AN INCENTIVIZED HONOR SYSTEM FOR GRADING PREPARATORY ASSIGNMENTS OF BUSINESS GAMES AND CASES: THEORY AND IMPLEMENTATION

by Precha Thavikulwat, Towson University | Rebecca Schmeller, Ashland University

pthavikulwat@towson.edu
rschmell@ashland.edu
Full Paper
Innovations Track

ABSTRACT

We present a theory and propose an empirical test of an incentivized honor system for grading preparatory assignments. Under the system, students apply an instructor-supplied rubric to self-scored submissions ex-ante, such that self-scores are delivered before or at the same time as submissions. The students' self-scores are the students' grades for their submissions unless the assignment is audited. When an assignment is audited, the instructor scores every submission and applies a veracity incentivized grading formula (VIGF) that includes a reward for veracity and a penalty for self-scores higher than instructor's scores. From expected-value analysis, we derive fixed VIGF parameters for randomly auditing half of a series of about five assignments. We derive adaptive parameters for auditing a series of about 10 assignments, using the reversed harmonic series to decide on audit frequencies. Manual and computerized implementation mechanics are discussed, and remedies for students gaming the system are explored. We caution on using ex-post self-scoring, whereby students deliver self -scores following their submissions after a duration that allows review of teaching materials and training in scoring, because the interest of instructors in saving time may conflict with the argued pedagogical advantage to students of the ex-post procedure. We end with a suggestion for testing the theory with data.

Keywords: engagement, honor system, incentivized, preparatory assignments, self-grading, self-scoring.

INTRODUCTION

When a group of students must decide on a set of decisions for a round of a group-scored business game or on an optimum course of action for the principal party of a business case, the group meeting for arriving at the group's position will be more productive if every student prepares for the meeting by studying the facts of the game or case beforehand. To ensure that every student arrives at the meeting so prepared, the instructor might ask each student to submit a pre-discussion essay written in response to a question that can apply to any round of any business game or to any business case. For the business game, the question could be, "What are the best options for the company in this round of the business game?" For the business case, the question could be, "What is the primary issue faced by the principal party of the business case?"

Preparatory assignments of this kind are instrumental to the educational activities that follow. When an assignment is preparatory, its purpose is to incentivize the student to perform an activity so that the student's experience in a subsequent activity will be more productive. Thus, the utility of pre-discussion essays arises from the students' better engagement in subsequent group discussions. The conundrum for instructors is that students may not enjoy preparatory work and that instructors may dislike grading them. But if instructors do not grade them, students would not do them.

We propose that the instructor should resolve the conundrum by requiring preparatory work that students would self-score, which the instructor would accept as the submission's grade when the instructor does not audit the assignment. When the instructor audits the assignment, however, the instructor would score submissions blind to the students' self-scores, before applying a mathematical formula to arrive at each submission's grade.

GRADING AND STUDENT ENGAGEMENT

The terms *grading, assessment*, and *evaluation* often appear interchangeably in the literature on educational measurement. We therefore make no distinction among them in our review. The distinction commonly made is between formative and summative, a distinction that goes back at least to Cronbach's (1963) argument that evaluation should be distinguished by purpose. For Lindvall and Cox (1970), purpose-based evaluations fall into three categories: individual pupil monitoring, to adapt instruction to individual needs; formative evaluation, to aid in developing an educational program; and summative evaluation, to make judgments about a program's value. Similarly, Bloom, Hastings, and Madaus (1971), arrived at three similar categories that they labeled initial evaluation, formative evaluation, and summative evaluation, each centered respectively on the student, the instructional process, and the desired outcomes.

For Yorke (2003, 2011), however, only two categories are necessary. Whereas formative evaluation “contribute to student learning through the provision of information about performance” (Yorke, 2003, p. 478), summative evaluation involves other “signals or indicators of performance” (Yorke, 2011, p. 251). Scherpereel (2010, p. 239) adds precision by identifying time, “early during the task/project” (p. 239) as characteristic of formative evaluation.

Following Yorke (2003, 2011) and Scherpereel (2010), the evaluation of preparatory assignments of which this study is concerned occurs early and is intended to contribute to student learning, so preparatory evaluation is a kind of formative evaluation. The particular purpose of a preparatory assignment is to advance learning by improving student engagement in the subsequent learning activity. Thus, the learning that is targeted by a preparatory assignment is less the learning that might result from the preparatory activity, more the learning that should arise from the targeted activity.

With respect to student engagement, the measurement of student engagement in games and other technology-mediated educational settings has been reviewed by Henrie, Halverson, and Graham (2015), who noted that student engagement has been defined as “investment or commitment, participation, and effortful involvement in learning” (p. 37); examined different approaches to its measurement; and identified their strengths and weaknesses. More recent additions to the literature on student engagement in business games include Zhang’s (2015), who found that login consistency, an objective measure of student engagement, correlated more with student peer evaluations, a subjective measure, than login frequency, another objective measure. Most recent is Schmeller’s (2019) finding that number-of-reports-opened correlate better with student peer evaluations and business game scores than login frequency and duration. These studies show that objective measures of student engagement closer to data analysis, namely login consistency and number-of-reports-opened, correlate better with student peer evaluations and game scores than measures less close. More generally, the studies show that students who are more engaged in class activities do better in business games.

Yet, considering that business games have been recommended for their apparent advantage in increasing student engagement (Wankel & Blessinger, 2012), the finding of Rogmans and Abaza (2019) that students report less engagement on average with business game activities than with traditional case discussions suggests that games themselves do not necessarily induce better engagement than traditional teaching. In the same vein, Wolfe (2016) found that engagement in an online business game was minimal for many enrolled students, and that about 27% of enrollees did not even purchase the course-required license necessary to access the game, its blogging feature, and ancillary teaching materials. Thus, business games may not be especially engaging, but those who are more engaged do perform better in business games.

THEORY

Having students perform preparatory work, such as reading instructions and studying reports for a business game or reading the narrative and studying accompanying exhibits of a business case, is one way of increasing their engagement with business games and case discussions. To incentivize students to do the preparatory work, this study would have students self-score submissions resulting from the preparatory work. As Sadler and Good (2006) have noted, the potential advantages of student self-scoring are logistical (saves instructor’s time), pedagogical (better subject-matter learning), metacognitive (better non-subject-matter learning), and affective (better attitude). From a study of middle school science classes, the researchers concluded that lower performing students tended to inflate their self-scores and that whereas self-scoring appeared to result in increased student learning, peer-scoring did not (Sadler & Good, 2006, p. 1).

To be sure, our argument is not that preparatory work should be a learning activity that would benefit from student self-scoring, although we do not deny the possibility. Rather, our argument is that preparatory work can give rise to a better learning experience in the subsequent learning activity for which the preparatory work has fortified the student. We define preparatory work broadly, as any assignment for which credit towards grades is given that precedes a learning experience, including showing up for classes on time.

Thus, agnostic to Sadler and Good’s (2006) second conclusion on the positive effect of self-scoring on student learning, our theory addresses the logistical advantage of self-scoring and is attuned to their first conclusion: that lower performing students tend to inflate their self-scores, a conclusion consistent with Boud’s (1989) review of the literature on self-scoring. Inflating one’s own score is an instance of self-enhancing bias, itself a robust psychological tendency also known as illusory superiority, the better-than-average affect, and the Lake Wobegon effect (Keillor, 1989; Zell, Strickhouser, Sedikides, and Alicke, 2020).

In the exposition that follows, we start by considering grading options available to the instructor. We then present our theory and its implementation in a software application, followed by discussing a possible area of conflict between student and instructor, and concluding with suggestions for and concerns about broader usage. For clarity, we shall refer to the requirements for the preparatory work as *assignments*, the gradable work that students deliver in response as *submissions*, the marks that students and instructors give to the submissions as *scores*, and the credits that students receive for their efforts as *grades*. Thus, assignments give rise to submissions, followed by scores and ending with grades.

GRADING OPTIONS

On the preparatory assignment preceding the group meeting, if the students share their submissions with members of their groups at the start of the meeting, their submissions will be helpful to the group's decision-making process to the extent that the submissions are substantive. To promote substantive submissions over cursory ones, the instructor could give students credit towards grades for submissions by one of the following methods:

1. Equal credit. Credit is the same for every submission.
2. Instructor credit. Credit is based on the instructor-judged quality of each submission.
3. Peer credit. Credit is based on scores supplied by the non-author members of each student's group.
4. Author self-credit. Credit is the score that each student gives to the student's own submission.
5. Incentivized honor credit. Credit is author self-score that is randomly audited and adjusted for veracity.

Equal credit incentivizes free riding, instructor credit burdens the instructor, peer credit commingles honest scores with dishonest ones, and author self-credit rewards self-enhancing bias. Incentivized honor credit (IHC) modifies author self-credit by an auditing procedure that counters bias and rewards veracity. This last option is the subject of our investigation.

INCENTIVIZED HONOR CREDIT

We define IHC as a method of grading whereby the student's self-score for the student's own submission is the instructor-accepted grade of the submission unless the instructor audits the assignment. We operationalize IHC by a procedure composed of the following sequential steps:

1. Every student sets the credit score of the student's own submission based on an instructor-supplied grading rubric.
2. To decide if the instructor should audit the assignment, the instructor applies a random process, such as flipping a coin or taking a card from a shuffled deck.
3. If the random process determines that the assignment is not to be audited, the grade of every submission is the score set by the student who authored the submission.
4. If the random process determines that the assignment is to be audited, the instructor reads every submission, scoring each based on the grading rubric supplied to the students. Then the instructor applies our veracity incentivized grading formula (VIGF) to arrive at the grade of each submission.

In specifying Steps 2 and 3, we set aside the mathematically equivalent alternative process of randomly choosing among submissions of every assignment to audit for two reasons. First, assuring students that the alternative process is truly random is more difficult. Second, the alternative process can give rise to runs wherein some students' submissions are frequently audited while other students' submissions are rarely audited that students may view as unfair.

VIGF is based on the position that students are fully informed risk-neutral rational actors who seek to maximize expected credits toward grades. The position is a reference point; some students will always fit the position better than others. Even so, students whose behavior diverges from the reference point will likely receive lower credit towards grades than other students.

The foundation of VIGF is a zero-unless-selected formula whereby assignments are randomly selected for grading with a set frequency (ν, $0 < \nu < 1$). All submissions of unselected assignments receive zero credit, and every submission of the selected assignment is credited with the instructor's score (x^*) for that submission.

An assignment where the zero-unless-selected formula is commonly applied is taking attendance. In taking attendance, the instructor effectively asks each student the question, "Are you present?" The student who answers affirmatively receives one point towards grades, so $x^* = 1$. If the instructor flips a coin to decide if attendance should be taken at a class meeting, $\nu = .5$.

$$y^0 = x^*\nu + 0(1 - \nu) = x^*\nu \quad (1)$$

Thus, for each class meeting, each attending student's expected attendance grade is $1(.5) + 0(1 - .5) = .5$. Formally, the formula for deriving the zero-unless-selected expected grade (y^0) is as formulated in Equation 1.

IHC applies to assignments whose scoring requires instructional resources, but one that students can perform reliably by themselves. For these assignments, students would self-score their submissions, each self-score (x) being the grade for each submission when the assignment is not selected for instructor grading, that is, not audited. When the assignment is audited, VIGF computation of the grade (y) includes a self-enhancement-countering multiplier (κ) when $x > x^*$, and a bonus (β, $0 \leq \beta$) that rewards veracity when $x = x^*$ and $x \neq 0$. Applying the notations $[x > x^*]$ and $[(x = x^*) \& (x \neq 0)]$ to mean that the logic-test result is 1 if true and 0 if false, the VIGF for y is as formulated in Equation 2. That is, provided $x \neq 0$, when $x = x^*$, $y = x^* + \beta$; when $x > x^*$, $y = x^* - \kappa(x - x^*)$; and when $x < x^*$, $y = x^*$. Effectively, β is reward for veracity; κ is penalty for self enhancement.

$$y = x^* - [x > x^*]\kappa(x - x^*) + [(x = x^*) \& (x \neq 0)]\beta \quad (2)$$

We shall show that the VIGF is constructed such that expected values cannot be raised by the student submitting a self-score (x) different from the instructor's score (x^*) of the audited submission. Following the reasoning of Equation 1 and incorporating Equation 2, should the student submit a self-score that is unbiased, $x = x^*$, the student's self-unbiased expected grade ($y^=$), accounting for the frequency of audits, would be as formulated in Equation 3.

$$y^= = y\nu + x^*(1 - \nu) = (x^* + \beta)\nu + x^*(1 - \nu) = x^* + \beta\nu \quad (3)$$

Should the student submit a self-diminishing score ($x < x^*$), the student's self-diminishing expected grade (y^-), accounting likewise for the frequency of audits, is as formulated in Equation 4. Should the student submit a self-enhancing score ($x > x^*$), the student's self-enhancing expected grade (y^+), accounting again for the frequency of audits, is as formulated in Equation 5.

$$y^- = y\nu + x(1 - \nu) = x^*\nu + x(1 - \nu) \quad (4)$$

$$y^+ = y\nu + x(1 - \nu) = x - (x - x^*)(1 + \kappa)\nu \quad (5)$$

To simplify Equation 5, we set κ relative to ν as given in Equation 6. Simplified, Equation 5 becomes Equation 7. Comparing Equation 3 with Equation 4, $y^= > y^-$ because $x^* > [x^*\nu + x(1 - \nu)]$ since $x < x^*$. Comparing Equation 3 with Equation 7, $y^= \geq y^+$ because $\beta\nu \geq 0$. Thus, the VIGF rewards veracity over self-diminishment even when $\beta\nu$ is zero and rewards veracity over self-enhancement to the extent that $\beta\nu$ exceeds 0.

$$\kappa = \frac{1}{\nu} - 1 \quad (6)$$

$$y^+ = x^* \quad (7)$$

As to the boundary conditions of $\nu = 0$, when the instructor never audits submissions, and $\nu = 1$, when the instructor always does, asking students to self-score their submissions would not be useful. For these conditions, the equal-credit option suits $\nu = 0$; the instructor-credit option suits $\nu = 1$.

SETTING PARAMETRIC VALUES

We turn next to the question of appropriate values for β relative to ν and κ. Although the incentive for veracity of the VIGF rises with increases in both β and ν, increasing β reduces the incentive for better performance relative to veracity and increasing ν increases the instructor's grading burden, both undesirable consequences.

To aid in setting β relative to ν, we link these two parameters to the number of assignments (n) by considering that the parameters should be set such that the student who is audit-penalized on the first assignment for a one-point self-enhancing bias, $x - x^* = +1$, should expect to recover the penalty by consistently being veracious for the remaining assignments. The penalty is κ and the expected recovery for veracity on each subsequent assignment is $\beta\nu$. There being $n - 1$ assignment after the first one, the penalty-to-recovery condition, incorporating Equation 6, is expressed by Equation 8.

$$(n - 1)\beta\nu = \kappa = \frac{1}{\nu} - 1 \quad (8)$$

For ν in terms of n and β, we rewrite Equation 8 in quadratic form without κ, resulting in Equation 9. Applying the quadratic formula gives rise to Equation 10.

$$(n - 1)\beta\nu^2 + \nu - 1 = 0 \quad (9)$$

$$\nu = \frac{-1 + \sqrt{1 - 4(n - 1)\beta}}{2(n - 1)\beta} \quad (10)$$

Table 1 gives the grading frequencies resulting from applying Equation 10 and Figure 1 graphs the results. The drop in grading frequency as β rises decelerates, so much so that an arguably reasonable set of settings for $n \approx 5$ should be $\beta = 0.5$ and $\nu = .50$, marked with a circle in Figure 1, implying (Equation 6) that $\kappa = 1$. For this set of settings, the instructor flips a coin to decide if the submissions of each assignment, from the first assignment to the last assignment, should be audited. If the decision is to audit, the instructor applies the fitted VIGF, Equation 11, to every submission of the assignment.

$$y = x^* - [x > x^*]1(x - x^*) + [(x = x^*) \,\&\, (x \neq 0)]0.5 \tag{11}$$

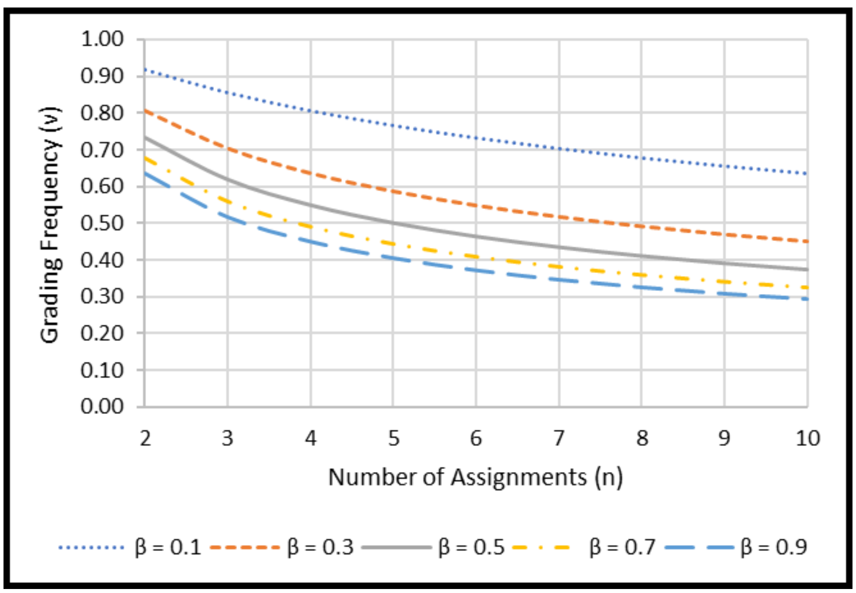
TABLE 1

Grading Frequency (ν) for Expected Complete Recovery of Audit Penalty for a One-Point Self-Enhancing Bias

Number of Assignments (*n*)	Bonus (β)				
	0.1	**0.3**	**0.5**	**0.7**	**0.9**
2	0.92	0.81	0.73	0.68	0.64
3	0.85	0.70	0.62	0.56	0.52
4	0.81	0.64	0.55	0.49	0.45
5	0.77	0.59	**0.50**	0.45	0.41
6	0.73	0.55	0.46	0.41	0.37
7	0.70	0.52	0.43	0.38	0.35
8	0.68	0.49	0.41	0.36	0.33
9	0.66	0.47	0.39	0.34	0.31
10	0.64	0.45	0.37	0.33	0.30

FIGURE 1

Grading Frequency (*ν*) for Expected Complete Recovery of Audit-Penalty Loss



For a better fit over the range $2 \leq n \leq 10$, ν could be changed with every assignment, rising from the first to the last item of the series, provided the mean ν approximates the values of Table 1. The ν could be the items of the harmonic series (1 + ½ + ☐ + ¼ …) reversed, whose means from the first to *n*th item is shown in the last column of Table 2. Table 3 shows the absolute differences between the harmonic-series values of Table 2 and the grading frequencies of Table 1, where the smallest absolute difference of each *n* is shaded. Thus, Table 3 shows that for the best fit when applying the reversed harmonic series from $2 \leq n \leq 4$, β should be set to 0.5; from $5 \leq n \leq 7$, β should be set to 0.7; and from $8 \leq n \leq 10$, β should be set to 0.9.

TABLE 2

Means of Harmonic Series From the First to *n*th Item of the Series

Item Number (n)	Harmonic-Series Value of nth Item ($1 / n$)	Mean of Harmonic-Series Values From First Item
1	1.00	1.00
2	0.50	0.75
3	0.33	0.61
4	0.25	0.52
5	0.20	0.46
6	0.17	0.41
7	0.14	0.37
8	0.13	0.34
9	0.11	0.31
10	0.10	0.29

TABLE 3

Absolute Differences Between Harmonic-Series Values and Grading Frequencies

Number of Assignments (n)	Bonus (β)				
	0.1	0.3	0.5	0.7	0.9
2	0.166	0.055	0.018	0.072	0.114
3	0.243	0.092	0.007	0.051	0.093
4	0.285	0.115	0.028	0.029	0.070
5	0.309	0.130	0.043	0.012	0.051
6	0.324	0.140	0.055	0.002	0.035
7	0.333	0.147	0.064	0.013	0.023
8	0.338	0.152	0.071	0.021	0.013
9	0.342	0.156	0.076	0.028	0.005
10	0.343	0.158	0.080	0.034	0.002

Operationally, applying the reversed harmonic series, the instructor acquires a pack of cards of a number equal to the number of assignments, designating one of the cards as the audit-decision card. Then for the first assignment, the instructor randomly selects one card from the set, auditing the first assignment if and only if the selected card is the audit-decision card. For the next assignment, the instructor removes one non-audit-decision card from the pack and repeats the random selection, continuing in this fashion for the remaining assignments.

For example, for 10 assignments, the instructor would have 10 cards for the first assignment, one of which would be the audit-decision card. For the second assignment, 9 cards; for the third, 8 cards; and so forth. Thus, the likelihood of selecting the audit-decision card for the first assignment is 1 in 10, so $\nu = 1 / 10 = 0.10$; for the second assignment, 1 in 9, so $\nu = 1 / 9 = 0.11$. These values are shown in the second column of Table 2. By using this reversed harmonic-series procedure, the mean ν of the 10 assignments will be 0.29, the value in the cell of the last row of the last column of Table 2.

Although the instructor could compute and apply a different κ (Equation 6) and β (Equation 8) for each ν, a simpler procedure is to fix κ and β to the mean ν of 0.29, giving $\kappa = 1 / 0.29 - 1 = 2.448 \approx 2.5$ and $\beta = 2.448 / [(10 - 1) \times 0.29] = 0.938 \approx 1.0$ for all assignments. The simpler procedure, however, affects the incentive for veracity when ν is not at its mean of 0.29. To assess the strength of the effect, we examine how the expected advantage of bias changes with each assignment for these settings of this procedure.

For self-diminishing bias ($x < x^*$), the expected advantage ($y^- - y^=$) is derived by subtracting Equation 3 from Equation 4, arriving at Equation 12. The right-hand side of Equation 12 is never positive, therefore self-diminishing bias is never advantageous. For self-enhancing bias ($x > x^*$), the expected advantage ($y^+ - y^=$) is derived by subtracting Equation 3 from Equation 5, arriving at Equation 13.

$$y^- - y^= = (x - x^*)(1 - \nu) - \beta\nu \quad (12)$$

$$y^+ - y^= = (x - x^*)[1 - (1 + \kappa)\nu] - \beta\nu \quad (13)$$

To relate Equations 12 and 13 to assignments, the reversed harmonic-series method of setting ν implies that ν is related to the number of unprocessed assignments (n'), the total number of assignments (n), and the to-be-processed assignment number (i), as given in Equation 14.

$$\nu = \frac{1}{n'} = \frac{1}{n+1-i} \quad (14)$$

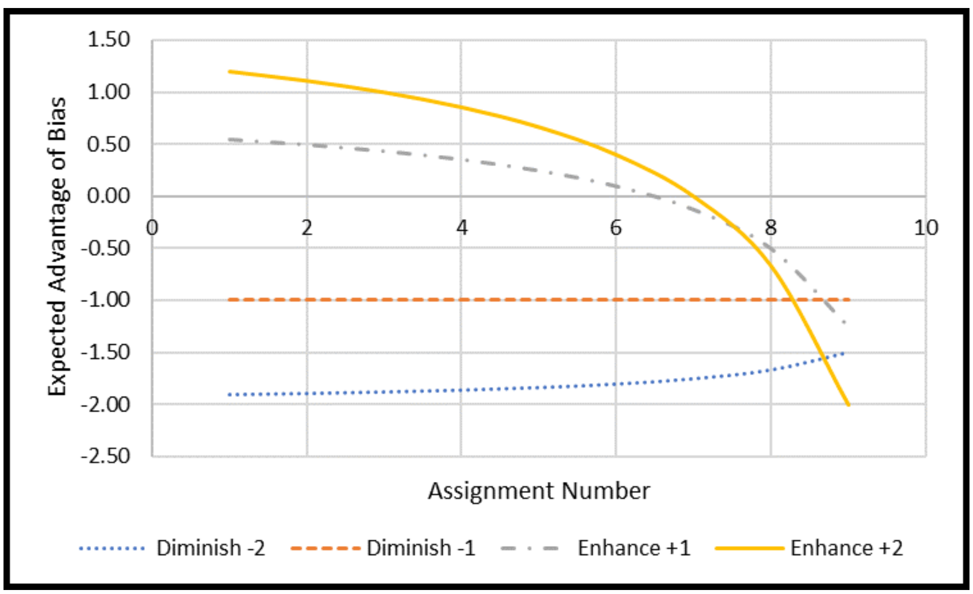
Thus, for the first ($i = 1$) of ten ($n = 10$) total assignments, $\nu = 1 / (10 + 1 - 1) = 1 / 10$; for the second, $\nu = 1 / (10 + 1 - 2) = 1 / 9$; and so forth until the tenth assignment, when $\nu = 1 / (10 + 1 - 10) = 1$. Applying Equation 14 to Equations 12 and 13 yields Equations 15 and 16, respectively.

$$y^- - y^= = (x - x^*) - (x - x^*)\nu - \beta\nu = (x - x^*) - \frac{(x-x^*)-\beta}{n+1-i} \quad (15)$$

$$y^+ - y^= = (x - x^*) - (x - x^*)(1 + \kappa)\nu - \beta\nu = (x - x^*) - \frac{(x-x^*)(1+\kappa)-\beta}{n+1-i} \quad (16)$$

Figure 2 is a graph that combines Equations 15 and 16 for four bias-point levels ($x - x^*$), from most self-diminishing (−2 points) to most self-enhancing (+2 points). As the graph shows, self-diminishing is never advantageous; self-enhancing is advantageous up to the sixth assignment.

FIGURE 2
Expected Advantage of Bias With κ = 2.5 and β = 1 by Assignment Number for Bias Points From Most Self-Diminishing (Diminish -2) to Most Self-Enhancing (Enhance +2)



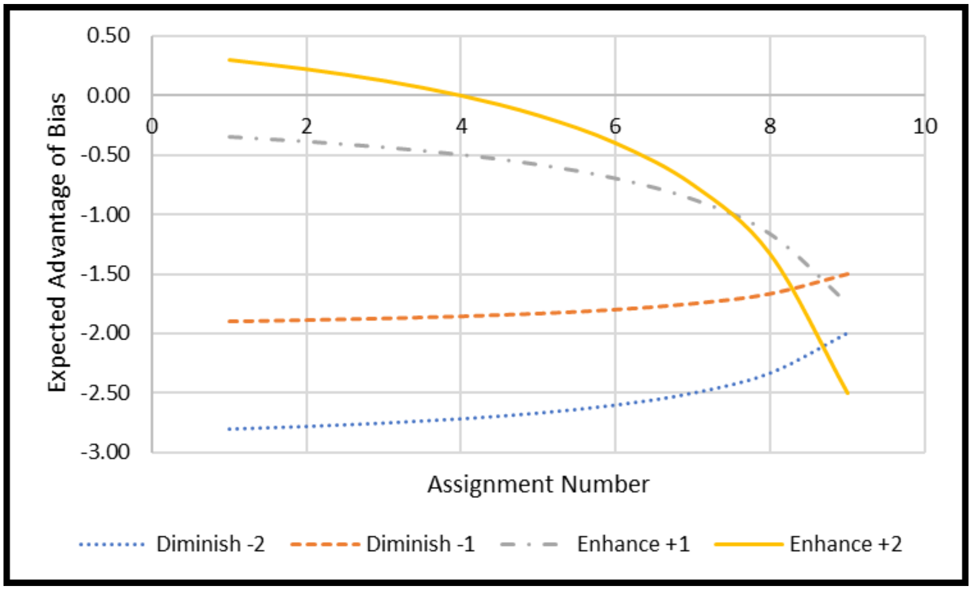
To reduce the expected advantage of self-enhancing bias, the slightly more complicated procedure of adapting β to the number of remaining assignments (n') could be used. Following Equation 8 and incorporating Equation 14, the adaptive β would be as given in Equation 17.

$$\beta = \frac{\frac{1}{\nu}-1}{\nu(n'-1)} = \frac{n'(n'-1)}{n'-1} = n' = n + 1 - i \quad (17)$$

Thus, as the assignments proceed from $i = 1$ to $i = 2$ and so forth, until $i = 9$; ν rises from $\nu = 1 / 10 = .10$ to $\nu = 1 / 9 = .11$ and so forth, until $\nu = 1 / 2 = 0.50$; and β falls from $\beta = 10 + 1 - 1 = 10$ to $\beta = 9$ and so forth, until $\beta = 2$. For the last assignment, when $i = 10$, the boundary condition of $\nu = 1$ applies, which call for the instructor-credit option, so $\beta = \kappa = 0$. Figure 3 applies the adaptive β to the data graphed in Figure 2. As Figure 3 shows, with adaptive β, self-enhancing bias is never advantageous at the +1-point level and only advantageous for the first three assignments at the +2-point level.

FIGURE 3

Expected Advantage of Bias With κ = 2.5 and Adaptive β by Assignment Number for Bias Points From Most Self-Diminishing (Diminish -2) to Most Self-Enhancing (Enhance +2)



To summarize, for $n \approx 5$, fixing the parameters to $\nu = .50$, $\kappa = 1$, and $\beta = 0.5$ should be adequate. For $n \approx 10$, applying the reversed harmonic series to ν, fixing κ to the mean ν, and adapting β so that it falls by one point after each assignment from the first one at $\beta = n$ to the second to last one at $\beta = 2$ enables a lower ν without a higher κ, and no positive expected advantage of self-enhancing bias after the first few assignments. In short, a lower instructor burden without a higher penalty is gotten for a small window of advantaging self-enhancement. Finally, because random processes sometimes give rise to runs, the instructor should delimit the sum of all graded submissions to reasonable bounds, such as not below zero and not above $n(\hat{x} + \beta\bar{\nu})$, where $\hat{x}$ is the maximum x and $\bar{\nu}$ is the mean ν of the n assignments.

IMPLEMENTATION MECHANICS

IHC can be implemented on an electronic spreadsheet. When an assignment is audited, the instructor would enter instructor scores in one column, student self-scores in another column, and apply the VIGF to compute the grade.

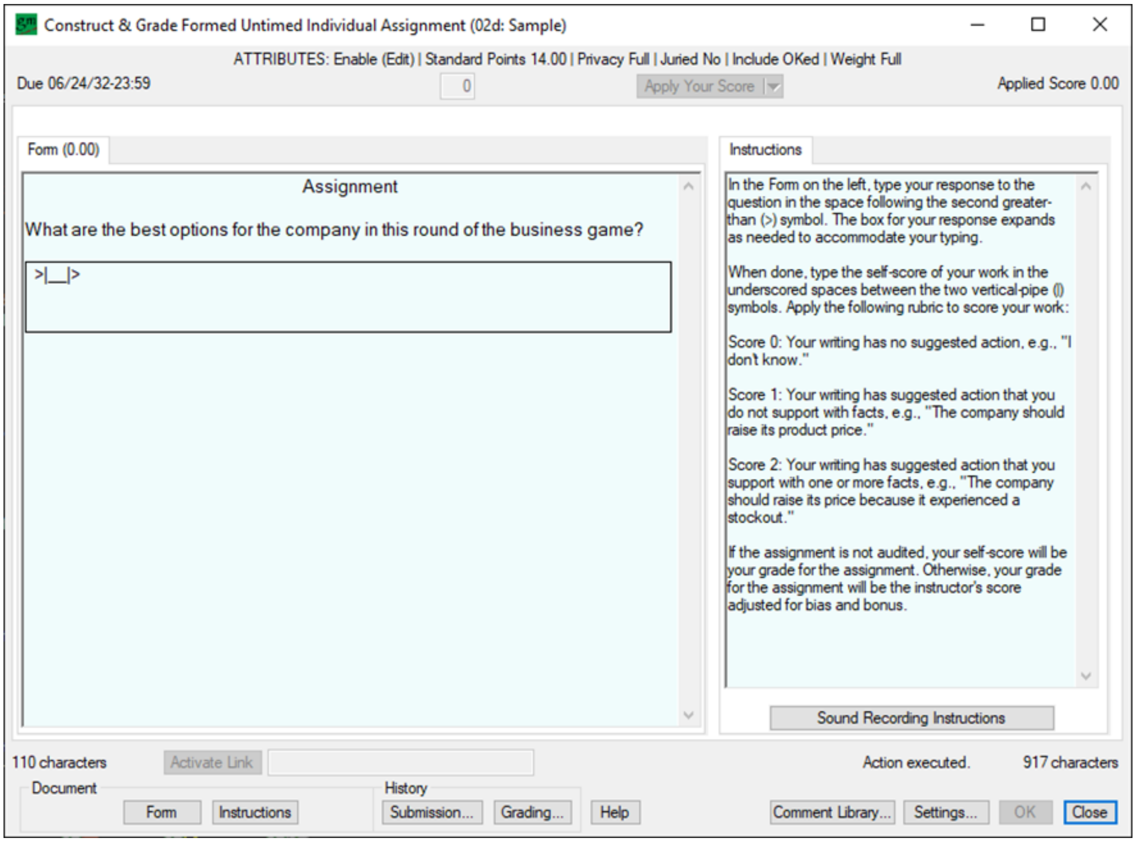
Even so, an IHC-cognizant application that administers assignments would reduce the likelihood of error and save time. We developed an internet application, GroupMaker, for the purpose because we could not adapt common learning management systems (LMS), such as Blackboard, Canvas, and Moodle, to our requirements.

GroupMaker is a Windows application that can access data from either a distant server or a local file. Thus, when teaching is active, the data resides on a distant server. When teaching has ended, the data can be removed from the server and saved in a local file that the application accesses directly. The same is not viable with web services, for downloaded files cannot be accessed without server software. Consequently, our application is more responsive, because it bypasses the web service requirement of common LMSs; and more supportive of the instructor’s intellectual property rights, because it allows the instructor to own the data with access when teaching has ended.

A unique feature of our application is that instructors transmit instructions on assignments to the application through Assignment Markup Language™ (AML), a text-based language we developed that enables the instructor to embed commands in the document file of any word processor whose contents can be copied and pasted through Microsoft Windows’ clipboard or saved in a text-format file such as plain text, rich text, and HTML. Thus, the text containing the open-ended question we posed for a business game could be part of a form constructed on Microsoft Word, advantageous because Microsoft Word has tables and other graphical features that can be included in the form. Figure 4 shows a screenshot of the assignment-construction panel of the application containing a Word-created sample assignment form clipboard-pasted into the left area of the panel.

FIGURE 4

Sample Screen Shot for Constructing and Grading Assignment



The AML command of the sample assignment is >|__|>, meaning that this is a self-scored, open-ended item (Command 4a of the Appendix). The student responds to the question by typing text after the ending greater-than (>) character. The student enters the student's numeric self-score in the underscored spaces between the vertical-pipe (|) characters.

When the instructor audits the assignment, the instructor is shown the student's response but not the student's self-score. After entering the instructor score, the instructor clicks a button to apply the VIGF to grade the submission.

GAMING THE SYSTEM

Should a student consider engagement to be undesirable, the student may attempt to secure a higher score at a lower cost by interpreting IHC's instructor-supplied grading rubric as a rule that binds rather than a principle that guides. For example, the grading rubric of the Instructions tab shown in Figure 4 advises that in answering the question "What are the best options for the company in this round of the business game?", the answer should receive the score of 0 if it does not suggest an action; 1, if it suggests action that the student does not support with facts; and 2, if it suggests action that the student supports with one or more facts.

Interpreting the rubric as rule, the student's reply could be that the best option is to repeat unchanged the decisions of the previous round (an action) because the facts have not changed sufficiently to make a difference (supported with facts), for a rule-based score of 2. The student could repeat the reply in every round and insist on the score of 2 every time, for in truth repeating previous decisions is an action and supporting it by asserting that facts have not changed sufficiently is supporting the action with facts. Thus, the student games the system by repeating a cursory answer.

The instructor could discourage cursory answers by impressing on students that rubrics are not rules, but principles to be applied with judgment. The instructor also could use the first assignment of the series for training in self-scoring by auditing it, ignoring the students' self-scores, and applying instructor-credit grading. For the remaining assignments of the series, the instructor would either switch to fixed-parameter IHC or continue with unreversed harmonic-series settings by applying $\nu = ½$

to the second assignment, $\nu = ⅓$ to the third assignment, and so forth, with or without adapting β from assignment to assignment. In fact, for students whose desire for engagement rises from assignment to assignment, the unreversed harmonic-series procedure may fit better than the reversed harmonic-series procedure, because auditing their submissions becomes less useful with increased engagement.

Still, the student who games the system may be engaged but struggling. With patience and foresight, the instructor might channel that engagement to a better learning experience.

CONCLUSION

Although IHC has been presented in the context of preparatory assignments for business games and cases, IHC should be useful for any set of equally weighted assignments in which the instructional resources of scoring each assignment are high relative to the learning that the instructor's scoring may induce in the student. Thus, IHC should be useful also for scoring problem sets assigned as homework in many fields. Supporting these broader applications are Sadler and Good's (2006) findings on the positive learning effects of self-scoring.

In Sadler and Good's (2006) self-scoring study, as in Laufer's (1986) report of self-scoring in an accounting course, the students self-scored their work ex-post their submission of the assignment, allowing time between assignment submission and self-scoring for students to review teaching materials and for the instructor to train the students on how to score their submissions. In self-scoring preparatory assignments, however, review and training should not be necessary, because the preparatory assignment, which precedes the learning assignment, is not an assignment from which much learning is expected. Thus, for preparatory assignments, students can be asked to submit their self-scores ex-ante, that is, before or simultaneous with their preparatory-assignment submissions, saving time and simplifying administration.

Although ex-post self-scoring (EPSS) may have pedagogical advantages over ex-ante self-scoring (EASS), promoting EPSS for its pedagogical advantage may be unconvincing to those attuned to conflicts of interest, for EPSS also saves instructor's time, which benefits the instructor but not necessarily the students every time. To be safe, EPSS should be reserved for assignments where the pedagogical advantages of self-scoring are beyond question.

In contrast, EASS of preparatory assignments is an application with no conflict of interest between students and instructor, for students cannot learn without engagement. Preparatory assignments enhance engagement. EASS reduces the burden that preparatory assignments impose on the instructor, a burden that instructors might otherwise avoid by not assigning preparatory assignments, to the detriment of the students' engagement but not necessarily a loss to their satisfaction. For there is some truth to the observation that education is a business where the customers are often happy to get less than what they pay for.

The follow-up question to ask is "Does it work?" The answer depends on criterion and data. A reasonable criterion would be that the observed self-enhancing bias should not rise from the first to the last submission of the set. If the observed self-enhancing bias should fall, then the data would substantiate the effectiveness of the κ parameter of VIGF (Equation 2) in countering self-enhancement. But if the observed self-enhancing bias should not change across assignments, then κ must be ineffective, in which case setting it to zero would remove an apparently unnecessary parameter. Answering the question with data, however, is a project for another day.

REFERENCES

Bloom, B.S., Hastings, J.T. and Madaus, G.F. (1971). Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill.

Boud, D. (1989). The role of self-assessment in student grading. Assessment and Evaluation in Higher Education, 14, 20–30.

Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record, 64*, 676-683. DOI: 10.1177/016146816306400802.

GroupMaker. [Developed by P. Thavikulwat] Towson, MD (601 Worcester Road, Towson, MD, USA). Available from http://wp.towson.edu/precha/

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education, 90*, 36-53.

Keillor, G. (1989). Faith: Stories from the collection more news from Lake Wobegon [CD sound recording]. St. Paul, MN: High Bridge.

Laufer, D. M. (1986). Self-Grading of examinations by students: A viable alternative to traditional procedures. *Developments in Business Simulation & Experiential Exercises, 13*, 47-51. Available from the *Bernie Keys Library*, https://absel.org.

Lindvall, C. M., & Cox, R. C. (1970). *Evaluation as a tool in curriculum development: The IPI evaluation program*. American Educational Research Association Monograph Series on Curriculum Evaluation. Chicago: Rand McNally.

Rogmans, T., & Abaza, W. (2019). The impact of international business strategy simulations games on student engagement. *Simulation & Gaming, 50*, 393-407. DOI:10.1177/1046878119848138.

Sadler, P. M., & Good, E. (2006). The impact of self-and peer -grading on student learning. *Educational Assessment, 11(1)*, 1-31.

Scherpereel, C. M. (2010). Peer to peer: A new tool for student peer evaluation. *Developments in Business Simulation & Experiential Learning, 37*, 238-250. Available from the *Bernie Keys Library*, https://absel.org.

Schmeller, R. (2019). In strategy simulations, data analysis matters most (more than number of log ins and more than time spent logged in). *Simulation & Gaming, 50*(1), 62-75. DOI: 0.1177/1046878118821402.

Wankel, C., & Blessinger, P. (Eds.) (2012). *Increasing student engagement and retention using immersive interfaces: Virtual worlds, gaming, and simulation*. In collaboration with J. Stanaityte and N. Washington. Created in partnership with Higher Education, Teaching and Learning. Cutting-Edge Technologies in Higher Education, V. 6C. UK: Emerald.

Wolfe, J. (2016). Assuring business school learning with games. *Simulation & Gaming, 47*, 206-227. DOI: 10.1177/1046878116632872.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*, 477-501. DOI: 10.1023/A:1023967026413.

Yorke, M. (2011). Summative assessment: Dealing with the 'measurement fallacy'. *Studies in Higher Education, 36*, 251-273. DOI: 10.1080/03075070903545082.

Zell, E., Strickhouser, J. E., Sedikides, C., Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin, 146*, 118-149. Available http://dx.doi.org/10.1037/bul0000218.

Zhang, M. J. (2015). Using login data to monitor student involvement in a business simulation game. *The International Journal of Management Education, 13*(2), 154-162. DOI:10.1016/j. ijme.2015.02.003.

APPENDIX

Commands of Assignment Markup Language

<table>
<tr><th>Command*</th><th>Description</th></tr>
<tr><td>--|…|--</td><td>1. Comment. The in-between text (…) is neither read by software nor shown to students before the assignment's graded date.</td></tr>
<tr><td>++|…|++</td><td>2. Screen. The in-between text (…) is read by software but not shown to students after the assignment's graded date.</td></tr>
<tr><td>></td><td>3. Editable line. If appearing as the first character of a line or cell of a table, the student can type freely after the greater-than character (>).</td></tr>
<tr><td>|</td><td>3a. Transformable editable line. The student deleting the vertical-pipe character (|) causes the editable line to either be appended to the immediately preceding editable line or become un-editable if not so preceded.</td></tr>
<tr><td>>||></td><td>4. Open-ended item. The student can type freely after the ending greater-than character. The instructor will grade the submitted item. The student's response can span multiple lines.</td></tr>
<tr><td>>|__|></td><td>4a. Self-scored, open-ended item. The student can enter a numerical self-score between the two vertical-pipe characters. The number of characters typed cannot exceed the number of placeholder underscore characters (__).</td></tr>
<tr><td>>|__|x|></td><td>4b. Limit-specified, self-scored, open-ended item. The student cannot enter a self-score exceeding x points.</td></tr>
<tr><td>>|__|x|y|></td><td>4c. Multiplier-specified, limit-specified, self-scored, open-ended item. When instructor graded and the student's self-score exceeds the instructor score, the software will subtract from the instructor score the excess of the student's self-score over the instructor score multiplied by the self-enhancement-countering multiplier, y. Thus, when Self Score > Instructor Score, Net Score = Instructor Score – [y × (Self Score – Instructor Score]), otherwise Net Score = Instructor Score.</td></tr>
<tr><td>>|__|x|y|z|></td><td>4d. Bonus-specified, multiplier-specified, limit-specified, self-scored, open-ended item. The bonus of z points will be added to the item grade when the self-score matches the instructor score. Thus, when Self Score = Instructor Score, Grade = Net Score + Bonus, otherwise Grade = Net Score.</td></tr>
<tr><td>>|__|x|y|z|e|></td><td>4e. Tolerance-specified, bonus-specified, multiplier-specified, limit-specified, self-scored, open-ended item. The bonus of z points will be added to the item grade when the self-score matches the instructor score within the tolerance of $\pm e$ points. Thus, when the | Self Score – Instructor Score | $\leq e$, Grade = Net Score + Bonus, otherwise Grade = Net Score.</td></tr>
<tr><td>>|__ab__|<</td><td>5. Closed-ended item. The submitted item will be credited with the default point value if the characters typed conform, case insensitive, to the ab key.</td></tr>
<tr><td>>|__ab__|x|<</td><td>5a. Points-specified, closed-ended item. The submitted key-correct item will be credited with x points.</td></tr>
<tr><td>>|__ab__|x|abc|<</td><td>5b. Character-specified, points-specified, closed-ended item. The student can type only the abc characters.</td></tr>
<tr><td>>|__ab__|x|abc|aa|ac|<</td><td>5c. Multiple-correct, character-specified, points-specified, closed-ended item. The submitted item also will be graded as key correct if the characters typed conform to alternative keys aa, ac, and others that may be specified between the next to last and last vertical-pipe characters.</td></tr>
<tr><td>|>|x…</td><td>6. Extra credit points. The submission will be extra credited with x points, explanation (…) allowed.</td></tr>
<tr><td colspan="2">*The letters a, b, and c represent any case of any string of characters other than the vertical-pipe character; the letters x, y, z, and e represent numbers. All commands must be contained within a single line.</td></tr>
</table>