Journal of the Scholarship of Teaching and Learning, Vol. 13, No. 2, May 2013, pp. 1 – 22.

Standards-based grading in introductory university physics

Ian D. Beatty[1]

Abstract: Standards-based grading (SBG) is an approach to assessment and reporting in which scores are attached to the specific learning objectives of a course, rather than to assignments or tests. Each score represents a student's mastery of that learning objective, and may change over time in response to evidence that her level of understanding has changed. SBG is increasingly popular in K-12 education, but has been poorly documented and studied in a university context. I explored the practicality and effects of using SBG in a moderately large university class by incorporating it into the two successive courses of an introductory physics sequence. Although design flaws and logistical difficulties plagued these attempts, most students responded positively to the basic intent and elements of the approach. Our experiences revealed likely implementation errors and suggested some wise design choices. More interestingly, I found that SBG foregrounds and forces us to confront some fundamental tensions present but latent within most or all teaching.

Keywords: grading, standards-based grading, assessment, mastery learning

I. Introduction.

"Standards-based grading" (SBG), also called "standards-based assessment and reporting," is an alternative approach to assessing, tracking, reporting, and grading student learning in a course. It has garnered growing attention in recent years, as K-12 schools seek grading methods consonant with standards-based curriculum frameworks and assessment systems (Guskey, 2001), and as reformers seek to avoid the drawbacks of traditional grading (Wiggins, 1998).

Despite its growing popularity at the K-12 level, SBG seems largely neglected within higher education. A notable exception is Rundquist (2011), who reported a very positive experience implementing a pure SBG design SBG in a small upper-level physics course, with the novel feature that all assessment evidence had to include the student's voice. Thus, students were required to demonstrate their mastery of learning objectives via oral exams in class, one-on-one discussions with the instructor, or the submission of "pencast" videos in which they narrated a proof or a problem solution as they wrote it out. While intriguing, such an approach is clearly impractical in a large-enrollment course; Rundquist admitted that his grading load was heavy with only nine students.

During the spring and fall of 2012, I implemented SBG in the two semesters of an introductory calculus-based university physics sequence. In this article, I describe my SBG design and implementation, summarize student reactions to the approach, reflect upon the successes and difficulties we encountered, and draw some general lessons from the experience. My aim is to stimulate conversation about the benefits and drawbacks of SBG in higher education, and to assist instructors who might be inclined to try SBG.

[1] Department of Physics & Astronomy, University of North Carolina at Greensboro

II. Background and Motivation.

In a typical university course, each student is awarded a score, grade, or point total for assignments, tests, and other course components, and her overall course grade is determined from a sum or average of her component scores. In SBG, scores are not attached to specific assignments and tests. Instead, the instructor identifies a set of learning objectives or "standards" for the course and the student receives a mastery score for each standard. Every such score represents how well the student has mastered the standard, based on evidence from one or more assignments, test questions, in-person interactions, or other sources. Thus, as her "result" on an assignment or exam, a student would receive a set of scores for the standards it addressed rather than a single overall grade or point total (Marzano & Heflebower, 2011). At the end of a course, her scores for all standards can be combined to yield an overall course grade if one is required.

In an ideal SBG implementation, a student's score on any particular standard indicates how well he's mastered the standard *at that point in time*. This score may increase over time as he demonstrates increasing understanding and skill (an expected result of learning), and it may decrease if new evidence reveals previously-overlooked flaws in understanding. Thus, his set of standard scores provides a real-time snapshot of his skill and knowledge. In reality, each standard can be assessed only intermittently, but students still have some opportunity to persevere on standards until they reach an acceptable level of mastery.

Philosophically, SBG is predicated upon three principles. First, feedback from assessments should be linked to specific learning objectives in order to help students know and target what they need to learn (Marzano & Heflebower, 2011). Second, students should be permitted to remedy deficiencies in their learning when an assessment reveals them (Dueck, 2011). Third, standard scores should communicate students' degree of mastery of the course learning objectives, and not be confounded with other variables such as effort, good behavior, or rate of progress (Brookhart, 2011; O'Connor & Wormeli, 2011). The second and third principles, taken together, imply that poor performance on an early assignment or test should not forever weigh down a student's course grade, but should be completely overwritten by later evidence that the learning objectives have ultimately been met. Some SBG implementations step back from these ideals, for example by including an effort/behavior component in the calculation of the final course grade, by using a decaying average calculation for the various measurements on each standard so that later scores do not completely overwrite earlier scores, or by setting a deadline for the first attempt at a standard in order to discourage excessive procrastination.

SBG is intimately related to *mastery learning* (Block, 1971; Bloom, 1985; Guskey & Gates, 1986). In conventional learning, students study a particular topic or skill for a pre-determined time window, and then move on to subsequent topics regardless of their progress. Over the span of a course, all students attempt the same set of topics, and are differentiated by their average progress on each. In mastery learning, students persevere on each topic or skill until reaching a specified threshold of competence. Over the span of a course, all students achieve comparable levels of proficiency on those they study, and are differentiated by how many topics they master. Mastery learning designs may be sequential, with students focusing on one topic at a time; or overlapped, with students continuing to seek mastery on older topics while encountering new ones (e.g., Leonard, Hollot, & Gerace, 2008). Proponents of mastery learning claim that solidly mastering a core subset of a course's material is preferable to incompletely learning the entirety, and meta-analyses of mastery learning implementations often show positive impacts on student outlook variables (see discussion in Leonard et al., 2008).

To date, much of the argument in support of SBG is theoretical. Of the evidence that exists, much is anecdotal (e.g., Erickson, 2011). However, given the criticisms leveled against traditional grading practices—such as conflated variables and uninterpretable grades, inadequately timely or specific feedback, minimal incentive to learn from mistakes, and damage to intrinsic motivation (see discussions in Docan, 2006; Wiggins, 1998)—considering alternatives such as SBG seems worthwhile. Finding the principles underlying SBG persuasive, I sought to appraise SBG in introductory university physics. Rather than investigating specific research questions, I conducted a preliminary exploration of the terrain, guided by the general hypothesis that SBG can be practically implemented in a moderately large university physics course and provide an overall positive experience for the students and instructor.

III. Narrative: Course Design and Execution.

A. Physics 1: Diving in Deep.

At my university, Physics 291 and 292 constitute a two-semester sequence of *Introductory Physics with Calculus*, taught in the spring (291) and subsequent fall (292) semesters. The course serves physics, chemistry, computer science, mathematics, pre-engineering, and biochemistry majors. In recent years, typical initial enrollments have been 50–60 students for 291 and 30–40 for 292. I first taught the sequence in 2011, with conventional grading and a highly interactive pedagogical approach centered on "clicker"-mediated classroom discussion (Beatty, Gerace, Leonard, & Dufresne, 2006; Beatty & Gerace, 2009). I taught 291 again in 2012, aspiring to a full, pure SBG scheme but otherwise altering the curriculum and pedagogy as little as possible in order to observe how the switch to SBG altered the experience.

Initial Course Design. My 291 SBG implementation design had three components: a list of standards, a grading scheme, and an assessment plan. I retained the syllabus of topics and approximate schedule of coverage from the previous year, which organized sixteen chapters into five units. To develop a list of standards, I studied the textbook and identified specific skills or competencies that could be articulated, understood, and assessed more or less independently, and that seemed to form the essential core of the material. I sought a compromise between overly coarse-grained, general standards (which would be little better than topic headers) and overly fine-grained, specific standards (which could be logistically and administratively impractical). Following the general wisdom for SBG practice, I phrased each standard as an "I can…" statement in order to frame students' learning as the development of capacities rather then the retention of declarative knowledge. Before the course began I articulated 28 standards for the four chapters of the first unit, corresponding to an average pace of 3.5 standards per lecture meeting. Table 1 lists the standards for Chapter 2 as an example.

For the grading system, I chose a four-point scale to represent each student's mastery on each standard, as shown in Table 2. A student's score for a standard could change over time due to evidence from reassessment attempts, with a later score completely overwriting any earlier ones. At the conclusion of the semester, a student's latest scores for all standards would be averaged, combined with the lab score (see below), and the result mapped to a letter grade such that 4.00 yielded an A+, 3.75 an A, 3.50 an A–, and so on, with 1.00 or lower yielding an F. The syllabus asserted my right to adjust the thresholds, should the system prove too lenient or harsh in practice.

I scheduled five unit exams as the primary assessment of standard mastery. I intended brief in-class quizzes every few days on some of the more basic standards, hoping to reduce the

Table 1. Learning standards for Chapter 2, "Kinematics in One Dimension."

Standard
1. I can use the uniform motion model to analyze physical situations.
2. I can determine or reason about an object's instantaneous velocity at various instants during its motion, based on position vs. time information.
3. I can determine or reason about an object's position and displacement, based on velocity vs. time information.
4. I can use the constant acceleration model (including constant-acceleration kinematics formulae) to analyze physical situations.
5. I can explain and analyze cases of free-fall using the constant acceleration model.
6. I can use the inclined plane model (a special case of the constant acceleration model) to analyze situations involving a sloped surface.
7. I can determine or reason about an object's instantaneous acceleration.
8. I can produce, interpret, and interrelate graphs of position vs. time, velocity vs. time, and acceleration vs. time for various motion scenarios.
9. I can use basic calculus (derivatives and integrals) to interrelate expressions for position, velocity, and acceleration as a function of time.

number of standards covered by the exams. My plans for reassessment after unit exams were less clear. I intended to reassess a few of the most difficult earlier-unit standards on later-unit exams. I also hoped to let students reassess some standards via one-on-one oral quizzing outside of class, but due to the large enrollment I was reluctant to promise this. I reserved the final exam for last-chance reassessment. In the syllabus and in class, I stated that reassessment was not a guaranteed right, that not all standards would be available for reassessment, and that reassessment must be earned by demonstrating remedial work done (such as re-working an earlier exam for homework and articulating what learning had resulted).

Table 2. Rubric for assigning a mastery level score to each standard.

Mastery level	*Score*
Got it solidly!	4
Mostly got it. (Understand the idea well, but sometimes make small mistakes or get confused by subtleties.)	3
Making progress. (Definitely understand it somewhat, but still have misconceptions, gaps in knowledge, or make serious mistakes.)	2
Starting out. (Know a wee bit about this, but not enough to really use it for anything.)	1
Nothing yet. (Have no idea yet what this is.)	0
Not yet assessed.	—

The course included assigned textbook reading, corresponding sections of the accompanying *Student Workbook*, and homework problems drawn from the textbook and automatically evaluated by an online homework system. Although I stressed that completing these earnestly was "essential" to learning, they were not checked, did not contribute towards the course grade, and had due dates that meant nothing more than "keeping up with the course."

Using points and grades to coerce student behavior clashes with the SBG philosophy, so I relied on students appreciating the connection between doing the work and performing well on exams.

This course includes a one-credit laboratory section, with a single overall grade given to the combined lecture and lab. Thus, integrating the lab into the SBG system was essential. My undergraduate teaching assistants and I developed a list of 35 lab standards, divided into three categories: measurement and analysis, scientific communication, and experimentalism. Lab standards used the same four-point mastery scale as lecture standards, and at the end of the course would be averaged into an overall lab score that would then be combined with the overall average lecture score (weighted 1:3), with the result mapped to a letter grade as indicated above. Lab standards would be assessed through a variety of means—lab quizzes, lab notebooks, lab reports, and TA observation/interaction—with the details to be worked out as the semester unfolded.

Initial Course Execution. As the course began, students revealed no particularly strong reaction to the SBG aspect. My general teaching practice is to meta-communicate heavily about my pedagogical principles and tactics, so I spent considerable time explaining and justifying SBG, clicker question discussion, and group whiteboarding. Though few students appeared to grasp the specifics, all seemed amenable to the general idea. The presence of "reassessment" seemed reassuring to many, even though details were lacking. I suspect many interpreted it as "free second chances."

During the eight lecture meetings of Unit 1, I gave two quizzes, but scores were too low to omit the corresponding standards from the first unit exam. (I gave only one more quiz during the remainder of the term.) That exam assessed 22 of the 28 Unit 1 standards; one more was deferred to the second exam, and five were quietly dropped from the course. Even so, the exam was too long for the three hours provided.

Scores on the first exam were low. Of the 1254 scores (57 students times 22 standards), 22% were 4s and 14% were 3s. Eighteen students earned a 3 or 4 on at least half the standards, and eleven of these earned a 4 on at least half. Thus, most students needed reassessment on a significant fraction of the material, and a majority needed reassessment on almost everything. Had I given overall exam grades by averaging all of a student's standard scores, the course median would have been 1.95, corresponding to a low C–.

I generally teach demanding courses using unorthodox methods, and require an unusual level of engagement and responsibility from students; thus, the first exam is very often a wake-up call, with a lower average than subsequent exams. Additionally, this exam had been longer and harder than intended. As a result, I was not overly worried by these poor results, and I reassured the students that with some recalibration on all our parts, we had no need to panic.

However, the results of the second exam, three weeks later, were similarly dismal. Many students were becoming openly hostile or discouraged, and the class mood was souring badly. I realized that the course needed major intervention. After a frank discussion in class, I posted an extensive online questionnaire for students to complete anonymously. I motivated it by saying that I was seeking input in order to make significant improvements to the course.

The questionnaire contained 50 items: 23 short-answer responses, 21 scale ratings, and 6 numerical responses. 44 students started it and 34 completed it, although many who completed it did not respond to every item. Six of the items directly requested students' opinions about aspects of SBG, and a few others also provoked comments about the grading scheme.

Mid-Course Adjustment. In response to the survey responses, in-class discussions, and attendant one-on-one conversations with students, I made three major modifications to the course design, which I announced during spring break (the approximate midpoint of the semester). The first was to include more worked problem examples and short lecture segments in class, in response to complaints that clicker-based discussions showed them what they didn't understand but "left them to learn the material on their own." The second was to change the course exam schedule by introducing two "reassessment exams" for Units 1 and 2, displacing the Unit 3 exam until later in the semester, abbreviating Units 4 and 5 due to schedule slippage, and merging the Unit 4 and 5 exams into one. The third modification was to designate fewer, broader standards for units 3–5.

The re-exams for Units 1 and 2 occurred on the first and second Wednesday evenings after spring break, respectively. Both consisted of a set of problems analogous to the original unit exam problems, each targeting one or a few related standards. Students were given in advance a list of the standards targeted by each problem, though not the specific problems. During the exam, they could choose which ones they wanted to try. Once they took and read a problem, they would be scored on it no matter how well or poorly they did, with those scores replacing the corresponding scores from their prior exam. In other words, either they believed they had improved their understanding on those standards and were willing to bet on themselves, or not. I dropped as logistically unmanageable the requirement that students rework and turn in the first exam in order to qualify for reassessment.

On both re-exams, almost all students chose to reassess on at least a few problems, and many tried most or all. Overall, scores increased significantly. For Unit 1 standards, the median of students' average scores was 2.91 (up from 1.95), on track for a final course grade between B– and B. 26 (up from 11) students earned a 4 or better on at least half of the 22 standards, and 39 (up from 18) earned a 3 or better on at least half.

As a result of these mid-course adjustments, the class mood improved noticeably. The primary drawbacks of the reassessment exams were the degree to which they distracted students from learning the Unit 3 material being taught concurrently, and the time they required of me to create and grade. Consequently, I told the class that giving reassessment exams after each future exam was impractical. Instead, I provided a realistic practice exam before each of the remaining two exams, since familiarity was likely one reason students improved on the re-exams. I also used the final exam period as a reassessment opportunity for the most-needed standards from Units 3–5, following the same choose-your-problems approach as for the Unit 1 and 2 re-exams.

The remainder of the course passed without major incident in a blur of practice exams, unit exams, and the reassessment final. By the end, the course had assessed 69 lecture and 35 lab standards. After submitting final grades, I created a second anonymous questionnaire soliciting students' reflections upon the mid-course adjustment and the course as a whole, repeating some of the questions from the mid-course questionnaire. I sent the invitation and questionnaire link by email after students had left campus, and received only 12 responses despite sending follow-up reminder messages.

B. Physics 2: Simplifying Greatly.

The following fall, I taught Physics 292, the second course in the sequence. Of the 46 students who had completed Physics 291, 23 continued into 292, joined by nine additional students for a starting roster of 32. My SBG design for the first course had proven to be barely manageable, and due to a double teaching load in the fall term I had to drastically simplify my SBG implementation for the second course. In hindsight, this simplification resulted in a fatally thin, incomplete version of SBG.

Initial Course Design. I made six primary changes. First, I established fewer, broader learning standards: only 25 across the lecture and lab. Examples include “I can describe the particle-ray model of light, define and explain its elements, and cite empirical evidence to justify it”; “I can deploy the particle-ray model of light to explain phenomena, qualitatively and quantitatively analyze novel physical systems, predict their behaviors, and calculate the values of physical quantities”; and “I can interpret calculus and connect it to physical situations, for example by constructing integrals to mathematically represent infinite sums of infinitely small contributions in a physical context, or by interpreting the gradient of a scalar field.”

Second, I replaced the large online homework sets with a much smaller number of challenging, extended homework problems and mini-projects—some from the textbook, and some of my own creation—to be turned in on paper. I provided written feedback on these, but no grades, and merely said that they would be “taken into account” when determining standard scores from exams.

Third, I included only two exams in the course: one midterm and one final.

Fourth, I allowed students to take each exam twice: once as a three-hour closed-book individual exam, and over the subsequent days as an untimed, open-book, collaborative take-home exam. (This is a strategy pioneered by my colleague William Gerace. We have found that it dramatically increases student learning from the exam process, and also helps fuse the class into a tightly-knit peer support cohort.) I assigned scores for each standard by looking at a student’s in-class and take-home responses together and inferring what level of mastery they revealed. In practice, the in-class response usually determined the score, with a notably stronger or weaker take-home response nudging that score up or down.

Fifth, I provided no reassessment mechanism for standards assessed on the exams. Except for standards assessed in lab (through a mix of quizzes, hands-on practical challenges, lab notebooks, lab reports, and instructor observation), the exams provided the one and only score for each standard (with an occasional upward nudge in response to unusually convincing work on a homework problem or relevant lab report).

Sixth, I instituted a “brutally strict attendance policy.” Three lecture absences or one lab absence put a student “on probation” (unofficially and only for purposes of this course), which means I’d give the student a dire warning and try to discuss the causes with them. Three more lecture absences or one more lab absence would, in principle, automatically cause the student to fail the course. In practice, I planned to offer exceptions to students who would promise to reform. My intent was to establish an expectation of 100% attendance and have some leverage to lean on students with an attendance problem, without doing violence to the spirit of SBG.

Course Execution. The course ran as designed, with only minor adjustments during the term. Homework completion was spotty, though it improved after the mid-term exam when I commented that several students had seen a slight improvement in one of their mastery scores based on their homework. I had great difficulty providing timely feedback on homework submissions. By the end of the course, the mood of the class seemed generally positive, with a highly interactive classroom dynamic, good discussion, and frequent jokes and laughter. Despite this, the course evaluations completed during the last class meeting contained many negative comments and lower ratings than I have previously received.

IV. Results.

This section summarizes students' reactions to the SBG aspect of the two courses (based primarily on their responses to the two questionnaires in the first course and their end-of-course evaluation forms), as well as my experiences as an instructor implementing SBG (based on notes I made during the courses and post-course reflections). Although this article focuses on SBG, the courses contained many other unusual elements, such as clicker- and whiteboard-based active engagement tasks during lecture, a strong focus on deep conceptual understanding and transfer, labs taught by undergraduate teaching assistants and focusing on open-ended experimental challenges rather than well-defined procedures, and take-home second chances on exams. Students are generally poor at discriminating between such things when they react and opine, so conflation is likely. Also, since questionnaire completion was voluntary and anonymous, some self-selection bias is possible. This is especially true for the post-course questionnaire, which had a response rate below 25%.

A. Student Opinions.

In general, many but not all students liked the SBG approach in Physics 291 despite the difficulties we encountered. Typical comments on the mid-course and post-course questionnaires and the course evaluation forms include:

> [I liked] The standards based grading
> I love the grading system.
> I love it, but sadly I know its hell on you when grading.
> the standards based grading is a little weird and trying to get use to it. may not like it because its different but i am trying to figure it out.
> [I dislike and would want to get rid of] Standard base grading

The questionnaires distinguished between three aspects of SBG: attaching scores to learning objectives rather than assignments, using grades only to represent achievement of learning objectives and not to coerce behavior, and permitting reassessment to change standard scores. For the first aspect—attaching scores to learning objectives rather than assignments—both the mid-course and post-course questionnaires asked:

> This semester, I'm trying a new grading system where I try to indicate and record how well you've mastered each "skill" or "piece" of the topic, rather than simply giving points and adding them up on each assignment. I know it's a bit more confusing and harder to answer the question "Overall, how well am I doing?" Aside from that, how do you like this approach?

Of the 32 responses on the mid-course version, 20 chose "I really like it," 6 chose "I like it a little," 2 chose "I'm indifferent; doesn't make much difference," 3 chose "I slightly dislike it," and 1 chose "I really dislike it!" On the post-course version, the corresponding response counts were 6, 3, 0, 1, and 0 (10 total). Representative comments are:

> I think that it makes it easier for a student to assess what he/she needs to work on. I know that for me, depending on the scores I receive for an exam, I will go back and decide whether or not I understand that standard and decide if I need to work on it more

> It kinda lets you know what you are golden on and what you suck at. It tells me what I need to look back at. If I bomb standard 2.1 but ace standards 2.2-2.4 then obviously there is something that I'm not getting from the earlier stuff. Using this approach I can go back and specifically target what I need to study much easier than if I were just handed a paper back with a 93% scrawled at the top in some undergrad/grad student's handwriting.

> I feel like it puts alot more pressure because there are many many standards to learn and each one could potentially mess up your score. If you try to reassess and don't do well then you get the worse score and there's not another chance to fix it so it puts alot of pressure instead of just getting an overall percentage of what you understand.

Following this, both questionnaires posed the open-ended prompt "Any other thoughts about scoring 'by standard' (by learning objective) rather than simply counting up points for each assignment?" 17 students offered comments on the mid-course version, and 7 did so on the post-course version. I aggregated those 24 aggregated responses, and coded them according to their essential point(s) with an emergent coding scheme. A few comments contained multiple distinct ideas and were multiply-coded. Of the positive comments, 8 expressed a non-specific liking (for scoring by standard); 6 appreciated the formative value of precise feedback; 2 appreciated the ability to focus their studying on learning, rather than jumping through hoops for points; and 2 thought the resulting assessment were more accurate indicators of understanding. Of the negative comments, 3 thought the approach was more stressful; 3 thought the standards and/or their scoring levels were unclear; 3 thought the scoring system was less precise than traditional percentage scores; 2 thought the effect was less motivating; 1 thought it was more demanding; and 1 disliked the possibility of having a standard score go down when reassessed.

For the second aspect of SBG—using grades only to represent achievement of learning objectives and not to coerce behavior—both questionnaires asked:

> As a matter of principle, I don't use points and grades to coerce you into doing things for your own good (like coming to class, completing the workbook, or doing the homework). I think grades should indicate how well you're learning the course objectives, and nothing else. I think it's insulting to treat students like they won't do any work that's not directly graded, and that they're not mature enough to recognize that they need to attend class, read the text, do the homework, etc. in order to get a decent grade on the eventual exams. Do you like this policy?

Of the 32 responses on the mid-course version, 18 chose "I very much like it," 7 chose "I kinda like it, I think," 2 chose "I'm neutral or ambivalent," 5 chose "I dislike it a bit," and none chose "I really don't like it." On the post-course version, the corresponding response counts were 4, 4, 1, 4, and 0 (10 total). Representative comments are:

> I study better figuring out the actual topics and focusing on those instead of completing homework just for the grade. the work book can help some times but i am a visual and auditory learner. I find videos online to drive the topics home. i think that's more beneficial than the work book for me.

> I have a hard time doing work that is not helpful to me learning regardless of whether or not it's graded. I'm here to learn, not to make straight A's (though obviously that would be nice) by doing pointless work that doesn't ultimately contribute to my education. I really appreciate that you trust us to make the decision of what is and isn't helpful for us and I really like your grading policy. I wish every class could be like that.

> I like being given several options / approaches and being able to find what works for me and not having to do something that isn't helping just for a grade, however there was obviously an adjustment period. At the beginning, attempting to go through all of the material to find what was going to help the most was tedious, but in the end worked out.

Despite generally liking this freedom, students reported mixed success at self-motivating themselves to complete ungraded work. The mid-course questionnaire asked, "How good are you at motivating yourself to do work well and on time without grades and deadlines to put pressure on you?" Of the 31 responses, 5 chose "no trouble, don't need grades and deadlines," 16 chose "a bit of difficulty, but overall do okay," 6 chose "struggle quite a bit, do significantly less work," 1 chose "won't do it if not due and graded. Period," and 3 chose "Something else…" In the free-text response for the "Something else…" option, one student referred to difficulty prioritizing physics over other courses with graded deadlines, and another claimed he or she "ALWAYS" does the work before the exam, but probably doesn't leave enough time to it adequately.

Several students wanted some kind of grade credit or reward for effort, separate from the benefits of learning the material and doing better on exams:

> I believe attendance policies like that, per me… they don't coerce me into doing well, I think you will find most students at this level truthfully want to do well. However, it's more about maintaining confidence, keeping your chin up, if all we have are tests, and everyone keeps bombing them (mostly) it will wreck moral. Additionally, no curve (not that I believe in them because I don't either) kinda does not help replace the void where typical "hoops" are used to pad grades.

> I am a self-motivated student and I don't need to be "bribed" with grades every step of the way. What I do need, is the feeling that my hard-work and dedication will pay off. I put more effort into this one class than into any of my other classes, and yet I am getting worse results than I ever have in my entire collegiate career. It would be nice to see some pay-off from that work, whether from homework, or quizzes or doing the work book. Perhaps those who complete assignments can have their work factored in when grading the exams, and those who don't do the hw/workbook will only be graded on their exams.

> I know you don't like bribing us with grades, but getting credit for getting all of our assignments done would be nice.

A related theme was that many students wanted some lower-stakes assessments in addition to the exams, both to provide earlier feedback about how well-prepared they were for the exams, and to reduce the stress of having "everything ride on exams."

> Essentially, the only time that our attempts "count" is during the exam, and that is very stressful.

Both the mid-course and post-course questionnaires included the open-ended prompt "Any other thoughts about using or not using grades and deadlines to 'make' you do the work of learning in a timely fashion?" Of the 23 relevant comments aggregated from the two rounds, 4 expressed a non-specific positive response to course's approach; 2 liked the freedom to choose what study tactics worked best for them; and 2 liked the flexibility to choose when to do physics work. On the other hand, 5 disliked the fact that physics studying took a back seat to other courses with homework deadlines and grades; 3 claimed they need deadlines to make them get work done; 2 missed the formative feedback that graded homework provides; 1 wanted extrinsic rewards for homework, 1 disliked the fact that the lack of any homework contribution towards grades made the exams higher-stakes, 1 wanted to see more direct payoff for doing homework, and 1 simply found the approach uncomfortably unfamiliar.

For the third aspect of SBG—permitting reassessment to change standard scores—students generally appreciated the "no history" aspect of SBG, where later evidence completely overwrote earlier evidence of mastery for each standard:

> It is the best method of evaluating student performance, in my mind, that I have ever encountered. Though the amount of reassessment we have is extreme now, I have always felt that it is wrong to penalize a student for doing poorly early in the semester and then succeeding very well by the end. The standards-based system takes the high-school "point system" mentality away from the learning experience so that you can focus more on getting the material down, and less on "how do I get a 90 or better"

> I like that we can reassess if we do poorly on an exam. A lot of the stress of grades comes from doing poorly on one exam and then having to hit a home-run on subsequent exams or the final.

> Also, the reassessing helps if you are having a bad day. Most teachers just drop one test if you have a bad day. But that keeps the student from learning from it.

On the other hand,

> I think the amount of reassessment is unnecessary. I think in the future, it would be fair to state that it was not a guaranteed right, but that should a student demonstrate significant improvement or effort, it may be granted as a privilege. And then try to include as much and as broad a selection of content on the final as possible, and use that as the "primary" opportunity to demonstrate that knowledge was, in fact, gained…

Despite my fears to the contrary, students claimed that the prospect of reassessment did not impact how well they prepared for the initial exams. The mid-course questionnaire asked:

> How has the possibility of "reassessing" standards later in the course, with later scores replacing earlier scores, impacted how you prepare for exams?

Of the 31 responses, 9 chose "It hasn't changed what I do at all," 18 chose "I'm a bit more relaxed and less stressed, but I still prepare and study just as much as I would if there weren't any reassessments," none chose either "I prepare and study a lot less than I would without reassessments, because I know these exams don't matter all that much" or "I don't bother preparing for the exams at all, since they don't really 'count'," and 2 chose "Something else…"

B. Student Difficulties.

In their questionnaire and course evaluation comments, students reported several common difficulties with the Physics 291 implementation of SBG. One difficulty was operationalizing standards:

> Also if the standards were explained more and feed back was given as to what could be worked on more to get a far better understanding would make this way of grading far better.
>
> The labs also had standards that were not clearly explained…

After the mid-course adjustment, I made a point of linking standards to specific textbook sections, and of listing representative problems for each standard.

> I liked how each standard had corresponding questions to it so we knew exactly which ones we needed to practice… I also liked [that] we either knew specifically or we generally knew which standards would be tested. That really helps in the preparation of exams.
>
> [GOOD changes to the second half of the course included] Having practice tests available with answers before taking the exams, as well as showing what each standard meant by giving page numbers from book as well as suggested practice problems.
>
> Instead of giving the practice exam key… maybe work one problem from practice exam per class period. Sometimes it is hard to know exactly what a standard wants until seeing a problem. Giving book problems for specific problems was good, but I feel I benefitted more when you wrote the questions.

Related complaints involved dislike of or difficulty interpreting the 1–4 mastery score scale.

> Another concern that I have is what a 4, 3, 2, or 1 for each standard is defined as. It was not clearly explained how exactly the numbers were given.
>
> If you want to grade by standard, then make it clear what each standard means. Answer each test with a 4, 3, 2, 1, and 0. Then compare the answers of our test to each test. Do we give a 4 pt answer or do we give a 2 pt answer? They use very similar type of grading in the essay writing portion of the SAT's.

> I think there should be more partials like 2.5 or 3.25 to really pinpoint our positions and if it's split into sub quarters determining our grade would be much easier because there would be a letter grade that corresponds to each quarter (including pluses and minuses).
>
> I do really enjoy the idea, but I have two main stipulations with it. The first is that this scale has the potential to be susceptible to more subjectivity than a traditional 0-100% grading scale. The area between a two and a three or a one and two can get rather gray sometimes…
>
> It seemed like there was kind of an overly-significant drop-off (4 to 3) for any small mistake, but in general it does give a pretty nice picture.

Several students requested more frequent feedback.

> I like being graded on assignments and having homework that counts for a grade. To me, I don't see this as carrot-dangling or demeaning or anything like that. I see grades as a running indicator of how well I'm doing in a class, and how well I am prepared for exams.
>
> I think there should be more deadlines and grades because if a student didn't do so well on the workbook then you can help them figure out what they did wrong and then they will know how to do it on the exam.
>
> At the same time, however, it is stressful that these standards, the only real "grades" we have in the course, are tried almost exclusively on the exams. It would be nice if we could knock out some of the standards with some sort of consistent homework or quiz schedule. Personally, I think that this would encourage and enable me to set a better pace for learning the material.

Often, students need more than feedback on specific standards or topics: They need an overall assessment of how well they are doing in the course and what sort of final grade they are headed towards. Some students are concerned that they are not ready for the course, and want an overall grade estimation in order to decide whether to drop the course. Others need a partial-term grade reported for athletic eligibility or fraternity/sorority reasons. Yet others simply want to know whether their current approach to the course is adequate or needs to be rethought, and cannot interpret an assembly of standard scores well enough to answer the question "Am I doing okay?"

C. Instructor Experience.

In the previous section, I summarized students' most significant reactions to my SBG implementations. In this section, I summarize major elements of my experience as instructor.

At the conclusion of the course, I followed my usual process of calculating prospective course grades according to the formula published in the syllabus; generating a list of students ordered by overall score; using my personal knowledge of several specific students' degree of understanding as points of reference to adjust the mapping from mastery score averages to letter grades until most grades seemed appropriate; fine-tuning the cutoff thresholds between letter-grade bins to avoid dividing students based on meaninglessly small numerical differences; and, finally, adjusting the grades of any specific students with relevant extenuating circumstances.

During this process at the end of Physics 291, I found that although I did not have to re-tune the mapping drastically, the grading system made the A-level grades perhaps a bit too difficult to earn and a respectable C or C+ rather too easy to attain. Many of the students in the C regime were ones I knew to be relatively clueless. At the end of Physics 292, I again had to bump the top students up a bit and the bottom student down a touch to align with my sense of what grades ought to represent. In the future, I could alter the advertised rubric, or perhaps be stinger about awarding 1s and 2s and more generous with 4s while grading. However, my personal observation is that this particular cohort was light on A-level "star" students, which leaves me unsure of how well the grading scheme would work at that level. Similarly, most D-level students did not survive to Physics 292, so my data on how well my second implementation worked with low-end students is weak.

I discovered that I very much like the process of grading exams within SBG. In fact, that is perhaps my favorite aspect of the entire experiment. SBG allows me to scrutinize students' exam responses with one question in mind: "How well have they shown that they understand X?" The four-point mastery scale allows me to answer that question quickly. I don't need to ponder how many points I should take off for each error, or wonder how many points a sound solution to an incorrectly interpreted problem is worth. Given a suitable set of standards and exam problems, grading is relatively fast, easy, and communicative.

While grading, I did discover a need for more miscellaneous, crosscutting standards. For example, a standard for units and dimensions fluency would allow me to ding students for sloppy units and emphasize their importance, while still giving full credit for understanding the physics of a problem. Similarly, standards for using conventional notation, performing algebra reliably, and sanity-checking results could be helpful. Instead, my policy was to drop students from a 4 to a 3 for such errors, conflating imperfections of understanding with weaknesses of process.

From my experiences in prior courses, I have come to believe that the take-home open-book collaborative exam retake is perhaps the most powerful instructional innovation I have ever tried, and in neither semester did I find a satisfactory way to integrate it with SBG. In non-SBG courses, I simply average a student's in-class and take-home scores, which gives the take-home all the gravitas of a full exam. This is part of what makes the approach successful. (Since most students earn high scores on the take-home, and the class' take-home scores show little variance compared to their in-class scores, averaging the two rescales but does not significantly reorder student scores. Since I control the overall course grading scale, I can adjust for this effect. Thus, the practice is instructionally beneficial without being unfair.) During Physics 291, I indicated that a correct redo of the exam would be required for eligibility to reassess, and then backed off from that requirement for logistical reasons. During Physics 292, I found myself making difficult and unsatisfactory inferences about a student's understanding based on a weak in-class attempt and a strong take-home quite possibly copied from a peer's solution.

I discovered a steep learning curve to creating good SBG assessments. One challenge was "factoring" the standards: developing exam questions that isolated specific standards. While grading, simply giving half-credit to a partially correct problem is inadequate; I must identify what component knowledge or skill was lacking, and score the corresponding standard appropriately. I discovered that most "interesting" physics problems—meaning problems that are not simple isomorphs of the examples used in class or the textbook, and that demand some robustness of knowledge rather than memorized procedures—require multiple skills and knowledge elements. Targeting more than one standard per problem (or section of a problem) does more than slow down grading: It also causes difficulty for reassessment, when one wants to let students reassess on specific standards.

I reacted to this in the second course by choosing very coarse, broad standards that encompassed almost any problem within a broad swath of content. Basically, my standards represented topic areas rather than specific competencies. This made exam creation and grading easier, but was unsatisfactory because it sacrificed much of the benefit SBG can offer by providing specific feedback to students and supporting targeted remediation and reassessment.

Within typical courses, few of us try to test every possible skill and piece of knowledge covered. Instead, we rely on a representative sampling to gauge student learning and determine grades. A second challenge I discovered with SBG was that a fine-grained standard system forces me to assess every bit of content and skill articulated in the standards list, rather than sampling a subset of learning objectives in order. This can lead to impossibly long or impractically frequent tests.

During both semesters, my solution (in addition to giving painfully long exams) was to quietly ignore some standards, neither assessing them nor including them in grade calculations. In the future, I would want to leave them in the official list of course standards, so students know that they are "responsible" for that material. I might explicitly state that students will be assessed on many but not all standards.

Wrestling with this did highlight a fundamental tension in instruction: We want to teach, and have students take seriously, more things than we can practically assesses. This problem is not unique to SBG, although the need to assess and reassess all standards exacerbates it.

While SBG considerations dominated my assessment and grading practices, I found that they had little effect on my actual teaching. I occasionally displayed the wording of a standard during class, in the context of clicker questions and whiteboard problems relevant to that standard, but otherwise I didn't use course standards to direct, organize, or frame class activity. In hindsight, that may have undermined the impact SBG could have had on student learning.

V. Reflection and Synthesis.

This section will distill general lessons about SBG from my experiences as summarized above. First, it will present three fundamental tensions that any instructor seeking to implement SBG will have to confront. Following that, it will list some suggestions that should help an instructor avoid or minimize at least some of the potential difficulties.

A. Three Fundamental Tensions.

Reassessment. I claim that reassessment is the heart of SBG: It is the most difficult aspect of SBG to implement sustainably and to get right, but is also the most crucial. Reassessment gives SBG the power to steer student learning, encourage perseverance, and improve learning outcomes. When I eliminated reassessment from my SBG design for Physics 292, I found that the result was little different than a traditionally-graded course.

On the other hand, the logistics of reassessment pose many challenges, as my Physics 291 course narrative demonstrated. These challenges include allocating sufficient class time, developing multiple assessments for each learning standard, managing the grading load, and ensuring that students take initial assessments seriously. An additional difficulty is getting students to pay adequate attention to new material while concurrently preparing for reassessments of older material. Although the introduction of two reassessment exams after spring break may have rescued the first course and helped many students avoid disastrously low grades, it significantly distracted from Unit 3, slowed the pace of coverage, and ultimately caused me to abbreviate or cut several topics from later units. As a student commented on the Physics 291 post-course questionnaire:

> It wasn't really a change exactly, but the entire last half of the course was plagued with a problem of multi-tasking. I could barely get myself to care about the new material constantly being introduced, when I was faced with what could be my last assessment on earlier standards. I realize this wasn't entirely intentional, but it was my main difficulty.

I see this as more than a mere logistical difficulty. Rather, it reveals a fundamental tension between students' need to spend time and attention revisiting "past" material that they have not yet solidly mastered, and our need to drive course coverage forward at a rate sufficient to address what must be addressed in the time allotted.

Grain Size and the Dead Frog Problem. A second tension arises in the choice of learning standards for a course. This is a critical step, since the choice of standards will shape assessment requirements and (hopefully) direct students' learning efforts. The tension is one of "grain size": choosing many fine-grained, specific standards vs. choosing relatively few, coarse-grained ones encompassing much subsidiary knowledge and skill.

Fine-grained standards help students know exactly what they should be learning, can be linked very neatly to highly-targeted exemplar problems that help operationalize them, provide specific, diagnostic feedback to support remediation, and allow efficient reassessment of only what needs reassessing. On the other hand, fine-grained standards are inevitably numerous, which creates administrative headaches and can be difficult for students to digest and track. Creating assessment questions that target single standards without being trivial or repetitive is difficult. Also, a large number of explicit standards inhibits an instructor's ability to assess efficiently by sampling a subset of the covered knowledge. These are all difficulties I experienced during Physics 291.

Coarse-grained standards have the opposite benefits and drawbacks. They are easy to track, allow great flexibility in exam questions, and enable sampling for efficient assessment. However, they tend to represent topic areas rather than well-defined competencies, conflating multiple component skills and knowledge elements. As such, they are not easily operationalized, provide little specific guidance for students seeking to remedy deficiencies in their knowledge, and force reassessment of large swaths of content at a time. I encountered or hid from all of these difficulties during Physics 292.

Finer-grained standards would seem more in keeping with the spirit of SBG, suggesting that we aim for as many standards as we can practically manage. However, grain size is not just a practical issue. The more finely we factor a subject into distinct learning objectives to be separately assessed, the more strongly we encounter what I call the "dead frog problem." The name comes from a witty response a friend once gave when he made a joke and I, failing to understand it, asked him to explain. He said, "A joke is like a frog. You can dissect it, but it dies."

I maintain that physics, too—and almost any other academic subject—is like a frog. Physics is more than a collection of distinct knowledge bits and skills, and critical aspects of "knowing" and "doing" physics do not reside in any specific bits. Rather, much of physics expertise resides in having a richly interwoven, multi-scale mental map of how the pieces of content interconnect. More resides in the capacity to identify which pieces of content are appropriate to any given situation and integrate them at need. I fear that SBG fails to be faithful to the interconnected nature of physics or amenable to assessment via rich, authentic, integrative problems and tasks (Wiggins, 1998).

One might imagine having a set of fine-grained learning standards for the "elements" of physics and augmenting them with additional standards articulating higher-level skills and perspective. I am not convinced that one could adequately capture these higher-level pieces in standards explicit enough to be reassessed fairly and to be grasped and operationalized by students. Additionally, such standards are likely to be broad and nebulous enough that factoring them into valid, repeatable assessment items seems daunting. While SBG may be a good approach to teaching lower-level knowledge and skills, a complete course may need to fuse it with another system for assessing and grading higher-level learning, one that honors the value of "putting it all together" while being consonant with the mechanisms and core principles of SBG.

Dissecting the physics content into focused, factored learning objectives can do more harm than just preventing us from assessing the holistic aspects of learning. It also plays into students' inclination to reduce a course down to an explicit list of relatively simple, clear things to learn how to do: a catalog of specific kinds of problem to solve in a standard way. I see a great danger that SBG could unwittingly and implicitly, but strongly, frame physics and learning undesirably.

As one facet of this, consider the question: Whose responsibility is it to make sense of the material as an organic whole by breaking it down, studying the bits, and then building it back up into a knowledge structure? The students', or the instructor's? Perhaps the work of unpacking this big thing called "physics" into separate pieces is something the students should be doing, to improve both their learning of the content and their ability to proactively learn future topics and subjects. One way to address this concern within a SBG approach might be to involve students in the identification of learning standards for the course. Clearly, such a change would require a drastic shift in class power structures and responsibilities, but the result might be beneficial in many ways (Rothstein & Santana, 2011).

The Attention Economy. A third tension unavoidable in any SBG implementation, at least in most current postsecondary environments, is between the SBG principle that grades only report mastery of learning objectives and the unavoidable fact that many students will not complete homework, attend lab meetings, or otherwise put in the requisite work without the carrot-and-stick of points and grades. Most college students perceive themselves as busy and stressed, constantly juggling many demands on their time, and they operate in an "attention economy" mediated by grades and deadlines. As one student put it on the Physics 291 mid-course questionnaire:

> The primary issue I have is that other classes don't have this mentality. So there are a large number of assignments due, that are required in order to get good grades. Which means that when I have big assignments due (which is, it seems, ALWAYS the case), it makes it very difficult to convince myself that my time would be better spent studying physics than working on some assignment that is due. I know that that is not the absolute truth, but I hope you understand how hard it is to put off something that's being graded, with the potential of not completing the assignment, or completing it at a sub-standard level, in order to study/work on

> something that isn't going to "directly" impact my grade. (Though I know that, in a sense, it will directly impact my physics grade, it is not an "immediate" impact, like the other assignments.)

Many other students expressed similar sentiments, appreciating the principle of "no coercion points" but finding difficulty with the reality. Overall, they seemed to be divided about whether they considered the SBG approach a net positive or negative. As another student candidly observed:

> I can honestly say I would get more done if it were graded, however I need to learn how to motivate myself and I believe this is helping.

Ultimately, this may pose the single biggest practical obstacle to successfully implementing SBG in an individual university course: realigning students' motivational structures in opposition to a conflicting environment.

B. Experience-Based Suggestions.

1. Develop assessments before, or with, standards. Committing to a set of standards before developing specific assessments for them invites anguish. Some standards sound great in the abstract, but are murder to develop practical assessments for (and even harder to develop reassessments for). Some groups of standards that sound distinct in the abstract may be difficult to separate on an exam. A set of standards that seems ambitious but not unreasonable can easily lead to a prohibitively long exam. Additionally, in the process of creating exam questions, I have invariably discovered requisite student knowledge or skills that I overlooked while designing standards. In the future, I shall discipline myself to develop at least one round of assessments (and preferably one or two rounds of reassessment as well) before finalizing the standards for a course.

2. Attend to topic weighting. Unless a grading system weights standard scores unequally for final grade calculation, topics will be weighted by the number of standards they contain. That may be undesirable, as one topic may naturally unpack into more or fewer learning objectives than another, equally-important topic. During both courses, this fact pushed me towards bifurcating or merging standards somewhat arbitrarily in order to get a comparable number per chapter or general topic. This is a tension between standards as tools for communicating and tracking desired learning outcomes, and standards as elements of a quantitative grading system. One solution is to assign each standard a weighting factor. Another is to group standards by larger topics, use the standard scores to determine an average score for each topic, and then combine those topic scores for an overall grade.

3. Include standards for crosscutting skills. Some skills cannot be factored out of assessment questions for other standards, but necessarily appear throughout exams. Examples include arithmetic, algebra, trigonometry, and other calculation skills; fidelity to units and dimensions; checking answers; adopting clear and conventional notation; communicating solutions clearly; and anything else an instructor might like to "ding" a student for erring on, but which is not the core content targeted by the question. Creating separate standards for these eases and clarifies grading and feedback, and also communicates to students that developing reliability in these skills is an important course objective. Any original or reassessment exam can then be used as a (re)assessment of such crosscutting standards, with a mastery score assigned holistically based on the entire exam.

4. Name standards usefully. If we want students to use our standards to organize and direct their thinking about the subject, we should give them conveniently short yet meaningful names. Neither my students nor I could reliably remember what "standard 6.02" referred to.

5. Keep (re)assessment efficient. The majority of current SBG activity seems to be in K-12 schools, and I've gleaned many good ideas from the weblogs of high school teachers (e.g., Noschese, 2012; O'Shea, 2012). However, the university teaching context differs from the K-12 context in some very profound structural ways, with implications for practical SBG implementations. Perhaps the most salient difference is that a university physics course is expected to cover more material than a high school course, in greater depth and with greater rigor, with far less in-class time. Consequently, much of the burden of learning is correspondingly shifted to the student outside of class. With class time at a premium, we cannot afford to dedicate much to assessment and reassessment; yet with a faster pace of coverage, we must assess material more rapidly. Thus, efficiency of assessment and especially of reassessment is critical. Many of the strategies that K-12 teachers employ, such as reassessing via one-on-one oral interviews or requiring students to qualify for reassessment by explaining what has been learned from homework or re-worked test questions, will be impractical for many of our courses. We may also have to eschew foundational, low-level skills in favor of more integrative competencies for our standards.

6. Invest in quizzing. My students' comments and my own reflections indicate that I should have persevered with my initial intention to give short, frequent, low-stakes quizzes in class. Such quizzes could have helped meet my students' needs for more frequent feedback and guidance, operationalization of standards, and familiarization with my exam question style. They might have allowed me to remove at least some standards from the overly-long unit exams. Additionally, they could very plausibly have helped motivate students to keep up with reading and homework, in the same way that graded homework would but without doing violence to the SBG philosophy. The precious class time devoted to the quizzes could be at least partially recouped by using the quiz questions as fodder for class discussion, in much the way that I use clicker questions and group whiteboarding tasks.

7. Schedule regular reassessment times. As an alternative to disrupting the course and distracting students with full-scale reassessment exams, one could designate one or two time windows a week for reassessment, and allow students to show up and reassess on a few standards at a time (perhaps with an appointment). By making any given standard available for reassessment only for one or a few consecutive days, an instructor could limit the opportunity for diffusion between students, especially if she has a few different questions ready for each standard.

8. Require students to qualify for reassessment. One way to motivate the students most in need of homework to complete it, without resorting to "coercion points," is to require relevant homework be completed to a high level of proficiency in order to qualify for reassessment on a standard. This could be implemented efficiently by means of an online homework system that allows students to redo problems until correct. Two additional benefits of doing this would be stressing the link between homework and assessment success, and reducing the number of "hail Mary" reassessment attempts by students who have not significantly improved since the original assessment.

9. Link standards to instruction. A carefully engineered set of standards can shape how students conceptualize the subject. Explicitly organizing our instruction around those standards reinforces that conceptualization and helps students understand the implications of each standard.

For example, if we present physics as a "toolkit" for thinking and analyzing, and we align standards with the various "tools," then we can structure class as an inventory of the tools and an exploration of how various tools apply to different problems. Teachers using both SBG and the *Modeling Instruction* pedagogy (Brewe, 2008; Hestenes, 1992) seem to do this particularly well, linking standards to the physics "models" (e.g., Noschese, 2012; O'Shea, 2012).

10. Build catch-up time into the syllabus. Learning rarely happens linearly, with each successive chunk of material learned adequately on first exposure and then never revisited. Rather than plan a traditional, linear course syllabus and then try to wedge SBG reassessment into and around it, an instructor could set aside two weeks or so—perhaps at the end of the term, perhaps distributed throughout it—for remediation and reassessment. Students who need additional help on specific topics could seek it, and reassessment could be carried out without competing with ongoing new instruction. Those students not needing significant reassessment could pursue advanced and/or elective topics semi-independently. (One could attach additional standards to those advanced topics, necessary to reach the very highest course grades.) Additionally, such a strategy could help us avoid the chronic trap of optimistically trying to cram just a little too much content into each course.

11. Don't overemphasize grades. As Docan (2006) observed, increasing students' awareness of course grades and grading can increase their stress. During Physics 291, my attempts to help students understand the novel grading system and allay their fears about it may have backfired by keeping them more conscious of the grading aspect of the course than they would otherwise have been.

VI. Summary and Discussion.

In this paper, I have described a teaching experiment in which I implemented standards-based grading differently in each course of a two-semester introductory calculus-based physics sequence. I've candidly narrated my design decisions, difficulties, and adjustments. I've summarized the students' reactions and my own experiences as an instructor. Based on that, I've identified three fundamental tensions that I claim must be negotiated by any instructor attempting to implement SBG at the university level, and I've made several suggestions for others who might wish to try SBG in their own courses. Overall, I have found the principles of SBG compelling and some of my experiences with it promising, but I have encountered serious practical challenges and pedagogical dilemmas.

I claim that many of the tensions we encounter when implementing SBG are not actually created by SBG itself. Rather, they are inherent within all our teaching, but are brought to a head by SBG because it forces us to be explicit about many things often left implicit: our learning objectives, how we prioritize remediation of older material vs. coverage of new, the purpose of homework, whose responsibility various aspects of learning are, and what grades communicate. For example, any instructional approach or assessment strategy can suffer from the dead frog problem, but SBG forces us to articulate our learning objectives precisely and link them to assessment, thus making our deconstruction of the subject clear. Similarly, we always hope our students will heed the feedback they get on assessments and remedy their knowledge of past topics, but SBG forces us to budget time and attention for that and to value it in the grading system.

Because of this, I have found that my explorations of SBG have clarified my thinking about teaching in general. I am confident that my future teaching will be better, whether or not it includes "standards-based grading" per se. I am optimistic that consideration of SBG by reflective university instructors and education researchers will provoke valuable discussions about the more difficult and elusive aspects of the art of teaching.

Acknowledgements

While the professional literature is relatively silent on SBG in physics and in higher education, the weblogs and *Twitter* feeds of innovative instructors are not. I owe a great intellectual debt to SBG pioneers such as Andy Rundquist (2013), Joss Ives (2011), Frank Noschese (2012), Jason Buell (2013), and Kelly O'Shea (2012), who use these channels to chronicle their SBG-related thoughts and experiences openly and candidly. I am also grateful for the always-sage advice of my colleague William Gerace, and for the helpful suggestions of Ellie Sayre, Warren Christensen, and Sissi Li.

References

Beatty, I., Gerace, W. J., Leonard, W. J., & Dufresne, R. J. (2006). Designing effective questions for classroom response system teaching. *American Journal of Physics*, *74*(1), 31.

Beatty, I. D., & Gerace, W. J. (2009). Technology-enhanced formative assessment: A research-based pedagogy for teaching science with classroom response technology. *Journal of Science Education & Technology*, *18*(2), 146–162.

Block, J. H. (Ed.). (1971). *Mastery Learning: Theory and Practice*. Holt, Rinehart & Winston.

Bloom, B. S. (1985). *Developing Talent in Young People*. New York: Ballantine Books.

Brewe, E. (2008). Modeling theory applied: Modeling instruction in introductory physics. *American Journal of Physics*, *76*(12), 1155. http://dx.doi.org/10.1119/1.2983148

Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership*, *69*(3), 10–14. http://goo.gl/pKsL7

Buell, J. (2013). Standards-based grading implementation. *Always Formative* (weblog). http://alwaysformative.blogspot.com/p/standards-based-grading-implementation.html

Docan, T. N. (2006). Positive and negative incentives in the classroom: An analysis of grading systems and student motivation. *Journal of Scholarship of Teaching and Learning*, *6*(2), 21–40. http://josotl.indiana.edu/article/view/1668/1666

Dueck, M. (2011). How I broke my rule and learned to give retests. *Educational Leadership*, *69*(3), 72–75. http://goo.gl/GqupA

Erickson, J. A. (2011). How grading reform changed our school. *Educational Leadership*, *69*(3), 66–70. http://goo.gl/KufkC

Guskey, T. R. (2001). Helping standards make the grade. *Educational Leadership*, *59*(1), 20–27. http://goo.gl/nbOze

Guskey, T. R., & Gates, S. L. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary classrooms. *Educational Leadership*, *43*(8), 73–80. http://goo.gl/lCYLR

Hestenes, D. (1992). Modeling games in the Newtonian world. *American Journal of Physics*, *60*(8), 732–748. http://dx.doi.org/10.1119/1.17080

Ives, J. (2011). Standards-Based Grading. *Science Learnification* (weblog). http://learnification.wordpress.com/category/course-components/standards-based-grading/

Leonard, W. J., Hollot, C. V., & Gerace, W. J. (2008). Mastering Circuit Analysis: An innovative approach to a foundational sequence. *38th ASEE/IEEE Frontiers in Education Conference*. F2H3–8. http://dx.doi.org/10.1109/FIE.2008.4720568

Marzano, R. J., & Heflebower, T. (2011). Grades that show what students know. *Educational Leadership*, *69*(3), 34–39. http://goo.gl/fXfMn

Noschese, F. (2012). Standards-Based Grading. *Action-Reaction* (weblog). http://fnoschese.wordpress.com/category/standards-based-grading-2/

O'Connor, K., & Wormeli, R. (2011). Reporting student learning. *Educational Leadership*, *69*(3), 40–44. http://goo.gl/Czm1m

O'Shea, K. (2012). SBG. *Physics! Blog!* (weblog). http://kellyoshea.wordpress.com/category/sbg/

Rothstein, D., & Santana, L. (2011). *Make Just One Change*. Cambridge MA: Harvard Education Press.

Rundquist, A. (2011). Standards-based grading with voice: Listening for students' understanding. (N. S. Rebello, P. V. Engelhardt, & C. Singh, Eds.) *AIP Conference Proceedings*, *1413*, 69–72. http://link.aip.org/link/doi/10.1063/1.3679996

Rundquist, A. (2013, January 6). SBG. *SuperFly Physics* (weblog). Retrieved February 11, 2013, from http://arundquist.wordpress.com/category/sbg/

Wiggins, G. P. (1998). *Educative Assessment*. San Francisco, CA: Jossey-Bass.