GOTTA CATCH 'EM ALL:

SKILLS GRADING IN UNDERGRADUATE LINGUISTICS

Kie Zuraw[1], Ann M. Aly[2], Isabelle Lin[1], Adam J. Royer[1]

[1]University of California, Los Angeles
[2]American Council on the Teaching of Foreign Languages

Email and mailing address of first author:

kie@ucla.edu

UCLA Department of Linguistics

3125 Campbell Hall, Box 951543

405 Hilgard Ave

Los Angeles, CA 90095-1543

GOTTA CATCH 'EM ALL:

SKILLS GRADING IN UNDERGRADUATE LINGUISTICS

ABSTRACT

Dissatisfied with traditional grading, we developed a grading system to directly assess whether students have mastered course material. We identified the set of skills students need to master in a course, and provided multiple opportunities for students to demonstrate mastery of each skill. We describe in detail how we implemented the system for two undergraduate courses, Introductory Phonetics and Phonology I. Our goals were to decrease student stress, increase student learning and make students' study efforts more effective, increase students' metacognitive awareness, promote a growth mindset, encourage students to aim for mastery rather than partial credit, be fairer to students facing structural and institutional disadvantages, reduce our time spent on grading, and facilitate complying with new accreditation requirements. Our own reflections and student feedback indicate that many of these goals were met.

1. INTRODUCTION

A few years ago, the first author of this article had a revelation, while teaching Phonology I for the first time. She was compiling a review handout for the final exam, and made a list of the fifty-odd skills students should prepare to be tested on. She immediately regretted not having drawn up the list for the students at the beginning of the term—that would be easy to fix next time. But then came a more difficult regret: if the final exam was testing this set of skills, shouldn't students' grades in the course directly reflect whether they had mastered them? This article lays out the system we developed for two linguistics courses in our attempt to achieve this, based on concepts from standards-based grading studies that we'll review in the next section.

All teachers have some dissatisfaction with traditional grading. Most of us want our grades to be an objective measure of how well students have mastered the material by the end of the course (or, even better, how much mastery they retain years later). But in reality, our grades assess some combination of mastery, compliance, and luck. When a student is penalized for a missing or late assignment, this is a penalty for non-compliance, not non-mastery; but to have no such penalty is unfair to students who turned in the assignment and did poorly, or who could have done better if they'd had extra time. To truly assess final mastery, we could rely solely on final exams and final projects. But this puts too much stress on students and gives too much incentive to cheat. And, it penalizes students who have the bad luck to be tired, sick, or distracted on the day of the final exam, or who get a question wrong because they happen to have forgotten the definition of the feature [continuant], even though they could have gotten an equivalent question right if it was about [sonorant] instead.

We have all devised work-arounds, like dropping the lowest couple of homework or quiz grades, allowing the final exam to count for more if the student does well, or bumping a borderline student up to the next letter grade if they show an upward trend over the term. But these seem like patches on a fundamentally flawed system. And then there is the daily tedium of adding up points, tracking how many points we are deducting for each type of error, and deciding which error types are better or worse than others.

In developing our grading system, we followed a list of requirements. First, grades should reflect whether students eventually mastered the material, not how long it took them to master it. Accordingly, students should have multiple opportunities to demonstrate their mastery of each skill. Second, grading should be easy. Each question would be marked either right or wrong, with no partial credit. However, students who are struggling should still have an opportunity to demonstrate and receive credit for what they *do* know. Third, there should be a meaningful distinction between simple mastery—a B grade—and going beyond that level—an A grade.

As detailed below, our system in the phonology course is to tag each quiz or exam question for the skill it tests. If a student gets that question right, they receive credit for having mastered the skill; if they get it wrong, they can try again on a subsequent assessment. In order to distinguish A and B grades, there are opportunities for students to demonstrate advanced mastery of skills. In principle, a student could master and then later forget a skill, and our system would not penalize this—wrong answers later in a course don't undo the earlier credit. But in practice this has not been a problem, because the skills themselves are mostly cumulative. For example, if a student forgot basic distinctive-feature skills, they would not be able to apply phonological rules correctly; if they forgot how to apply phonological rules, they would not be able to argue for a rule ordering. In our phonetics course (section 4), students did need to demonstrate mastery of the same skill group multiple times, so it would be difficult to attain full credit for proficiency if a student forgot a skill mid-way through the course.

Phonology homework assignments work similarly to quizzes and exams, except that it is up to the student to tell us which skill they are using in which part of their solution. We also allow students to mop up credit for some additional skills in their individual final project, by pointing out to us which skills they are using, or using at an advanced level. We detail in the following section some results we hoped to achieve.

2. BACKGROUND

There were several results we hoped our grading approach would achieve. First, student stress and temptation for cheating should be reduced, and student motivation increased (Buckmiller et

al. 2017), because of the low stakes on each assignment or test: until the final exam, there is always another opportunity to demonstrate mastery on some skill. Assessing skills frequently would benefit students (National Research Council 2000, Brown, Roediger & McDaniel 2014), both by directly increasing their learning and by making both students and instructor(s) more aware of student progress, with clear criteria and objectives (Buckmiller et al. 2017). These criteria would tell students what areas they needed to work on, and how to focus their studying and office-hours visits. If students want to compare their grades-in-progress to their peers', then rather than just seeing a number, they can see how many peers have mastered which skills.

Students' meta-cognitive awareness should increase, as should overall comprehension of material (Iamarino 2014). They should become more aware of the ingredients that go into solving a phonology problem, for example. Additionally, a growth mindset (Dweck 2008) should be promoted: students can see that they are learning to do things they couldn't previously do, and to treat proficiency in each skill as something that can be worked on and improved.

This focus on a growth mindset is important when taking into account the different reasons students may be taking the course. While some students may be taking it because of academic interest, other students will be taking it for more practical reasons like major requirements. Students with weak motivations for taking the course may be easily discouraged by poor performance early in a points-based course, but a skills-based course could avoid that issue. Since students have multiple chances to obtain a skill, early failures are to be expected instead of dreaded. Students' attention is focused more on their progress in obtaining skills and using past failures to improve their performance.

While partial credit can be appealing as a way to encourage students to attempt an answer, Nilson (2014) found that giving partial credit for wrong answers can encourage students to just write down something—anything—in hopes of scoring some points. The lack of partial credit in a skill-based system would encourage students to aim to understand a concept and correctly apply it.

In terms of equity, grades derived from a skill-based-system should be fairer to students who enter the class with inadequate background, or whose performance is hampered at some point by medical, family, personal, or financial problems. Given the structural factors that make students from underrepresented minority groups, first-generation college students, and students

with disabilities particularly subject to such external stressors, it is important to ensure that students who eventually master the material are not penalized for temporary setbacks.

Following Schinske and Tanner (2014), we hoped that spending less time grading would give us more time to dedicate to pedagogical reflection, development, and innovation.

Furthermore, having a tally of how many students achieved proficiency in each skill makes it easy to comply with new accreditation requirements that we report how well students have achieved our stated learning outcomes.[1] Most of the course skills ended up being straightforward instances of our department's stated curricular goals (e.g., the ability to write up technical material); if we were designing a new course, we could also operate in a more top-down fashion, with curricular goals driving the choice of skills to include.

What we have done is not new. Similar approaches go under the names *standards-based grading* (see Schimmer 2016 for an overview and guide), *mastery-based grading* (e.g., Armacost and Pet-Armacost 2003; Brackett and Reuning 1999), and *specifications grading* (Nilson 2014). There are many variants, but all these approaches have in common that students' final grades should reflect how well they have mastered the course material, not how long it took them to get there. This type of grading is used more often in K-12 districts seeking school- or district-wide standardized criteria (such as Spokane Public Schools, as detailed in Iamarino 2014). Implementation of standards/skills-based grading in higher education is growing, but nascent; the majority of post-secondary institutions in the U.S. using this type of grading (Nodine 2016) have done so for five years or fewer.

The following section discusses in detail how our system worked for three offerings of Phonology I, and section 4 describes how we adapted it to an introductory phonetics class.[2] Section 5 provides student reactions, 6 provides discussion, summary, and future directions, and in 7 we offer suggestions on how others can adapt our system. We also include, as supplementary materials, course syllabi and instructions for three sample phonology assignments.

3. DESCRIPTION OF INNOVATION, PART 1 GRADING SYSTEM FOR PHONOLOGY I

At our university, students in Phonology I have already taken a phonetics course (see section 4), so phonetics is not covered except as review. We used as our textbook Hayes's *Introductory Phonology* (Hayes 2011), and covered the chapters on phonemic analysis, features, basic morphology, phonological alternation, morphophonemic analysis, productivity, the role of morphology and syntax, diachrony and synchrony, abstractness, syllables, stress, and tone/intonation. About a third of the students go on to take Phonology II, which covers topics including Optimality Theory, Lexical Phonology and Morphology, autosegmentalism, and prosodic constituency.

3.1. SKILLS FOR PHONOLOGY I

We identified skills that students should master coming out of Phonology I, listed in Appendix A. In the first iteration of using skills grading for Phonology I, we started with 54 skills, and abandoned five of them. (In the following, numbers in parentheses refer to the skills described in the table in Appendix A. Curly brackets (3.2), Greek letter variables (3.4), and optionality (3.13) ended up not being used enough times, and we did not have room in the course to add more teaching and assessment of those skills. Rule ordering for explaining distribution (5.7) and for explaining alternation (6.6) were found to be redundant with general rule ordering (4.1).) After eliminating these five skills and adding one new one (7.4, invent data that would decide between two analyses), there were 50 skills in the subsequent iterations of the course.

Appendix A gives each skill's definition, as well as the code number and nickname that we used for ease of reference in grading and in instructions to students. The table also lists the percentage of students who achieved proficiency and advanced proficiency (described below) in the first two iterations of the course—this information was provided to students after the first iteration. We don't include percentages for the third iteration, because it was a summer course with only 20 students, and these numbers would be less reliable.

For our non-phonologist readers, we offer here brief explanation of a few example skills. FeaturesToSubset (1.1) is an example of a very mechanical, low-level skill: given a phone

inventory such as {a, e, i, o, u} and a set of feature specifications such as [+round, -high], the student must give the subset of phones that these features pick out ({u}). This is an essential part of applying a phonological rule—and therefore for reading a phonological analysis or for proposing one, since the student must be able to test whether their own rule works. UndergoersNonundergoers (2.3) is a skill needed for developing and testing an analysis (or taking an active approach to reading an analysis): given a rule, such as [+high, +syllabic] → Ø / __]word (high vowels delete at the end of a word), the student must invent forms that would be subject to the rule (e.g., /falu/, which would change to /fal/) and forms that would not (e.g., /hala/, /pik/). FeedingBleeding (4.2) is an example of a high-level skill that depends cumulatively on many others: given two rules, the student must say whether their order is feeding, bleeding, etc. This requires understanding how the rules apply, considering actual and counterfactual derivations, and thinking of hypothetical input forms that might not be in the data provided.

The skills were not limited to practical aspects of solving phonology problems, but also covered conceptual learning. For example, classifying the distribution of sounds (skill 5.4) asks students to compare two sounds in a data set and determine whether they are separate phonemes that contrast in all contexts, separate phonemes that neutralize in some context(s), or allophones of the same phoneme. Achieving a correct answer depends on practical skills such as extracting (5.2) and summarizing (5.3) the environments of the sounds, but coming to a correct conclusion requires conceptual understanding. Some skills were expository, such as explaining why a rule is needed (7.2).

3.2. EARNING CREDIT FOR SKILLS IN PHONOLOGY I

There are four ways for a student to demonstrate proficiency on each skill: weekly quizzes, midterm and final exams, near-weekly homework problems, and a final paper.

Each week we have a short quiz, made up of mechanical problems that can be solved quickly. Figure 1 shows part of a typical quiz from towards the beginning of the course, and Figure 2 shows a quiz question from towards the end of the course. Hypothetical correct answers are added in a different font. The boxes on the right show students which skills they can demonstrate proficiency on, and provide a convenient spot for the instructor or teaching assistant

to check off how the student did, for later entry into the grade spreadsheet. The "developing" option does not affect student grades, but is there to let the student know that their answer is on the right track but not quite correct.

Advanced proficiency can be demonstrated in several ways: (1) Some questions are difficult enough that we judged a correct answer as evidence of advanced mastery. Examples of this can be seen in the homework assignments in the supplemental materials, where the basic data to be explained are followed by advanced data requiring the solution to be modified or extended. Inherently advanced questions also occurred in midterm and final exams (and were flagged to students as such). (2) For some questions, two answers may both be correct but one is advanced by virtue of being more explicit, fuller, or better (as in the Turkish example in Figure 2, discussed below). (3) Less commonly, an answer could be advanced because it surprised us by showing more conceptual mastery than we were requiring at that point in the course. Because these cases were not expected or defined beforehand, identifying them was more subjective and sometimes required further discussion between instructor and teaching assistant.

<INSERT FIGURE 1 ABOUT HERE>

To be counted as proficient in either skill in Figure 1, a student must give an answer that is completely correct. No partial credit is awarded. The principle is that students can try multiple times to demonstrate that they can do something, but must eventually do it correctly. This means that for skills like CombineRules, where there are many ways an answer could go wrong, we need to offer students many attempts.

<INSERT FIGURE 2 ABOUT HERE>

For the question in Figure 2, we found that students were having trouble articulating exactly how domain edges work in a rule, so we decided that a fully explicit answer like the hypothetical one given above would count as advanced. (We also allocated more time to this concept in response.) A correct but less-explicit answer, like "the rule only applies across a stem boundary", would count as proficient.

Midterm and final exams are in a similar format, but more challenging, with more opportunities to demonstrate advanced proficiency. The exams aim to provide opportunities to demonstrate all skills seen so far; this makes them appear long, but students are not expected to

answer all the questions. The first part of each exam is a series of short, stand-alone questions. Students are encouraged to bring in a list of which skills they still needed to demonstrate proficiency on, so that they can choose which questions to spend time answering. The second part of each exam is an extended problem that builds up piece by piece, becoming more advanced as it goes on. A student can stop working on the extended problem at any point, but it would be difficult to jump into the later parts of the problem without having first done the earlier parts. Rarely, a question's skill label must be redacted because it would give away the answer.

Nearly every week there is a homework assignment. This is where most of the opportunities for advanced proficiency occur. Data to be analyzed in these assignments are usually divided into a basic portion and an advanced portion. Students are encouraged to first solve the basic portion, and then attempt the advanced portion if they have time. We believe this is more encouraging to students than our former practice of giving all the data, and assigning higher grades to the students who managed to solve the more difficult aspects. That practice left most students feeling frustrated and wondering if something was fundamentally wrong with their solution, when perhaps they merely needed to add an extra rule. With the data divided into two parts, it is usually clear to students whether or not they have a solution that works for the basic data; and if they do have one, they receive a good amount of credit even if they haven't solved the advanced data.

Homework instructions tell students which skills they will be using. But, students must indicate in the margin of their paper which skill they use where. This is partly to make grading feasible (the grader does not have to search the paper for possible evidence of each skill), and partly to encourage students to reflect on their own problem-solving and realize which parts of class material they are applying to which parts of the problem. Students are free to incorporate a skill not mentioned in the instructions. Supplementary files include complete instructions for three Phonology I homework assignments.

Students' final opportunity to earn credit for skills is on the final individual paper. The paper is worth 25% of the final grade (including stepping-stone assignments such as an elicitation plan), and is graded in a traditional way, with a grading rubric. But if there were skills students were still missing that they realized they were using in their papers, they could indicate

this in the margin. Most students availed themselves of this opportunity at least for advanced proficiency in a couple of skills.

Once a student demonstrates proficiency in a skill, they are not penalized for previous failures on that skill, but they are also not penalized for later failures. In other words, all that mattered for each skill is whether there was at least one successful attempt (as well as whether advanced proficiency has been demonstrated). We run the risk that a student could forget a skill and their grade would not reflect this. As discussed in the introduction, this turned out not to be a problem, because skills are fairly cumulative. To make progress on later-in-the-course skills, the student needs to retain mastery of earlier skills. If we had unlimited opportunities for assessment we could use a criterion like "has the student succeeded on at least four out of the five last attempts?" to ensure that mastery is retained, but this wasn't feasible in a one-term course.[3]

3.3. ASSIGNING GRADES IN PHONOLOGY I

Skill proficiency was worth 75% of the final course grade. (The other 25% was for a final project.) The basic standards for the skills portion of students' final grade are given in (1).

(1) Basic grade standards for Phonology I

   i.   To earn an A (i.e., middle of the A range: 95%), demonstrate proficiency on all 50 skills, and advanced proficiency on at least 15 skills (reduced from 20 in the first iteration)

   ii.   To earn a B (85%), demonstrate proficiency on all 50 skills

   iii.   To earn a C (75%), demonstrate proficiency on all 33 core skills

   iv.   To earn a D (65%), demonstrate proficiency on at least 25 skills

This was a simple scheme, but where it became a bit complicated was assigning numerical scores to intermediate situations, such as proficiency on all 50 skills but advanced proficiency on just 10; or to mixed situations, such as proficiency on only 48 skills (does not meet standard for B), but advanced proficiency on 15 skills (exceeds standard for B). We used the following formula, admittedly more complex than we would like:

(2) Final grade = whichever is higher of the following two

i.   grade from (1)

ii.   50% [*baseline]*

+ 1%*minimum_of(25, number_of_core_skills_proficient) [*1% each for first 25 core skills]*

+ 0.5%*maximum_of(0, number_of_core_skills_proficient-25) [*0.5% each for add'l core skill]*

+ 0.5%*minimum_of(number_of_skills_proficient-number_of_core_skills_proficient, 20) [*0.5% each for non-core]*

+ 0.5%*minimum_of(number_of_skills_advanced,15) [*0.5% each for first 15 advanced]*

+ 0.33%*maximum(0, number_of_skills_advanced-15) [*0.33% each for add'l advanced]*

Thus, the hypothetical student proficient on all 50 skills and advanced on 10 would earn the greater of 85% or 50% + 1%*25 + 0.5%*8 + 0.5%*17 + 0.5%*10 = 92.5%. And the student proficient on 48 skills (including all the core skills) and advanced on 15 would earn the greater of 75% or 50% + 1%*25 + 0.5%*8 + 0.5%*15 + 0.5%*15 = 94%.

In order for this formula to correspond exactly with the scale in (1) (especially at the lower ranges), the number of core skills will need to be adjusted in future. Although the grading formula was complicated, students were able to try out different scenarios in the grading spreadsheet to see how their grade would change. In the phonetics course, discussed next, we came up with a simpler formula, and will devise a simpler formula for phonology in the future.

4. DESCRIPTION OF INNOVATION, PART 2: GRADING SYSTEM FOR INTRODUCTORY PHONETICS

The challenge with adapting our grading system to Introductory Phonetics was that the number of skills to be mastered was either smaller or *much* larger. It was smaller in the sense that we identified just 15 skill groups (listed in Appendix B), such as defining a phonetic symbol. But it was much larger in that each skill group could include hundreds of skills, such as defining each phonetic symbol, including combinations with diacritic marks. It is not feasible to give students multiple opportunities to define *each* phonetic symbol, so we needed to arrive at a compromise between skills grading and traditional point accumulation.

4.1. SKILLS FOR INTRODUCTORY PHONETICS

Reluctant to give up the benefits we had experienced in Phonology I, we arrived at a compromise system where students increase their grades by demonstrating proficiency in each skill group a certain number of times. The idea is that if a student can correctly define 15 IPA symbols, this represents a reasonable sample of all the symbols. Appendix B shows a table of skills used in Introductory Phonetics. Of course, it would be possible for a student to systematically misdefine, say, the front rounded vowels, and still be counted as proficient. But it would not be possible for a student who systematically misdefined, say, manner of articulation to achieve proficiency, because they would have errors on all consonant definitions—as in the phonology course, an answer must be completely correct to earn credit.

4.2. EARNING CREDIT FOR SKILLS IN INTRODUCTORY PHONETICS

Students can demonstrate skill proficiency in weekly quizzes, a small number of homework assignments, and an in-class listening and transcription exercise (listed in Appendix B as *language demo*). Because there is no midterm or final exam, we found that towards the end of the course we have to offer a quiz twice a week, and each quiz towards the end of the course includes a large number of questions that students can choose from. (The questions they choose not to answer or don't have time for can be used as study material after the quiz is handed back.)

Unlike in the phonology course, a skill is not all-or-nothing (proficiency achieved or not). Rather, students receive some credit for how many times they demonstrate proficiency in each skill, up to full credit for succeeding the required number of times. Because it takes some time to accumulate the required number of successes, this means that students can't, for example, segment one spectrogram correctly and then forget how—or rather, if they do, they will receive a lower grade than if they continue to succeed on nine more spectrograms to reach the required ten. Requiring multiple successful demonstrations of proficiency for crucial or complicated skills

is a good option for instructors who are concerned that a single demonstration may allow skill attrition or incomplete skill mastery.

Another difference from the phonology course is that earning an A on the skills here does not depend on achieving advanced proficiency in some skills—we do not employ a concept of advanced proficiency in this course. Instead, quizzes include a question that invites students to apply their knowledge to a situation or problem that we have not seen in class. Figure 3 gives a sample of quiz questions, with a hypothetical answer added in a different font. As in Phonology I, boxes in the margin tell students which skill was being tested, and provide a spot for the grader to record a correct answer, to be added to the gradebook.

<INSERT FIGURE 3 ABOUT HERE>

The skills grading system allowed us to try something new with a homework assignment. The students' first assignment is to download sound files of an English speaker saying words and sentences, and use the TextGrid function of Praat Boersma and Weenink 2017 to transcribe and segment it. This is difficult for beginners, and students make a lot of mistakes. We decided to divide the assignment into two parts. Students first turn in their segmentations of individual words, then get feedback on their errors. Then, students repeat the task with full sentences. This allows students to put into practice what they've learned from their first try. Students are told that they shouldn't worry about accumulating credit for the first assignment—though they will receive credit if they get items correct—but should rather treat it as a practice run to learn from. Students ended up doing a better job on the sentence transcription and segmentation than they had in the past. (Those who haven't mastered the task in the second half of the assignment, but go on to master it later, can still demonstrate their proficiency in later quizzes.)

4.3. ASSIGNING GRADES IN INTRODUCTORY PHONETICS

Skill proficiency was worth 60% of the final grade. The remainder was 25% for an individual project and 15% for an individual production exam.

(3) Basic grade standards for Introductory Phonetics

    i.   To earn an A (i.e., middle of the A range: 95%), demonstrate proficiency in all 15 skills

    ii.   To earn a B (85%), demonstrate proficiency in skills except skill 5, *Apply*

    iii.   To earn a C (75%), get three-quarters of the way to proficiency in all skills except skill 5

    iv.   To earn a D (65%), get half-way to proficiency in all skills except skill 5

To assign numerical grades for mixed or intermediate scenarios, we used the following formula:[4]

(4) Final percentage grade = whichever is higher of the following two

    i.   grade from (3)

    ii.   45 [*baseline*]

        + for each skill other than Apply

        (40/14)*minimum_of(1, proportion_ attained) [*up to 2.86% for each skill besides* Apply]

        + 2*number_of_successes_on_skill_5 [*2% each for successful* Apply]

As in the phonology course, students could use the grading spreadsheet to try out different scenarios and see how their final grade would change.

5. STUDENT REACTIONS

We invited students enrolled in the four courses to participate in an anonymous online survey,[5] and 24 students responded. They were invited to offer comments on the following questions—we summarize themes that were mentioned by more than one student and quote some notable responses.

Looking back over the course, what are some of aspects of the grading system that worked well for you?

    (5)   Less worry and stress about initially not understanding a topic; there is time to master material gradually. Students felt that they had more control over their grade. Benefitting from reduced stress led to greater enthusiasm and enjoyment of the

course. There was more focus on learning rather than testing. "I felt less pressured/demotivated by failure, because I could see it as a missed opportunity rather than a detriment to my grade"

(6) It is clear what students are supposed to be learning, and what they should study

(7) The system encouraged spread-out studying rather than cramming. Students retained more material than in other courses. "It prompts me to keep studying the parts I don't understand till I really mastered them"

(8) Liked being able to track one's improvement and progress

(9) It was enjoyable to have the opportunity to work on more-advanced data in the problem sets when possible

(10) Frequent quizzes led to greater learning (e.g., successfully memorizing whole IPA chart)

What are some aspects of the grading system that worked badly for you?

(11) 7 respondents answered "None," "NA," or similar

(12) It took time to get used to the system.

(13) (For Phonology I) It was stressful trying to achieve enough advanced proficiencies, and what counted as advanced was often unclear or subjective. There should be more opportunities to demonstrate advanced proficiency

(14) (For Phonology I) It was disappointing to work hard on an assignment and still not get it right

(15) (For Phonology I) It was overwhelming how many skills there were to master

(16) (For Introductory Phonetics) Need additional opportunities to take quizzes, such as outside of class. One respondent, however, commented that it was stressful to have so many quizzes, and parts of them were too hard

(17) "The amount of control a student has can sometimes be overwhelming. I remember spending countless hours on assignments from both 103 and 120A. As I am writing this I am not so sure that it is something "bad" about the system."

(18) "I found myself skipping out on some homework assignments because I relied on the quizzes, midterm, or final to satisfy the rest of the skills needed."

(19) "When I couldn't understand or acquire a skill, it was sometimes hard to go back and work on it."

If the system is used in the future, what changes would you recommend?

(20) 10 respondents answered "None," or similar

(21) (For Phonology I) Provide clearer criteria for what will count as an advanced answer

(22) (For Phonology I) Make numerical formula for final grade simpler; make the grade spreadsheet easier to read

(23) Change how gradebook works, e.g. to just show what percentage of the class is proficient so far: it could be stressful to see that other students are doing better (See the question after next for more responses on this point)

What advice would you have to a fellow student wondering whether they should take a class that uses skills grading?

(24) Would recommend it

(25) This is less stressful. "You are an active participant in the learning process". Your grade is in your hands — directly proportional to effort

(26) You will learn the material thoroughly because you can focus on areas where you need more work rather than those you understood easily. "I had the opportunity to focus on specific goals rather than just a grade. The system made learning challenging material significantly less stressful."

(27) "Try to get ahead ASAP"

(28) "I would simply explain to them how the system works, whether or not they like it is their own choice."

(29) "Don't stress early if things are tough, just continue learning and focus where you are struggling"

(30) Don't forget past material

(31) "It is a good way to really be forced to learn something lol. You will not be safe to just learn before [the] exam and forget everything right after."

In this class, you had access to a coded gradebook that included everyone's grades. Although you could not see their names, how did you feel about being able to compare progress with your peers?

Posting an anonymous gradebook with codenames was really a technical decision for us, but, as discussed in section 6, we hope to have other options in the future. In order to make good decisions about how to exercise those options, we asked students what they thought of the current gradebook.

(32)   17 respondents like being able to see how others were doing. It was useful to see where the student was ahead or behind compared to others—helped to focus studying. It was useful for group study: helped identify skills where many people were struggling. For students with a competitive personality, it was motivating.

(33)   2 respondents disliked being able to see how others were doing. "I felt bad about it sometimes, when I was struggling. I think it would be better to not show relative comparison since there isn't a curve anyway."

(34)   4 respondents felt ambivalent, mostly along the lines of "it was fine for me, but could have been stressful for others" (1 respondent skipped this question).

We asked respondents to rate their level of agreement with a series of statements; results are shown in Figure 4.

<INSERT FIGURE 4 ABOUT HERE>

We did not carry out a comparable survey of students who'd taken the courses in earlier years under traditional grading, but we did examine anonymous student evaluation forms from those students, looking for comments that related to the assessment structure of the course. In the phonology course, two students commented that there was not enough time for exams or quizzes, one that it was stressful to have a final exam, and one that homework should have been weighted lower or the lowest homework grade dropped. In two offerings of the phonetics course, four students commented that quizzes were too difficult, too numerous, or too long for the time allotted; one expressed a wish for grading on a curve. We then looked at student evaluation forms from the skills-grading offerings of the courses and didn't find these comments, even though the quiz and exam content were very similar. (As noted above there was one student who

responded in the survey that having so many quizzes in phonetics was stressful.) Presumably this difference in reaction to frequent quizzes arises because under skills grading, a student's grade can only benefit from numerous quiz and exam opportunities, and because students did not need to answer all questions on a test and thus felt less time pressure.

As a reviewer suggests, rather than only comparing student reactions we would ideally compare student *performance* in the two grading systems, say by administering the same exam to both groups. (And ideally students and instructors would be randomly assigned to the two conditions.) This will probably never be feasible for upper-level linguistics courses at one university—there aren't enough students—but could be done for introductory linguistics courses that enroll hundreds of students each term.

6. DISCUSSION AND FUTURE DIRECTIONS

We have been happy with our grading experiment, and don't plan to go back to traditional grading. Although we lack quantitative data that would make it possible to measure the success of our goals, our and our students' subjective sense is that many of our goals were met.

In terms of pedagogy, this grading method decreased student stress over grades and increased student motivation, including maintaining motivation to overcome difficulties. Knowing which skills each student was struggling with allowed for better focus in students' study and office-hours visits.

In terms of grades, it was easier to comply with accreditation requirements to report results on desired learning outcomes. The final grades were also fairer. We observed several students who struggled early on, and whose poor early grades would have doomed them to a C at best under traditional grading, but who eventually mastered the material and received final grades in the B range, which we believe fairly reflected their final mastery. We can compare this to our conventional grading in previous iterations of the courses. There, each quiz, exam, problem set, etc. was worth a certain number of percentage points towards a final grade. We often would drop the lowest weekly quiz grade, and calculate the average quiz grade from the remainder. How each item was graded was also conventional: each question (or aspect of a problem set) was worth a certain number of points, with a fully correct answer earning all of

those points. Grading this way was tedious and introduced a great deal of subjectivity. Deciding ahead of time how many points each question should be worth was a subjective process, as was deciding which types of errors should result in deducting how many points. We would have to grade a few papers, noting common error types, and develop an informal point-deduction guide: for example, a feature set that is correct but redundant gets two points off, omitting a needed feature gets three points off. Especially on problem sets, where student responses are more open-ended, even classifying an error type was hard to do objectively: for example, if the student's solution doesn't produce correct results, should we blame incorrect underlying representations (worth 10 points) or incorrect rules (worth 20 points), given that the two sets of analytic decisions are interdependent? If two or more graders were at work, we had to coordinate to make our point deductions consistent. If we altered the rubric mid-way, we had to go back and re-tally. After the papers were returned, sometimes a student would successfully argue that they shouldn't have been docked so many points for a certain error type, in which case we would have to ask everyone who made that type of error to turn their papers back in for re-grading. Because we follow the typical US grading standard of 90-100% = A, 80-89% = B, 70-79% = C, 60-69% = D, and 0-59% = F (a failing grade that means the student needs to repeat the course in order to get credit for it), when we used to grade conventionally we had the awkward result that a missing homework assignment or quiz often hurt a student's final grade far worse than turning in work of failing quality. If an assignment was worth 10 percentage points of the final grade, not turning it in meant the student loses all 10 points; doing extremely poor work might still, with partial credit, result in an F grade that lost only 5 or 6 points. Effectively, we assumed that if the work was not turned in, the student had no mastery whatsoever of the material. This assumption is almost always false, but it's hard to see a fair way around it under conventional grading.

It was harder to assess, even subjectively, whether other goals were met (increased student learning from frequent assessment, increased meta-cognitive awareness, and aiming for full understanding because of lack of partial credit).

The grading process was faster, because it wasn't necessary to keep track of or readjust how many points were being deducted for which type of error. On the other hand, it could be time-consuming to maintain the gradebook, which we devised in Excel with multiple worksheets and complicated formulas. We have talked to the information-technology staff who run our

university's online gradebook service, and they expressed enthusiasm about working with us in the future to add functions to the standard online gradebook that would make it flexible enough to do what we need. This would save us from needing to set up the elaborate Excel spreadsheet, generate code-names and send them to students, and anonymize and re-post the gradebook every week. We also hope to let students see which proficiency credit comes from which test or assignment, so that if they think there is an error they can track it down easily. And ideally we would like the gradebook to give the students virtual badges to commemorate achievements such as "Proficient in 10 skills," "First advanced skill," or "All feature skills achieved." A modest amount of gamification here (i.e., applying design elements from games: Deterding et al. 2011; Kapp 2012; Hamari et al. 2014 could stimulate student motivation without compromising educational goals. A more ambitious extension would be to include the option of "leveling up": when a student has demonstrated certain proficiencies, they could unlock access to more-advanced questions. Leveling up would be more workable in online quizzes, or an online course.

One goal of our grading system was to not put students at risk for failing if they missed a class or assignment, or otherwise got off to a rocky start. We especially didn't want to further disadvantage students when systemic discrimination makes it harder to get to class or prioritize class every week. By providing multiple attempts to demonstrate proficiency in skills, with no penalty for earlier incorrect or incomplete work, we allow students to catch up (to the extent that the term allows).

Marbouti et al. (2016) develop models to identify, early in a school term, students who are at risk of failing a course, based solely on their performance in that course so far. The results suggest that standards-based grading is helpful, because in selecting the best model (using past-year grades as training data), there is a possibility of identifying skills that are particularly predictive. We would like to explore this in the future, because having a good tool to identify students at risk would allow us to intervene, perhaps providing those students with extra drill or tutorial material on crucial skills. We could also simply encourage these students to visit us in our office hours, and then focus our conversation there on the crucial skills. Besides improving students' performance in our course and subsequent courses that draw on it, we would hope to improve student retention by avoiding the poor grades that can set a student on the path to

dropping out. Again, this is particularly important for students already facing systematic or institutional disadvantages.

At our university, phonetics and phonology courses are set in a long chain of prerequisites: introductory linguistics is a prerequisite for introductory phonetics, which is a prerequisite for Phonology I, which is a prerequisite for Phonology II. With some extra work to identify the skills from one course that are most needed for the next, we could implement better (ungraded) pre-tests at the beginning of the term. A better pre-test could tell us which skills generally need review because many students have forgotten or never mastered them, and could tell individual students which skills they need to review on their own.

In traditional grading using, there are two ways students can look at their grade in the course while it is in progress. If a student has earned 60 out of 100 possible course points, they can look at this as "my grade will be at least 60%." Or, they can see that the assignments and tests so far total to 70 possible points, and look at the result as "my grade so far is 60/70, or 86%." Our system only allows students to say "my grade will be at least 60%," and some students have expressed that they would like a way to estimate how they are averaging so far. Now that we have used the system at least once in each course, we have been able to go some distance towards meeting this desire, by showing students an estimate of how many chances are left on each skill (in Introductory Phonetics), and how many students achieved proficiency in each skill in the previous year (in Phonology I). Student suggestions for what the gradebook could add include showing, for previous years, what percentage of students had achieved proficiency in each skill by each week.

We would like to offer students more chances to demonstrate their proficiency (and learn from previous mistakes), especially for material introduced later in the course, such as gestural scores in the phonetics course. But there is only so much in-class time that can be spent on quizzes, and we have only so much time for grading. We would like to add on-line quiz opportunities (that are graded automatically), if we can figure out a good way to increase security and prevent cheating. It would help to have a large bank of equivalent questions that the on-line quizzes can draw randomly from, so that a student can't get an advantage from watching a friend take the quiz. It would also help to assign questions to the on-line quizzes for which we don't care whether students consult other materials. For example, "define the following phonetic

symbol" would be a poor choice of question for an on-line quiz, because the student can simply look up the answer in an IPA chart; but multiple-choice spectrogram reading is a better question, because the answer cannot be found anywhere.

Something we need to work on in the phonology course is making it clearer to students what counts as demonstrating advanced proficiency. For homework assignments, this is not difficult, and we think we have already made progress. For quizzes and exams it is more difficult to tell students what will count as advanced without giving away too much of the answer. Possibly, we should stop trying to include opportunities for advanced proficiency in quizzes, and for exams, include more questions labeled as advanced. If we let students know ahead of time that this is what they should expect, and how many opportunities for advanced proficiency they can expect each week, this should reduce students' uncertainty and anxiety about how to achieve an A grade.

7. SUGGESTIONS FOR IMPLEMENTING SKILLS GRADING

For instructors considering implementing something like this system for the first time, we offer the following questions to consider in course design.

1) If the course has been taught before, look through previous assignments and tests to identify recurring question types. What skills are they testing? If the syllabus lists learning outcomes, which outcome(s) does each question type develop or assess?

2) If the course is part of a degree program with defined learning objectives, how can they be translated into concrete skills (whether low-level or high-level) relevant to this course?

3) How much assessment in the course is of discrete, easy-to-grade skills, and how much is more holistic and subjectively graded? Should that material (e.g., an individual project) be a separate part of the course grade or integrated into skills grading?

4) Is the structure of the course material more like our phonology course, where there is a longish list of distinct (even if cumulative) skills? Requiring students to demonstrate proficiency once in each skill might be suitable. Or is it more like our phonetics course, where there is a smaller list of skill groups, each of which contains effectively unlimited items? In that case multiple successes in each skill group might be required.

5) What kind of mastery should be necessary for a passing grade? For instance, are there certain core skills that are necessary to advance to subsequent courses?

6) What kind of performance should differentiate a B, indicating mastery of the course material, and an A, indicating outstanding performance? As discussed above, in the phonology course we distinguished proficiency in a skill from advanced proficiency; in the phonetics course, we devised a whole different category of application questions.

7) How many opportunities will students realistically have to demonstrate their mastery of each skill?

8) How will the gradebook be managed, and how will students be able to track their grades?

9) What kind of grading formula will be used to translate skills mastery into a percentage or letter grade?

As the course is in progress, we recommend the following practices:

1) Communicate with students, often, about the rationale for the new grading system. Encourage students to use the system to their advantage, for instance coming to office hours knowing which skill(s) they want to work on.

2) Solicit student feedback early on, from technical issues (are students have trouble using the gradebook?) to design issues (is it clear what counts as advanced proficiency?)

3) Especially the first time using skills grading in a given course, be prepared to adjust expectations—for instance dropping a skill if it won't be covered sufficiently, or reducing the number of successes needed if it's clear students won't have enough chances.

4) Frequently monitor how many students have achieved proficiency in each skill. This information can be used to decide what to emphasize in classes and review sessions. Choose quiz or exam questions to allow opportunities for students to demonstrate the most-needed skills. Offer students a choice of which questions to answer, according to which skills they still need to pick up credit for.

5) If the course is small enough to allow individual attention, reach out to students who have not yet mastered foundational skills that are prerequisites for other skills to let them know what they should focus on. Make extra practice on those skills available.

6) Although grading should be easier, it also needs to be prompter, so that students know where to direct their efforts on the next assignment or test.

8. APPENDICES

APPENDIX A

List of skills used in Introduction to Phonology Course (with student attainment analytics). Code numbers marked with * were considered core skills (explained below). Code numbers marked with (*) were considered core skills in the first iteration of the course only.

category	*number and nickname*	*description*	% of students proficient in 1st iteration (% advanced) $n^6 = 38$ students	% of students proficient in 2nd iteration (% advanced) $n = 41$ students
Features	**1.1*** **FeaturesToSubset**	Given a phone inventory and a set of feature specifications, give the set of phones picked out by those feature specifications	97% (0%)	97% (0%)
	1.2* **SubsetToFeatures**	Given a phone inventory and a subset of that inventory, give the (smallest) set of feature specifications needed to pick out that subset	100% (3%)	94% (0%)
	1.3* **FeaturesToChange**	Given a set of phones and a set of feature changes, say what each phone changes to	100% (0%)	92% (0%)
	1.4* **ChangeToFeatures**	Given a set of phone pairs, give the (smallest) set of feature changes needed to take the first member of every pair to the second	92% (0%)	92% (0%)

	1.5* CombineRules	Use features to combine subrules into a single rule	82% (3%)	92% (0%)
Single rules	**2.1* NotationToProse**	Translate a rule into prose	100% (3%)	100% (0%)
	2.2* ProseToNotation	Translate prose into rule notation	100% (3%)	97% (0%)
	2.3 Undergoers Nonundergoers	Given a rule, invent some forms that would undergo it (and what they would change to) and some that wouldn't	100% (5%)	94% (6%)
	2.4 BeforeAndAfter	Given a set of "before" forms and a set of "after" forms, write a rule that could have changed the "before"s into the "after"s	100% (0%)	97% (0%)
Apply the single-rule skills to rules that include…	**3.1* Parentheses**	Parentheses, including expanding a rule with parentheses into its component subrules	97% (37%)	94% (28%)
	3.2 CurlyBrackets	Curly brackets	*abandoned*	*not used*
	3.3 Subscript Superscript	C_0 and other subscripts and superscripts	92% (0%)	83% (3%)
	3.4 GreekLetter	Greek letter variables	*abandoned*	*not used*
	3.5* SyllableNode	Syllable nodes	100% (0%)	100% (3%)
	3.6 StraySegment	C' and V'	97% (0%)	92% (0%)
	3.7* LightHeavyNotation	σ, ˘, and ‾	92% (0%)	89% (0%)
	3.8* StressFeatures	The features [stress] and [main]	100% (0%)	97% (0%)
	3.9 ToneInProseRules	Tone, but only in prose form	84% (0%)	89% (0%)
	3.10* DomainEdges	Domain edges (e.g., [word,]phrase)	100% (66%)	92% (31%)

	3.11* Bounding Domains	Bounding domains	100% (76%)	92% (44%)
	3.12(*) Iterative	A note that a rule is iterative	82% (24%)	100% (39%)
	3.13 Optional	A note that a rule is optional	*abandoned*	*not used*
	3.14 MinorRule Features	Exceptionality features, as in " → [– long]", which means the rule is a "minor rule" that applies only to forms that have the "opt-in" feature [+Shortening].	97% (0%)	94% (0%)
	3.15 OptOutFeature	A lexical entry can also have an "opt-out" feature like [–Vowel Deletion] ("The rule of Vowel Deletion doesn't apply to me.")	97% (0%)	89% (0%)
Rule interaction	**4.1* Order**	Determine what order two rules need to be in, or what changes if an order is reversed	95% (47%)	94% (50%)
	4.2 FeedingBleeding	Say whether two rules' ordering is feeding, bleeding, counterfeeding, counterbleeding, or none of these but the order does matter, or order doesn't matter	92% (34%)	92% (25%)
	4.3 Hasse	Draw a Hasse diagram of crucial orderings	95% (29%)	89% (14%)
	4.4* FillInDerivation	Fill in a blank derivation, which may include both morphological and phonological rules	100% (61%)	97% (50%)
Solving phonology problems: identify and analyze	**5.1* MinimalPairs**	Find minimal pair(s) in a data set	97% (18%)	100% (0%)

phonotactic patterns				
	5.2* Extract Environments	Extract a sound's environments from a list of data (e.g., t__a, etc.)	100% (3%)	97% (0%)
	5.3* Summarize Environments	Summarize those environments in prose (e.g., "before a vowel")	97% (13%)	97% (22%)
	5.4* Classify Distribution	By inspecting two or more sounds' environments, determine that they are… (and justify your answer): (i) allophones of the same phoneme (complementary distribution), (ii) different phonemes (similar distribution, maybe minimal pairs), (iii) different phonemes but in contextually limited contrast (distribution is partly similar but there are some environments where only one occurs)	100% (32%)	97% (11%)
	5.5 Phonemicization Diagram	Draw a phonemicization diagram	97% (50%)	94% (19%)
	5.6* RulesFor Distribution	Write one or more rules (in both notation and prose) to explain two or more sounds' distribution	82% (32%)	94% (6%)
	5.7 RuleOrderFor Distribution	Order rules if necessary	*abandoned—folded in to 4.1*	*not used*
Solving phonology problems: Identify and	**6.1* Morpheme Boundaries**	Locate morpheme boundaries in a set of data—be aware of alternate possibilities in case your first doesn't work out.	97% (47%)	97% (67%)

analyze alternations				
	6.2* MorphologicalRule	Write a simple morphological rule	100% (21%)	97% (39%)
	6.3* IdentifyAllomorphs	Identify the allomorphs of a morpheme (affix or root), and what phonological environments they occur in. Extract general phonological alternations. (E.g., if 'dog' has allomorphs [sat] and [sad], and 'cat' has allomorphs [kib] and [kip], you can say that morpheme-final stops seem to alternate in voice, even if you don't yet know what conditions the alternation, or what the underlying forms are.)	100% (63%)	97% (44%)
	6.4* UnderlyingForm	Determine a morpheme's underlying form—be aware of alternatives in case your first guess doesn't work out, including abstract underlying forms (that aren't identical to any of the surface allomorphs, or, as a last resort, that contain sounds not seen in any surface allomorphs)	100% (55%)	94% (53%)
	6.5* RulesForAlternation	Write rules (including bounding domain, if any) to explain alternations	89% (34%)	97% (53%)

	6.6 RuleOrderFor Alternation	Determine the ordering of your rules	*abandoned— folded in to 4.1*	*not used*
Illustrate/explain an analysis	**7.1* GiveDerivation**	Give derivations for suitably chosen examples	95% (53%)	94% (39%)
	7.2 ExplainRule	Explain why a rule is needed	79% (32%)	86% (22%)
	7.3 ExplainOrdering	Show why part of your ordering is necessary	87% (29%)	75% (22%)
	7.4 FurtherData	Explain (with invented examples) what data would be needed to decide between two analyses	*not used*	47% (36%)
Structure beyond the segment: syllables	**8.1 IdentifySyllable Rule**	Identify cases where syllable structure can improve a rule	58% (16%)	67% (8%)
	8.2* 3Part Syllabification	Apply the 3-part basic syllabification rule (Syllable Assignment, Onset Formation, Coda Formation)	97% (0%)	94% (3%)
	8.3 Persistent Syllabification	Apply syllabification persistently throughout a derivation	95% (0%)	92% (8%)
	8.4* Allowable Syllables	Determine what onsets and codas a particular language allows. Describe in terms of sonority sequencing, if that helps explain.	82% (29%)	89% (8%)
	8.5* Drawsyllable	Draw syllable structure with σs. If applicable, include the features [stress] and [main] on the syllable tier.	100% (0%)	100% (6%)

Structure beyond the segment: Stress	9.1(*) DescribeStress	If a language has predictable stress, state in prose where it falls (including possibly what counts as a heavy or light syllable)	92% (13%)	94% (22%)
	9.2(*) StressRule	Formulate a stress rule in notation (may include parentheses)	76% (58%)	97% (47%)
Structure beyond the segment: Pitch	10.1* DrawTone	Draw tones on the tone tier, with association lines.	74% (0%)	83% (0%)
	10.2 IPAToAutoseg	Translate IPA tone marks (e.g., ` and ´) into autosegmental tone representations (there may be more than one answer: e.g., does [ábó] have a H for each vowel, or a single H that the two vowels share?)	87% (0%)	97% (0%)
	10.3 AutosegToIPA	Translate autosegmental tone representations into IPA tone marks	87% (0%)	94% (0%)

APPENDIX B

List of skills used in Introductory Phonetics course (with student attainment analytics)

Category	Skill number and name	Skill description	# of times needed to show mastery	Where can you demonstrate it?
Phonetic symbols (IPA)	**1.1** DefineSymbol	Define a phonetic symbol	15	quizzes
	1.2 GiveSymbol	Give the phonetic symbol that fits a definition	10	quizzes
Transcription	**2.1** TranscribeMulti	Answer a multiple-choice transcription question	25	quizzes
	2.2 TranscribeC	Free-transcribe consonants	10	quizzes
	2.3 TranscribeV	Free-transcribe vowels	5	quizzes
	2.4 TranscribeTone	Free-transcribe tones	5	quizzes
	2.5 TranscribeKnown	Free-transcribe whole words in a known system: nonsense English words, or words in a language you've had a chance to hear baselines for	30	quizzes homework 3 language demo
Understanding articulation	**3.1** ReadVocalTract	Identify what a drawing of the vocal tract depicts (which place, manner, or IPA symbol)	5	quizzes
	3.2 GesturalScore	Translate a gestural score into IPA, or match it to a drawing of the vocal tract	5	quizzes
Acoustics	**4.1** SegmentSpectro	Segment a waveform/spectrogram	10	quizzes homework 1A&B homework 3
	4.2 ReadSpectro	Read a waveform/spectrogram	10	quizzes homework 2

	4.3 IdentifyDisplay	Identify whether an acoustic display is a waveform, spectrogram, spectrum; or, identify dimensions of a display	2	quizzes
	4.4 ReadFrequency	Use an acoustic display to determine or compare frequency (and explain how you did it)	2	quizzes homework 2
	4.5 HarmonicsFormants	Correctly use harmonics and formants: read them from a display, convert from one to the other, use the relationship, make or read a formant chart	2	quizzes homework 2
Application	**5** Apply	Apply your phonetics knowledge to a novel problem	5	quizzes

9. REFERENCES

ARMACOST, ROBERT L; and JULIA PET-ARMACOST. 2003. Using mastery-based grading to facilitate learning. *33rd ASEE/IEEE Frontiers in Education Conference*, T3A20-T3A25. Boulder, CO.

BOERSMA, PAUL; and DAVID WEENINK. 2017. Praat: doing phonetics by computer (Version 6.0.29) [Computer program]. Retrieved 2017 from http://www.praat.org/.

BRACKETT, CAROLYN C; and RICHARD H REUNING. 1999. Teaching phramacokinetics using a student-centered, modified mastery-based approach. *American Journal of Pharmaceutical Education* 63.272–277.

BROWN, PETER C.; HENRY L. III ROEDIGER; and MARK A. MCDANIEL. 2014. *Make It Stick: The Science of Successful Learning*. 1 edition. Cambridge, Massachusetts: Belknap Press: An Imprint of Harvard University Press.

BUCKMILLER, TOM; RANDAL PETERS; and JERRID KRUSE. 2017. Questioning points and percentages: Standards-Based Grading (SBG) in higher education. *College Teaching* 65.151–157.

DETERDING, SEBASTIAN; MIGUEL SICART; LENNART NACKE; KENTON O'HARA; and DAN DIXON. 2011. Gamification. Using Game-design Elements in Non-gaming Contexts. *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, 2425–2428. CHI EA '11. New York, NY, USA: ACM. doi:10.1145/1979742.1979575. http://doi.acm.org/10.1145/1979742.1979575.

DWECK, CAROL S. 2008. *Mindset: The New Psychology of Success*. Ballantine Books.

HAMARI, J.; J. KOIVISTO; and H. SARSA. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. *2014 47th Hawaii International Conference on System Sciences(HICSS)*, 3025–3034. doi:10.1109/HICSS.2014.377. doi.ieeecomputersociety.org/10.1109/HICSS.2014.377.

HAYES, BRUCE. 2011. *Introductory Phonology*. John Wiley & Sons.

IAMARINO, DANIELLE L. 2014. The Benefits of Standards-Based Grading: A Critical Evaluation of Modern Grading Practices. *Current Issues in Education* 17. https://cie.asu.edu/ojs/index.php/cieatasu/article/view/1234.

KAPP, KARL M. 2012. *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. John Wiley & Sons.

MARBOUTI, FARSHID; HEIDI A. DIEFES-DUX; and KRISHNA MADHAVAN. 2016. Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education* 103.1–15. doi:10.1016/j.compedu.2016.09.005.

NATIONAL RESEARCH COUNCIL. 2000. *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. National Academies Press.

NILSON, LINDA B. 2014. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*. Reprint edition. Sterling, Virginia: Stylus Publishing.

NODINE, T. R. 2016. How did we get here? A brief history of competency-based higher education in the United States. *The Journal of Competency-Based Education* 1.5–11. doi:10.1002/cbe2.1004.

SCHIMMER, TOM. 2016. *Grading From the Inside Out: Bringing Accuracy to Student Assessment Through a Standards-Based Mindset*. Bloomington, IN: Solution Tree.

SCHINSKE, JEFFREY; and KIMBERLY TANNER. 2014. Teaching More by Grading Less (or Differently). *CBE—Life Sciences Education* 13.159–166. doi:10.1187/cbe.cbe-14-03-0054.

10. NOTES

[1] We would like to thank one of our anonymous reviewers for this observation.

[2] The authors are (i) the instructor for two of the phonology courses and the phonetics course, (ii) the graduate teaching assistant (TA) for one of the phonology courses, (iii) the TA for the phonetics class, and (iv) the TA for the second phonology course and the instructor for the third (which had no TA).

[3] We also didn't want grades to depend only on the final attempt at each skill, for two reasons. First, it reintroduces the element of luck that we were trying to avoid: if a student's final attempt at rule-ordering falls on a day they were feeling unwell, or involves a problem that stymied them for a reason unrelated to the target skill, they will be penalized. Second, we would have to abandon our practice of allowing students to choose which questions to answer on exams (otherwise they would simply never make another attempt once they had succeeded on a skill), and lose the associated benefits (less stress for students on exams, being able to offer a full range of questions even though no one would be able to answer all the questions in the time allotted).

[4] We actually used something a little more complicated, but then devised this improved formula for the subsequent offering of the phonetics course.

[5] We applied for and received approval from UCLA's Institutional Review Board system to conduct this survey, as well as to summarize comments from anonymous student evaluation forms from earlier offerings of the courses.

[6] *ns* exclude students who dropped the course or took a grade of Incomplete, since there were many skills they never attempted.

1. FIGURES

6. Assuming the inventory shown, **use features to combine** these three rules into a single rule (your answer should have only features, no IPA symbols).

Skill 1.5, **CombineRules**
☐ approaching
☐ proficient

p
b   t   k
d   g   i   u
f   θ   s   x   e   o
v   ð   z   a
m   n   ŋ
ɹ
l

b → m / {ɹ, l} ___
d → n / {ɹ, l} ___
g → ŋ / {ɹ, l} ___

[+voice, -cont] → [+nasal, +son] / [-syll, +approx.]

7. **What feature changes are needed** to change the sounds on the left to those on the right? (Just fill in the empty double lines—ignore the rest of the rule)

Skill 1.4, **ChangeToFeatures**
☐ approaching
☐ proficient

b → p
d → t
ʒ → tʃ
g → k

.... → ‖ -voice
         -continuant ‖ / ...

*Figure **Error! Main Document Only.**: Quiz questions from early in the course (Week 3 of a 10-week course)*

1. Here is a rule from Turkish (from Inkelas 1988). Based on the data below, **fill in what's missing in the oval** (the label for the domain edge).

3.10, **DomainEdges**
☐ developing
☐ proficient
☐ advanced

Velar deletion → Ø / __]

( ]stem)

Data (all of these are surface forms):

[[bebek]stem]word   'baby'   [[bebe]stem i]word   'baby-*accusative*'
                                      [[bebe]stem e]word   'baby-*dative*'
[[sokak]stem]word   'street'   [[soka]stem i]word 'street-*accusative*'

Explain why 'street' isn't *[soak] (you shouldn't need all this space–a sentence or two will suffice)

Although the middle /k/ in /sokak/ is [-syll, +DORS], and it is followed by a [+syll] sound, there is no intervening stem boundary, as the rule requires.

Figure Error! Main Document Only.: Quiz question from late in the course (Week 7 of a 10-week course)

2. Warm-up: Write the English word *taxes* in IPA: _____ 'tʰæksɨz _____ Here is a waveform of *taxes*–segment it by drawing vertical lines and writing the IPA symbols underneath in the box
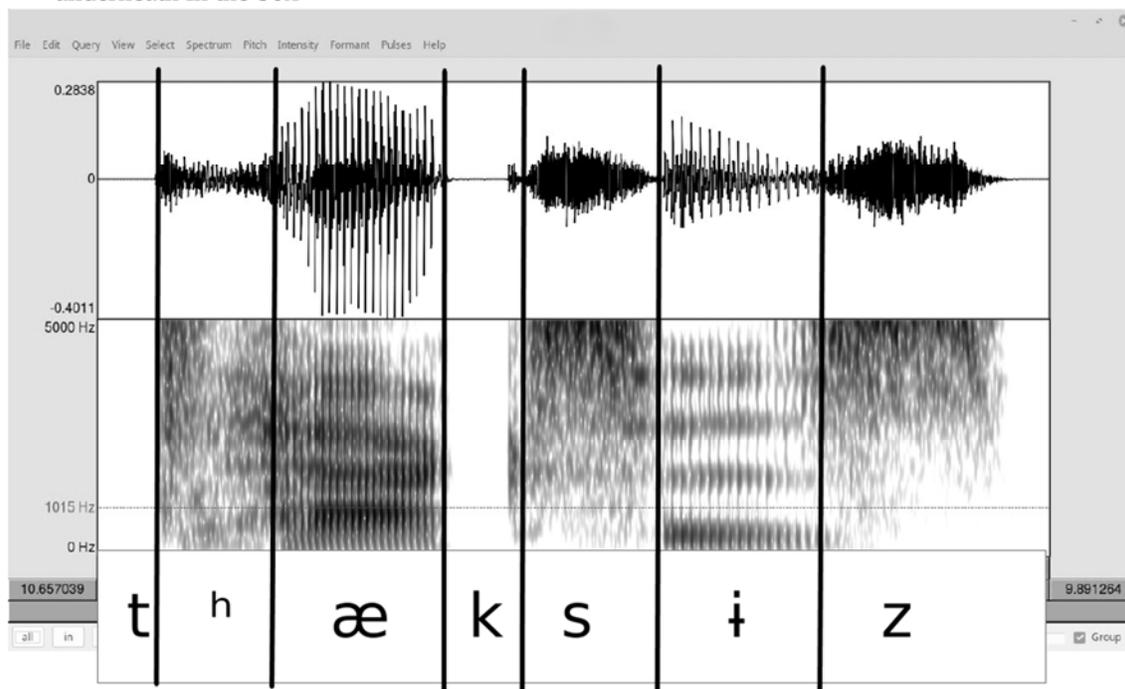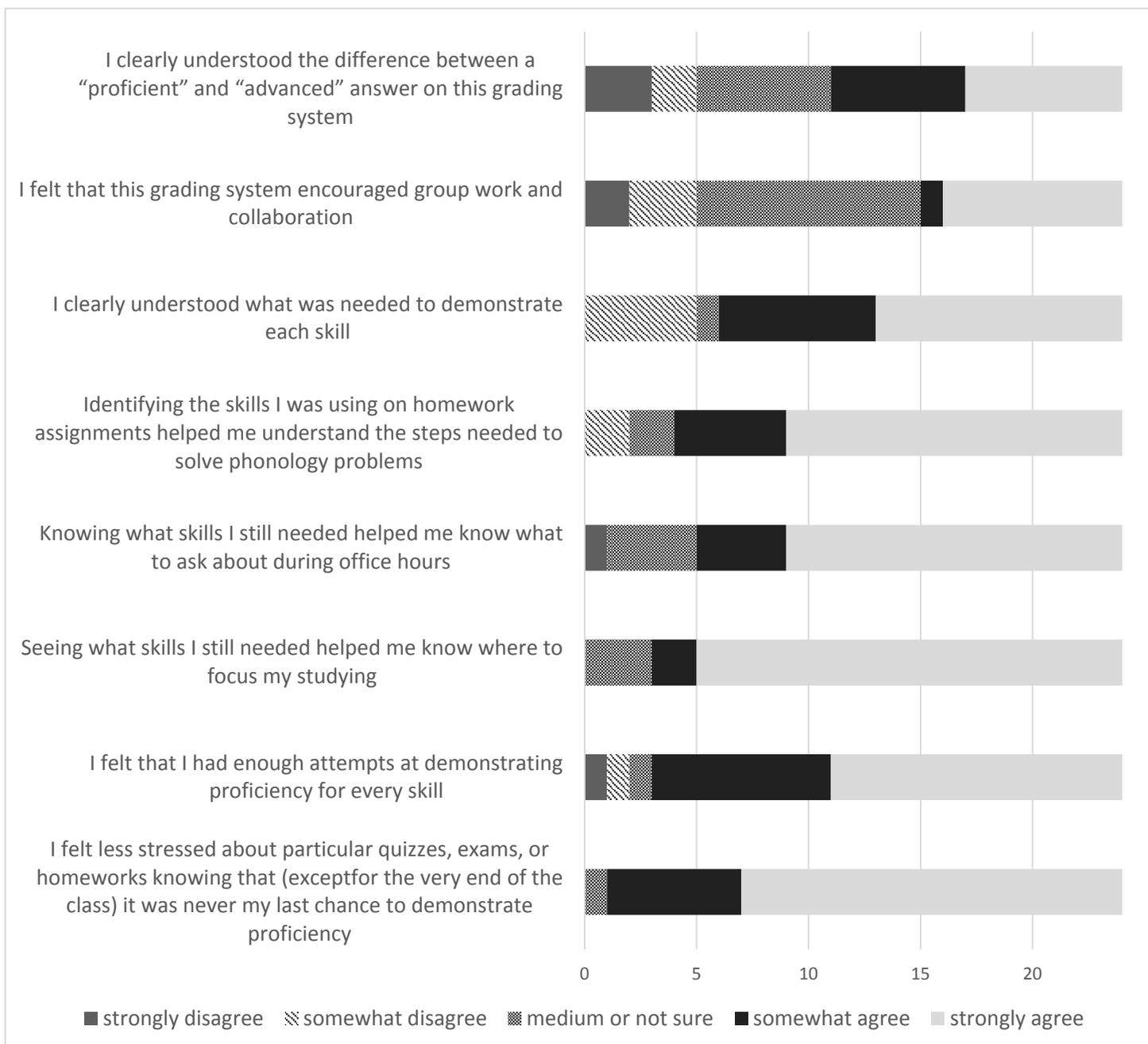

Figure Error! Main Document Only.: question from Week 3 of a 10-week course

*Figure **Error! Main Document Only.**: Degree to which respondents agreed with statements*