

Quality or Quantity: Completion Rewards and Formative Assessments in Flipped Instruction Classes

Kurt Schmitz

Georgia State University

Received 10 February 2019; Accepted 3 September 2019

Flipped instruction shifts the burden for engaging course content to the students. Moving these activities outside the classroom creates motivational challenges. This study investigates the role of formative assessments and completion rewards. Definitions are provided for flipped instruction and formative assessments. A classification of reward scores is offered to guide data collection in two field experiments. The study seeks to provide an empirical basis to guide use of completion rewards for flipped instruction classes. This study finds that completion rewards can increase *quantity* of formative assessment engagement. However, this change in quantity does not improve exam scores. The data suggests completion rewards may undermine the *quality* of engagement.

INTRODUCTION

Flipped instruction approaches are increasingly being applied to University curriculum (Akçayır & Akçayır 2018). Flipping the classroom moves events that have traditionally taken place inside the classroom to outside the classroom and vice versa (Lage et al. 2000). Content dissemination moves away from face-to-face hours and into online delivery outside of class, while face-to-face class time is used for practice and application (Hill 2012).

A key factor influencing learning outcomes with flipped instruction is the amount of time students spend with the material outside of class (Lim & Morris 2009). Flipping the classroom assumes that students will take control of their learning in terms of pace of study, mastery of content, and coming to class prepared (Davies et al. 2013). Motivation plays a key role initiating and sustaining self-directed learning (Garrison 1997) and is positively associated with exam performance (Janssen & O'Brien 2014). In addition, lack of motivation is the major reason students drop out of online courses (Kim 2004). These self-directed learning insights have proven robust not only for MOOCs, but also online classes at community colleges and universities (Lee & Choi 2011; Levy 2007).

This study examines two techniques aimed at student motivation for the outside-the-classroom portion of flipped instruction. The first is the use of Formative Assessments. The cognitive development literature reports that formative assessment feedback motivates deep learning (Higgins et al. 2002). Beyond motivation, formative assessments are often used by students to adjust their study process leading to improved learning outcomes (Cauley & McMillan 2010). The second tactic is the use of completion rewards tied to a desired behavior. In this study rewards are scores in an online gradebook. Students consistently report that grades are an overarching concern, such that all other goals are secondary (Pressley et al. 1998). In addition, knowing their progress and grades in a class may provide students a sense of satisfaction that motivates their effort (Docan 2006).

There is very little rigorously designed research on flipped classroom approaches (Abeysekera & Dawson 2015). To partially address this gap, this study seeks to examine the effect of formative assessment completion scores for flipped instruction learning objectives. The following specific study questions are examined:

Is Formative Assessment engagement associated with improved learning outcomes for outside the classroom flipped instruction content?

Do completion scores increase Formative Assessment engagement?

Do completion scores increase test/exam performance?

This paper proceeds as follows: Section two of this paper provides literature review on formative and summative assessment and establishes definitions for flipped instruction and completion scores suitable for measurement and study. Section three provides the hypothesis which guide this study. Section four describes the design of a field study involving both medium size and large classrooms. Section five presents the findings followed by discussion.

THEORETICAL BACKGROUND

Flipped Instruction

In a traditional classroom a teacher directs instruction of material using lecture, demonstration, and sometimes discussion. In this model students perform a mostly passive role of watching, listening, taking notes, and sometimes speaking up with comments and questions. Students may do some follow-along and practice in the classroom, but homework is where most practice and skill refinement take place.

Flipping the classroom is an instructional technique inspired by the idea that experiential learning is highly effective for many learning objectives. In a flipped classroom, the limited time available for face-to-face interaction is allocated to coaching and guiding the application of skills to develop and mature competences. This technique often employs group-based interactive learning activities inside the classroom, with directed computer-based individual instruction outside the classroom (Bishop & Verleger 2013).

Flipped classroom instruction resides near the middle of an instructional modality continuum with traditional lecture on one end, and fully online at the other (Hill 2012). Variations of the flipped classroom are sometimes called “blended” or “hybrid”.

Where a hybrid model may allow a mixture of traditional in-class lecture with in-class practice, the fully flipped classroom does not use face-to-face time to deliver material. Terms, concepts, ideas and examples are available to students in materials they must explore on their own outside of class. Outside of class activities are often guided with schedules and a scaffolded path that leads students through the material as a planned journey. While the journey may be organized and planned, the burden is still on the student to take control of their learning (Davies et al. 2013).

A challenge for students in a flipped environment is their low level of self-regulation (Iwamoto et al. 2017). Self-directed learning is essential for students in a flipped classroom. While this may seem a high-risk proposition, today's students are surprisingly well equipped. Students already know how to navigate information portals such as Wikipedia and YouTube. They regularly employ internet search to find definitions and examples. With a few well-placed learning management system links and signposts, students can locate and engage terms, definitions, concepts and examples. A key success factor influencing learning outcomes for flipped instruction classes is average study time (Lim & Morris 2009), which is driven by motivation. Lack of motivation is the major reason students drop out of online courses (Kim 2004), an insight that should inform efforts to improve outside-the-classroom flipped instruction assignments. Motivation plays a significant role in not only the initiation of self-directed learning effort, but also the maintenance of that effort and the achievement of cognitive goals (Garrison 1997).

Summative Assessments and Formative Assessments

Providing students feedback during the learning process is one technique linked to learning motivations. Assessments have long been used by instructors to expose the gap between a desired target and a student's actual knowledge, skills and abilities (KSAs). All forms of assessment attempt to make a judgement involving a target standard, goal or criteria (Taras 2005). Assessments generate information about a KSA gap that can trigger feedback. Two forms of assessment have emerged in the literature.

A *Summative Assessment* is an evaluation conducted *after* completion of an instructional activity. Summative assessments involve an external evaluator for credibility and are administered for the benefit of an external decision maker (Scriven 1981), such as the instructor. These assessments involve evidence of student achievement (Black & Wiliam 2009). Example summative assessments include pop quizzes, end of activity scored quizzes, tests and exams, as well as graded essays and problems.

A *Formative Assessment* is an evaluation conducted *during* the development and improvement of KSAs (Scriven 1981). Formative assessments generate feedback to the learning that reveals a gap between actual KSA and the target or "standard" (Taras 2005). A best practice is for the feedback to include an indication of how to improve and eventually reach the standard. Example Formative assessments include prototypes, comment-only marking, peer assessment, self-assessment, practice questions and formative use of summative tests. A practice question is equivalent to golfing without a scorecard. The formative use of a summative test is equivalent to golfing then discarding your scorecard at the end of the round. Formative assessments minimize the anxiety of being judged by an external evaluator and fosters an environment that allows students to focus on the learning objectives. Formative

assessments allow the student to ask: "How *am* I doing?" Summative assessments reform the question to: "How *did* I do?" (Rolfe & McPherson 1995). A simple analogy is attributed to Bob Stake "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative." (Scriven 1981).

Formative assessments provide a valuable tool to guide outside-the-classroom learning in a flipped classroom. Formative assessment feedback facilitates intrinsic motivation, curiosity for a topic and "deep learning" (Higgins et al. 2002), which directly addresses the need for motivation during self-directed learning. In addition, Formative assessments are used by students to adjust their learning tactics (Cauley & Mcmillan 2010), which may facilitate learning objective achievement for outside-the-classroom material.

Participation scores and Completion Scores

The ability of formative assessments to motivate deep learning, highlights the importance of students engaging these assessments as part of their out-of-classroom time allocation. Most instructors learn quickly that assigning points to an activity will draw students' attention to that activity. In their review of metacognition literature, Pressley et al (1998) noted that "Obtaining good grades is an overarching concern... Students made it quite clear that all other goals were secondary." Many recent studies concur that knowing their grades and progress in class gives students a sense of satisfaction and motivation (Docan 2006).

Learning Management Systems such as Blackboard and Desire2Learn are common at many institutions. These systems provide online gradebooks that allow students to monitor, often in real time, the cumulative grade impact of each task, activity, quiz and test. More than a casual suggestion that behavior matters, a score in the gradebook provides hard evidence to the student that the requested activity does impact their final grade.

A score in the gradebook for formative assessments is a delicate matter. Measuring student performance shifts an assessment from formative to summative, and thereby fundamentally changes the student's relationship with the task. An alternative way to leverage the motivational role of grades and scores is the use of participation scores in an open gradebook. Three forms of participation scores are defined here. Each form manipulates the conditions necessary to obtain the reward and thereby makes a meaningful difference in the degree of student engagement induced. In addition to altering the behavioral profile needed by the student to achieve the reward, each is expected to influence student motivations differently.

Participation scores take a variety of forms in outside-the-classroom activities. Some instructors measure the quantity of participation in online message boards. Others collect peer review assessments for team projects. When considering formative assessments, a score in the gradebook for "doing" a task regardless of quality and completeness is a *participation reward*. This is labeled a type-I participation reward¹ with a true/false classification. A subtle aspect of type-I rewards is that there need not be a threshold for any specific quantity of participation. Returning to the golf course analogy, a golfer who tees off on the first hole could earn a participation reward even if they lose their ball and fail to complete any of the 18 holes in a standard round of golf. This reward serves a similar role to an attendance score. Students who abandon engagement after registering their attendance still qualify for the full type-I participation reward. Best

practice guidelines for educational settings discourage this type reward, noting that rewards should not be given “for mere participation.” (Akin-Little et al. 2004).

Establishing a quantity of participation threshold changes the type-1 participation reward into a type-2 *quantity reward*. Students receive a score reflecting the quantity of the requested activity attempted. In the context of formative assessments, this would require a student to answer a minimum number of formative assessment events (for example 10 multiple choice questions). This quantity reward would not measure accuracy for those assessment events, but instead measure the quantity of participation. In the golfing analogy, this score represents the number of holes finished in the allotted time, regardless of the number of strokes taken. Students attempting to “game” this reward can provide random responses to the questions and qualify for the full reward. Best practice guidelines for educational settings state that rewards should embrace both completion and quality (Akin-Little et al. 2004).

An alternative includes a quality dimension whereby students must accurately complete a minimum number of formative assessment events. This completion model transforms the reward into a type-3 *completion reward*. This score need not track how many attempts were required. In an assessment using multiple choice questions, failed attempts would draw another question from a pool for a new attempt. In the golfing analogy, this allows the golfer to take a “mulligan” whenever a poor shot is made, or restart any hole not performed to par. This score captures the number or percent of holes accurately completed in the allotted time, regardless of the number of attempts required to achieve the standard.

While none of these approaches capture the success or failure *rate* (i.e. performance), the type-3 completion reward supports meaningful engagement (students must strive for accuracy, even though accuracy *rate* is not calculated into the score) and provides feedback that guides students to close the gap between KSAs and the standard to facilitate learning. This alignment with formative assessment best practices guides the selection of type-3 rewards in this study.

Hypothesis Development

During this study formative assessments are “practice questions” where accuracy does not play a role in the student’s grade. Repeated assessment trials allow students an opportunity to practice knowledge recall. From a cognitive learning standpoint, the mental process associated with practice allows the brain to organize and store the associated information. The result is improved learning that can translate into higher scores and better grades (Richardson & Gropper 1969). The formative assessment method also includes immediate feedback to help the learner recognize gaps in understanding and capability, then guides efforts to close that gap. As students observe the feedback for each formative assessment question, they have an opportunity to gain additional insight to the meaning and use of the associated concepts. The mental act of trial-and-error with feedback can improve learning and lead to better student scores on summative assessments (Chi 2009, Michael 2006). Furthermore, to the extent that formative assessment feedback provides intrinsic motivation to learn (Higgins et al. 2002), we expect improved learning outcomes to be positively associated with formative assessment engagement. Therefore:

H1: Formative Assessment engagement in outside-the-classroom flipped content is positively related to increased performance on Tests/Exams.

It has long been known that extrinsic rewards work in the classroom. Techniques such as verbal praise, token economies, contracts and others have been advocated with effect since the 1950s (Akin-Little et al. 2004). A 40-year meta-analysis of motivation found that incentives are a good predictor of behavior *quantity* (Cerasoli et al. 2014). Experimental studies demonstrate that undergraduate students respond to rewards by spending more time on tasks (Pierce et al. 2003). Translating these economies to outside-the-classroom tasks associated with blended and flipped instructions necessitates identifying situationally appropriate rewards.

Drawing on the insight that students value grades above all other considerations (Pressley et al. 1998), assigning a gradebook score to any activity is expected to increase student engagement with that activity. Reward tactics can enhance student time on task (Akin-Little et al. 2004). During this study, the open gradebook is a continuously updated digital database that is always available to students. A type-3 completion score appearing in the open gradebook is a tangible “reward” for engaging formative assessment tasks. While in-classroom environments are different than outside-the-classroom, we believe the cognitive process of recognizing and responding to completion scores as an explicit reward for formative assessment engagement will transfer to flipped instruction settings. Therefore:

H2: Providing completion score rewards for flipped instruction formative assessments will increase the quantity of formative assessment engagement.

Formative assessments are often used by students to adjust their learning tactics (Cauley & McMillan 2010). How students treat learning activities such as Formative assessments may vary from cursory (going through the motions), to deep engagement where the feedback is assimilated and KSAs improved. Independent of any change in the *quantity* of formative assessment engagement, learning objective attainment indirectly reflects the *quality* of formative assessment engagement.

The real goal for most college classes is mastery of learning objectives. Summative assessments serve as a proxy for mastery of learning objectives and can provide measure of effectiveness for instructional treatments. Summative assessments that introduce an external evaluator for credibility and involve evidence of student achievement are common in university level classes.

Rewards are widely recognized in the behavioral psychology literature and economics literature to be an effective way to improve performance (Hendijani et al. 2016). The human learning literature is a bit more nuanced. Cognitive Evaluation Theory (CET) (Deci et al. 2001) suggests that external events, such as offering rewards, the delivery of evaluations, and the setting of deadlines, influence a person’s perceptions of competence and self-determination. Events that increase self-determination enhance intrinsic motivation, while those that advance perceptions of being controlled can undermine intrinsic motivation.

Rewards and other external events (such as formative assessment feedback) have information that conveys self-determined

competence that can enhance intrinsic motivation. A reward in the form of an explicit grade provides knowledge of progress and a sense of satisfaction that accentuates intrinsic motivation (Docan 2006). However, some aspects of completion rewards, such as deadlines, have a controlling effect that could undermine intrinsic motivation (Deci et al. 2001).

Completion rewards seeking to increase *quantity* of formative assessment engagement must navigate a delicate balance to protect intrinsic motivation and avoid undermining the *quality* of formative assessment engagement. Best practices have emerged from years of study on motivation to suggest that rewards should not be presented on a single occasion, but repeated. In addition, rewards should not be given for mere participation in a task but should recognize progress and quality (Akin-Little et al. 2004). Considering best practice advice for instructional rewards to the environment of blended and flipped instruction, leads to:

H3: Providing completion score rewards for flipped instruction formative assessments will increase performance on Tests/Exams.

RESEARCH DESIGN

Testing of these hypotheses is done in two field experiments involving an Introduction to Computer Information Systems course. This course provides a good setting for study as it is a flipped instruction class required by all Sophomores in the College of Business. This class involves a cross section of students from all degree programs, not just those pursuing degrees in the specific domain of computers, and therefore minimizes the effects of prior domain knowledge. Considering flipped instruction techniques, the material in this class is easily segregated into conceptual topics (vocabulary, definitions, etc.) that can be the distinctive focus of outside the classroom efforts, and skill development (building algorithms, organizing and manipulating data, and engaged problem solving) that is emphasized during inside the class hands-on coaching. The instructional setting for this study isolates outside the classroom learning domains to minimize potential carry-over effects from inside the classroom activities.

The first field experiment involves two sections of a medium size class (40-45 students in each section). The second experiment involves two sections of a large size class (300 students in each section). For each experiment a repeated measures design was employed.

Figure 1 depicts a repeated measures cross-over design, sometimes called a “switching replications design” (Cook et al. 1979; Trochim 2005). During one event cycle the treatment is applied to one group while it is withheld from the second group. In a succeeding event cycle, treatment and control conditions are swapped. The white boxes with a diagonal line represent event phases where the treatment is withheld. The dark boxes with the letter T indicate event phases where the treatment is applied. The Dark boxes labeled M_n are the measurement times that take place after each event phase.

A challenge for the cross-over design in a field experiment during a single semester (15 weeks) is the “wash-out period” required for the effects of a treatment to dissipate before entering the next event cycle. In the absence of a wash-out period, effects carry-over for subjects who have received a treatment in an earlier event phase. In a field experiment using real students, the semester cannot be suspended for a viable wash-out period.

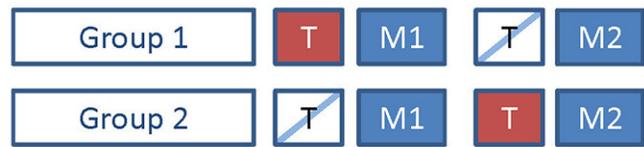


Figure 1. Repeated measures experiment with cross-over treatment

Instead of using a cross-over design, this study uses repeated measures with a baseline period where neither group receives the treatment as depicted in **Figure 2**. This design is favored for longitudinal studies as it captures within-unit change as well as inter-unit differences over time (Ployhart & Vandenberg 2010). The first event cycle establishes a baseline difference between groups. Event phase 2 applies the treatment to Group 1 but not Group 2. Data collected during event phase 2 exposes within-group effects of the treatment. This is measured as the within-group change in test scores from M1 to M2. Comparing the within-group change in test scores of the two groups (where both groups are taking the same test) controls for the difference in test difficulty and reveals the performance effects due to the treatment applied during T2.

Formative Assessments and Formative Assessment engagement

The formative assessment instrument for this study uses an eBook portal provided by a major textbook publisher. The portal allows establishing “assignments” that contain a set of reading material and a companion set of practice questions. These assignments constitute the major portion of out-of-classroom activities for this flipped instruction class. These assignments are intended to introduce students to concepts and terms, describe relationships between terms, and provide examples that allow students to define, assess and differentiate ideas fundamental to the knowledge domain of the class.

Each assignment has a target number of practice questions that constitute a formative assessment activity. Practice questions are multiple-choice, multiple-select, or fill-in-the-blank. Immediately upon answering, the student is given feedback on the correct answer with links to the text to help the student locate supporting and descriptive information on the associated concept. In addition to providing feedback on an individual question, the portal displays overall progress toward completing the set of assigned questions. When a question is answered correctly, progress is visually represented on a “progress bar”. When a question is answered incorrectly the progress bar does not advance, and a new question is drawn from the question pool (the question pool is not infinite, and eventually questions will repeat). A measure of formative assessment engagement (FAE) is maintained during all event phases of the study. The FAE score is calculated as the percentage of assigned practice questions that are accurately completed during the scheduled time interval (each week involves one assign-



Figure 2. Repeated measures study design without cross-over treatment

ment of between 13 and 25 practice questions). Incorrect answers do not penalize the student's FAE score or diminish their reward.

Completion Scores - the Reward Treatment

The reward treatment adopted in this study conforms to best practices advocated to mitigate threats to implicit motivation (Akin-Little et al. 2004). Rewards should be repeated, not isolated to one-time events. In addition, rewards should not be given for mere participation in a task but should recognize quality and progress. This study employs type-3 completion reward where scores reflect quantity of accurate progress toward a target quota. However, these scores do not track the number of trials and therefore do not measure performance.

All students have full access to the same formative assessment practice questions. When the treatment is applied, the FAE completion score is added to the online gradebook and visible to the student. Once the treatment is active the student can see completion scores added as a reward in the gradebook. The open gradebook is updated in real-time as engagement takes place. Students not receiving the treatment do not receive a completion reward in the gradebook. There is not even a grade category for formative assessment activities in the gradebook for these students.

Summative Assessment

After each event cycle a Unit Test is administered. This Summative assessment measures student mastery of learning objectives associated with outside the classroom assignments. Summative assessment measures establish a baseline performance for both groups in the absence of the treatment, then within-group and between-group change associated with various treatment combinations.

Study Design

Figure 2 depicts the treatment schedule for all groups participating in this study. The first experiment involved two groups of 45 students. All students in both groups receive the same assignments, the same weekly schedule, and the same unit tests. The students in Group 1 meet in a classroom together as a group and students in Group 2 meet in a classroom together as a group at a different time. At the end of the semester only students who completed the final exam and agreed to participate in the study were included for data collection. Group 1 ended up with 27 participants and Group 2 had 37 participants.

The second experiment involves new groups of students in large section classes (two sections of 300 students) taking place in a subsequent semester. Group 3 had 228 participants and Group 4 had 273 participants. The difficulty of assignments, material and tests used in experiment 2 is similar but different from that in experiment 1. However, Group 3 students all received the same assignments, the same weekly schedule, and the same unit tests as Group 4.

The semester course is divided into four units, each lasting 3 to 4 weeks. Each week students are assigned a module in the eBook for outside-the-classroom reading with the practice questions serving as a companion Formative assessment. During Unit 1, completion scores are calculated by the eBook portal, but not published to the online gradebook as a completion reward. While students can self-monitor their progress with the assignment (the progress bar is visible while working on formative assessment questions), this progress during the first unit is not

scored, measured or reflected in the online gradebook in any form. Unit 1 culminated in a Summative Unit test during week 3 of a 15-week semester. During Unit 2 completion scores are calculated and updated in the online gradebook in real-time for Group 1 but hidden and not published for Group 2. Unit 2 culminated in a summative assessment during week 6. Students sharing information about rewards is expected to increase as the semester progresses. Due to the threat of crossover effects, scores from Unit 3 and Unit 4 portions of the class are excluded from analysis in this study.

RESULTS

Data collected for all four groups is analyzed with OLS for H1 using the following regression equation:

$$UT = \beta * FAE + \epsilon$$

Table 1 reports the regression results for each experiment and group. In all cases Formative Assessment engagement scores are statistically significant predictors of Unit Test Summative Assessment scores, supporting H1. R^2 range from 0.75 to 0.92, suggesting that Formative Assessments explain a very large portion of the variance in Unit Test performance. Students who have high Formative Assessment engagement demonstrate higher learning objective mastery for this flipped instruction class.

Table 1: H1 OLS results

Panel	N	UT1: Baseline [M1]	UT2: Retest 1 [M2]
Exp 1	64	Mean UT1 =79.81 $\beta = 9.160^{***}$ Pvalue <0.000 $R^2=0.80$	Mean UT2 =85.13 $b = 9.292^{***}$ Pvalue <0.000 $R^2=0.91$
Exp 1 Group 1	27	Mean UT1 =77.48 $\beta = 9.328^{***}$ Pvalue =0.000 $R^2=0.75$	Mean UT2 =82.37 $b = 9.121^{***}$ Pvalue <0.000 $R^2=0.91$
Exp 1 Group 2	37	Mean UT1 =81.51 $\beta = 9.064^{***}$ Pvalue <0.000 $R^2=0.85$	Mean UT2 =87.14 $b = 9.411^{***}$ Pvalue <0.000 $R^2=0.91$
Exp 2	501	Mean UT1 =76.36 $\beta = 8.477^{***}$ Pvalue <0.000 $R^2=0.87$	Mean UT2 =81.80 $b = 9.703^{***}$ Pvalue <0.000 $R^2=0.91$
Exp 2 Group 3	228	Mean UT1 =77.00 $\beta = 8.552^{***}$ Pvalue <0.000 $R^2=0.87$	Mean UT2 =81.28 $b = 8.573^{***}$ Pvalue <0.000 $R^2=0.92$
Exp 2 Group 4	273	Mean UT1 =75.83 $\beta = 8.414^{***}$ Pvalue <0.000 $R^2=0.86$	Mean UT2 =82.25 $b = 8.819^{***}$ Pvalue <0.000 $R^2=0.90$

H2 is analyzed considering the within-group change in FAE between event periods T1 and T2. Results reported in **Table 2** show relevant metrics for each experiment collectively, and each group individually during event period 1 and event period 2. Column N contains the sample size, columns T1 and T2 show which groups received the treatment during each event period. Columns FAE1 and FAE2 report the Mean FAE scores for each group. Both the treatment group and the control group experienced an increase in FAE from period 1 to period 2. Within-group change is reported as the change in FAE between T1 and T2, which is the metric used to inform H2. Data collected for both experiments was analyzed using the following ANOVA equation:

$$\Delta FAE = \beta * T2 + \epsilon$$

Panel	N	T1	F AE1 Mean (StDev)	T2	F AE2 Mean (StDev)	ΔFAET2 -T1
Exp 1	64		70.19 (37.80)		83.96 (26.81)	+13.76
Exp 1 Group 1	27	-	62.32 (40.05)	✓	82.83 (27.14)	+20.51
Exp 1 Group 2	37	-	75.94 (35.52)	-	84.78 (26.91)	+8.84
ANOVA	Δ FAE = β*T2					F=12.05 Pvalue<0.000
Exp 2	501		78.63 (32.07)		86.44 (27.58)	+7.80
Exp 2 Group 3	228	-	78.70 (32,17)	✓	88.96 (25.17)	+10.26
Exp 2 Group 4	273	-	78.57 (32.04)	-	84.33 (29.31)	+5.76
ANOVA	ΔFAE = β*T2					F=25.38 p-value<0.000

As predicted by the hypothesis, FAE grows more in absolute terms for the treatment group. In Experiment 1 involving smaller class sizes, Group 1 receiving the treatment experiences a dramatic increase in FAE (Δ F AE = +20.51), whereas Group 2 experiences a more modest increase (Δ F AE = +8.85). Similar though less dramatic response is revealed for the larger class sizes in Experiment 2 Group 3 (Δ F AE = +10.26) and Group 4 (Δ F AE = +5.76). An ANOVA analysis comparing the within-group change in FAE reveals the effects of the treatment is statistically significant in both Experiment 1 ($F=12.05, P\text{-value}<0.000$) and Experiment 2 ($F=25.38, P\text{-value}<0.000$). The data supports H2, suggesting that completion rewards are associated with more formative assessment engagement.

While H2 assess the treatment effect of completion rewards on the *quantity* of FAE, H3 informs on the *quality* of FAE, and the effect of completion rewards on Exam/Tests. H3 is analyzed considering the difference of within-group change in Unit Test scores.

Results reported in **Table 3** show relevant metrics for each experiment. Column UT1 reports the Summative Assessment Unit Test score obtained at the end of event period 1. Column UT2

Panel	N	UT1	UT2	UT2-UT1
Exp 1	64	79.81	85.13	+5.32
Exp 1 Group 1	27	77.48	82.37	+4.89
Exp 1 Group 2	37	81.51	87.14	+5.62
ANOVA	UTn-UT1 = β*T2			F=0.05 p=0.82
Exp 2	501	76.36	81.80	+5.44
Exp 2 Group 3	228	77.00	81.28	+4.28
Exp 2 Group 4	273	75.83	82.25	+6.42
ANOVA	UTn-UT1 = β*T2			F=2.79 p=0.09

reports Unit Test scores obtained at the end of period 2. Change in within group Unit Test scores are calculated as a change from baseline (UT1). Within group changes are analyzed for statistically significant differences from the grand mean using ANOVA with the following equation:

$$\text{Performance_Change(UT2-UT1)} = \beta * T2 + \epsilon$$

Within group differences in SA scores from period UT1 to UT2 indicate that the group **not** receiving completion rewards improved their Unit Test scores more than the group that increased their FAE in response to the completion reward. Both the direction of change and the lack of statistical significance suggest H3 be rejected. While completion rewards may increase the quantity of FAE (H2), the quality of this engagement is not enhanced, and summative assessment test scores suffer for the treatment group relative to the control group. While Unit Test effects in this experiment are inconclusive, the data suggests that completion rewards may reduce learning objective mastery in flipped instruction settings.

DISCUSSION

A summary of findings is provided in **Table 4**. This study found strong support that formative assessment engagement is positively associated with student performance measured by summative assessments tests. This finding is a replication of other studies in non-flipped teaching environments and serves as a validity check for this study's sample and data collection methods. In addition, this study found statistically significant support that completion rewards can increase the *quantity* of FAE for the out-of-classroom material assigned in this flipped instruction course. However, completion rewards are not associated with an improvement in test/exam performance and may be responsible for a decrease in test/exam performance. While completion rewards motivate an increase in the *quantity* of FAE for this flipped instruction class, they may simultaneously be undermining the *quality* of FAE! In response to the completion rewards, these students appear to have migrated toward more cursory participation in the formative assessment task without assimilating the formative assessment feedback or translating the effort into learning. Completion rewards appear to motivate score accumulation behaviors at the expense of learning.

Implications for Theory

As a unique contribution to the literature we provide a classification of gradebook score reward types. By defining type-1, type-2 and type-3 rewards we establish criteria for aligning rewards with best practices for formative assessment engagement. Type-3 rewards are tuned to incent full engagement and thereby maximize our opportunity to induce real learning.

This study examined two aspects of flipped instruction. The first involves a demonstration that formative assessments are positively associated with learning measured by tests and exams. While this is a replication of existing studies in the general sense, this finding demonstrates that the same cognitive processes that make formative assessments effective for instructor led classes serve a similar role for self-led learning outside the classroom.

The unexpected finding that completion rewards do not induce learning objective performance suggests an opportunity for further investigation. Adapting the type-3 completion rewards to a type-4 quality reward that emphasizes quality formative assess-

Table 4: Hypothesis & Results

Hypothesis		Experiment 1	Experiment 2	Conclusion
H1	Formative Assessment engagement in outside-the-classroom flipped content is positively related to increased performance on Tests/Exams. + FAE → Performance	p-value < 0.001 (Table 1)	p-value < 0.001 (Table 1)	Accept
H2	Providing completion score rewards for flipped instruction formative assessments will increase the quantity of formative assessment engagement. + Reward → FAE	p-value < 0.001 (Table 2)	p-value < 0.001 (Table 2)	Accept
H3	Providing completion score rewards for flipped instruction formative assessments will increase performance on Tests/Exams. + Reward → Performance	p-value > 0.05 (Table 3)	p-value > 0.05 (Table 3)	Reject

ment engagement may be possible if rewards are tied to feedback engagement and not mere completion. Alternately the rewards could be adapted to target competency gaps uniquely for each student.

Cognitive Evaluation Theory suggests that learning efficacy can be influenced by both implicit motivations and explicit motivations. It is worth noting that the participants in this study are not exclusively Computer Information Systems majors, but rather represent a cross-section of business school students from all majors. The lack of a strong a priori interest in the material (i.e., implicit motivation) may be a relevant moderating factor worthy of further investigation.

Implications for Teaching

While explicit rewards in the form of completion scores in an online gradebook were demonstrated to increase the *quantity* of formative assessment engagement, they did not improve the *quality* of this engagement. Instructors seeking the benefits of formative assessments, should approach completion score rewards with caution. This study suggests completion rewards activate grade seeking behaviors of students at the expense of learning motivation.

An additional insight from this study is the relative effect differences for medium and large size classes. Completion rewards have larger absolute effects on FAE *quantity* in classes of 40-45 students (+20.51 basis points in column “Δ FAE T2 -T1”) than they do for classes of 300 students (+10.26 basis points). A challenge for faculty scaling up to large class sizes is the muted efficacy of many instructional techniques. This study suggests that completion rewards share this characteristic.

CONCLUSION

Flipped instruction classes require the student to take responsibility for learning material outside the classroom. While some tools are available to instructors attempting to guide outside-the-classroom learning, the efficacy of these tools may be different when the instructor is not present. This study defined three types of rewards and tested the efficacy of a type-3 completion rewards that adopt recommendations provided by meta-analysis of learning literature. This study demonstrates that type-3 completion rewards increase the *quantity* of formative assessment engagement but failed to confirm an increase in the *quality* of that engagement for outside-the-classroom material. The data suggests that type-3 completion rewards may decrease the quality of formative assessment engagement.

NOTES

1. Type-1, type-2, type-3 classification of rewards is original to this study. The distinction embraces the ideal of recurring reward events that are not given for mere participation but recognize progress and quality (Akin-Little et al. 2004).

REFERENCES

Abeyssekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research & Development*, 34(1), 1-14.

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, 126, 334-345.

Akin-Little, K. A., Eckert, T. L., Lovett, B. J., & Little, S. G. (2004). Extrinsic reinforcement in the classroom: Bribery or best practice. *School Psychology Review*, 33(3), 344.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. Paper presented at the ASEE National Conference Proceedings, Atlanta, GA.

Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability* (formerly: *Journal of Personnel Evaluation in Education*), 21(1), 5.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83(1), 1-6.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological bulletin*, 140(4), 980.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, 1(1), 73-105.

Cook, T. D., & Campbell, D. T. (1979). *Validity. Quasi-experimentation: Design and Analysis issues for field settings*, 37-51. Boston: Houghton Mifflin.

Davies, R. S., Dean, D. L., & Ball, N. (2013). Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. *Educational Technology Research and Development*, 61(4), 563-580.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of educational research*, 71(1), 1-27.

Docan, T. N. (2006). Positive and Negative Incentives in the Class-

- room: An Analysis of Grading Systems and Student Motivation. *Journal of Scholarship of Teaching and Learning*, 6(2), 21-40.
- Garrison, D. R. (1997). Self-directed learning: Toward a comprehensive model. *Adult education quarterly*, 48(1), 18-33.
- Hendijani, R., Bischak, D. P., Arvai, J., & Dugar, S. (2016). Intrinsic motivation, external reward, and their effect on overall motivation and performance. *Human Performance*, 29(4), 251-274.
- Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in higher education*, 27(1), 53-64.
- Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in higher education*, 27(1), 53-64.
- Hill, P. (2012, November/December 2012). *Online Educational Delivery Models: A Descriptive View*. Educause Review.
- Iwamoto, D. H., Jargis, J., Bordner, R., & Chandler, P. I. (2017). Self-Regulated Learning as a Critical Attribute for Successful Teaching and learning. *International Journal for the Science of Teaching and Learning*, 11(2), 10.
- Janssen, S., & O'Brien, M. (2014). Disentangling the Effects of Student Attitudes and Behaviors on Academic Performance. *International Journal for the Science of Teaching and Learning*, 8(2), 29.
- Kim, K.-J. (2004). Motivational Influences in Self-Directed Online Learning Environments: A Qualitative Case Study. Paper presented at the Association for Educational Communications and Technology, 27th, Chicago, IL.
- Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education*, 31(1), 30-43.
- Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, 59(5), 593-618.
- Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & Education*, 48(2), 185-204.
- Lim, D. H., & Morris, M. L. (2009). Learner and instructional factors influencing learning outcomes within a blended learning environment. *Journal of Educational Technology & Society*, 12(4), 282.
- Michael, J. (2006). Where's the evidence that active learning works? *Advances in physiology education*, 30(4), 159-167.
- Pierce, W. D., Cameron, J., Banko, K. M., & So, S. (2003). Positive effects of rewards and performance standards on intrinsic motivation. *The Psychological Record*, 53(4), 561-578.
- Ployhart, R. E. and Vandenberg, R. J. (2010). "Longitudinal Research: The theory, design, and analysis of change", *Journal of Management*, (26:1), pp94-120.
- Pressley, M., Van Etten, S., Yokoi, L., Freebern, G., & Van Meter, P. (1998). The metacognition of college studentship: A grounded theory approach. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 347-366). New York, NY: Routledge.
- Richardson, J., & Gropper, M. S. (1964). Learning during recall trials. *Psychological Reports*, 15(2), 551-560.
- Rolfe, I., & McPherson, J. (1995). Formative assessment: how am I doing? *The Lancet*, 345(8953), 837-839.
- Scriven, M. (1981). *Evaluation thesaurus* (third ed.). Pt. Reyes, CA: Edgepress.
- Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British journal of educational studies*, 53(4), 466-478.
- Trochim, W. M. K. 2005. Experimental design. *Research methods: The concise knowledge base*, 150-170. Atomic Dog Publishing.