Exploring the effect of standards-based grading on student learning

K. Clay McKell
Electrical Engineering Department
California Polytechnic State University
San Luis Obispo, CA, USA
https://orcid.org/0000-0001-6027-0641

Andrew Danowitz
Electrical Engineering Department
California Polytechnic State University
San Luis Obispo, CA, USA
https://orcid.org/0000-0002-4842-2005

***Abstract*—This is a Research to Practice Full Paper. Standards-based grading (SBG) is an assessment framework that rewards students for demonstrating mastery of course learning objectives. It also calls for feedback on formative assessments that allow students to target their deficiencies in a course area without being entangled with other confounding factors commonly found in summative assessments like test-taking aptitude or even class attendance. SBG is closely related to mastery learning and stands in contrast to traditional points-based grading methods that average a student's performance over time (and topics), and some research suggests that SBG-based courses, when ideally implemented, can lead to increased standardized test scores and enhanced retention of course learning objectives over time.**

In this work, we describe our experience attempting to implement SBG in a sophomore-level Continuous-Time Signals and Systems course offered by an Electrical Engineering Department. For our implementation, the data show that student learning of course topics is comparable between the experimental SBG section (N = 33) and a control traditional points-based grading section (N = 34) as measured by performance on a Continuous-Time Systems and Signals Concept Inventory pretest and posttest. Additionally, student opinion of the SBG class was mixed as reported in end-of-term evaluations; in these evaluations, students judged the SBG course to be less educationally effective than the traditional section.

In this paper, we present the full results of this experiment and explore reasons why our SBG implementation did not achieve the improved student outcomes found in other research. Through this work, we hope to help other educators avoid common pitfalls and mistakes that can hamper implementations of this promising classroom innovation.

***Index Terms*—standards-based grading, mastery, signals and systems**

I. Introduction

Standards-based grading (SBG) [1] is a type of formative assessment in which students must master learning objectives to successfully complete a course. As opposed to more traditional summative grading methods, where student success depends on the ability to achieve a passing score averaged over a number of fixed, single-attempt assignments, SBG allows students as many attempts as they need to demonstrate their grasp over each learning objective. While many implementations of summative grading allow students to make up for a deficiency in some learning objectives by showing strength in other areas, SBG requires students to gain a level of proficiency across all learning objectives to successfully complete the course.

Because it requires students to become proficient in all learning objectives, and because of increased opportunities for students to demonstrate mastery in each area, some research has shown that SBG can improve overall student learning and long-term student success. There is, however, relatively little research showing how SBG can be implemented in an engineering classroom, how engineering students respond to SBG techniques, and whether SBG always provides superior results to more traditional pedagogical methods. This work starts to address that knowledge gap by presenting a comparison-study of SBG and traditional points-based grading techniques in a Continuous-Time Signals and Systems course at California Polytechnic State University San Luis Obispo.

A. Background

The implementation of SBG described in this work has its origins in Bloom's "Learning for Mastery" [2] in which he advocates for frequent formative evaluation tests to help assess each student's mastery of individual course concepts. The argument goes that a student's grade in a course should reflect how much of the course's content he or she has mastered rather than how he or she stacks up against the other students. That is, grades should be *criterion-referenced* rather than *norm-referenced* [3].

There is a good deal of literature pertaining to standards and mastery in the K–12 context. Relatively few publications exist describing its efficacy in higher education and in university-level engineering in particular. Notable exceptions are the narrative by Beatty [4] and the qualitative study by Bekki, Dalrymple, and Butler [5].

B. Motivation

The apparent opening in the literature was one of the factors motivating this work. Another important motivating factor was to determine how students react to SBG, and to determine whether and to what extent learning benefits tied to SBG can be realized by an instructor who is new to the technique.

While previous literature has shown benefits for a number of non-traditional teaching techniques, literature has also shown that students can be resistant to, or feel like they learn less when these tools are employed [6]. Other research has shown that evidence-based learning methods, when implemented sub-optimally, can actually hurt student learning [7], [8]. Therefore, it is important to understand how SBG will be received by students, and the risks associated with SBG implementation.

The other motivating factors for this work were largely drawn from past student evaluations of the Continuous-Time Signals and Systems course at our university. Common complaints in these evaluations are that the homework assignments are too challenging or time consuming, and that exams are too time-constrained and too unlike preparatory exercises.

In the standards-based course implementation described in Section II, homework problems are assigned but not graded and the few high-stakes time-constrained exams are replaced with a larger number of assessments spread out over the entire term. We see giving students agency over when they are ready to be tested on a subject as an additional benefit of this structure.

Furthermore, we hypothesized that the criterion-referenced nature of SBG would benefit students by removing ambiguities about their grade in the course and removing the competitive nature that norm-referenced grading necessitates.

II. Methodology

The goal of this experiment was to determine whether and to what extent SBG improves student learning over a traditionally delivered continuous-time signals and systems course. To that end, in Spring 2019 one of the authors conducted one section with a "traditional" syllabus and the other section with the "SBG" syllabus.

The instructor determined which section would employ SBG techniques to simplify logistics; students did not register for the class knowing it would be non-traditional.

The instrument used to quantify student learning is the Continuous Time Signals and Systems Concept Inventory (CTSSCI) [9]. Students in both sections took this 25-question multiple choice exam on the first day of class as well as during their final exam 10 weeks later. Using the CTSSCI as a pre- and posttest to measure student gains is recommended by its authors as a valid use case [10].

A. Similarities Between Treatments

Despite the different grading methodologies, both course sections used in this study shared a number of similarities. Both sections met for 110 minutes two days per week and received identical lecture content. Both sections used the same textbook [11] with the same assigned readings every week. Both sections were eligible to attend the same office hours and had access to the same online discussion tool.

Finally, the sections were given the same weekly homework assignments, and received the same solutions to those homework assignments.

B. Elements Unique to the Traditional Section

The students in the traditional section were assigned a letter grade at the end of the course according to how many standard deviations their total score fell from the section average.

TABLE I
Enumerated learning objectives and the week (Wk) in which each was introduced.

LO#	Learning objective	Core?	Wk
1	Energy and Power signal classification	Bonus	1
2	Periodic signal classification	Bonus	8
3	Linear system classification	Core	2
4	Time invariant system classification	Core	2
5	Causal system classification	Bonus	2
6	Instantaneous system classification	Bonus	2
7	Stable system classification	Core	2
8	Convolution integral evaluation	Core	3
9	Sieving convolution property application	Core	3
10	Commutative convolution property application	Bonus	3
11	Associative convolution property application	Bonus	3
12	Linear convolution property application	Core	3
13	System analysis via convolution	Core	4
14	Differential equation solution of ZIR	Bonus	5
15	Differential equation solution of ZSR	Bonus	5
16	Laplace transform evaluation	Core	6
17	Laplace properties application	Core	6
18	System analysis via Laplace	Core	7
19	Fourier series evaluation	Core	8
20	Fourier series properties application	Core	8
21	System analysis via Fourier series	Bonus	8
22	Fourier transform evaluation	Core	9
23	Fourier transform properties application	Core	9
24	System analysis via Fourier transform	Bonus	10

The weekly homework assignments, which were graded for accuracy, constituted 20% of the total score. The majority of the grade depended on student performance on two midterm exams and one cumulative final exam (part of which comprised the CTSSCI posttest).

C. Elements Unique to the Standards-Based Section

The standards chosen for the SBG section were drawn from the Electrical Engineering Department's standardized learning objectives for the course as submitted to ABET. These 24 learning objectives are listed in Table I. The instructor designated 14 of these as "core" objectives and denoted the remaining 10 as "bonus" objectives. The core objectives were chosen such that any student that could demonstrate mastery of them would be comfortable in any of the sequel courses.

To demonstrate mastery of any of the learning objectives, a student had to score at least 90% on a short quiz tailored to that objective. The instructor put significant effort in to writing the quizzes for each objective such that they were as decoupled as possible from other learning objectives. The quizzes were available to be taken at any of the following times:

- During any of the instructor's five weekly office hours.
- During a special "quiz hour" where the instructor only proctored these quizzes.
- In class on the days when the traditional section took their midterm exams.
- During the university-scheduled final exam time.

The quizzes ranged in difficulty from identification by inspection (Fig. 1) to solving for outputs given a system description (Fig. 2).

In each part below, the operator form of a system is shown. For each part, determine whether the system is dynamic or memoryless. Explain your answer.

(a) $S_a[x(t)] = x(2t)$
(b) $S_b[x(t)] = \sin(x(t))$
(c) $S_c[x(t)] = \int_{t-1}^{t} x(\sigma)d\sigma$

Fig. 1. Sample quiz for LO06: Instantaneous system classification.

This quiz is intended to assess your mastery of Laplace analysis. Use of other methods to solve this question will not count.
Consider a system whose input/output relationship is described by $y'' + 4y' + 6y(t) = x(t)$. If the input to this system is a unit step, determine the zero-state output in the time domain.

Fig. 2. Sample quiz for LO18: System analysis via Laplace transform.

If a student failed to earn 90% on any quiz, that student was required to correct their errors and submit these corrections before being allowed to take a different quiz on the same learning objective.

To encourage students to take the quizzes seriously and to lighten the grading load on the instructor, the number of attempts on any learning objective was capped at three. If the student failed to earn 90% on three different quizzes on the same learning objective, that objective became inaccessible to them.

The grading scale for the SBG section was tied to the number and type of learning objectives of which students demonstrated mastery, as shown in Table II. To incentivize students to perform well on the CTSSCI posttest, a "plus" grade modifier would be given to those who outperformed their pretest score and a "minus" would be given to those who under performed. These rules were revealed to the students only after they had finished the pretest.

Homework assignments were available to students in the SBG section but were completely optional. Additionally, every homework question was tagged with the various learning objective numbers that it covered. This feature was added to ensure students in the SBG section were able to identify which homework questions should help them practice for each quiz.

When polled, every student reported that this was their first exposure to standards-based grading.

TABLE II
GRADE SCALE FOR SBG SECTION.

Letter grade	LO mastery requirement
A	All 14 Core and 10 Bonus
B	All 14 Core and at least 5 Bonus
C	At least 14 Core
D	At least 12 Core
F	Fewer than 12 Core

TABLE III
NORMALIZED GAIN COMPARISONS BROKEN DOWN BY ITEM CATEGORY.

	Math	LTI	Conv	Trans	Filt
Traditional	0.78	0.40	0.43	0.40	0.33
SBG	0.72	0.26	0.38	0.30	0.13

III. RESULTS

We define the raw (learning) gain, G_n, exhibited by student n as the difference between pre- and post-course test scores on the Signals and Systems Concept Inventory:

$$G_n = \text{post}_n - \text{pre}_n. \tag{1}$$

All scores are expressed as percentages.

Following Hake's lead [12], we define the normalized gain, g, of a population of N students as

$$g = \frac{\sum_{n=1}^{N} G_n}{\sum_{n=1}^{N} 100 - \text{pre}_n}, \tag{2}$$

which represents an unbiased estimator of the fraction of originally unknown content that a student learns.

Fig. 3 contains a scatter plot of the raw gains exhibited by the students in each of the two sections. The dashed lines are curves of constant g and the low, medium, and high-gain regions were identified by Hake [12].

The normalized gain for students in the traditional section was $g_T = 0.44$ and the normalized gain for the SBG section was $g_S = 0.35$. Using these quantities, the estimated raw gains for both sections are also plotted against the respective average pretest scores in Fig. 3. The observed value of g_T is in line with normalized gains the instructor has seen in previous offerings of this course in a traditional format.

The 25 CTSSCI version 5.0 questions—called "items"—fall into five categories. In the order that they were covered in this class, those categories (along with abbreviations and the number of items in each) are:

1) background mathematical concepts (Math, 4)
2) linearity and time invariance (LTI, 5)
3) convolution (Conv, 4)
4) transform representations (Trans, 11)
5) filtering (Filt, 1)

The normalized gains for each section within each category are shown in Table III. The traditional section maintains a higher normalized gain across all categories, but the difference appears more pronounced for categories of topics taught later in the term.

There is an important aspect of the posttest scores that factor in to all of these gain calculations: the motivation of the students when taking the posttest. Both sections were allotted one hour to complete the posttest after which they were allowed to leave the final exam. Grades for students in the traditional class section depended on how they performed on the posttest (among other things) in relation to their classmates. That is, the students in the traditional section

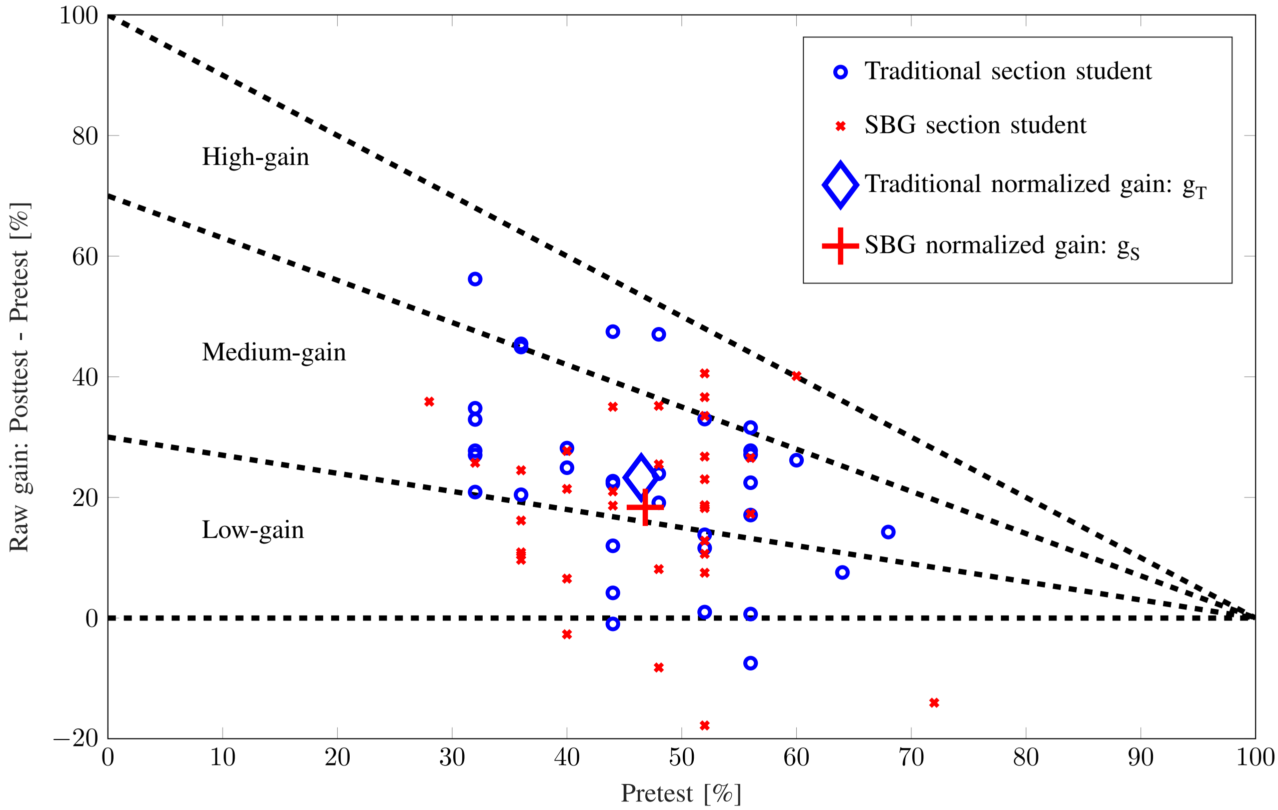

Fig. 3. Comparison of raw gains achieved by students in the traditional section ($N = 34$) with those in the SBG section ($N = 33$). Due to the discrete nature of the data, a small dither has been added to the vertical component of every data point to help distinguish between co-located data.

had extrinsic motivation to do well regardless of their prior performance in the class.

In contrast, there were at least two populations of students in the SBG section with an obvious lack of motivation to do well on the posttest. One population is those students who were assured an "A" grade based on their quiz performances before the final exam. Our institution does not offer "A+" grades, and the risk of under performing their pretest score was very low, so these students had limited extrinsic motivation to try their best on the posttest. The other population is those students on the other end of the spectrum: those whose prior quiz performances were so bad that they were nearly—but not absolutely—assured an "F." Students in this category who attended the final exam to complete the posttest may have had limited extrinsic motivation to give it their best shot as there is no discernible difference between an "F" and an "F+."

IV. Discussion

One of the motivations for attempting this experiment was to give students agency over when they were ready to demonstrate their knowledge of a subject. However, this agency was seen as a burden by some students who would rather have set deadlines and structure. These students' feelings are well summarized in the following anonymous end-of-term evaluation comment from the SBG section:

> "I think this system rewards good time management and punishes procrastinators, even if they know the same amount of material. As a procrastinator, I am definitely not a fan."

While time management is an important skill for students to develop, it appears that many students in the SBG section chose to delay attempting many of the quizzes until well after the topics were covered in lecture. On its face, increasing student agency is a positive feature of SBG, and for students who chose to delay assessment SBG should help ensure longer-term knowledge retention and review of earlier course materials. In a 10 week term, however, there is very little opportunity afforded students working within any grading scheme to take their time with the material. Also, the choice to delay assessment (and potentially understanding) of concepts that later course topics build on may hinder student learning learning of those later topics.

Fig. 4 shows the distributions of the dates that students first attempted each of the quizzes. The general trend is that students prioritized taking the core quizzes earlier than the bonus ones, but there are numerous instances of students tak-

TABLE IV
NORMALIZED GAIN COMPARISON BETWEEN SUB-COHORTS ZERO MONTHS AFTER COMPLETING THE SIGNALS AND SYSTEMS COURSE AND THEN AGAIN NINE MONTHS AFTER.

	After 0 months	After 9 months	Change
Traditional	0.51	0.83	+0.31
SBG	0.27	0.68	+0.41

ing their first attempt at a quiz at the last possible opportunity (especially core objectives 7, 12, 13, and 18).

While the results from the SBG course appear disheartening, nine months after the conclusion of the course in which this experiment was conducted, a subset of both cohorts took a sequel course from the same instructor. In the final exam of this class, the students were asked a subset of the questions from the CTSSCI—those that were pertinent to the sequel class—in order to track how the two delivery modes might affect long-term retention. The sample size is small: 11 students from the traditional section and 13 students from the SBG section answered four out of the 25 CTSSCI questions. The normalized gains for these sub-cohorts are displayed in Table IV, and appear to show higher change in gains in learning attainment for the SBG section. The normalized gains in the table are measured with respect to the original pretest scores.

One takeaway from this data might be that even if an SBG class does not produce immediate learning gains commensurate with a traditional lecture, it may facilitate relearning as posited by Gentile and Lalley [13]. The data is inconclusive, however, as the SBG initial gains indicate that these students had more room for conceptual growth compared to the cohort who took the traditional course.

V. CONCLUSION

The normalized results presented here did not support the idea that SBG is *always* superior to traditional grading methods regardless of implementation. There are, however, ample opportunities to improve on the implementation of standards-based grading presented here.

Many adjustments could be made to try and improve the outcomes of the standards-based section. Those adjustments fall into two phases: simple logistical changes specific to our implementation and larger structural changes that generalize to other possible implementations.

A. Logistical Changes

Upon reading student feedback after the course, it became apparent that the quiz hour that the instructor carved out did not fit all of the students' schedules. While it is difficult to find an hour each week that would fit everyone's schedules, it might be possible to find two half-hour slots that maximize the probability that each student can attend at least one quiz session per week without sacrificing lecture time.

The workload on the instructor was also daunting. This was partly due to the novel nature of the grading scheme, including creating an automated online "quiz request" system and writing 111 unique quizzes to accommodate the necessary number of retakes. However, the quiz writing and grading workload could be reduced if there were fewer learning objectives. This could be accomplished by simply cutting some of the bonus objectives. Another solution would be to merge some of the 24 learning objectives into broader objectives. For example, learning objectives 5–7 might be collapsed into "other system classifications," and 9–12 might be collapsed into "properties of convolution." Automated quiz creation from large, learning objective specific question banks could also help; especially if paired with an automated grading system.

Another recommended change relates to the process of requiring students to turn in their corrections to their previous quiz. For this implementation, these "corrections" did not have to be correct before students took their next quiz attempt. Many times a student would turn in their previous quiz corrections simultaneously with their second attempt only for the instructor to find that both submissions were inadequate and that the student was faced with the dreaded "third and final try." The logistical change required to avoid this scenario would be to prevent a student from requesting an additional quiz attempt until their prior attempt was regraded and deemed correct.

B. Structural Changes

Building on the suggestions of Gentile and Lalley [13], another version of this class would shift the assessment focus away from quizzes and on to a wider variety of demonstrations of knowledge. In this vein, within each core learning objective, a student would be required to pass a quiz with a 70% or 80% threshold to secure a "C" for that objective. Once a student has surmounted that hurdle, they would be presented with additional possibilities for demonstrating mastery to achieve a higher grade. *Enrichment activities* might include tutoring another student, writing a custom quiz problem (and solution), writing a short research paper, or some other assignment on that topic. Completing one or two of these would increase that learning objective's grade to a "B" or "A," respectively. Each student's term grade would then be the average of their learning objective grades.

This structure has the disadvantage that it lowers the standard for a "C" grade in the class from that in Table II. It also shares the flaw with the traditional grading scheme that a passing class grade can obscure gross deficiencies in one subject with excellence in others. At the same time, assessing the enrichment activities could lead to a dramatic workload increase for the instructor, especially if multiple attempts are allowed for enrichment activities.

There are, however, several advantages in this structure's favor over the SBG structure utilized in this paper. It reduces the number of quizzes that have to be written and graded, and it allows for students to demonstrate knowledge without the stress or time pressure of quizzes. Principally, it also encourages "overlearning" [13] whereas the SBG structure outlined in this work provides no incentives for students to ponder a topic once they had passed the quiz.

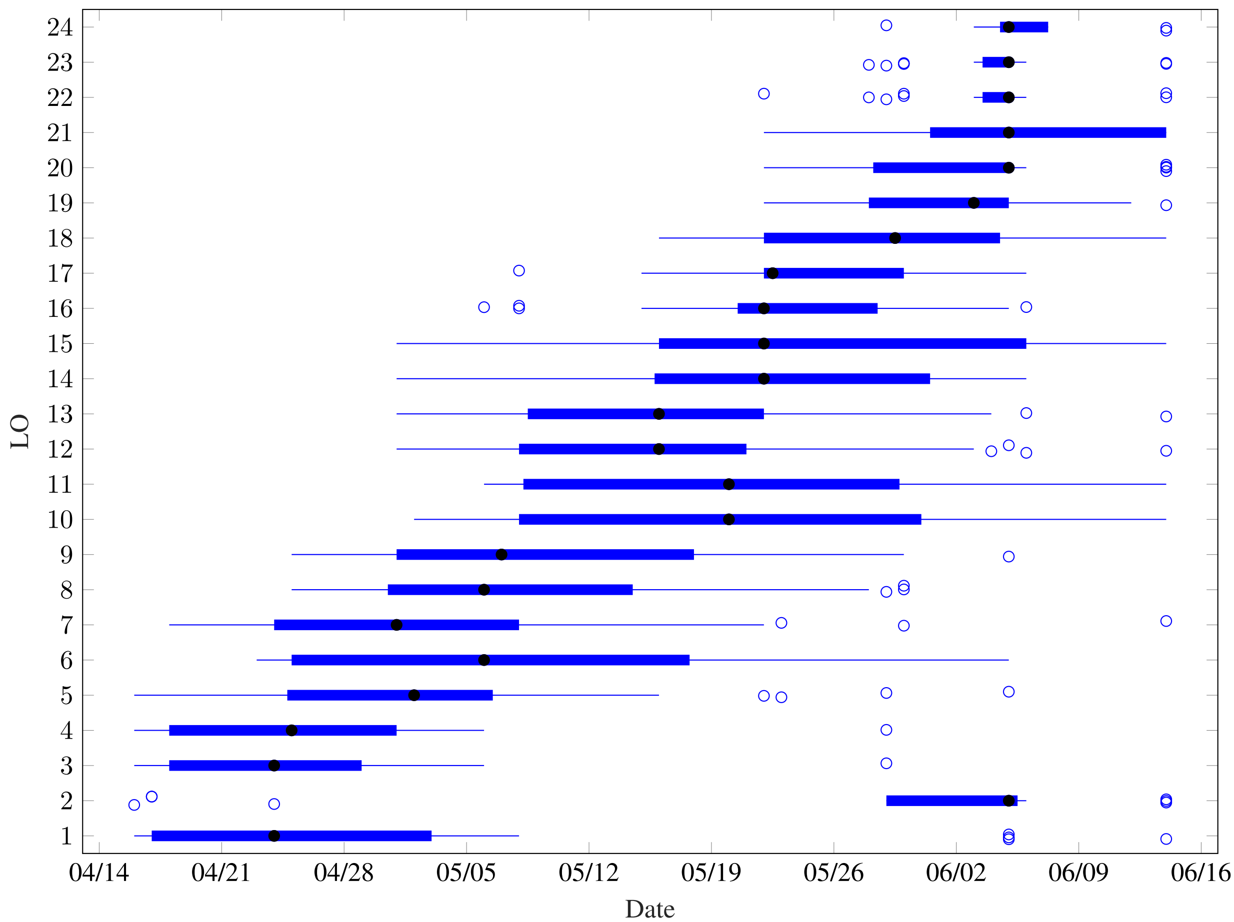

Fig. 4. Box-and-whisker plot showing the dates each learning objective (LO) quiz was first attempted by students in the SBG section. Black dots indicate the median first attempt date. The thick bars (boxes) indicate the interquartile range. The thin bars (whiskers) indicate the maximum and minimum first attempt dates, excluding outliers (which are represented by the discrete circles). The first day of the term was 04/02 and the final exam was on 06/14.

In the future, the authors plan to implement the structural changes outlined above and evaluate how those changes impact student learning gains. Simultaneously, the authors also plan to track the long-term effects of standards-based grading on retention of signals and systems knowledge. Finally, it would be interesting to run a statistical regression on the normalized gains exhibited by the two sections to try and control for externalities like grade point average, transfer student status, and other factors.

References

[1] R. J. Marzano, *Formative assessment & standards-based grading*. Solution Tree Press, 2011.

[2] B. Bloom, "Learning for Mastery.pdf," *UCLA CSEIP Evaluation Comment*, vol. 1, no. 2, may 1968.

[3] J. P. Lalley and J. R. Gentile, "Classroom assessment and grading to assure mastery," *Theory into Practice*, vol. 48, no. 1, pp. 28–35, 2009.

[4] I. D. Beatty, "Standards-based grading in introductory university physics," *Journal of the Scholarship of Teaching and Learning*, vol. 13, no. 2, pp. 1–22, 2013.

[5] J. M. Bekki, O. Dalrymple, and C. S. Butler, "A mastery-based learning approach for undergraduate engineering programs," in *Proceedings - Frontiers in Education Conference, FIE*. IEEE, 2012, pp. 1–6.

[6] L. Deslauriers, L. S. McCarty, K. Miller, K. Callaghan, and G. Kestin, "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom," *Proceedings of the National Academy of Sciences*, vol. 116, no. 39, pp. 19 251–19 257, 2019. [Online]. Available: https://www.pnas.org/content/116/39/19251

[7] K. A. Smith, "Cooperative learning: effective teamwork for engineering classrooms," in *Proceedings Frontiers in Education 1995 25th Annual Conference. Engineering Education for the 21st Century*, vol. 1, 1995, pp. 2b5.13–2b5.18 vol.1.

[8] A. Danowitz, "Group work versus informal collaborations: Student perspectives," in *ASEE 2017 Pacific Southwest Section Meeting*, 2017.

[9] K. E. Wage, J. R. Buck, C. H. G. Wright, and T. B. Welch, "The signals and systems concept inventory," *IEEE Transactions on Education*, vol. 48, no. 3, pp. 448–461, 2005. [Online]. Available: https://ieeexplore.ieee.org/document/1495653

[10] J. R. Buck, K. E. Wage, M. A. Hjalmarson, and J. K. Nelson, "Comparing student understanding of signals and systems using a concept inventory, a traditional exam and interviews," in *Frontiers in Education Conference*. IEEE, oct 2007, pp. S1G–1–S1G–6. [Online]. Available: http://ieeexplore.ieee.org/document/4418043/

[11] M. Nahvi, *Signals and Systems*. New York: McGraw-Hill, 2014.

[12] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *American Journal of Physics*, vol. 66, no. 1, pp. 64–74, 1998.

[13] J. R. Gentile and J. P. Lalley, *Standards and Mastery Learning: Aligning Teaching and Assessment so All Children can Learn*. Corwin Press, 2003.