Implementing Skills-Based Grading in a Linguistics Course

Maura O'Leary & Richard Stockwell*

Abstract. This paper discusses how to implement 'Skills-Based Grading' (SBG) in linguistics, with a university-level course in formal semantics as a case study. Particular focus is given to transitioning to SBG from traditional grading. Regardless, of the grading system, all courses start out from a set of desired learning objectives, with final grades intended to reflect the extent to which these objectives have been achieved. Traditionally, grades are determined by a series of mandatory assessments and tests, for which somewhere between partial and full credit is awarded. In this paper, we illustrate how mastery of learning outcomes can be more directly measured by re-packaging them as "skills". In SBG, students are given multiple opportunities to demonstrate full mastery of each skill. However, grading is all-or-nothing, with no partial credit awarded. SBG has been shown to improve student learning, encourage effective study, lower student stress, and achieve more equitable outcomes, and has been successfully adapted for linguistics courses in phonology (Zuraw et al. 2019) and semantics (O'Leary & Stockwell 2021). Here, we offer step-by-step instructions for creating an SBG course, covering skill types, skill groupings, opportunities, grading, and assessments.

Keywords. Skills-Based Grading; pedagogy; semantics

1. Introduction

This paper discusses how to implement 'Skills-Based Grading' (SBG) in linguistics courses, using our novel application of this grading system to a university-level introductory formal semantics course as an example (O'Leary & Stockwell 2021).

The rest of this introduction summarizes O'Leary & Stockwell (2021), giving a brief outline of SBG in opposition to traditional grading and discussing some of the recognized benefits. §2 steps through the process of preparing an SBG linguistics course, with an emphasis on transitioning from a traditionally graded course. §3 highlights two points to bear in mind during the course regarding grading and explaining SBG to students. §4 concludes.

* Thanks first and foremost to our students for their patience and feedback. Thanks also to Douglas Ezra Morrison for help handling the data; Kie Zuraw and Franny Brogan for their advice and encouragement with Skills-Based Grading; Jessica Rett for her advice throughout; Dylan Bumford, Ethan Poole, Yael Sharvit, Giuseppina Silvestri and Jos Tellings, for whose semantics courses we have been teaching assistants; Jeffrey Reaser and two anonymous reviewers for this journal; three anonymous LSA reviewers; and audiences at UCLA, LSA 2021, and the SALT 31 Workshop on Inclusive Teaching in Semantics.

1.1. Traditional vs. Skills-Based Grading

There is much ongoing discussion of grading systems that seek to improve upon traditional methods: portfolio grading (e.g., Metzger & Bryant 1993), ungrading (e.g., Blum 2020), specifications grading (e.g., Nilson 2015), among others. In O'Leary & Stockwell (2021), we contributed to this discussion with an evaluation of the benefits of Skills-Based Grading (SBG). This subsection defines SBG by describing how it differs from traditional grading.[1]

Grading in the North American education system is traditionally composed of a series of assessments – homework assignments, quizzes, midterm and final exams. Every part of these assessments contributes a proportion of the final grade, and generally lost points cannot be recovered. This grading structure requires students to learn and demonstrate mastery over material along a uniform timeline in order to keep up with assessment deadlines. To compensate for this inflexibility, partial credit is commonly awarded for answers that show partial understanding of the material.

In SBG, by contrast, each assessment offers *opportunities* to demonstrate mastery over material. No failed attempt counts against the student's grade – until the final, there are always more opportunities to try again. At the outset, students are provided with a list of 'skills' that make up the course. As will be discussed in §2, these 'skills' are effectively a detailed list of learning outcomes which the students should master by the end of the course. Assessments are still composed of exercises, but each exercise offers an opportunity to demonstrate proficiency in one or more skill rather than making a demand. No one exercise is required of the students and there is no penalty for incorrect answers. Attempts are graded as "not yet proficient", "approaching proficiency", or "proficient". Only "proficient" answers count towards final grades, adding to a running tally of 'skills demonstrated' that determines a letter grade at the end of the course. "Not yet" and "approaching" are for guidance only, as either way the student will have to try again in a later assessment in order to earn that credit. While there is no penalty for incorrect answers, only fully correct answers are marked "proficient" – there is no partial credit.

On the one hand, SBG is more generous than traditional grading: while students are encouraged to attempt skills as they encounter them, they are granted multiple opportunities to demonstrate proficiency. On the other hand, SBG is more rigorous: students are required to demonstrate full proficiency in a skill, without partial credit as a safety net.

Implementations of grading schemes along the lines of SBG have been reported in a variety of secondary and higher education settings; for example, standards based grading (see, e.g., Beatty 2013, Buckmiller et al. 2017, Schimmer 2016), mastery based grading (e.g., Armcost & Pet-Armacost 2003, Brackett & Reuning 1999), and specifications grading (Nilson 2015). Zuraw et al. (2019) were the first to report an application of SBG specifically to linguistics, with courses in phonology and phonetics. O'Leary & Stockwell (2021) followed with an application of SBG to formal semantics and an assessment of its benefits, as summarized in the next section.

[1] We caricature traditional grading somewhat so as to provide a clear contrast with SBG. In practice, there is a continuum of options, drawing from both traditional grading and SBG. For instance, we know of classes that have attributed 50% of the total grade to SBG and 50% to traditionally graded essays (p.c. Franny Brogan). We touch on other mixed options in §2.6.

1.2. Benefits of Skills-Based Grading

In O'Leary & Stockwell (2021), we evaluated the effectiveness of SBG by collecting quantitative and qualitative data during a university-level formal semantics course.[2] Our findings substantiated a number of benefits that have been claimed for SBG and similar systems in the literature.

First, SBG enables more effective study. Once students are proficient in a skill, there is no need to demonstrate that skill again. As Buckmiller et al. (2017) found, this encouraged students to direct their efforts away from skills they have already mastered towards those they still need to work on. In addition, SBG incentivizes reattempting skills in which they were not initially proficient. Because a failure does not affect the grade in any lasting way, students can treat unsuccessful attempts as a learning opportunity rather than a failure. Students are therefore more likely to adopt a growth mindset (Dweck 2008), in stark contrast to traditional, subtractive grading, where students are permanently penalized for mistakes.

Second, SBG lowers student stress (e.g. Buckmiller et al. 2017, Zuraw et al. 2019). No specific exercise is ever required, nor will it make or break a student's final grade. Until the final there is always another chance to demonstrate a skill. This means that each individual assessment brings less stress, fear, and time pressure.

Third, SBG offers a more equitable system of evaluation for a diverse student body.[3] There is no penalty for missing or underperforming on assignments due to external setbacks. There will always be factors that prevent students from giving their best to every assignment: illness, grief, global pandemics, etc. This leniency is particularly beneficial to students facing structural and institutional disadvantages, as noted by student responses in O'Leary & Stockwell (2021). In addition, SBG is beneficial for students who enter the class with a weaker background in the subject matter or study skills. In traditional grading, the grade for a student's first homework is tantamount to a grade for their background for the course. In SBG, students have until the end of the course to make that ground up.[4]

With a definition and summary of benefits in hand, the next section turns to the practicalities of creating an SBG course.

---

[2] The course was introductory formal semantics at the University of California, Los Angeles (UCLA Ling 120C: Semantics 1). Our 29 students had taken two prerequisite courses: an introductory linguistics course for intending majors and introductory syntax. While our case study draws on a rather technical course within a dedicated linguistics major, we believe that this grading system can be applied to any course that can be broken down into detailed learning outcomes.

[3] Our class was taught online during the covid pandemic. To the extent that inequalities are magnified in an online class (e.g., access to suitable workspace) and in a pandemic (e.g., sudden financial pressures), the equity benefits of SBG are also magnified.

[4] It might be possible to gain some of these benefits of SBG within the framework of traditional grading. For example, a common way to lower stress and improve equity is to drop the lowest quiz or homework score. However, such mitigations are liable to both under- and over-perform. They under-perform in being insufficient for late blooming students, or those who suffer setbacks across multiple weeks. SBG, on the other hand, accommodates students who are faced with outside factors that affect their performance over more than one homework or quiz, and those for whom the material 'clicks' later in the course. At the same time, the mitigation of dropping the lowest score over-performs in being overly generous to students who perform consistently throughout the course. These students are then free to (ab)use this mitigation and make no attempt at the last assignment, which may test material not covered elsewhere. Compare the discussion of 'early check-out' in §2.6, however.

2. Setting up a Skills-Based Grading Course

This section offers step-by-step instructions for how to transition a linguistics course to SBG, using our application to formal semantics as a case study. We cover how to identify skills of various types, how to determine skill groupings and the number of required demonstrations, how to create assessments, and whether to conclude the course with a culminative component.

2.1. Identifying Skills

Skills-based grading is, at its core, an identification of the skills that students should master by the time they complete the class. Thus, creating a 'skills list' is the most important step in setting up a skills-based course. It may seem daunting to compile and edit a long list of skills (our semantics course used 47 skills sorted into 12 course topics, or "skill groups"; for phonology, Zuraw et al. 2019 used 49 skills sorted into 10 groups). However, with a traditional syllabus in hand, instructors should rest easy knowing that they have already done most of the work needed to create a skills list.

For a broad picture of the skill groups you may want to require in your course, consult the topics and learning objectives/outcomes defined in the syllabus or course proposal. Required by many academic institutions, these provide a good starting point for identifying the content and skills that students should acquire within the course. Learning outcomes may include concrete items ("learn propositional logic"), as well as outcomes focused on student growth ("develop strong critical thinking skills").

To move from this broad picture to a more detailed list of skills, instructors should go through individual exercises from homeworks, quizzes, and exams in previous renditions of the course. In each exercise, what did the student need to be able to do? For instance, we found that in the process of teaching propositional logic, we asked students to complete four types of tasks. As shown in Table 1, each of these tasks makes up a single 'skill' that students needed to demonstrate mastery of in the SBG version of our course:

Skill identifier	Propositional Logic skills
5.1	Complete a truth table
5.2	Evaluate the truth of a formula
5.3	Translate from natural language to propositional logic
5.4	Translate from propositional logic to natural language

Table 1. The Propositional Logic skill group

2.2. Types of Skills

2.2.1 Concrete, Abstract, and Language Skills

The application of SBG to concrete skills, where exercises have a single correct answer, is relatively straightforward. For instance, in phonology, you might ask students to order two rules

or draw a Hasse diagram (as in Zuraw et al. 2019). In semantics, students may be asked to complete a truth table or to enumerate a set given a set description (as in O'Leary & Stockwell 2021).

In O'Leary & Stockwell (2021), we showed that in addition to concrete skills, SBG works equally well for the more abstract topics in semantics, encompassing skills requiring knowledge as a language user (e.g., creating original examples or identifying meaning patterns such as contradiction and ambiguity) and philosophical skills with infinitely many possible answers (e.g., explaining why a definition is too strict or too lax). The absence of partial credit, which rewards complete mastery rather than partial understanding, helped in the grading process, in that we did not have to award partial credit for answers that were overly vague or that used key terminology incorrectly; e.g., if a student confused "strict" and "lax" with respect to dictionary definitions. Instead, such answers could be marked as "not yet proficient" or "approaching proficiency", pending future opportunities to provide a "proficient" answer later in the course. The proficient-or-not grading scheme also did not inhibit demonstration of mastery over these subjects. Across all three types of skills – concrete, abstract, and language – grades showed that students benefited from having multiple opportunities to demonstrate their learning over the course of the term. On average, students mastered roughly the same proportion of each skill type by the end of the course (concrete: 85%, abstract: 76%, language user: 84%).

2.2.2 Activity Skills

If desired, SBG also allows easy integration of points for activities that are generally 'checked-off' in traditional grading, such as reading comprehension questions or participation. Including active engagement with the material and peers as part of the skills list indicates to students that these are skills work developing. In our class, points were available for active class participation under a skill group called "Explain". This skill group required participation in an online forum and in class, including presenting a real-world example illustrating a semantic concept. Offering credit for participation came with the usual benefits – encouraging students to talk and ask questions helps them engage with and learn the other (non-participation) skills.

2.2.3 Critical Thinking Skills

A key feature of specifications grading (e.g., Nilson 2015, Ludwigsen 2017), one of the systems upon which SBG is based, is a clear division between what constitutes an 'A' grade and a 'B' grade.[5] Zuraw et al. (2019) achieved this in their phonology course by adding a fourth level of proficiency to their grading scheme; further to "proficient", "approaching proficiency", and "not yet proficient", they added "advanced proficiency". Advanced proficiency was gained by answering a particularly difficult question or providing an answer that was "more explicit, fuller, or better" (2019, e410). Demonstrating ordinary proficiency in every skill amounted to 85% – a mid-level B grade. The remaining 15% was awarded for demonstrating advanced proficiency. However, this system led to some complications in how grades were calculated.[6]

[5] 'B', 'C' and 'D' grades remain grouped together as varying levels of 'satisfactory'.

[6] Each letter grade was earned by mastering a predetermined number of skills: A = proficiency on 50 skills and advanced proficiency on 15 skills, B = proficiency on 50 skills, C = proficiency on 33 skills, and so on. The extra level of advanced proficiency required for an A caused some issues in "mixed situations, such as proficiency on only forty-eight skills (does not meet standard for B), but advanced proficiency on fifteen skills (exceeds standard for B)" (2019, e412). Addressing these situations required formulae (e412, e415) that were "admittedly more complex than we would like" (e412).

We prefer the approach taken by Zuraw et al. (2019) in their phonetics course. Moving from a B grade to an A grade required five demonstrations of a skill called "Apply," for which students had to apply course concepts to problems that were significantly different from those seen in class.

Along similar lines, we achieved a division between A and B grades with a skill called "Scientific Analysis", which required 10 demonstrations. Scientific Analysis questions asked students to think beyond the lecture material, evaluating theories and applying them to further data. An example of a Scientific Analysis question is given in Figure 1. Notice how the question flows from application of the lecture material in part (a) to problematizing it in (b) and revising it in (c):

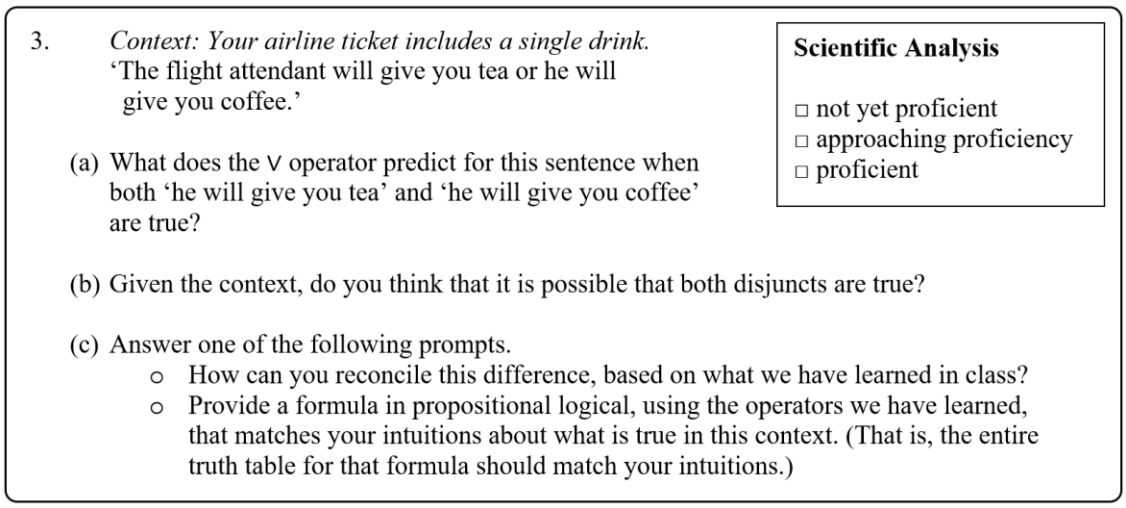

Figure 1. A question testing for the Scientific Analysis skill

2.3. Granularity of Skills

In creating a skills list, it is important to break down skills into the right amount of detail. Creating a skills list is a balancing act between specific and general. Skills need to be specific enough to ensure that students have to tackle each type of task. At the same time, the overall number of skills needs to be few enough that students are not overwhelmed, and that it will be reasonably possible to offer multiple attempts at each skill. We recommend starting with the learning outcomes identified in §2.1, and then mapping out all of the potential skills involved with varying levels of granularity.

For their phonetics course, Zuraw et al. (2019) chose a broad approach. Their course was built of just 15 skills, with each skill requiring between 2-30 demonstrations. They reasoned that "if a student can correctly define fifteen IPA symbols, this represents a reasonable sample of all the symbols" (2019, e413). Assuming that the number of exercises offered for each skill is controlled (see §2.4), the all-or-nothing grading system should prevent a student who exhibits significant and systematic mistakes on a subset of the material from checking off every demonstration of that skill.

This broad approach often made sense for our semantics course. For instance, returning to our Propositional Logic Skill Group in Table 1, consider skill 5.1, 'Complete a truth table'. As diagramed in Figure 2, this skill could conceivably have been broken down further into separate skills, one for each of the Boolean operators (¬, ∧, ∨, →). However, we would then have had to create multiple, separate questions for each operator, despite the principles of truth tables being the same. As such, we kept all the operators bundled as skill 5.1.

That said, we kept skill 5.2 separate from skill 5.1. We chose to do this despite evaluating the truth of a formula being tantamount to completing one line of a truth table. This allowed us to set more challenging formulae without forcing students to tabulate every possibility.

Skills 5.3-4 were likewise better kept separate. The alternative would have been to combine them into one 'translating propositional logic' skill – a broad skill which could be 'checked off' by completing either of the tasks that we separated into 5.3-4. As it was, their separation guaranteed that a student would have to demonstrate proficiency in each of those unique concepts. Marking skills as separate, identifiable activities also helps students' metacognitive awareness of the distinct skills that they are developing and the steps that are involved in solving problems or developing an analysis (Zuraw et al. 2019).

We could have further separated 5.3-4 into 'simple' translation tasks containing a single Boolean operator and 'hard' translation tasks with more than one. We did not take this tack for Propositional Logic, as we did not require enough demonstrations of 5.3-4 to warrant further division. We did take this tack later in the course, however, with the corresponding skills in Predicate Logic (one vs. two scopal elements) and Lambda Calculus (simple/complex denotations and derivations), where we wanted to offer many opportunities at a simple version of the skill without removing the need to attempt the difficult version.

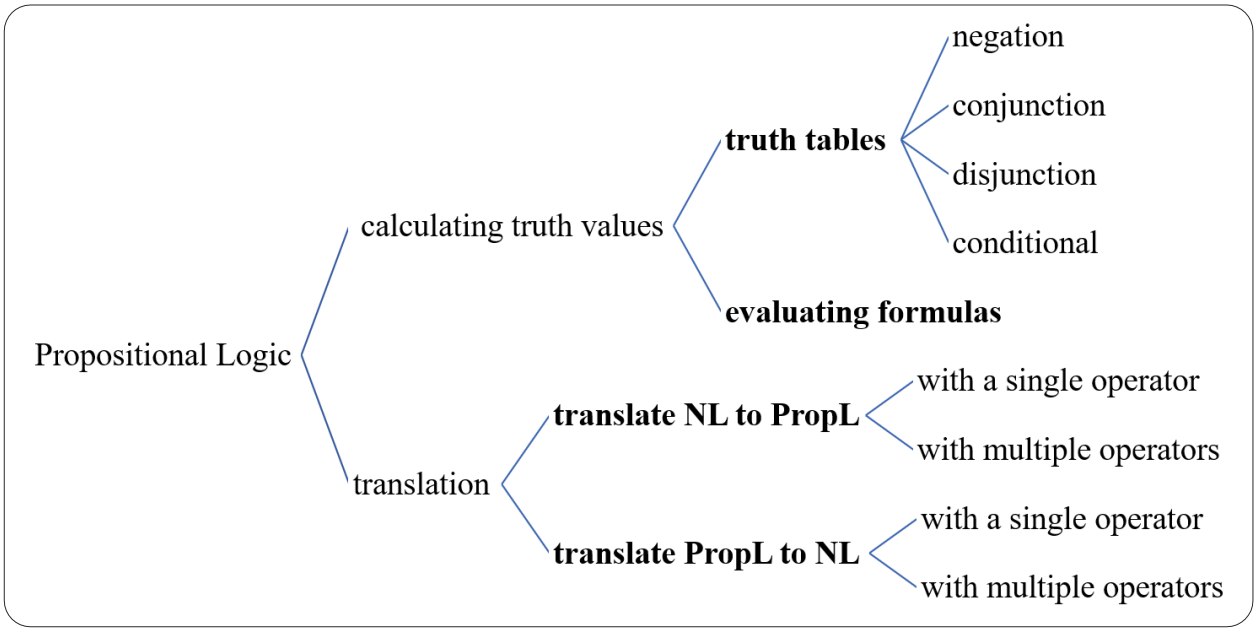

Figure 2. Potential breakdown of Propositional Logic Skill Group from Table 1, with the selected skills in bold

The overall level of granularity in your skills list will be a product of decisions regarding how much of the grade to attribute to each skill group and how many demonstrations should be required for each skill, as discussed in the next subsection.

2.4. Number of demonstrations and proportion of the grade

With a skills list in hand, the next step is to determine how many times students will need to demonstrate proficiency in each skill. For some skills, proficiency can reasonably be shown with a single demonstration; for instance, explaining how a dictionary definition is too lax. Other skills may need multiple repetitions to guarantee that the student has mastered the concept. This might be due to either the difficulty or the breadth of the skill. For example, our "truth table" skill needed multiple demonstrations so that students would interact with multiple Boolean operators. Multiple demonstrations might also be appropriate for foundational or algorithmic skills where proficiency requires practice; for example, simple denotations of predicates in lambda calculus.

The number of required demonstrations is intrinsically tied to the proportion of the grade attributed to each skill/demonstration. Therefore, after taking a first pass at the number of demonstrations from the bottom up, we recommend transitioning to figuring out the grade breakdown from the top down. That is, what proportion of the grade should be attributed to each skill group? Helpful heuristics include how much class time is devoted to each topic, the relative weights given to each topic on a previously used final exam, or the relative weights for each topic throughout all of the assessments in a previous rendition of the course.

From here, you will have to reconcile the number of skills or demonstrations with the weight attributed to each skill group in order to figure out how the grade will be calculated. Don't worry if this step doesn't sort itself into a clear pattern quickly; some amount of fiddling and negotiating is inevitable. The simplest option is to treat demonstrations as the unit of grading: each proficient demonstration earns one point. An alternative is to treat skills as the unit of grading: either holistically – one point for full proficiency in a skill, once all demonstrations are completed; or proportionally – e.g., half a point per demonstration of a skill requiring two demonstrations.

For the semantics course reported in O'Leary & Stockwell (2021), we adopted a blend of these two approaches. For most skills, students earned one point per demonstration. For five of our 47 skills, however, all of the demonstrations for that skill added up to one point. These "foundational" skills were ones that required comparatively little effort, but were valuable enough to warrant significant repetition. For example, giving the denotation for a proper name or predicate is formulaic but important preparation for lambda calculus.

In future, however, we would be consistent in using demonstrations as the unit of grading, awarding one point per proficient demonstration. This option is very intuitive, which is especially helpful early in the course for students new to SBG. It does mean that many points are given out for simple, repetitive skills, which earn the same credit as difficult skill demonstrations. However, grade boundaries can always be set in an appropriate way to account for this.

For final grades, it is helpful to talk in terms of points rather than percentages. Percentages are evocative of traditional grading, where every part of every assessment contributes its portion of the final grade. Using different terminology helps students grasp the distinctive additive nature of SBG. There is also no need for the number of available points to total 100, or to peg points to the traditional percentage grade boundaries. Instead, grade boundaries can be set wherever appropriate. For example, our course set 70/80 points as an A, 60/80 as a B, etc. Zuraw et al. (2019)'s initial approach to calculating grades also had this 'points not percentages' character.[7]

[7] See footnote 6 for details. Zuraw et al. (2019) subsequently framed their grading systems in terms of percentages rather than points, achieving the same effect as the system in the main text by awarding a baseline percentage 'for free' (50% for phonology, 45% for phonetics). In truth, our course in O'Leary & Stockwell (2021) was also framed in terms of percentages rather than points, achieving the effect of the grade boundaries in the main text by giving a

2.5. Creating assessments

Assessments (assignments, quizzes, exams) in an SBG course offer opportunities to demonstrate proficiency in some of the course's skills.[8] Often, SBG assessments take the form of long lists of exercises. However, students need not attempt all of the exercises; they can skip skills for which they have already demonstrated proficiency or skills they don't yet feel ready to attempt. Thus, most students only complete a subset of the available questions. Likewise, single exercises may involve more than one skill, thereby lowering the total number of exercises per assignment.[9]

Exercises should clearly label which skill is being attempted, as in the example in Figure 3. Notice also the box for the grader to indicate the demonstrated level of proficiency.

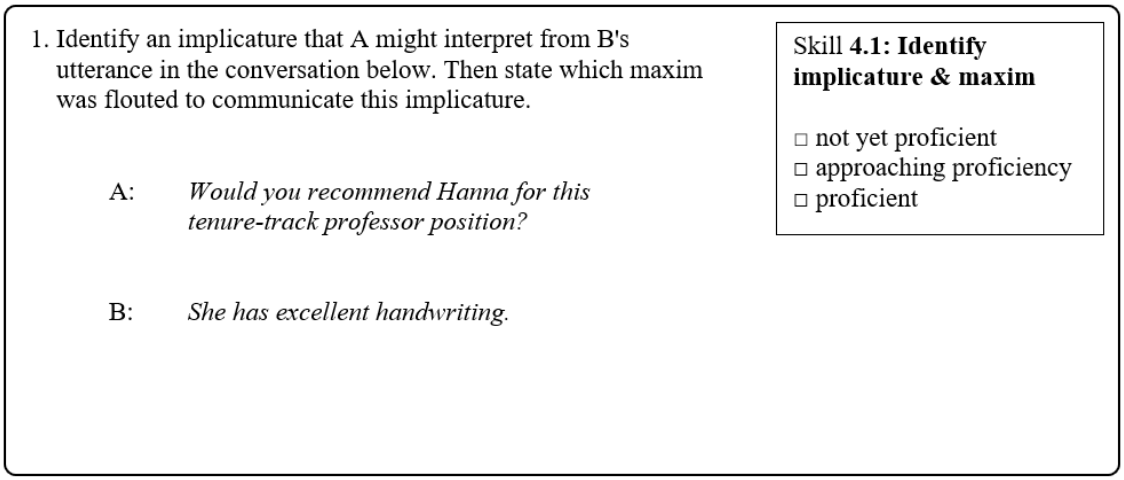

Figure 3. A question testing for a skill in the Sentence Relationships skill group

Exercises should be distributed throughout assessments based on (a) when students will first learn the material, (b) when they should review the material, and (c) when the end of the course is approaching. Of these, (b) is perhaps the most useful: by intentionally delaying second chances at certain skills, instructors can covertly encourage students to revisit early skills at a time that is most beneficial for their learning – for instance, right before they are taught a more complex skill which builds on the earlier one. Thus, even though SBG allows students to delay performing a skill until later in the term, they can be encouraged to review it at instructor-determined times.

It is valuable to decide ahead of time the (minimum) number of opportunities that will be provided to demonstrate each skill. Offering more opportunities helps to lower student stress, but knowing that there are limited opportunities overall encourages students to value each chance and disincentivizes guessing. We recommend publicizing a minimum number of opportunities to help alleviate student stress (see student comments in O'Leary & Stockwell 2021) while reserving the

baseline 20% 'for free'. We regret shoehorning our grading scale into the traditional percentage system in this way, and have framed subsequent renditions of the course exclusively in terms of points.

[8] In our semantics course, homeworks were assigned weekly with a fixed deadline for ease of grading workflow. See the discussion in §3.1, however, on the potential of question banks and automated grading to allow students to proceed entirely at their own pace.

[9] Students might also be allowed to 'write in' additional skills they think they have demonstrated, as in Zuraw et al. (2019). We made this option available to our students, but they rarely made use of it.

option to spontaneously add more opportunities for skills students are struggling with. In our course, we offered each required *demonstration* once on the final exam and at least twice throughout the rest of the assessments (e.g., a skill which required two demonstrations overall would be offered twice on the final exam and at least four times throughout the rest of the course), although this may have been slightly more than was necessary. Note that our breakdown allowed students to earn 100% in the course purely from the final exam, thus maintaining maximum flexibility about when students mastered a particular skill.

Finally, creating assessments presents some challenges. Providing multiple opportunities means creating multiple versions of each exercise. This quantity can be challenging of itself, especially for early career instructors who do not have multiple previous iterations of a course to draw upon. Moreover, it is far from straightforward to create multiple instances of adequately different yet consistently difficult questions. Contrary to the spirit of SBG, this meant that in our course, exercises testing a particular skill tended to grow gradually more difficult over time. Thus, students who achieved proficiency later in the term had to conquer more difficult exercises. An online assessment pulling exercises randomly from a question bank would offer a more equitable approach – see §3.1 for further discussion.

2.6. Culmination

A final consideration is whether to conclude the course with a culminative task, and whether this task is part of or supplementary to the SBG nature of the course. A final exam, paper, or project could provide the last chance to demonstrate SBG skills, or could be a traditionally graded assessment, which the SBG course builds towards.

One benefit of a culminative task is in providing a 'buffer' at the end of the course. SBG cannot fully eliminate the pressure to complete assignments along a particular timeline; in particular, skills learned in the last week of the course will need to be demonstrated on the last homework or the final exam. This issue might be solved by saving the final few weeks of the term to focus on a final project or paper. During those weeks, students could catch up on skills from earlier in the course without any new skills being taught.

A second benefit of a mandatory, culminative component is in guarding against 'early checkout' – an uncomfortably available option for students in SBG courses. Without a mandatory final assessment, we had one student who ceased participating in the course entirely during the penultimate week after earning their desired B grade. While they achieved their target grade, this student did not fulfil their potential and would be ill-prepared for follow-up semantics courses, having left before seeing a λ.

3. During the Course

The majority of the labor of an SBG course is in the preparation stage of transitioning from a traditional course, as discussed in the previous section. The relative brevity of this section is testament to how straightforward SBG courses are to teach, once designed. Still, this section discusses two points to bear in mind while teaching an SBG course: how to grade assessments, and how to help students transition to SBG.

3.1. Grading SBG assessments

If students are given adequate opportunities to demonstrate each skill, assessments will inevitably consist of quite a large number of diverse exercises, the sum of which may seem daunting to grade. However, each problem is ultimately graded on a binary: students either get the problem fully correct and check off proficiency for that skill, or they do not. In our experience, the high number of exercises graded in this more straightforward way equaled out to a 'normal' grading load in a traditionally graded course. Zuraw et al. (2019) reported a lighter grading load than in a traditionally graded course.

SBG saves a lot of time that is traditionally spent debating exactly how much partial credit should be awarded. In this regard, Zuraw et al. found it much easier to coordinate between graders (e.g., multiple teaching assistants) in SBG, since in traditional grading the exact amount of credit to award to a partially correct answer can be a rather subjective call.

The only remaining discretion is whether a non-proficient answer counts as "approaching" or "not yet proficient". While this choice does not have any effect on the grade, it can give students an indication as to whether their answer was on the right track, guiding further study. Instructors may choose whether to add additional commentary.

In order that students know which skills they have yet to complete, and thus which skills to revise and look out for on future assignments, it is important to track progress in a clear and timely fashion. Unfortunately, most learning management systems are not compatible with SBG. Thus, instead of using built-in grading tracking programs, we uploaded a spreadsheet weekly which tracked each student's successful demonstrations. Figure 4 gives a snapshot. For anonymity from their peers, each student was given a three letter code which labeled their demonstrations.

Skill #	Abbreviated Skill Name	Max #	PWG	FHZ	VNE	MTK
1.1	Ambiguity type	2	2	2	2	1
1.2	Ambiguity example	1	1	0	1	1
1.3	Pronominal ambiguity	1	1	0	1	0
2.1	Lax	1	1	1	1	0
2.2	Strict	1	1	1	1	1
2.3	Hyponym/hypernym	1	1	1	1	1
3.1	Intension/Extension	1	1	0	1	1
3.2	Sense/Reference	1	1	0	1	1
3.3	Indexical character/content	1	1	0	0	0
3.4	De re	1	1	0	1	0
3.5	De dicto	1	1	0	1	0

Figure 4. Snapshot of a grade spreadsheet partway through the term

To make maximum use of the binary credit system, the fastest option for grading would be to use automatically graded online exercises wherever possible. A "truth table" skill, for example, might be easily gradable by a computer, though perhaps not a full lambda calculus derivation. There are two additional benefits to computer graded exercises. First, students can complete them at any time and receive immediate feedback. Second, questions can be pulled randomly from a repository. The latter benefit would mean that two students working together would likely face different questions, and that slight variations in difficulty between questions would be randomly distributed.

3.2. Explaining SBG to students

A final and incredibly important aspect of running a successful SBG course is to explain the system to students clearly and often *throughout* the course. Be repetitive! Traditional grading has the benefit of familiarity; students are fluent in proportional assessments building towards traditional percentage grade boundaries. SBG, by contrast, is likely totally different from anything your students have experienced before, which some may find disconcerting – see O'Leary & Stockwell (2021) for student comments in this regard.

We found that students needed regular reminders in the following areas: how to read the grade spreadsheet, how to gauge their progress in the class, and how many opportunities they would have to demonstrate each skill. Students should always be aware that they do not need to complete the entirety of an assessment, especially as they are sitting down to a lengthy homework or exam. They might additionally be advised to begin assessments by browsing for which questions they want to answer. If you are allowing (or requiring) students to write in (additional) skills that they think they have demonstrated (see footnote 9), this option should be mentioned frequently. Finally, we recommend labelling skills on lecture slides and handouts as reminders of the system, and to make it clearer for students where to direct their attention in studying.

4. Conclusion

This paper discussed how to implement SBG in a linguistics course. SBG improves on traditional grading systems by balancing multiple opportunities against a lack of partial credit; full mastery of the material is the goal, without imposing a timeline for progress. Based on our experience of applying SBG to semantics (O'Leary & Stockwell 2021), the discussion centered on how to transition to an SBG course from a traditional one. The learning outcomes of a traditional course are repackaged as skills, which are grouped and weighted so as to fulfill those learning outcomes more effectively. We conclude with a summary of key considerations for creating an SBG course:

- Identify skills at appropriate levels of granularity, and group them thematically.
- Balance the number of skills and required demonstrations with the weight of each skill group (formerly 'topic') in the course.
- Distribute exercises throughout assessments based on (a) when students will first learn the material, (b) when they should review the material, and (c) when the end of the course is approaching.
- Decide whether to conclude the course with a culminatory final paper or exam, and whether it will be traditionally graded or a last opportunity to demonstrate skills.
- Publish grades promptly, and explain SBG clearly and often.

Maura O'Leary
University of California, Los Angeles
mauraoleary@ucla.edu

Maura O'Leary is a Ph.D. Candidate in Linguistics at the University of California, Los Angeles. Her pedagogical research focuses on equitable and inclusive methods for teaching and evaluation in university-level linguistics courses. Her dissertation provides a formal model of nominal evaluation times at the syntax-semantics interface, part of a larger research program on understudied effects of tense. She additionally conducts fieldwork on the critically endangered Alaskan Dene language Hän, and heads the Hän Revitalization Project.

Richard Stockwell
Christ Church, University of Oxford
richard.stockwell@chch.ox.ac.uk

Richard Stockwell is a postdoctoral Junior Research Fellow in Linguistics at Christ Church, University of Oxford. He teaches and researches in syntax and semantics. He received his Ph.D. from the University of California, Los Angeles in 2020 with a dissertation titled "Contrast and Verb Phrase Ellipsis: Triviality, Symmetry, and Competition". Richard's previous degrees are from the University of Cambridge.

References

Armacost, Robert L., and Julia Pet-Armacost. "Using mastery-based grading to facilitate learning." In 33rd Annual Frontiers in Education, 2003. FIE 2003., vol. 1, pp. T3A-20. IEEE, 2003. https://doi.org/10.1109/FIE.2003.1263320

Beatty, Ian D. "Standards-based grading in introductory university physics." Journal of the Scholarship of Teaching and Learning 13, no. 2 (2013): 1-22.

Blum, S. D. (Ed.). (2020). *Ungrading: Why rating students undermines learning (and what to do instead)*. West Virginia University Press.

Brackett, Carolyn C., and Richard H. Reuning. "Teaching pharmacokinetics using a student-centered, modified mastery-based approach." American Journal of Pharmaceutical Education 63, no. 3 (1999): 272-276.

Buckmiller, Tom, Randal Peters, and Jerrid Kruse. "Questioning points and percentages: Standards-based grading (SBG) in higher education." College Teaching 65, no. 4 (2017): 151-157. https://doi.org/10.1080/87567555.2017.1302919

Dweck, Carol S. Mindset: The new psychology of success. Random House Digital, Inc., 2008.

Ludwigsen, Daniel. "Effective and efficient grading: tips for assessing student learning." In Proceedings of Meetings on Acoustics 173EAA, vol. 30, no. 1, p. 025001. Acoustical Society of America, 2017.

Metzger, E., & Bryant, L. (1993). Portfolio Assessment: Pedagogy, Power, and the Student. *Teaching English in the Two-Year College*, 20(4), 279-88.

Nilson, L. B. (2015). Specifications grading: Restoring rigor, motivating students, and saving faculty time. Stylus Publishing, LLC.

O'Leary, Maura, and Richard Stockwell. "Skills-Based Grading: A novel approach to teaching formal semantics." Proceedings of the Linguistic Society of America 6, no. 1 (2021): 869-881. https://doi.org/10.3765/plsa.v6i1.5025

Schimmer, Tom. Grading from the inside out. Bloomington, IN: Solution Tree Press, 2016.

Zuraw, Kie, Ann M. Aly, Isabelle Lin, and Adam J. Royer. "Gotta catch'em all: Skills grading in undergraduate linguistics." Language 95, no. 4 (2019): e406-e427. https://doi.org/10.1353/lan.2019.0081