Use of Contract Grading to Improve Grades Among College Freshmen in Introductory Psychology

Dana F. Lindemann¹ and Colin R. Harbke¹

Abstract

The use of behavioral techniques in college teaching has declined during the past three decades. The purpose of this study was to compare a behaviorally based grading approach with a traditional point-based system. A total of 40 college freshmen were randomly assigned to a Traditionally Graded or Contract Graded Introductory Psychology course. Contract graded students were one third as likely to fail or withdraw, 3 times more likely to earn an A grade, and were more likely to perceive a high degree of control over their grade. These findings support use of a contract grading system in the contemporary college classroom.

Keywords

applied psychology, behaviorism, educational psychology, educational research, higher education, psychology, contract grading

Despite their effectiveness, the use of behavioral techniques in the classroom has decreased since their height of popularity in the 1970s (Saville, Zinn, Neef, Van Norman, & Ferreri, 2006). Bryan Saville and his colleagues identified four possible factors contributing to this declining use of behavior principles, including (a) an instructor’s resistance to change, (b) incompatibility between program requirements and the academic calendar, (c) the increased workload required by some behavioral systems, and (d) lower confidence in the success of behavioral systems following misuse or misapplication. Perhaps in response to these criticisms, researchers have started looking for ways to incorporate variations on behavioral principles to today’s classroom. Among these variations is contract grading.

Contract grading is a system in which students determine and specify at the beginning of a class the grade they would like to earn, from a set of instructor-defined parameters. Although there are a myriad of ways to implement contract grading, in general, the following components are employed: (a) assignments are graded pass or fail; (b) a mastery criterion (often between 80% and 100%) is determined by the instructor, and meeting this criterion is required to obtain a passing grade; (c) students are provided multiple attempts to earn a passing grade for each assignment; and (d) students choose from a variety of assignments.

The use of contract grading is supported by wide-ranging benefits. Students may be better able to monitor their progress, feel they have a greater role in their learning, and are more likely to perceive primary control over their grade (Grau, 1999; Kirschenbaum & Wetter Riechmann, 1975; Polczynski & Shirland, 1977). As such, contract graded students may be more motivated to perform well. In addition, when assignments are graded pass or fail, emphasis is placed on mastery of the material, as opposed to gaining a partial understanding of the material. And, despite its novelty, students may also prefer contract grading to traditional points-based grading systems (Bunn Hiller & Hietapelto, 2001; Kirschenbaum & Wetter Riechmann, 1975).

Among the few studies used to compare the effects of contract grading and traditional grading approaches, the data support using contract grading or its components (see Johnston & O’Neill, 1973; Kirschenbaum & Wetter Riechmann, 1975; Semb, 1974). Despite these promising findings, contract grading appears to be rarely used in classrooms today. In addition, relative to the 1970s, many professors employ technologies not previously available, such as computer-based lecture presentations and use of online course management systems. More important, however, is the impact technological changes, in general, occurring over the past decades (e.g., computers, cellular phones) have had on college students. For example, students today were born after widespread adoption of the Internet and are accustomed to receiving information and communications immediately via text messaging and Internet phones. Because of these technological and resulting cultural changes,

¹Western Illinois University, Macomb, USA

Corresponding Author:
Dana F. Lindemann, Department of Psychology, Western Illinois University, 1 University Circle, Macomb, IL 61455, USA
Email: DF-Lindemann@wiu.edu
and their influence on young adults (see, for example, Twenge & Campbell, 2001), there exists a need to systematically evaluate and confirm the effectiveness of contract grading approaches in the contemporary classroom. In particular, the question remains, “Is contract grading, as a teaching technology, as effective and applicable in today’s classrooms, and today’s students, as it was 40 years ago?” To that end, the purpose of this research is to evaluate and compare contract grading and traditional points-based grading systems on course performance and preferences among college freshmen in a contemporary classroom.

Method
Participants and Setting
Participants were 16 male and 24 female college freshmen attending a rural, state university. Participants were enrolled in the survey course, “Introduction to Psychology,” which was also designated a First Year Experience (FYE) course. The FYE program is designed to ease students’ transition from high school to college and to increase retention between the freshmen and sophomore years. FYE courses have small class sizes and in addition to the required curriculum, emphasis is placed on learning outside of the classroom through service and cocurricular events. All students attending their freshmen year were required to register for at least one designated FYE course.

Contract Grading Components and Course Materials
Each of the common components of contract grading were utilized: (a) assignments were graded pass or fail, (b) a mastery criterion of 85% was required to earn a passing grade, (c) students were allowed up to two submissions to earn a passing grade for each assignment, and (d) students choose from a variety of assignments.

Course assignments included writing selections, activities, and exams. Each writing selection equaled approximately three pages of text and required students to engage in some task and then write about their experiences (e.g., create and present a poster describing a published research study). Activity selections included in-class activities as well as participating in university events or psychology-related research outside of class. Four of the exams covered three chapters of material each and consisted of matching, fill-in-the-blank, multiple-choice, short-answer, and essay questions. The fifth and final exam was cumulative over the semester and consisted of matching and multiple-choice questions.

Requirements for each grade were explained on the grade contract (see appendix). Students identified the grade they wanted to contract for and then selected the writing and activity assignments they would complete to accomplish this goal. The appropriate criterion for passing exams was marked, and student and instructor signed the contract.

Measures
Student retention and course grades were recorded as primary dependent variables. As both these variables were based on the official university roster at the beginning of the semester, these dependent variables were available for all 40 students involved in the study.

Secondary measures came from three self-report questionnaires designed to evaluate the course, instructor effectiveness, and student perceptions of their own effort. For the first questionnaire, students were asked to “grade” their instructor (A, B, C, D, or F) on 9 items, such as “Grade the instructor’s ability to provide a challenging course,” and “Grade the clearness of the course requirements and grading procedures.” The second questionnaire (23 items) asked students to rate the frequency of specific behaviors by the instructor or student as it pertained to the course (response scale ranged from 1 = hardly ever to 5 = almost always). Example items include, “The instructor clearly defined students’ responsibilities,” and “I worked hard for the grades I received in this class.” For the third questionnaire, students also rated their agreement with 12 statements about the course format and content (response scale ranged from 1 = disagree strongly to 5 = agree strongly), such as “I enjoyed the format of the course” and “The overall workload for this course was too much.”

An additional outcome measure was taken from the average response to a self-report question regarding students’ perceived control over their grade: “On a scale from 1 (not at all in control) to 10 (almost entirely in control), how much control do you believe you personally had over your grade in this class?”

Procedure
At the outset of the semester, the instructor randomly assigned students enrolled in one course section to the Contract Grading Group (n = 20) and those enrolled in the other section to the Traditional Grading Group (n = 20). Both sections were taught by the same instructor, who at the time had 6 years of college teaching experience. With the exception of the contract grading elements (including the ability to resubmit an assignment), all of the lectures, writing and activity assignments, and exams were identical in both sections. Requirements for each grade were determined such that any grade was comparable between sections (i.e., an A in the contract grading course was comparable with an A in the traditional grading course).

The grading structures were described in the syllabus corresponding to each section and reviewed in class during the 1st week of the semester. Students in the Contract Grading Group submitted a signed contract during the 1st week of class and, if necessary, were allowed to modify their contract during Week 10 of the semester. Students in the Traditional Grading Group received points for each assignment, and
grades were based on the percentage of total points earned (an A for 90% or higher, B for 80%-89%, etc.). Grades for students in the Contract Grading Group were based on the number of writing assignments, activities, and exams that received a passing grade (see appendix). When a writing or activity submission did not receive a passing grade, the student was informed of what changes were needed, and one resubmission was allowed to reach the mastery criterion. For exams, students who did not pass the first attempt viewed their graded exam for 10 min and were allowed a second attempt 2 days later to reach the mastery criterion (i.e., 80% correct responses). All assignments were graded by a graduate teaching assistant, who was blind to the manipulation.

During the penultimate week of the semester, students were asked to complete the three self-report measures and the perceived control item by a graduate student who was blind to the study and manipulation. To complete these self-report measures, students needed to be still enrolled in the course and present in class on the day they were administered. As such, self-report data were available from a total of 28 students, with 15 contract grading (75% of initial enrollment) and 13 traditional grading students (65% of initial enrollment) completing these measures. Responses to the self-report measures were anonymous and the instructor was not present during administration. Aggregate mean scores for each self-report measure as well as inspection of representative items from the student and instructor behavior measure and the course format and content measure were used to compare student opinions and experiences at the end of the semester.

Results and Discussion

The purpose of this research is to evaluate and compare contract grading and traditional points-based grading systems on course performance and preferences among college freshmen in a contemporary classroom. Categorical data for both groups were summarized with percentages and compared between groups using the relative risk (RR) ratio. Students in the Contract Grading Group (60%) were 3 times more likely than those in the Traditional Grading Group (20%) to earn an A grade in the course, RR = 3.00, 95% confidence interval (CI) = [1.16, 7.73], p = .01. In addition, Contract Grading students (15%) were one third as likely (RR = 0.33, 95% CI = [0.11, 1.05], p = .04) to attrit through course withdrawal or failing grades than Traditional Grading students (45%).

For the nine-item teaching evaluation for which students “graded” the instructor’s performance and the course, the average grade was approximately one half of a letter grade higher for the contract grading class (M = 3.61 on a 4-point scale) than for the traditional grading class, M = 3.21.

Quantitative data were summarized with group means and compared using independent-samples t tests. Students in the Contract Grading Group consistently rated the instructor’s and their behaviors toward the course more favorably. All 23 behaviors described on the first questionnaire received higher average endorsements by the Contract Grading Group than the Traditional Grading Group. And, averaged across the 23 items, this difference amounted to a mean rating that was 1.45 standard deviations higher for the Contract Grading Group (M = 4.51, SD = 0.58) than for the Traditional Grading Group (M = 3.80, SD = 0.38), t(26) = 3.74, p < .001. The upper half of Figure 1 displays the mean ratings for six representative items and an aggregate score for the entire scale. Representative behavior items were selected based with emphasis on likely areas of both similarity and differences between the Contract Grading and Traditional Grading Groups. As both groups received the same exams, for instance, little differences were expected to the item “Course exams related to the material”; in contrast, larger differences were expected for items related to clarity of expectations (e.g., “Student responsibilities clearly defined” or “Class and assignments clearly organized”). The average item-total correlation, using Fisher’s (1921) r-to-z conversion, for the six representative items were higher (mean item r = .82, range = .28-.93) than those for items that were not selected (mean r = .67, range = .04-.91).

A similar pattern emerged for the 12 statements regarding the course format and content (see bottom half of Figure 1) where the average difference was nearly one standard deviation (d = 0.97) higher for Contract Grading Group (M = 4.05, SD = 0.76) than for the Traditional Grading Group (M = 3.36, SD = 0.65), t(26) = 2.54, p = .017. Figure 1 also includes scores for six representative format and content items. As before, representative items were selected based on likely sources of difference between the Contract Grading and Traditional Grading Groups, and the average item-total correlations, following Fisher’s (1921) r-to-z conversions, were higher among the representative items (mean r = .79, range = .50-.90) than nonrepresentative items (mean r = .73, range = .07-.89).

When asked about the extent to which they felt in control over their grades, 67% of students in the Contract Grading Group reported a 9 or 10 (the maximum on the response scale was 10, “almost entirely in control”) compared with only 23% of those in the Traditional Grading Group, RR = 2.89, 95% CI = [1.01, 8.30], p = .02. Viewed across the full range of the scale, however, the mean ratings were similar for both groups (Contract Grading: M = 8.5, SD = 2.18; Traditional Grading: M = 8.04, SD = 1.27), d = 0.26, t(26) = 0.67, p = .51. Inspection of the response distributions for both groups revealed that, although the majority of Contract Grading students indicated a maximum or near maximum amount of control of their grades, two contract grading students (13%) indicated less control (i.e., 3 or 4) than any of the traditional grading students, for which a response of 6 was the lowest.

To summarize briefly, students in the Contract Grading Group were one third as likely to fail or withdraw, 3 times more likely to earn an A grade, and, in general, were more likely to
Figure 1. Mean student ratings of instructor and student behaviors (a) and course format and content (b) in the Contract Grading and Traditional Grading Groups for six representative items and aggregate scores for each measure. Note: Responses were made on a 5-point Likert-type scale. Error bars represent ± one SE.

perceive a high degree of control over their grade. In addition, those in the Contract Grading Group consistently rated their effort, the course, and the instructor more favorably.

These findings indicate that contract grading is applicable to and effective in a contemporary college classroom, among today’s college students. In addition, contract grading was implemented with few changes to the preexisting material. The methodology employed allows a direct comparison between contract grading and traditional points-based grading systems. The nature of contract grading required
variation in key course and evaluation components (e.g., student selection of assignments). Most notably, contract grading permitted multiple attempts to achieve the mastery criterion on assignments. Although this presents a fundamental methodological confound, the reliance on mastery grading criteria is a key feature of contract grading, and one-shot assessments are ubiquitous in traditional grading approaches. As such, for the purposes of this evaluation, we elected to implement both grading systems as they are most typically employed in the classroom (i.e., mastery grading with multiple attempts for contract grading students and single-shot assessments for traditional grading students). The findings as described above should be viewed in consideration of this methodological limitation.

The use of both grading systems during the same semester introduces a possible alternative explanation for the present findings. It is possible that traditional grading students became aware that contract grading students were able to select their own assignments, held to a mastery criterion, and were permitted to resubmit course assignments, including retaking exams. Awareness of the evaluation differences between course sections may have demoralized those students assigned to the Traditional Grading Group and manifested in reduced motivation and effort toward the course, course assignments, or in their evaluation of the course on the self-report measures. As assignment to the Contract or Traditional Grading Groups was determined by course section and the instructor did not share the details of the other grading system with the students, methodological efforts were employed to minimize this and other forms of participant reactivity. Moreover, as FYE courses maintain small class sizes, the 40 students from the present evaluation represent a small proportion of the total introductory psychology enrollment that semester (<5%). Because of the low proportion of total introductory psychology students who were enrolled in the contract grading section (<2.5%), it is unlikely, albeit still possible, that contract grading students shared the details of their grading system with the other 2.5% of students who were assigned to the Traditional Grading Group in this evaluation. Despite some inherent differences between the evaluation components of the contract grading and tradition grading systems, it should be noted that instructor, graduate student grader, class day, course materials, assignments and exams, and college year were all held constant.

These findings echo those from previous contract grading studies. For example, consistent with Kirschenbaum and Wetter Riechmann (1975), higher course evaluation scores were reported by participants in the contract grading course. And, consistent with past research on performance criteria (see Johnston & O’Neill, 1973; Semb, 1974), participants in the Contract Grading Group earned higher grades than those in the Traditional Grading Group; as such, the contract grading students appeared to have matched their performance with the higher criteria for earning a passing grade on each assignment (i.e., 85%). Contract graded students in this study also self-reported greater perceived control over their own grade, which follows a myriad of studies (Grau, 1999; Kirschenbaum & Wetter Riechmann, 1975; Polczynski & Shirland, 1977). The secondary measures (i.e., self-report measures related to the course, instructor, and student effort) were summarized with averages for brevity in the analyses. Nonetheless, some key differences found in the literature between traditional and contract grading are evident from comparing responses to the individual items described in Figure 1. For example, contract grading students indicated higher ratings for working hard for their grade, enjoying the course format, and for enhancing independent thinking. In addition, other items, not displayed in Figure 1, show a similar pattern as depicted by the aggregate score.

The outcome measures for the current comparison were limited to course grades and retention, students' self-reported perceived control over their grades, and course and teacher evaluations. There are a variety of potentially interesting dependent variables suitable for future research. One example is student effort, perhaps operationalized as time spent preparing for exams. Although there is evidence that contract grading may in fact increase motivation and effort (e.g., Bunn Hiller & Hietapelto, 2001), in the present study, some students in the Contract Grading Group may have prepared less for the first attempt on each exam knowing that they could repeat the exam a couple of days later. Although the majority of contract grading students perceived more control over their grade than their counterparts in the Traditional Grading Group, this perception was not universal. In fact, two contract grading students reported less control than any of the traditional grading students. This difference suggests important differences in individual student’s reactions to and experiences with contract grading. As such, analysis of the individual differences in the effect of contract grading and its impact on individual student performance is needed to further explore this. The benefits of contract grading among upper division students should also be explored, as it is possible that contract grading is even more beneficial among older students, given the growth that often occurs during college. And, a meta-analysis of research on contract grading, which has not yet been done, may draw much-needed attention to the effectiveness of this and other alternative grading procedures.

In conclusion, the authors encourage faculty to consider incorporating contract grading into their college courses. Based on the present study, contract grading systems may lead to higher grades, increased sense of control, and more positive experiences with the course from the student’s perspective.
Appendix

Introductory Psychology Grade Contract

Name: ___________________________  Student ID#: _______________________

Please circle the grade you would like to earn in this class: A  B  C  D  F

If you want to earn an A: Pass 4 exams (80% or higher), 3 writing selections, and 3 activity selections.
If you want to earn a B: Pass 4 exams (80% or higher), 2 writing selections, and 3 activity selections.
If you want to earn a C: Pass 4 exams (75% or higher), 1 writing selection, and 3 activity selections.
If you want to earn a D: Pass 4 exams (75% or higher), 1 writing selection, and 2 activity selections.
If you want to earn a F: Fail to meet the minimum criteria for a passing (D) grade.

Writing Selections
(choose 3 for an A, choose 2 for a B, choose 1 for a C or D; please check the box next to your choices)

☐ Complete 3 “Applying Psychology” reaction papers (1 page per reaction paper) ___________________________       ☐

☐ Complete 3 “Applying Psychology” reaction papers (1 page per reaction paper) ___________________________       ☐

☐ Complete a poster presentation of a psychology research study (done with a partner) ___________________________       ☐

☐ Design and describe a psychology study (3 page paper) ___________________________       ☐

Assignments Completed
(Only list assignments submitted on-time and receiving a passing grade)

Check When Done

Activity Selections
(choose 3 for an A, B, or C, or choose 2 for a D; please check the box next to your choices)

☐ Complete 10 in-class activities ___________________________       ☐

☐ Make a 10-min in-class presentation (done as a pair or small group) ___________________________       ☐

☐ Attend two out-of-class First Year Experience events ___________________________       ☐

☐ Complete 4 research experience credits ___________________________       ☐

Exams (Circle each exam meeting the 80% or 75% pass criterion)

E1  E2  E3  E4  E5

☐

Signed by: ________________________________  ________________________________
(Student)     (Instructor)

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research and/or authorship of this article.
References

Bunn Hiller, T., & Hietapelto, A. B. (2001). Contract grading: Encouraging commitment to the learning process through voice in the evaluation process. *Journal of Management Education, 25*, 660-684.

Fisher, R. A. (1921). On the “probable error” of a correlation coefficient deduced from a small sample. *Metron, 1*, 3-32.

Grau, H. J. (1999). “Streamlined” contract grading—A performance measuring alternative to traditional evaluation methods. *Journal of College Science Teaching, 28*, 254-258.

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*, 261-268.

Kirschenbaum, D. S., & Wetter Riechmann, S. (1975). Learning with gusto in introductory psychology. *Teaching of Psychology, 2*, 72-76.

Polcynski, J. J., & Shirland, L. E. (1977). Expectancy theory and contrast grading combined as an effective motivational force for college students. *Journal of Educational Research, 70*, 238-241.

Saville, B. K., Zinn, T. E., Neef, N. A., Van Norman, R., & Ferreri, S. J. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis, 39*, 49-61.

Semb, G. (1974). The effects of mastery criteria and assignment length on college-student test performance. *Journal of Applied Behavior Analysis, 7*, 61-69.

Twenge, J. M., & Campbell, W. K. (2001). Age and birth cohort differences in self-esteem: A cross-temporal meta-analysis. *Personality and Social Psychology Review, 5*, 321-344.

Bios

**Dana F. Lindemann** earned her doctorate in experimental psychology at Washington State University and her bachelor’s degree from the University of Alaska Anchorage. She is currently an assistant professor of psychology at Western Illinois University. She is a Board Certified Behavior Analyst and teaches in the areas of applied behavior analysis and experimental psychology.

**Colin R. Harbke** earned his doctorate in experimental psychology at Washington State University and his bachelor’s degree from the University of Alaska Anchorage. He is currently an associate professor of psychology at Western Illinois University. He teaches in the areas of statistics and research methodology. His primary research interests pertain to students’ growth in self-efficacy for math, science, and statistics.