment, saying, "I just kind of wanted to push myself and see if I could do it even during my season. I still did it." Under the contract, only 67% (*n*=143) earned an A, meaning some did not reach their goal, but this was expected, following the syllogism. Alexander, for example, contracted for an A but earned a B, saying, "I felt like it was well deserved. I mean, you know, I've never really done a research paper before. It was something new, and I did my best."

Conversely, Ainsley (PE 11th regular), the only interviewee who earned an F, never submitted her paper. After two previous poor experiences with the assessment, she described herself as

"extremely" likely to avoid tasks at which she might not succeed and said, "I have previous bad experiences with research papers that interfere with my ability to tell myself that I'm going to do good on this one." Despite this reflection during the interview, she described "start[ing] off pretty well with gathering sources," continuing—

> But that's really kind of where it stopped, and after I missed one or two assignments, that's where something in my brain just clicks to where I didn't think that. . . [pause] I didn't think that I could catch up, so I just kind of stopped.

Ainsley said the contract made the task seem smaller, but that perception shifted as she found herself unable to meet a progress check, which led her teacher to zero into the grade book. She said, "It was very scary because it brought down my grade a lot, and it just continued. It felt hard to get up and start again." Representing the 2.5% (*n*=7) of students who earned failing grades, Ainsley's experience breaks down the myth of the motivational F and the power of grade applications to trigger self-worth protection behaviors, shutting down learning and motivating students away from learning.

Fortunately, under the contract, the majority of students were less prone to procrastination, the focus of the next section.

Quantitative Findings on Reducing Academic Procrastination

Prior Experiencers were significantly less likely to engage in maladaptive avoidance behaviors and "play it safe" under the contract, although the effect size was small (see Table 6). The quantitative analysis revealed that students with PE were significantly less likely to avoid challenges that could end in failure under the contract (M=3.82, SD=1.468) compared to their prior experience (M=4.22, 1.53) 3.598(283), p=.000. They were also significantly less likely to play it safe by "choos[ing] goals that were within my reach" under the contract (M=4.41, SD=1.412) compared to their prior experience (M=4.74, SD=1.262) 3.495(284), p=.004. They were also less likely to choose "safe goals," as 75% of students with PE contracted for an A while 25% contracted for a B, and thus they were significantly more likely to report that it was "easy to put my best effort into the research paper because there was little risk of failure" (from M=3.96, SD=1.493 to M=4.22, SD=1.433), -2.47(284), p=.014; and that they were "able to try my hardest because failing the research paper would not reveal low ability" under the contract (M=4.22, SD=1.433; M=4.08, SD=1.33) compared to their prior experience (M=3.96, SD=1.493; M=3.82, SD=1.432), -2.452(284), p=.000.

Qualitative Findings on Reducing Academic Procrastination

At all levels, the qualitative analysis found that the majority (57%, *n*=23) experienced more confidence under the contract, as Madison (FT regular 9th) revealed:

> With the contract, it's not like you're guessing to yourself, like, "Did I do this right?" It's like, "Did I do everything I needed to do because there's a list right there?" And it's also kind of like you're getting your self-confidence that you're saying you're gonna get an A. . . . I think their mindset changes and they convince themselves they're going to get an A.

Many Prior-Experiencers (39%, *n*=7) shared that under the analytic rubric, lack of clarity or purpose in expectations led to maladaptive task avoidance that results from a student becoming debilitated by indecisions and then failing to complete tasks on time (Lay & Schouwenburg, 1993). Under the contract, however, Evan (PE AP 11th) explained:

Table 6

Matched-Pairs t-Test Results for Prior-Experiencers Avoidance Orientation

Statement	Grading Rubric		Grading Contract		*n*	*t*	FDR-adjusted *p*-values	Cohen's *d*
	M	SD	M	SD				
It was easy to try hard because my performance wasn't being judged.	3.54	1.523	3.9	1.415	284	-3.852	.000	0.245
I avoided challenges with the research paper that could end in failure.	4.22	1.53	3.82	1.468	283	3.598	.000	0.266
I felt like my performance was being judged, so it was hard to try my best.	3.63	1.721	3.2	1.553	284	3.852	.000	0.262
I tended to play it safe with my research paper and choose goals that were within my reach.	4.74	1.262	4.41	1.412	284	3.495	0.004	0.237
I found it easy to put my best effort into the research paper because there was little risk of failure.	3.96	1.493	4.22	1.433	284	-2.47	0.014	0.178
I was able to try my hardest because failing the research paper would not reveal low ability.	3.82	1.432	4.08	1.33	284	-2.453	0.024	0.188
I underachieved relative to my level of ability, choosing easy goals in order to ensure success on the research paper.	3.82	1.495	3.35	1.458	279	2.277	0.033	0.183

Note. This table reveals the significant t-test results for avoidance orientation among Prior-Experiencers.

> I spent less time procrastinating on it—I was able to just jump in and do what I had to do for that day and just be done, and I wasn't like wasting time being worried, like, "How do I write this?"

In previous years, task delay was often a result of ambiguity of purpose, which led students to dread the task, but this year, Riley (PE regular 11th) said:

> I was more motivated because sometimes in the past, I had difficulty with the research paper, so it's almost like that dreading feeling, not wanting to get into it because I knew I was struggling, but this year, I did want to get into it because it was just simple, and I wanted to get it done. . . . For me, when it's less work, I'm more likely you just do it immediately. If it's like a three-hour assignment that's due the following day, I'm going to push that back to later in the night, so for me, that's horrible.

Even Sara (PE honors 11th), a high-achieving student with a history of success on the assessment, described a difference in task-oriented motivation under the grading contract:

> This year, I felt it was better. In previous years, maybe it felt more like a negative stress was put on me, like, 'Oh, I don't want to do this. I don't want to do this.' But this year, I felt like it was more positive in motivation, like, 'Let's just get it done. Let's just get it done' kind of thing. Yeah, this year was that: 'Let's just get it done.'

As a result of minimizing task delay by "just get[ting] it done," she found that she had more "more time to relax" because she "got everything done so early." The ability to 'check' off the requirements also increased personal satisfaction with her work and move on. As she reflected on her prior experience, she observed little of the before-the-due-date stress and pressure this year because of her new attitude:

> Last year, I felt like I was procrastinating a little bit. I feel like last year, it was like two days before it was due and I was still editing things, like fixing them and doing like my revisions from the Writing Center. Whereas this year, I went there and edited early, and I got it done, and then I had like a week and a half where I barely touched my paper.

Sara earned a 95, saying, "It was comforting to know that this grading was a two-way street: if I completed all the work I needed to, at a high level, I would be rewarded with that grade." To Sara, contract grading is a reciprocal and fair grading system, in which both student and teacher enter into a mutual agreement, as Inoue (2019) noted. The implication is that traditional grading practices are less rewarding or fair for students as the expectations are often obscured.

Some (n=10, 25%) participants also described enjoying their topic more this year than previously. For example, William (PE AP 12th) revealed that the grading contract curbed procrastination by helping him think less about evaluation and judgment and more about writing, saying:

> I would procrastinate or kind of push the work to a later time because I wasn't enjoying myself while I was writing, because I was always being a mathematician in my head, always calculating points. I love being an artist, and I always love being free, and I always love being very expressive. And that's what I could do with this year's research paper, so I got excited about writing, so I would get to it more quickly than in the past. . . . I wasn't thinking that I was writing for a school. It was more about writing this to share to other people and show that I really care about this, because since I'm not caring about grades as much, I was caring more about content.

Enjoyment of learning may be a positive consequence of experiencing less worry and doubts about performance outcomes, as James (FT honors 9th) explained:

> I think in a way it helps you to focus less on all the individual points that you should have and allow you to just write and be more open to that. Because in the past I've found people are stressing, making sure they check every single bullet point so they can get the best grade possible. And in a way, I think it could be helpful to have a specific rubric so that

> you know exactly what you should be doing. But at the same time, it's important to have those liberties of knowing. You know, I can just do these and get a good grade and then still write the way I feel fits the topic.

The grading contract liberated some students to take risks by writing in 'the way [they] feel fits the topic' while still having the assurance of 'get[ting a good grade.'

Quantitative Findings on Reducing Social Comparison

Under the grading contract, students were also significantly less likely to engage in upward and downward social comparison, a coping mechanism for managing demands (see Table 7). More specifically, Prior Experiencers were significantly less likely to say that they "compared [their] ability to those around [them]" under the grading contract (M=4.74, SD=1.777) compared to the previous year (M=5.21, SD=1.695), 3.976(284), p=.000, with an effect size of .270.

Table 7

Social Comparison t-Test Results for Prior-Experiencers

Statement	Grading Rubric		Grading Contract		*n*	*t*	FDR-adjusted *p*-values	Cohen's *d*
	M	SD	M	SD				
I compared my ability to those around me.	5.21	1.695	4.74	1.777	284	3.976	.000	0.270

Note. This table reveals the significant *t*-test results for social comparison among Prior Experiencers.

Qualitative Findings on Reducing Social Comparison

The interviews revealed that social comparison often arises from doubts about one's work and unclear expectations. In the absence of these fears, the contract reduced social comparison behavior:

> Sara (PE honors 11th): I feel like in previous years I've been more eager to ask people like, "Oh, where are you? How many words do you have?" And this year, I was more just focused on myself and getting my own work done.
>
> Interviewer: Why do you think that was?
>
> Sara: I don't know. I just kind of felt like if I did those 10 steps on the contract, I would be fine, and I don't worry about what anyone else is doing.

Students relied on social comparison as a means of self-evaluation. With a reduction in their perception of stress and fear of failure, social comparison significantly decreased as they did not need "to worry about what anyone else is doing" (Sara, PE honors 11th). Students in this study described prior experiences in which social comparison served as a way to check their progress against their peers and ensure their work was complete and correct. Social comparison theory, put forth by Festinger (1954) and cited in Taylor et al. (1990), states that individuals prefer objective, non-social criteria to evaluate themselves, yet in its absence, they will use "social information—

namely, other people" (p. 75). Strong emotions, such as fear, can also lead to social comparison. Combined with previous findings (Ward, 2021b), perceptions of stress and evaluative threat significantly reduced under the contract, thereby decreasing social comparison.

Quantitative Variables Impacting Performance

Across course types (i.e., regular, honors/AP, and accommodated for learning disabilities), the academic performance of students enrolled in regular English courses improved most under the contract system (see Table 8).

Table 8

Results of Academic Achievement t-Test by Demographic Variable (Grades 10-12)

Variable	Grading Rubric		Grading Contract		*n*	*t*	FDR-adjusted *p*-value	Cohen's *d*
	M	SD	M	SD				
1 All Students	85.49	9.34	88.26	9.90	283	-4.718	.000	.288
2 Accommodated Students	88.17	7.5	84.30	14.11	23	1.610	.144	.342
3 Regular Students	81.16	10.01	86.96	10.18	147	-6.934	.000	.574
4 Honors Students	90.62	4.98	90.78	7.77	112	-.220	.895	.024
5 Sophomores	84.21	12.09	89.89	8.97	56	-4.166	.000	.534
6 Juniors	86.66	9.3	86.58	11.36	126	.078	.938	.008
7 Seniors	84.66	7.217	89.35	8.67	101	-5.866	.000	.588
8 In-Season Athletes	85.83	8.69	87.75	8.5	126	-2.179	.040	.223
9 Non-Athletes	85.21	9.85	88.67	10.91	156	-4.401	.000	.332
10 Ill During Unit	85.31	9.63	87.86	9.66	96	-2.359	.028	.264
11 Healthy During Unit	85.59	9.2	88.47	10.05	186	-4.145	.000	.300
12 Males	83.55	9.68	86.74	9.18	125	-3.569	.001	.338
13 Females	87.32	8.36	90.04	7.4	154	-3.858	.000	.345

Note. This table reveals statistically significant and non-significant findings of the matched-pairs *t*-test, which analyzed changes in returning students' academic performance from 2019 to 2020.

Qualitative Variables Impacting Performance

Honors students described prior situations in which they had to ascertain their teacher's implicit expectations and thus were able to meet high standards without explicit directions. For example, Joshua (PE AP 11th), noting similar grades all three years, explained:

> I liked that [the contract] laid out very clearly what was expected, but it wasn't that much different than the rubric for me, but, overall, I feel like my response to it was maybe slightly more like I knew what was happening, but there wasn't much of a difference for me.

As measured by outcomes, "there wasn't much of a difference"; he had earned an A on the assessment each year. Prior-Experiencers enrolled in honors and AP courses earned equivalent grades under both the grading contract and the rubric: M=90.62 with the rubric and M=90.78 under the contract.

Adolescents enrolled in regular courses described situations in which they worked hard and earned low grades, as Eric (11th grade, regular) observed:

> If you don't have a set requirement or what you have to do to get to a certain grade, then, who knows? Like, maybe put in all your effort, but you still get a D or something like that, but now I know exactly what I have to do.

Teachers may assume that the academic performance of middle-ability students is low due to a lack of effort; however, to 'put in all your effort' and then 'still get a D' reveals misdirected effort. The contract enabled more students to focus on beneficial tasks and find newfound success. For example, Prior Experiencers enrolled in regular courses also had the most significant grade increase under the contract (M=86.96, SD=10.18) compared to last year (M=81.16, SD=10.01), t(147)-6.934, p=.000.

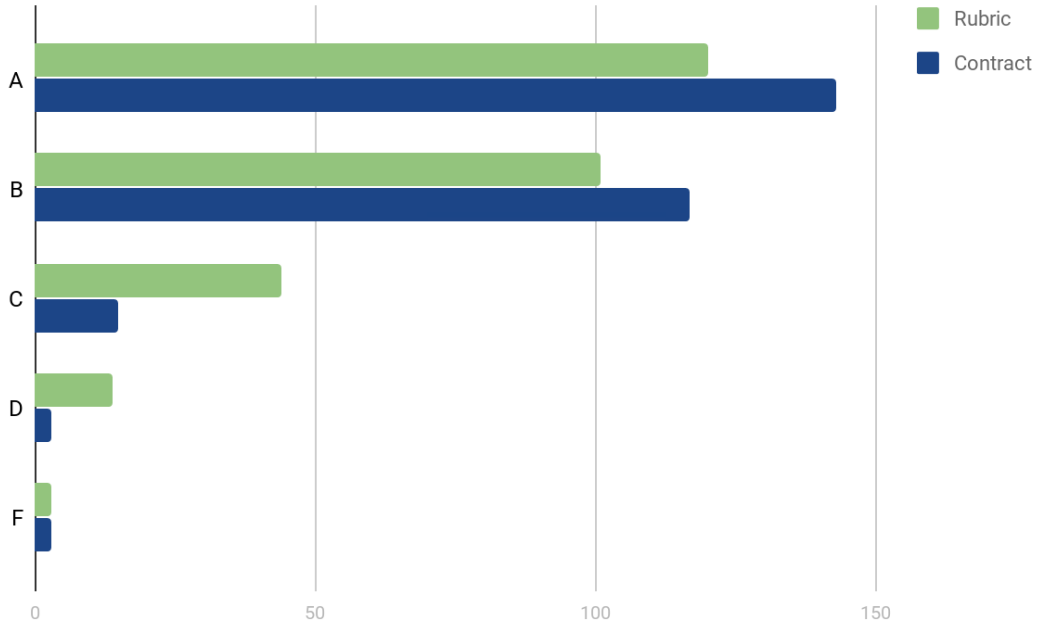
First-Timers enrolled in regular English courses also had the highest percentage of students (87%, *n*=82) who fulfilled the grading contract, compared to 81% (*n*=39) of honors courses and 75% (*n*=9) of accommodated courses. Finally, while regular students were identified as a high-need group due to their low or failing grades from year to year, 87% of regular 9th graders earned a B or higher under the contract, along with 91% (*n*=50) of regular 12th graders, only 64% (*n*=35) of whom earned a B or higher the previous year under the rubric.

Discussion

The quantitative and qualitative analysis reveals that contract grading, as utilized by 12 different teachers in 13 courses, reduced self-worth protection behaviors, facilitated high goal setting, and helped students reach higher academic standards, as measured by the rate of contract fulfillment, particularly for typically middle-ability students enrolled in regular courses. As previously mentioned, academic growth may account for some variance in students' performance, but it is unlikely to be statistically significant given increased task demands. Under the contract, however, the number of Prior Experiencers who earned As (*n*=143) increased by 19% and Bs (*n*=117) by 16% (see Figure 2). Additionally, 19% of Prior Experiencers (*n*=53) earned a C or lower the previous year compared to 7% (*n*=20) under the grading contract. Ultimately, the result was a 62% decrease in grades C or lower and an 11% increase in grades B or higher. Supporting the grade improvements observed in college students under the contract (Fairbanks, 1992; Lindemann & Harbke, 2011), the findings of this study also reveal that decreased workload alone did not improve academic performance, as the pilot study suggested (Ward, 2021a).

Figure 2

Academic Performance of the Cohort with Prior Experience



Note. The bar chart shows the improvement of students' grades from 2019, with the grading rubric, to 2020, under the grading contract.

Overall, 90% (n=390) of students fulfilled the high expectations of the grading contracts to earn either an A or B on the assessment. As Danielewicz and Elbow (2009) suggested, the goal of "badger[ing] and cajol[ing] every student into a getting a B—that is, into doing everything we specified in the contract" was overwhelmingly successful with the students in this study (p. 254). In this way, the criteria did appear to control performance, as Johnston and O'Neill (1973) reported, particularly because the majority of students found the expectations for quality work more clear than other assessment tools they received. As a result, they set high goals, which corroborates a finding that researchers (Koenig et al., 2016; Phan, 2009; Schippers et al., 2020) have observed in elementary school-aged children to college students: "explicitly setting goals can markedly improve performance at any given task" (Morisano et al., 2010, p. 256). Beyond academic performance, the clarity of purpose also led to a statistically significant decrease in fear of failure, thus improving the psycho-emotional environment and making it safe to put forth task-oriented effort.

The clarity and presentation of workload demands alleviated evaluative threat and reduced task avoidance. Solomon and Rothblum (1984) found that procrastination results from having too many things to do, viewing the task as unpleasant, feeling overwhelmed by the task, and feeling afraid to fail, all of which were partially alleviated under the grading contract, as Riley (PE regular 11th) revealed:

> Riley: This year was straightforward. I had a general path in my mind of where to go, and that gave me comfort, if you will, and how I should do my paper and less risk of failure.

Interviewer: You just used the word "comfort." What gave you comfort?

Riley: The contract's just, like, a piece of paper, but in a way, what it gives me as a student—the guidelines for the paper—gives me comfort.

While the expectations and word count increase with each grade level, students described prior experiences in which their workload was amplified by first having to ascertain the expectations, a laborious and uncertain task that led to academic procrastination. Defined as task avoidance or delay, academic procrastination is a form of maladaptive coping that can have consequences, including negative performance outcomes (Moore, 2008) and lower self-esteem (Owens & Newbegin, 2000). As a result, missed deadlines (Aitken, 1982), as well as lower grades and higher levels of stress (Tice & Baumeister, 1997), are common among procrastinators, yet the contract shifted students' appraisal of workload demands from threatening to challenging (Ward, 2021b). This suggests students are less likely to delay tasks over which they perceive a sense of control or manageability.

Additionally, prior experience with the assessment did not impact the rate of low or failing grades on the assessment, but this analysis suggests that the contract system combined with prior experience with high school writing expectations led to higher performance outcomes. First-Timers in Grades 10, 11, and 12 (88%) had a higher rate of contract fulfillment than 9th graders (84%) with varying prior experiences with academic writing before high school; however, Prior Experiencers had the highest rate of contract fulfillment (94%), pointing to the benefit of spaced practice and teacher feedback on durable learning that improves educational outcomes.

The academic performance of typically middle-ability students enrolled in regular courses most improved under the contract, which may be a result of improved psycho-emotional conditions. These students, however, were the most likely to contract for an A: compared to just 58% of honors students and 66% of students in accommodated English, 90% of students in regular English courses contracted for an A. This finding suggests that their ambitions may have increased their effort in the direction of success. As Fishman (2014) concluded, "those who perceived the capability to achieve academic outcomes were more likely to feel internally obligated to produce such outcomes" (p. 685). While fear of failure or disappointment held some students back, the majority of students were motivated to increase effort toward success, corroborating Polczynski and Shirland's (1977) findings that the grading contract promoted self-belief that led to more task-oriented motivation to achieve the desired results.

Limitations

One design limitation is the absence of a control group, which can lead to internal validity threats. Since 10-12th graders were asked to reflect on their academic stress during last year's research paper, they could have matured in a year's time; however, as previously mentioned, the project demands increase with each grade level, making it more likely that academic performance will remain steady, yet grades improved across multiple grade levels and demographics. Additionally, each scale adapted for the study was established with test-retest validity, and p-values were corrected for Type 1 Error as a result of multiple testing using the Benjamini-Hochberg procedure. Finally, the historical event of the global COVID-19 pandemic, which was on the rise at the time of the study, could explain a decrease in academic performance but not an increase.

Additionally, a data limitation is the participation rate of 45%, which, though lower than desired, is acceptable. I offered no incentives and took great care to exert no informal or formal

pressure on the students to mitigate the influence of my positionality as a teacher at the school at which I was researching. In the pilot study (Ward, 2021a), which had a similar participation rate, more students supplied their final grades than completed the surveys, yet grades were comparable in both groups.

Finally, one limit to the generalizability of this research is the institutional setting, which is a college-preparatory private school. While the survey participants were more ethnically diverse than the predominantly White (72%) institution, the upper-middle-class culture around the school gives students in this study access to privileged discourses (Shor, 2009), which may limit the generalizability of this research in more socio-economically diverse contexts.

Future Research

The findings of this study reveal many areas for further investigation. First, as previously mentioned, follow-up qualitative research is needed for students who earned low and failing grades to understand the factors that influence achievement. Unfortunately, only one participant who earned a low or failing grade volunteered for an interview, but he did not respond to emails to schedule the interview. Additional work can then follow students who have experienced writing under the contract into college-level writing experiences to examine their adjustment to college-level expectations, which may not be clearly communicated to students (Thomas & Rohwer, 1986).

Conclusion

The findings of this study reveal that contract grading serves as a promising alternative for secondary classrooms that use the letters and numbers of the conventional grading system (and thus satisfy institutional requirements to assign grades) while also restoring students' agency by allowing them to participate meaningfully in their assessment. Additionally, combined with the findings on perceptions of stress and evaluative threat (Ward, 2021b), the research reveals that reducing stress and improving strong academic performance are not mutually exclusive; in fact, the analysis of quantitative and qualitative data revealed that the improved psycho-emotional conditions under the grading contract had a direct and significant impact on self-worth protection behaviors and academic achievement. The limited use of contract grading generated buy-in for alternative practices that can be built on for larger institutional changes.

The findings of this research can have rebounding effects on writing assessment in secondary schools, where contract grading is underused but beneficial as a holistic tool for teaching and learning that can reduce evaluative threat during high-stakes writing assessment to improve psycho-emotional well-being and academic performance. This is not to say that five weeks of contract grading eradicated students' concern with grades—students, after all, exist in complex social environments with profound influence on their goals and behavior—but it does appear to allow them to work without being fueled by stress. As Litterio (2018) discovered with her college students, the students in this study were still performance-oriented, but importantly, many of the negative outcomes associated with such an orientation diminished, as students reported improved psycho-emotional well-being (Ward, 2021b) reduced self-worth protection behaviors, and high academic achievement. The findings of this study provide empirical support for the use of contract grading in secondary classrooms.

Disclosure statement

The author received no funding for this research and has no conflicts of interest.

References

Aitken, M. E. (1982). *A personality profile of the college student procrastinator*. [Doctoral dissertation, University of Pittsburgh]. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 43(3-A), 722–723.

Ainslie, R. C., Shafer, A., & Reynolds, J. (1996). Mediators of students' stress in a college preparatory environment. *Adolescence*, 31(124), 913–924.

Amsden, S. (1970). Have you ever tried contracting for grades? *English Journal*, 59(9), 1279–1282.

Arsenio, W., & Loria, S. (2014). Coping with negative emotions: Connections with adolescents' academic performance and stress. *The Journal of Genetic Psychology*, 175(1), 76–90.

Bedewy, D., & Gabriel, A. (2015). Examining perceptions of academic stress and its sources among university students: The Perception of Academic Stress Scale. *Health Psychology Open*. https://doi.org/10.1177/2055102915596714

Bowers, G., & Howard, M. (1975). Sameness is not a virtue. *English Journal*. 64(8),49-52.

Breland, H. M. (1983). The direct assessment of writing skills: A measurement review. *ETS Research Report Series No. 83-32*, 1–23.

Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research & Evaluation*, 16(17), 12.

Cowan, M. (2020). A legacy of grading contracts for composition. *Journal of Writing Assessment,* 13(2).

Covington, M. V. (1984). The self-worth theory of achievement motivation: Findings and implications. *The Elementary School Journal*, 85(1), 5–20.

Covington, M. V. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. Cambridge University Press.

Covington, M. V., & Beery, R.G. (1976). *Self-worth and school learning*. Holt, Rinehart, and Winston.

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Sage Publications.

Danielewicz, J., & Elbow, P. (2009). A unilateral grading contract to improve learning and teaching. *College Composition and Communication*,61(2), 244–268.

de Anda, D., Baroni, S., Boskin, L., Buchwald, L., Morgan, J., Ow, J., Gold, J. S., & Weiss, R. (2000). Stress, stressors and coping among high school students. *Children and Youth Services Review*, 22(6), 441–463.

De Castella, K., Byrne, D., & Covington, M. (2013). Unmotivated or motivated to fail? A cross-cultural study of achievement motivation, fear of failure, and student disengagement. *Journal of Educational Psychology*, 105(3), 861–880.

Eiland, L., & Romeo, R. D. (2013). Stress and the developing adolescent brain. *Neuroscience*, 249, 162–171.

Fairbanks, P. J. (1992). Treating mathematics anxiety: The optional contract. *Mathematics Teacher*, 85(6), 428–430.

Feldman, J. (2019). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin.

Fishman, E. J. (2014). With great control comes great responsibility: The relationship between perceived academic control, student responsibility, and self-regulation. *British Journal of Educational Psychology*, 84(4), 685–702.

Gaab, J. (2009). PASA – Primary Appraisal Secondary Appraisal. *Verhaltenstherapie*, 19(2), 114–115.

Guskey, T. R. (2013). The case against percentage grades. *Educational Leadership*, 71(1), 68.

Hancock, D. R. (2001). Effects of test anxiety and evaluative threat on students' achievement and motivation. *The Journal of Educational Research*, 94(5), 284–290.

Inoue, A.B. (2019). *Labor-based grading contracts: Building equity and inclusion in the compassionate writing classroom* (Perspectives on Writing Series). The WAC Clearinghouse, University Press of Colorado.

Jackson, C. (2010). Fear in education. *Educational Review*, 62(1), 2010, 39–52.

Jackson, K., & Bazeley, P. (2019). *Qualitative data analysis with NVivo*. Sage Publications.

James, S. M. (1977). The evolution of a language arts program for pre- and early adolescent students. *English Journal*, 66(4), 47–51.

Johnston, J. M., & O'Neill, G. (1973). The analysis of performance criteria defining course grades as a determinant of college student academic performance. *Journal of Applied Behavior Analysis*, 6(2), 261–268.

Koenig, E., Eckert, T., & Hier, B. (2016). Using performance feedback and goal setting to improve elementary students' writing fluency: A randomized controlled trial. *School Psychology Review*, 45(3), 275–295.

Kohn, A. (2006). The trouble with rubrics. *English Journal*, 95(4), 12–15.

Kokus, N., & Mussoff, L. (1975). Assessing English: An alternative. *English Journal*, 64(3),73–76.

Lay, C. H., & Schouwenburg, H. C. (1993). Trait procrastination, time management, and academic behavior. *Journal of Social Behavior and Personality*, 8(4), 647–662.

Lazarus, R., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer Publications.

Lindemann, D., & Harbke, C. (2011). Use of contract grading to improve grades among college freshmen in introductory psychology. *SAGE Open*, 1(3), 1–7.

Litterio, L. M. (2018). Contract grading in the technical writing classroom: Blending community-based assessment and self-assessment. *Assessing Writing*, 38, 1–9.

Martin, A. J., Marsh, H. W., & Debus, R. L. (2001). Self-handicapping and defensive pessimism. *Journal of Educational Psychology*, 93(1), 87–102.

Mazur, R. (2018). Backward design. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation*. SAGE Publications. https://doi.org/10.4135/9781506326139

McLaughlin, J. G. (1961). Contracts help solve project problems. *The Clearing House*, 35(7), 418.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. Jossey-Bass.

McArthur, J. (2018). *Assessment for social justice perspectives and practices within higher education* (1st ed.). Bloomsbury Academic.

Moore, R. (2008). Academic procrastination and course performance among developmental education students. *Research and Teaching in Developmental Education*, 24(2), 56–67.

Morisano, D., Hirsh, J., Peterson, J., Pihl, R., & Shore, B. (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. *Journal of Applied Psychology*, 95(2), 255–264.

Nicholls, J. (1979). Quality and equality in intellectual development: The role of motivation in education. *American Psychologist*, 34(11), 1071–1084.

Owens, A., & Newbegin, I. (2000). Academic procrastination of adolescents in English and mathematics: gender and personality variations. *Journal of Social Behavior and Personality,* 15(5), 111–124.

Phan, H. (2009). Relations between goals, self-efficacy, critical thinking and deep processing strategies: A path analysis. *Educational Psychology*, 29(7), 777–799.

Polczynski, J., & Shirland, L. (1977). Expectancy theory and contract grading combined as an effective motivational force for college students. *Journal of Educational Research*, 70(5), 238–241.

Putwain, D. W., Nicholson, L. J., Nakhla, G., Reece, M., Porter, B., & Liversidge, A. (2016). Fear appeals prior to a high-stakes examination can have a positive or negative impact on engagement depending on how the message is appraised. *Contemporary Educational Psychology*, 44-45, 21–31.

Putwain, D., & Remedios, R. (2014). Fear appeals used prior to a high-stakes examination: What makes them threatening? *Learning and Individual Differences*, 36, 145.

Putwain, D. W., & Roberts, C. M. (2012). Fear and efficacy appeals in the classroom: The secondary teachers' perspective. *Educational Psychology*, 32(3), 355–372. doi:10.1080/01443410.2012 .659845.

Rachal, J.R. (1984). Community college and university instructor consistency in the evaluation of freshman English themes. *Community Junior College Quarterly of Research and Practice*, 8, 127–140.

Schippers, M., Morisano, D., Locke, E., Scheepers, A., Latham, G., & De Jong, E. (2020). Writing about personal goals and plans regardless of goal type boosts academic performance. *Contemporary Educational Psychology*, 60.

Shor, I. (2009). Critical pedagogy is too big to fail. *Journal of Basic Writing*, 28(2), 6–27.

Starch, D., & Elliott, E. (1912). Reliability of the grading of high-school work in English. *The School Review*, 20(7), 442–457.

Solomon, L. J., & Rothblum, E. D. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology*, 31, 503–509.

Taylor, S. E., Buunk, B. P., & Aspinwall, L. G. (1990). Social comparison, stress, and coping. *Personality and Social Psychology Bulletin*, 16(1), 74–89.

Thomas, J., & Rohwer Jr., W. (1986). Academic studying: The role of learning strategies. *Educational Psychologist*, 21(1-2), 19–41.

Thompson, T. (1999). *Underachieving to protect self-worth: Theory, research and interventions*. Avebury.

Thompson, T., & Dinnel, D. (2007). Is self-worth protection best regarded as intentional self-handicapping behaviour or as an outcome of choking under pressure? *Educational Psychology*, 27(4), 1.

Thompson, T., & Parker, C. (2007). Diagnosing the poor performance of self-worth protective students: A product of future outcome uncertainty, evaluative threat, or both? *Educational*, 27(1), 111–134.

Thompson, T., & Richardson, A. (2001). Self-handicapping status, claimed self-handicaps and reduced practice effort following success and failure feedback. *British Journal of Educational Psychology*, 71(1), 151–170.

Tice, D. M., & Baumeister, R. F. (1997). Longitudinal study of procrastination, performance, stress, and health: The costs and benefits of dawdling. *Psychological Science*, 8(6), 454–458. https://doi.org/10.1111/j.1467-9280.1997.tb00460.x

Ward, E. (2021a). An integrated mixed-methods study of contract grading's impact on adolescents' perceptions of stress in high school English: A pilot study. *Assessing Writing*, 48, 100508.

Ward, E. (2021b). Easing stress: Contract grading's impact on adolescents' perceptions of workload demands, time constraints, and challenge appraisal in high school English. *Assessing Writing*, 48, 100526.

Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education*, 34(3), 14–27.