

Improving Student Learning with Aspects of Specifications Grading

SARAH E. VITALE
DAVID W. CONCEPCIÓN
Ball State University

Abstract: In her book *Specifications Grading*, Linda B. Nilson advocates for a grading regimen she claims will save faculty time, increase student motivation, and improve the quality and rigor of student work. If she is right, there is a strong case for many faculty to adopt some version of the system she recommends. In this paper, we argue that she is mostly right and recommend that faculty move away from traditional grading. We begin by rehearsing the central features of specifications grading and providing two examples of how to implement it in philosophy classes. In light of the examples, we argue that specifications grading fulfills two of Nilson's central desiderata (increasing rigor and motivating students) but not the third (saving faculty time). Since specifications grading generates two benefits that when combined increase student learning, without adding or increasing burdens, we conclude that student learning increases when courses are revised to include aspects of specifications grading.

1. What Is Specifications Grading?

“Specifications grading” is the name Linda Nilson gives to a grading system described in her book of the same title.¹ Specifications grading did not appear *de novo*. It “borrows elements of pass/fail grading, competency-based education, and classic contract grading.”² But its particular combination of elements is distinctive enough to merit its own name. Nilson stipulates that in a pure specifications grading system, pass/fail or binary evaluation occurs within and at the assignment level. As opposed to the A through F system with twelve grading tiers (A, A- . . . C+, C . . . D-, F), in pure specifications grading there are only two grading tiers for any assignment, pass and fail. Unlike other pass/fail grading systems where C, or even D-, performance is passing, to earn a passing mark in Nilson's system a student's work must

achieve a level of quality typically associated with B, B+, or even A- performance. Additionally, in pure specifications grading, students are allowed to attempt some assignments more than once.

Nilson argues that specifications grading will save faculty time, increase student motivation, and improve the quality and rigor of student work. We argue that she is mostly right and recommend that faculty move away from traditional grading. We begin by rehearsing the central features of specifications grading and providing two examples of how to implement it in philosophy classes. In light of the examples, we argue that specifications grading fulfills two of Nilson's central desiderata (increasing rigor and motivating students) but not the third (saving faculty time). Since specifications grading generates two benefits that when combined increase student learning, without adding or increasing burdens, we endorse versions of specifications grading. We conclude that student learning increases when courses are revised to include aspects of specifications grading, but we do not believe that pure specifications grading (Nilson's preferred version) is necessarily superior to synthetic versions.

I.1. Criterion and Assignment-Level Marking

At the individual criterion level, partial credit is not given in pure specifications grading. In Nilson's system, students are provided with a list of "specifications," which are descriptions of the characteristics their work is to exhibit. A typical specification or evaluative criterion for an argumentative philosophy paper is "has a clear thesis." With regard to this criterion, the grader determines whether the paper has or does not have a clear thesis. At the individual criterion level, specifications grading is binary. Consider a cookie.³ If a cookie recipe calls for walnuts, then a cookie (no matter how otherwise delicious) that does not contain walnuts fails to be the cookie that was called for. Even if the cookie has scrumptious macadamia nuts, it fails when the specification is for walnuts. Analogously, a paper without a clear thesis fails no matter what its other merits. Additionally, even if coarsely chopped walnuts make for a better cookie than very finely ground walnuts, if the walnuts are there, the cookie meets the specifications. While some clear theses are better than others, as long as a clear thesis is present, the criterion is met.

At the assignment level, a passing mark is given only when a student's work exhibits all of the specifications. If there are six specifications for an assignment and a student's work meets five of them, the assignment is marked "fail." By failing to meet one expectation, the work as a whole does not exhibit the totality of what it should. While meeting five criteria demonstrates more mastery than meeting three, and improvement from meeting only two to meeting four standards is

indicative of learning, if all of the specifications have not been met, then the work does not meet the expectations.

I.2. Rubrics vs. Specifications

We have been asked, “What is the difference between a rubric and a list of specifications?” The short answer is not much. Indeed, Nilson notes that lists of specifications “might be thought of as one-level rubrics.”⁴ Both rubrics and lists of specifications describe the features that student work is to demonstrate; they are both collections of the criteria by which student work will be evaluated. Nilson refers to an individual evaluative criterion as a specification or “spec.” In all but the simplest assignments, both rubrics and lists of specifications contain more than one criterion or spec.

The difference between pure specifications grading and traditional grading with a rubric is not the nature of the evaluative criteria, but what a teacher does with them. Because partial credit is available, in traditional grading teachers evaluate the degree to which student work meets each individual expectation. A student may earn seven of ten possible points for the “clear thesis” expectation and eleven of the fifteen points available for “quality argumentation,” and so on. In traditional grading, there are many possible marks for each criterion. In contrast, in pure specifications grading, only two marks, pass and fail, are possible at the criterion level. Partial credit is not given because “achievement is not a matter of degree; a student either achieves an outcome or does not achieve it.”⁵ This should not suggest that a teacher will not receive a paper that has a “somewhat clear” thesis. In such cases, the grader must make a judgment. Either the thesis is sufficiently clear such that is marked “pass” or it is not.

Traditional and specifications grading also differ at the assignment marking level. In pure specifications grading, only “if a student’s work has all of [the qualities specified] is it acceptable/satisfactory.”⁶ In specifications grading, it is the combination of all of the specifications being met that results in the assignment passing. To pass, a student’s work must meet *all* of the specs. In traditional grading, assignment grades are determined by tallying all of the (partial) points earned relative to individual criteria. As such, it is possible for a student to receive a passing grade when their work as a whole does not demonstrate that they learned what they should have in traditional grading.

I.3. Rigor

Nilson argues that specifications grading restores rigor by setting the standard for a passing mark quite high. To earn a passing mark, students’ work must be of the quality typically associated with at least

B-level performance. The evaluative criteria are such that if they are all met, the student's work is high quality.

As Nilson explains, "If students do not achieve the highest level, they do not pass."⁷ She continues to argue that setting standards at one's current B-level or higher raises academic standards because students must submit at least B-level or better work to get credit. Grade inflation stops dead in its tracks. All the problems inherent in the point systems—in particular, the time it takes faculty to decide on and justify partial credit and students' attempts to argue more points out of faculty—disappear. In addition, the evidence from faculty who have graded assessments pass/fail indicates that students perform at a higher level. They are more motivated and specifically more motivated to excel.⁸

In courses with traditional grading systems, some students may "skate by," and even earn a B for the term, without achieving functional proficiency or even minimal competency with respect to any learning outcome.⁹ In specifications grading courses, students are held to higher standards.¹⁰

Of course, there are some assignments where the performance level is justifiably set even higher than B level. To set a higher level of performance a teacher may add additional evaluative criteria, modify existing specs so that meeting them is more difficult, or both.¹¹ For a first paper in an introductory-level philosophy course, one might specify that a thesis be presented in the introductory paragraph. For the second paper, one might require that the introductory paragraph present a valid argument and that the thesis be presented as the conclusion of the valid argument (more specifications) or that the thesis be very clear (more difficult specifications).

I.4. Retries

Students are allowed a limited number of retries in a pure specifications grading system. Retries allow students to learn from failure. In philosophy classes, retries often take the form of students being allowed to revise and resubmit a paper. That student work must meet all of the specifications to pass and that the specifications are set such that typical B work is required to meet them result in many students earning a failing mark on their first attempt. Some students will need to retry an assignment more than once before they can achieve a passing mark, before they learn what they should.

Because retries are employed, pure specifications grading shares much with mastery-based grading, "a system that allows a student to resubmit work as often as desired until the student is satisfied with his or her grade, they cannot improve their understanding of the material, or time runs out in the course."¹² However, Nilson's promise that

pure specifications grading will save faculty time is in tension with the availability of (nearly) unlimited retries because (nearly) unlimited retries means (nearly) unlimited grading. This is why Nilson recommends a limited number of retries. It is helpful to characterize pure specifications grading as a version of mastery grading that allows only a limited number of retries.

1.4.A. Tokens

While there are other ways to facilitate retries, we find it helpful to introduce students to the notion of a token. Tokens are not physical objects. Rather, they are a way of naming and tracking how many retries a student has available. We keep track of tokens in a spreadsheet. Tokens are under student control; students tell us whether and when they want to use a token, that is, when they want to retry an assignment that has not met all the “specs.” In the formative feedback on the assignment, the teacher would remind the student of the missed specification. The student then could inform the teacher that they would like to use a token to retry the assignment. If the revised work now meets all of the specifications, the instructor would mark the assignment as passing.

It is important to limit the number of tokens or retries students have for several reasons. For example, when students can make changes and resubmit every piece of work many times, they may not work hard on initial attempts. Beyond disavowing an unlimited number of tokens, we believe there is no one right answer regarding how many tokens students should be allotted at the beginning of the semester. The appropriate number depends on many factors, including the size of the class, the number of assignments, and the amount of time it takes to evaluate each assignment. In the absence of relevant details, we suggest one token per high stakes assignment, plus or minus one.

The principle merit of employing tokens, as opposed to just having a column in one’s gradebook for retries, is that an instructor can allow them to be used for things others than retries. For example, a student may use a token to submit work late. Or, in a course that has an attendance requirement, an instructor might allow students to use a token to “erase” an absence. This has an inclusivity advantage since it is the students, not the teacher, who decides which reasons for absence are acceptable. With tokens, multiple areas of flexibility in how students navigate the semester can be managed under one heading.

1.5. Term-Level Grading: Points, Bundles, and Modules

All but the rarest of faculty are required to report term grades in a way that at least approximates the twelve-tiered letter-grade system, and so all of the pass/fail assignment-level work completed by students in a pure specifications system needs to be converted into a letter grade for

the term. In pure specifications grading there are three ways to make the conversion: points, bundles, and modules.

In a points system, each assignment is accorded a point value. If a student passes—meets all the specs of—an assignment, they receive all the points for that assignment. A small assignment might be worth five points. If the student passes, they get five points. Otherwise, they do not earn any points. A paper might be worth fifty points. If a student passes the paper, they receive fifty points. If they fail, they receive no points. The term grade is derived from the number of points earned out of the total points possible. A student who earns 475 of a possible 500 points, for example, would receive an A as their term grade. We discuss the use of a points system in an upper-level course in more detail below.

Alternatively, term letter grades may be assigned relative to bundles or collections of expectations. One bundle might be homework and another in-class writing. To pass the homework bundle, a student might be asked to pass 90 percent of all assignments. At the assignment level, work continues to be marked pass/fail. At the term grade level, if a student passes 90 percent of assignments, they have passed that bundle. A bundle might also include various assignments. To receive an A in a course with bundles, one may be required to pass both papers, the final exam, 90 percent of daily homework, and the attendance requirement. In such a course, a bundle for the term grade of B might be to pass both papers, the final exam, and 75 percent of the daily homework. We discuss the use of a bundle system in a core curriculum course in more detail below.

Bundles may also track content units. A unit bundle might include passing five of the six small assignments and the unit exam. At the end of the term, if the student passes a certain number of bundles, they will receive a specific grade. In a course with, say, five bundles, a student passing all five will receive an A for their term grade, a student completing four will receive a B for their term grade, and so on. In such a course, a student who completes bundles one, three, and five would earn a term grade of C, as would a student who completes bundles one, two, and three. Bundles are particularly useful in courses that have numerous sub-units that need not be completed in any particular order.¹³

Finally, term grades may be determined through modules, which Nilson defines as bundles that are to be completed in a particular order.¹⁴ Modules are especially helpful when it is important for students to master one skill or body of content before advancing to a second. In a course with four modules, for example, students would be required to complete the first module before attempting the second, the second before attempting the third, and so on. In such a course, term grades correspond to where in the sequence the student finishes the

semester. If a student completes only the first module by the end of the term, they would earn a D in the course. Completing modules one and two would result in a C; completing modules one, two, and three would result in a B; and completing all four modules would result in an A. A more detailed discussion of module-based term grading in a critical thinking course is Cahill and Bloch-Schulman's helpful article "Argumentation Step-by-Step."¹⁵

I.6. Pure vs. Synthetic Specifications Grading

To extend Nilson's work and clarify the choices faculty have, it is important to understand the range of specifications grading systems that are possible. Nilson stipulates that a pure, as opposed to synthetic, specifications system has six distinguishing features:¹⁶

- (1) Pass/fail assignment (and bundle or module) grading: Instructor employs binary criterion evaluation and conjunctive assignment-level evaluation. Either the work fulfills all of the expectations or it does not.
- (2) Clarity: Instructor provides very clear specifications, descriptions of what constitutes an acceptable piece of work.
- (3) Rigor: Student work passes only if it is of B-level quality or better.
- (4) Retries: Students are allowed at least one opportunity to revise unacceptable work.
- (5) Term Grades: Bundles and modules that earn higher term grades require students to demonstrate mastery of more skills and content, more advanced/complex skills and content, or both.
- (6) Relation of Term Grade and Learning Objectives: Bundles and modules used to determine term grades are tied directly to the learning outcomes of the course.

This is not a list of necessary (and jointly sufficient) conditions for a grading system to qualify as specifications grading. This list defines *pure* specifications grading. There is also synthetic specifications grading.

Nilson develops her definition of specifications grading simpler (either pure or synthetic) by contrasting it to traditional grading. Traditional grading is "reliant on partial credit and hair-splitting point allocation."¹⁷ In traditional grading, "instructors are expected to give partial credit for almost anything correct a student submits, including largely wrong or vague responses."¹⁸ Traditional grading occurs, then, when (i) partial credit is given at the criterion level and (ii) assignment grades are determined by summing the (partial) points earned. When

criterion-level assessment eschews partial credit or assignment grades are not determined by summing criterion-level points, then one is not engaged in traditional grading. The necessary and sufficient condition for being an instance of *synthetic* specifications grading is that binary criteria-level or non-tallying assignment-level grading occurs (at least some of the time). To be clear, meeting conditions two, three, and four but not one means that the system is traditional according to Nilson. If the system exemplifies all six of the features Nilson describes, and abides by 1–4 in *all* evaluations, then it is pure specifications grading.¹⁹

Our finer grain distinction between criterion- and assignment-level grading turns Nilson’s first criterion into two separate criteria. Accordingly, a grading system is aptly called (at least synthetic) specifications when at least one of two conditions is met:

(i) Binary Grading at the Criterion Level: Whether an expectation has been met is an either/or proposition. No partial credit is given for “somewhat” meeting a standard.

(ii) Conjunctive Grading at the Assignment Level: To pass an assignment a student’s work must meet *all* of the specs. Assignment grading does not involve tallying points.

Of course, a grading system that meets only these two criteria is unlikely to have the motivational influence on students that Nilson promises. We, and Nilson, believe any grading system, whether specifications or traditional, should also include:

(iii) High Standards: The quality of work required for a passing mark on an assignment is high. What is commonly taken to be B-level work or higher is required for a passing grade.

(iv) Clarity: Evaluative criteria, and the method of determining assignment and term grades, are easily understood by students.²⁰

(v) Retries: A limited number of retries are allowed. Students may revise and resubmit some of their work. (Tokens may be involved in keeping track of retries.)

However, since traditional systems may set high standards, be clear, and allow retries, it is not traits iii, iv, and v that distinguish specifications from traditional grading. Since we are comparing good traditional systems to good specifications systems, we assume high standards, clarity, and retries are in place in both cases. The issue between specifications and traditional grading hinges on whether binary and conjunctive marking is superior to issuing partial credit and summing points.

On Nilson’s taxonomy, then, there are two types of grading schemes: traditional and specifications. Within specifications there are two versions: pure and synthetic. The difference between pure specifications

grading and synthetic specifications grading are threefold. First, in pure specifications there are additional criteria to be met if bundles or modules are used to determine term grades. Second, the system is synthetic if criteria i and ii are met, but at least one of criteria iii, iv, or v are not. Third, the system is synthetic if not *all* student work is evaluated in a binary and conjunctive fashion. There is, then, quite a variety of synthetic specifications systems.

One might grade some but not all assignments, papers, exams, projects, etc. pass/fail.²¹ Perhaps papers are graded pass/fail but partial credit is possible on exams. One might grade all student work pass/fail but set the bar for passing on some work (e.g., daily homework) below B level. Or the evaluation of an assignment might be mixed. One might grade all work pass/fail at the B level, but award grades in a traditional fashion above the passing mark: a B+ or A- is issued for partially, and an A for completely, meeting standards above those required to pass.²² One might employ binary grading at the criterion level (e.g., 5 points for fully meeting a criterion, 0 for not fully meeting it) but arrive at the assignment grade by tallying the points garnered at the criterion level.²³ One might assess some items in a test pass/fail and others with partial credit. If an exam has ten short-answer questions, perhaps the first five questions are awarded full or no credit, while partial credit is available on the answers to the second five questions.²⁴

One, possibly synthetic, specifications grading system that is not included in Nilson's survey of synthetic systems incorporates three-tier assignment marking. In William Rapaport's system, while criteria-level evaluation is binary, assignment-level evaluation is trinary.²⁵ Rather than the two possible results (pass or fail) in Nilson's system, in Rapaport's there are three: "clearly adequate," "clearly inadequate," and "neither clearly adequate nor inadequate." Importantly, no partial credit is given, nor is the assignment grade determined by tallying points. If a piece of work meets all the specifications, it is clearly adequate. If it meets none of them, it is clearly inadequate. If it meets some of them, it is neither clearly adequate nor inadequate.

A different three-tier system might issue the assignment marks of "incomplete," "pass," and "high pass." In such a system, an instructor constructs two sets of specifications, one for pass and one for high pass. At the passing level, the specification regarding the thesis statement of an argumentative paper might be "thesis presented in the introductory paragraph." At the high passing level, the specifications might be "thesis presented as the conclusion of a valid argument in the introductory paragraph" and "thesis is very clear." With a full list of specifications for "pass" and "high pass" delineated, the teacher uses binary evaluation at the criterion level and trinary evaluation at the assignment level. Work that meets all of the specifications for high

pass is marked high pass; work that does not meet all of the specifications for high pass but does meet all of the specifications for pass is marked pass; and work that does not meet all of the specifications for pass is marked incomplete.

We are unsure whether Nilson would categorize three-tiered systems as pure or synthetic. The examples she provides of synthetic systems all involve issuing partial credit at the criterion level or arriving at an assignment grade by tallying points. The trinary systems do neither. Binary marking occurs at the criterion level (i.e., no partial credit is given); assignment grades are not determined by summing points; and the bar for “pass” is set high (at B level). However, Nilson also says that all assignments are graded pass/fail in a pure specifications system. Does this mean that to be pure specifications a grading system may not issue more than two marks—pass and fail—for assignments? In the second trinary system all assignments are indeed graded pass/fail (although the term “incomplete” is used in place of the term “fail”). The difference is that there are two ways to pass and thus three possible marks.

Noticing the range of synthetic systems brings valuable questions into focus. “Should I replace my traditional grading system with a pure specifications system?” is the wrong question to ask. Rather, there are many questions to address. Is there a good reason not to evaluate all student work in a binary (or trinary) and conjunctive manner? If not all, which assignments (tests, papers, etc.) should I definitely approach as a specifications grader? In which assignments, if any, should a mixed (some specs and some traditional) evaluation be used? Is it okay to set the pass/fail bar for daily homework low? Should I determine term grades via points, bundles, or modules? Should I use tokens, and if so how many should be issued? Taken together: which aspects of specifications grading should I employ to help motivate students to improve their learning?

II. Specifications Grading in Philosophy Courses

There has been a very positive response to specifications grading in the scholarship of teaching and learning literature in general, but the work tends to focus on math, science, and social science classrooms.²⁶ To our knowledge, this is the first article-length treatment of specifications grading in philosophy. William Behun authored a short piece that appeared in the 2017 issue of the *APA Newsletter on Philosophy in Two-Year Colleges*. Behun asserts that in two-year colleges, “it is perhaps more important for students . . . to achieve a more general familiarity with basic principles of philosophical enquiry and clear argumentation” than to “demonstrate a high degree of philosophical

rigor.”²⁷ Additionally, Behun reports that the term-grade distribution in his classes remained the same after he switched to specifications grading, which was not the case in our classes.²⁸ In 2019, Rebecca Scott discussed her use of specifications grading in an APA blog post.²⁹ In 2020, Jennifer McCrickerd positively reviewed Nilson’s book in *Teaching Philosophy*.³⁰

In this section, we extend Nilson’s work by applying it to philosophy courses. First, we argue that the benefits of specifications grading are amplified when how-to instruction is provided in philosophy courses. Of course, there is nothing about specifications grading that prohibits placing importance on how-to instruction, and so we conceive of this as a friendly amendment, not a criticism. Second, we detail two quite different examples of specifications grading in philosophy courses. The first is an instance of pure specifications grading that uses a points system to determine term grades in an upper-level course. The second is an instance of a synthetic specifications grading system that uses bundles to determine term grades in a core curriculum course. These two examples demonstrate different answers to the questions faculty should ask themselves regarding their grading system.

II.1. How-To Instruction

We believe student learning is enhanced when students are given “how-to” instruction. How-to instructions, descriptions of what students should do to create an academic artifact, are not the same as grading criteria, which are descriptions of characteristics that student work should exhibit. Nor is providing how-to instruction the same as providing models. Examples of cookies and the list of specifications that are to be exemplified in a good cookie are not the same thing as instructions for how to make cookies. “Fold the walnuts into the batter” is a how-to instruction, while “cookie has walnuts in it” is a criterion of evaluation or a specification.

How-to instruction provides a roadmap for students. It is difficult for experts to remember the extent to which novices need how-to instruction. As experts, professors are no longer conscious of all of the distinct steps involved, for instance, in writing a philosophical essay. To develop good how-to instruction, we need to slow down and note all of the things we do to write a good paper. Then we need to describe these activities and the order in which it is best for a beginner to proceed in plain language.³¹

Where specifications need to be sufficiently detailed for students to know what constitutes a quality artifact (e.g., a paper), how-to instructions need to be sufficiently detailed so that students know what to do to create an artifact that meets the standards. But how-to instruction should not be so detailed as to discourage creativity and individuality.

Additionally, when instructions become too detailed, novices have a difficult time understanding them. This tension between enough and too much detail must be managed by each teacher on each assignment. When done right, how-to instruction ensures that students understand how to create artifacts that demonstrate that they have met the grading criteria. When both clear evaluative criteria and how-to instructions are provided to students, it is very unlikely that an instructor will ever again hear a student say, “I didn’t know what you wanted” or “I didn’t know how to do it.” In both of the courses outlined below, we guide students with how-to instructions. For more on how-to instruction, see “Philosophers Folding Origami” by Jennifer Wilson Mulnix and Alida Liberman.³²

II.2. Example 1: An Upper-Level Philosophy Course

In an upper-level Critical Theories course, one of us uses a two-tier, points-based specifications system. One of the advantages of determining the term grade with points is that students can choose which assignments they wish to complete to achieve a specific grade.³³ The point categories include a final paper (25 points), presentation (15 points), four self-assessments (each worth 3.75 points for a possible total of 15 points), four unit assignments (each worth 3.75 points for a possible total of 15 points), and twenty-four reading question posts (1.25 points each for a possible total of 30 points). The key learning objective of the course is that students learn to defend an original thesis in an argumentative essay that engages with course material. Students are invited to engage with the authors they read throughout the semester. The reading questions require students to ask questions about the text and reply to other students’ posts. Each unit assignment is content-based, such that the student demonstrates that they are proficient with the course materials. The presentation covers a secondary article, which students will use in their paper, thus requiring students to examine the literature in the field. In addition, the self-assessments provide regular opportunities for meta-cognitive reflection. The self-assessments follow each content unit and require that students answer simple content questions as well as reflect on what they learned, found confusing, and wish to think more about.

For the essay, students are given a set of prompts from which to write, and they are allowed to create their own topic so long as it is appropriately related to the course material. The specifications for the argumentative essay are the following. To receive credit, the paper must

- engage with at least one text from the syllabus;

- engage with at least two secondary texts (from authors who are not on the syllabus, but instead who are writing about figures on the syllabus);
- demonstrate relevant and cogent argumentation in defense of position;
- show some originality and insight;
- explain the philosopher's argument in a generally accurate and fair manner;
- be organized in a cogent manner; and
- display proper citation.

The specifications were created using Nilson's advice to first determine what B-level work (or higher) is. If you already have a solid rubric for an assignment, try collapsing the descriptions for the top two levels of work into one. Or you might consider just tweaking the top-level description. Another good idea is to recall the student errors that have resulted in point deductions in your traditional grading courses or assignments and transform those errors into specifications that you can explicitly counsel your students to avoid.³⁴

A typical A-level paper would explain the author's argument in a fair manner, demonstrate originality and insight, and adduce strong argumentation in defense of the author's position. But a B-level paper may make some mistakes regarding arguments in a philosopher's text and still demonstrate general understanding. So long as students have met the specifications, they have passed the paper.

In this class designed for upper-level students, the instructor did not find that only offering two tiers was demotivating for strong students. Because the assignment was at the end of the semester and worth 25 percent of the final grade, it would be difficult to excel without writing the final paper. Strong students appeared to be motivated by their interest in the material and success with earlier assignments, such that they performed well on the final paper. Students in danger of failing the final paper were notified at the draft stage, such that they had the opportunity to revise. In addition, because of the high-stakes nature of the assignments, particularly the presentation and the paper, each student was given five tokens, which they were able to use to resubmit an assignment or receive an extension on an assignment.

In a previous iteration, before the instructor employed specifications grading, the course requirements were two papers (the first worth 20 percent of the overall grade and the second worth 25 percent), a presentation on a secondary text (20 percent of final grade), and a self-assessment at the end of the semester (15 percent). A weakness of this system was that the first paper did not prepare students to write

the second paper. Students focused on a paper early in the semester before demonstrating mastery of course material. There were no adequate checks on student progress to ensure that students were prepared to write an upper-level philosophy paper. In addition, students were unclear as to which characteristics an upper-level philosophy paper should have.

In the specifications grading iteration of the course, students received exemplars and how-to instructions to help them with the paper, beginning with collaborative work on a thesis. They then shared with each other bibliographies and outlines, before submitting a draft. Because the paper was graded on a pass/fail basis at the end of the term, students had ample opportunity to receive formative feedback on early steps of the paper.

II.3. Example 2: A Core Curriculum Course

In a core curriculum course, one of us uses a four-tiered assignment grading system where the evaluations possible for papers are high pass, pass, incomplete, and fail. (See Table 1.) The term grade is determined by bundles. (See Table 2.) At the assignment level, a passing mark is the same as what would have earned a B in the pre-specifications grading version of the course. The level of performance required to earn “high pass” is what would have earned an A in the pre-specifications version of the course. To earn “high pass,” a student paper must meet more and more difficult specifications than are required to earn “pass.” A mark of “incomplete” means the student did not meet all of the expectations required to earn a passing mark, but they did meet at least the most basic requirements of the assignment (traditional D- through B-level performance). Indeed, it is not disingenuous, although perhaps not sufficiently exacting, to describe this as a three-tier system, since the marks available for students who make an on-time, good faith attempt are high pass, pass, and incomplete. Students receive a mark of “fail” only when they do not turn in a paper, turn the paper in late, violate the length requirement, or submit work that is not remotely close to successful.

Student behavior under this system suggests that it induces significant motivational benefits. Students who enter the class with well-developed writing skills might “phone in” their papers if all they need to do is produce B-level work. The addition of the “high pass” mark ensures that all but the most extraordinarily prepared students have a challenging goal. The difference between “fail” and “incomplete” is that students receiving a failing mark do not get formative feedback on their initial attempt, if such an attempt had been made. Nevertheless, students receiving a “fail” still may use a token to submit a paper late and receive formative feedback on it or resubmit a paper that initially

Table 1: Specifications for Paper 1

Grade is earned when the paper has *all* of the qualities described.

Grade	Organization	Rigor	Intellectual Struggle	Well-written, Significant, and Original
High Pass	<ul style="list-style-type: none"> • introductory paragraph that conforms completely to the template • paper follows outline presented in the introduction 	<ul style="list-style-type: none"> • demonstrates accurate understanding of material from texts and class discussions • paraphrases or quotes sources to increase precision 	<ul style="list-style-type: none"> • has a relevant criticism • relevant criticism is difficult but not impossible to overcome (the “Goldilocksish” effect) • charitably explains the justification of the criticism, with an illustrative example or nuanced distinction • has a relevant response to the criticism that appears to overcome the criticism and is supported by a new insight • has additional rejoinders and further criticisms and replies as necessary 	<ul style="list-style-type: none"> • has no more than a few minor spelling or grammatical errors • writing is vibrant, clear, and engaging • addresses an important aspect of the issue being discussed • offers an argument, criticism, or response that was not in the reading or part of the class discussion • makes use of self-created philosophical distinctions that advance an argument, criticism, and/or response
Pass	<ul style="list-style-type: none"> • introductory paragraph that conforms completely to the template • paper follows outline in the introduction 	<ul style="list-style-type: none"> • demonstrates accurate understanding of material from texts and class discussions 	<ul style="list-style-type: none"> • has a relevant criticism • charitably explains the justification of the criticism • has a relevant response to the criticism that is supported by a new insight 	<ul style="list-style-type: none"> • has no more than a few minor spelling or grammatical errors • addresses an important aspect of the issue being discussed • offers an argument, criticism, or response that was not in the reading or part of the class discussion
Incomplete	<ul style="list-style-type: none"> • introductory paragraph conforms completely to the template 		<ul style="list-style-type: none"> • has a relevant criticism • has a relevant response to the criticism 	
Fail	<ul style="list-style-type: none"> • Does not meet the standard for “incomplete,” is not turned in, is not turned in on time, or violates the length requirement 			

violated the length requirement or was otherwise grossly inadequate and receive feedback on it. The possibility of a “fail” as opposed to “incomplete” is to motivate students to make a good faith initial effort.

Since instituting this system, nearly every student who earns an “incomplete” on their first paper has chosen to rewrite. (Approximately 3 percent, or one student per term, has elected to drop the class. Everyone else retries.) All who retry eventually pass the first paper, although some need to retry more than once. Additionally, most of the students who “pass” on their first attempt also rewrite in pursuit of a “high pass.” As a result, under this system, the number of students receiving a term grade of D or F is near, and typically is, zero, and the number of students earning a term grade of C is very low. In short, the students who used to earn Fs or Ds are now earning Cs or Bs. The students formerly earning Cs are earning Bs or As.

Three additional features of the system in this course likely contribute to student motivation and ultimately learning. First, students are required to have the introductory paragraph of their first paper approved by the instructor before they continue writing. Approximately 25 percent of students receive approval on their first attempt. Approximately 50 percent receive approval on their second or third attempt. The remaining 25 percent may need up to eight attempts. This process provides students who need formative feedback opportunities to receive it as they are working so they can avoid common mistakes and not waste time writing an ill-fated paper. Since the introductory paragraph serves as an outline of the paper, instructors can see where students are headed and make corrections without having to read an entire paper. This process moves the provision of formative feedback to the front end of the writing process when students need it most.

Second, the way the term grade is determined strongly incentivizes students to retry the first paper if they receive a mark of “incomplete” on their first attempt. If a student does not pass the first paper, they can earn no higher than a D for the term. As Nilson states, “No longer can [students] submit half-hearted attempts and hope to get a ‘good enough’ grade to maintain the end of the semester grade that they want.”³⁵

Third, students may retry the first paper without using a token. In this course, tokens may be used to retry a paper, have three missed daily assignments counted as complete, or erase an absence. There are two papers assigned during the semester, and students begin the term with two tokens. There are two advantages to building a “free” retry into the first paper assignment. First, students who will need three retries to succeed have three retries available. Second, students who may be leery of using a token (perhaps because they anticipate needing to use a token to erase an absence) to attempt to move from “pass” to “high pass” have no reason not to make the attempt. Additionally, along

Table 2: Term Grade Bundles

A	<ul style="list-style-type: none"> • Successfully complete 90% of daily assignments • No more than 3 absences (NB: This is a Tuesday/Thursday class) • High Pass on Paper 1 • High Pass on Paper 2 • 90 points or better on comprehensive final exam
B	<ul style="list-style-type: none"> • Successfully complete 80% of daily assignments • No more than 6 absences • Pass or higher on Paper 1 • Pass or higher on Paper 2 • 80 points or better on comprehensive final exam
C	<ul style="list-style-type: none"> • Successfully complete 60% of daily assignments • Pass or higher on Paper 1 • “Incomplete” or higher on Paper 2 • 70 points or better on comprehensive final exam
D	<ul style="list-style-type: none"> • Pass or higher on Paper 1 or Paper 2 • 60 points or better on comprehensive final exam
F	<ul style="list-style-type: none"> • Fail to meet the specifications for a D or higher <p>You must pass (60 points or higher) the comprehensive final exam to pass the course You must earn a Pass or higher on at least one paper to pass the course</p>
“+”	<p>If you earn 15 points higher on the final exam than the minimum needed to earn the term grade you are otherwise qualified to earn, you will earn a plus on your term letter grade. Example: If you have earned a “pass” on both papers, have completed 80% or more of daily work, and have 6 absences or fewer going into the final, you are qualified to earn a B for the semester. If you earn 95 points or more on the final exam, your semester grade will be a B+.</p>
Tokens	<p>Each student begins the course with two tokens. With them you may “buy” the opportunity to retry a paper, have three incomplete homework assignments counted as complete, or erase an absence. No extra tokens may be earned.</p>

with the revision, students must submit a “wrapper” or metacognitive reflection to earn the privilege of having their rewrite reviewed. The wrapper in this case is a short description of how the student used the formative feedback they received to improve their second attempt.

This course is currently an instance of synthetic specifications grading. In one sense, it might be synthetic in Nilson’s estimation because it has more than two grading tiers for the papers. But it definitely is synthetic in another way because the final exam is graded traditionally. Partial credit can be earned on each of the short-answer questions and the points for each answer are tallied to arrive at the final exam grade. The rationale for this concession to traditional grading is twofold. First, because the central learning objectives of the course are to help students develop perspective taking and critical thinking skills through paper writing, nowhere during the semester are students provided with how-to instruction or multiple opportunities to build their short-answer question answering skills while being provided with formative feedback. Insisting on a high bar without showing students how to do what you expect them to do is problematic. Second, given

that the final exam is high stakes and there are no retires because the semester is over, setting a passing bar at B-level and assessing the final in a binary, conjunctive manner would result in many students failing. However, while working on this paper the one of us using this final has realized that this rationale is not sufficient to justify the continued use of a traditionally graded, content-regurgitation final exam. In part, the disconnection between the course learning objectives and this assessment now seems unjustifiable. Next semester will be different!

In this system, paper evaluation is binary at the criterion level; the criteria are clear; at the assignment level, there are three possible marks for good faith efforts; the bar set for passing is high; and retries are available. This paper grading system appears to have an advantage over those that have only two possible marks at the assignment level, at least in a core curriculum class. Half-hearted initial attempts are discouraged, and already accomplished students have a challenging, even higher bar towards which to reach. The bundled term grade system further incentivizes students who want an A to produce “high pass” papers, and the term grade requirement that students pass the first paper to earn higher than a D for the semester ensures that the first paper of every student who wishes to pass is of B-level quality.

II.4. Different Approaches

Our courses differ in many ways. For example, the students in the upper-level course begin the semester with fairly high interest, motivation, skill level, and content knowledge. This difference goes a long way toward explaining why a separate bar for “high pass” is unnecessary in the upper-level course and why attendance is factored into the term grade only in the core curriculum course.

One might argue that there should not be an attendance requirement even in the core curriculum course. Daryl Close, for example, argues that student behavior or comportment (e.g., attendance) should not influence grades, except perhaps in courses that explicitly have professionalism as a learning objective.³⁶ Regarding tardiness, he notes, “the time of day that a student enters the classroom has no intrinsic connection with her understanding of the course material.”³⁷ Contra-Close, John Immewahr and Jennifer McCrickerd have argued that motivational grades are appropriate.³⁸ Motivational grades are those that induce student effort by providing (or withholding) course credit to students when they do certain things that promote student learning, such as coming to class. Perhaps the right conclusion to draw is that, barring a powerful justification for an exception, Close is correct that term grades should reflect only the extent to which a student has demonstrated learning. However, in some (usually core curriculum) courses, the need to incentivize student presence is just such a power-

ful justification. We disagree on this point, and the ways in which our term grades are determined reflect this disagreement.

III. Evaluating Specifications Grading

Nilson argues that specifications grading is superior to traditional grading in its ability to increase student motivation, improve the quality and rigor of student work, and save faculty time.³⁹ Additionally, Nilson appears to believe that specifications grading is clearer than traditional grading. We add fairness to this list of considerations. If one system is fairer than another, then it is to that degree superior. We argue for two conclusions in this section. First, some versions of specifications grading can satisfy all five of these desiderata (motivation, rigor, time, clarity, and fairness), but neither of us has experienced a time savings. Second, to achieve significant gains in student motivation, one need not institute pure specifications grading; one of the many synthetic versions of specifications grading can garner enough improvement in student learning that a good deal of traditional grading should be replaced.

III.1. Clarity

Nilson stops short of explicitly asserting that evaluative criteria and term grading systems in specifications grading are clearer than those in traditional grading, noting that traditional grading “does not obscure faculty standards and expectations, but it does not induce instructors to write particularly detailed directions either.”⁴⁰ She does say that in traditional grading “the descriptions are usually quite brief—short enough to fit in a tabular cell—and minor homework assignments may or may not have instructions and a rubric.”⁴¹ It is likely true that switching to specifications grading involves reflection such that many instructors improve the clarity of their evaluative criteria. But we think it is the reflection, not the switch, that generates the improvement in clarity. With constructive reflection an instructor could improve the clarity of their evaluative criteria without transitioning from traditional to binary and conjunctive grading. Additionally, as we noted above, the way in which evaluative criteria are written is not what differentiates specifications grading from traditional grading. The difference is in how assignments are graded, in whether partial credit at the criterion level and point tallying at the assignment level occurs. We also stipulated that we are considering good versions of each type of system; we are assuming that the evaluative criteria are clear regardless of which system is being discussed. It is not the case that performance criteria are necessarily clearer simply because an instructor evaluates whether they are met in a binary fashion.

Nevertheless, evaluations made in two-tier systems are likely to appear starker than those made in a traditional twelve-tier system. Because students know exactly what they must do to pass, which criteria their assignment must exhibit, they are encouraged to pay closer attention to the specifications than they may to a specific level on a rubric. But starkness by itself does not eliminate borderline cases. Some exegetical student work will be quite accurate but perhaps not sufficiently so, such that one evaluator might judge that it meets an accuracy criterion while another judges that it does not. The advantage of binary criterion-level evaluation and limited-tier assignment grading is not that there are no borderline cases. Rather, the advantage is that usually there are fewer borderline cases. This reduction in borderline cases is likely to decrease the number of instances where a student might think that the evaluation they deserve is one point or tier better. Consequently, students are less likely to contest evaluations, which is tantamount to saying that students will experience specifications grading as being clearer in one sense. Insofar as these differences are issues of clarity, specifications grading is clearer than traditional grading.

In addition, at the term level, specifications grading may be clearer than traditional grading because it allows students to know what assignments they need to do throughout the term to pass. At the term-grade level, bundles have a simplicity that may make the path to a desired grade clearer to some students. While traditional grading systems may clearly demarcate how a student may achieve an A, B, C, and so forth, students typically have to do some calculations that are absent in the specifications system. The information may be present, and even clear, in traditionally graded classes, but some students may not as easily draw the inference from this information to the performance they need on a particular assignment to achieve the grade they want. A student who knows that they must pass papers one and two to earn a B for the term and that to pass the paper they must meet all of the evaluative criteria should know precisely what they need to do with regard to paper one to keep alive the possibility of a B (or higher) for the term. In a course where partial credit is available on every assignment and points are tallied to arrive at a term grade, it is possible to earn less than a B on the first paper and still earn a B for the term. In such a course, there are many paths to a B for the term, and students must make calculations of points or percentages to determine what they need to earn on particular assignments to achieve the term grade they desire. Insofar as an absence of the need to perform calculations makes something clearer, students should experience courses that use bundles as clearer than those which tally points. However, since point tallying is one of the ways one may determine a term grade in a specifications

system, specifications grading is not necessarily clearer than traditional grading on this count.

We conclude that (i) at the term level, specifications grading might be but is not necessarily clearer than traditional grading; (ii) at the assignment level, specifications grading produces starker results (and thus fewer grade contestations) than traditional grading, and if starkness is part of clarity then specifications grading is clearer than traditional grading; and (iii) with regard to the descriptions of evaluative criteria there is no reason to think specifications grading is clearer than traditional grading. However, the fact that it is difficult to decisively claim that specifications grading is significantly clearer than traditional grading should not cause one to lose sight of the fact that many versions of specifications grading are exceedingly and valuably clear once students become familiar with how they work.

III.2. Fairness

There is nothing inherent to specifications grading that makes it fairer than traditional grading. However, some specifications systems are likely to be fairer than most traditional systems. Among the many factors that influence whether course grades are determined in a fair manner are the connection between term grades and assignment grades, the degree to which the ability to attempt an assignment is equal among all students, the clarity of the evaluative criteria, and how consistent the grader's marks are. With regard to the first three, we see no reason to think that any grading system is inherently superior to another. But perhaps there is a difference on the fourth measure. Consider arbitrariness.⁴²

At the criterion and assignment levels, pure specifications grading appears to have an advantage over traditional grading because the two-tier nature of pure specifications grading is likely to generate fewer borderline cases than a twelve-tier system, and thus there are fewer opportunities for arbitrary human judgments to enter the evaluation process. The degree to which there are fewer opportunities for human weaknesses (e.g., implicit biases, the tendency to give lower grades when tired or hungry) to enter the grading process, the fairer the system. However, there could be fewer moments of discernment in a traditional system. An assignment might be evaluated relative to one criterion for which partial credit (0–5 points) may be earned. Alternatively, an assignment targeting the same learning objective could deploy six evaluative criteria and assess each of them in a pass/fail manner. There are fewer human judgments to make in the partial credit system. So, while there is a tendency for specifications systems to have less arbitrariness than traditional systems, we cannot conclude that specifications grading simpliciter is necessarily less arbitrary, and

thereby fairer, than traditional grading. Nevertheless, that we cannot definitively conclude that all versions of specifications are superior to all versions of traditional grading with regard to fairness, we should not lose sight of the fact that specifications grading typically has low levels of arbitrariness.

III.3. Motivation and Rigor

Susan A. Ambrose et al. explain that motivation depends on three main factors: the value factor, the value students attach to a goal; the expectancy factor, the degree to which students believe they are capable of achieving a goal; and the environmental factor, how supportive the environment is for students as they pursue a goal.⁴³ Under the value factor, Ambrose et al. note that people work harder when

- i. they want to achieve a goal (intrinsic value);
- ii. they find pursuing the goal enjoyable (intrinsic value);
and
- iii. the achievement of the goal helps the person achieve their further goals (instrumental value).

With regard to the expectancy factor, people tend to put in considerable effort when they

- iv. understand what their goal is,
- v. believe they are capable of achieving it, and
- vi. know how to achieve it.⁴⁴

Finally, motivation increases when people pursue goals in an environment that includes

- vii. helpful, approachable mentors, supervisors, or teachers,
- viii. few distractions, and
- ix. sufficient time and materials to pursue it.

The interaction among these factors is complex, and sometimes one factor is in tension with another. Nevertheless, in general, motivation, and thus time spent on the task, increases as each of these nine subfactors is more fully realized.

Some of these factors are not usually amenable to much teacher influence, especially environmental factors viii and ix. And, while a teacher can influence value goals (e.g., by selecting topics that tend to be of interest to the students that tend to take the course and designing learning activities that the students tend to enjoy) and environmental factor vii, the grading system does not appear to be the origin of the

influence. It is in the expectancy factors where a grading system can have a significant impact on student motivation.

Clear evaluative criteria are crucial for students to understand what their goal is. The availability of retries is closely tied to a student's belief that they can meet a high standard. How-to instruction has everything to do with whether a student knows how to achieve a learning objective. However, as we argued above, since a course using a traditional grading system might have clear grading criteria, retries, and how-to instruction, none of this yet shows that specifications grading is superior to traditional grading with regard to student motivation.

With regard to expectancy factors, rigor and motivation are connected. When a person knows what a 100-meter dash is, what Olympic sprinters can do, and what one's own running abilities are, they are unlikely to believe they can reach the goal of running a sprint faster than an Olympic sprinter. No amount of how-to instruction and practice will cause anyone (except perhaps for an already elite runner) to believe they could win such a race, and no amount of how-to instruction will increase their motivation to try. On the other hand, if the goal is to run faster than a typical two-year old, any person who knows they can run much faster than a two-year old will not be motivated to improve their running performance by how-to instruction and practice. However, when a person is given a goal that they want to achieve and the goal is (known by that person to be) achievable with a little help and practice, that person will likely be motivated to train. For expectancy factors to enhance motivation, it is crucial that students have achievable but difficult goals, be made aware of the fact that they can achieve them (through how-to instruction), and be allowed to practice achieving them (through retries) with regular formative feedback until they do achieve them. It is the binary (or trinary) evaluation and high, conjunctive bar for passing, along with the fact that one earns no credit if they do not get over it (even if they are close), that are doing much of the motivational work in specifications grading. Even if we assume that the individual evaluative criteria are equally clear, insofar as students experience the high binary (or trinary) bar as starker than the high bar needed for a B when it is embedded within a multi-layer rubric, specifications grading should improve student motivation.⁴⁵ Most importantly, by removing the possibility of C- and D-level assignment performance, by insisting that students achieve B-level performance or fail, the high bar in specifications grading does motivate many students to work harder than they might otherwise.

III.4. Faculty Time

Finally, Nilson asserts that faculty will save time by using a specifications grading system.⁴⁶ First, she argues, assignment grading will take

less time because there are fewer judgments to be made when criterion- and assignment-level marking are binary. Marking can often be as simple as indicating which specification was not met. But we have argued that some traditional marking may involve fewer judgments. Second, the time spent tallying points at the assignment level is eliminated. Third, if any piece of an assignment “is missing or falls short of requirements, [teachers] can stop reading. The [work] is not acceptable and may be returned to the student for revision and resubmission.”⁴⁷ But we have argued that providing good formative feedback requires pointing out where improvement is needed, and some areas would be noticed only by reading past the first mistake. So, reading entire assignments remains necessary to provide top-notch formative assessment. Finally, Nilson asserts, in courses where term grades are determined by bundles or modules, some students not wishing to pursue an A for the term will see clearly which assignments they simply need not turn in, reducing the overall grading load. We argue below that this fourth reason does not convince us that Nilson is correct. Finally, switching to specifications grading takes additional time. While this initial startup cost is not dispositive of total time savings, it does need to be factored into the analysis.

We have not experienced Nilson’s promised time savings. If many students take advantage of the availability of retries, the total amount of work to be graded increases unless other adjustments are made. Because of this, one of us switched from assigning four papers with no retries available to assigning two papers with three retries available. The initially lower performing students now tend to make five total attempts by using all of their retries. The strongest students, however, submit only two papers, passing or high passing both on the first try. However, the introduction of the high pass tier resulted in many students who passed their first attempt to retry in an attempt to improve from pass to high pass. Almost every student retries the first paper at least once. In the end, the total number of papers to be graded did decrease just slightly. However, the fact that students were actually using the feedback they received on their papers as they worked on revisions was so exciting that more time was spent providing feedback for each paper. In the end, the amount of time spent grading papers stayed roughly the same.

We are not disappointed that we have not experienced time savings by adopting new grading systems. Faculty certainly need to protect their time and balance their efforts in multiple areas to have successful careers and meaningful lives. But the notion that one should switch grading systems because it will reduce their total amount of work is not particularly compelling to us. Rather, the impact one’s grading time has on student learning is what matters more. Because students carefully

use teachers' feedback in the retry process, the feedback is valuable to them. Giving students formative feedback that they use is rewarding work. Even if specifications grading does not reduce "clock time," the time spent grading feels less tedious and onerous. Even if replacing tedious effort with valuable effort is not a quantitative savings, the qualitative difference is a significant merit of specifications grading.

IV. Conclusion

While we have not saved much time by using specifications grading, we have not lost any either, and the time we do spend grading is more meaningful to students and enjoyable to us than it used to be. The clarity of binary and conjunctive assignment grading where the bar for passing is high appears to motivate our students since their grades have been much better since we have moved away from traditional grading. We have a high level of confidence in our claim that the move away from traditional grading is the cause of improved student learning because all of the other potential causes of the increase that we can think of remained unaltered as we changed our grading systems. We teach at the same institution; the backgrounds of the students in our classes are essentially unchanged; the class size has not changed; the course material has not changed; the standard for B-level performance has not changed; and so on. We welcome further research using social science methodologies to confirm or disconfirm this reasonable hypothesis. Our aim here has been to clarify what exactly specifications grading is and argue that most faculty will better serve their students by adopting at least some of the practices of pure specifications grading. Students learn more in our courses when we employ binary and conjunctive assignment grading.

Notes

We are grateful to Tim Berg, Leslie Burkholder, Melinda Messineo, Jen Rowland, and the anonymous reviewers of *Teaching Philosophy* for many very helpful criticisms and suggestions that led to improvements in this paper.

1. Nilson, *Specifications Grading*.
2. *Ibid.*, 14.
3. Thanks to Leslie Burkholder for the cookie analogy.
4. Nilson, *Specifications Grading*, 58.
5. *Ibid.*, 25.
6. *Ibid.*, 58.

7. Ibid., 51.
8. Ibid., 55.
9. Ibid., 5.
10. Ibid., 58.
11. Ibid., 37–39.
12. Armacost and Pet-Armacost, “Using Mastery-Based Grading,” T3A-222.
13. Nilson, *Specifications Grading*, 69.
14. Ibid.
15. Cahill and Bloch-Schulman, “Argumentation Step-by-Step.” This work was published prior to the popularization of the moniker “specifications grading.”
16. Nilson, *Specifications Grading*, 128.
17. Ibid., 6.
18. Ibid., 5.
19. It is worth noticing that conditions five and six refer only to courses where term grades are determined via bundles or modules. A pure specifications system does not require that term grades be determined with bundles or modules; a point system may be used to determine term grades in pure specifications grading (Nilson, *Specifications Grading*, 67–68). A system that satisfies conditions one through four but not five and six remains “pure” when term grades are determined via a point system.
20. We discuss clarity in detail later in the main text, but it is helpful at this point to distinguish between familiarity and clarity. Specifications grading is unfamiliar to most students. Instructors switching to specifications grading need to explain it numerous times to ensure that students understand how they will be evaluated. Initial unfamiliarity is entirely compatible with clarity.
21. Nilson, *Specifications Grading*, 120.
22. Ibid., 121.
23. Ibid., 122.
24. Ibid.
25. Rapaport, “A Triage Theory of Grading,” 353. Rapaport also reports that circa 1785, Yale used a four-tiered system: best, second best, inferior, and worse.
26. See, for example, Blackstone and Oldmixon, “Specifications Grading in Political Science”; Roberson, “Techniques for Using Specifications Grading in Computer Science”; Martin, “Introducing Components of Specifications Grading to a General Chemistry I Course”; Elkins, “Grading to Learn”; Bonner, “Grading Rigor in Counselor Education”; Rylaarsdam and Heinz, “Specifications Based Grading”; Jones, “Experimenting with Specifications Grading”; Talbert, “Observations and Changes”; Mittell, “Rethinking Grading”; Talbert, “Putting Specs Grading to Work”; Talbert, “If You Don’t Succeed”; and Talbert, “Specifications Grading.”
27. Behun, “Specifications Grading,” 3.
28. Ibid., 4.
29. Scott, “Syllabus Showcase.”

30. McCrickerd, "Review of *Specifications Grading*."
31. For more on breaking down philosophical skills to show students how to approximate expert performance, see Marcus, "Scaffolding for Fine Philosophical Skills."
32. Mulnix and Liberman, "Philosophers Folding Origami."
33. Nilson, *Specifications Grading*, 112.
34. *Ibid.*, 60–61.
35. McCrickerd, "Review of *Specifications Grading*," 117.
36. Close, "Fair Grades," 385.
37. *Ibid.*, 384.
38. Immerwahr, "The Case for Motivational Grading," and McCrickerd, "What Can Be Fairly Factored into Final Grades?"
39. Nilson's complete list of desiderata of a grading system has fifteen entries, but as the subtitle of her book indicates, the top three are restore rigor, motivate students, and save faculty time. The fifteen desiderata include uphold high academic standards, reflect student learning outcomes, motivate students to learn, motivate students to excel, discourage cheating, reduce student stress, make students feel responsible for their grades, minimize conflict between faculty and students, save faculty time, give students feedback they will use, make expectations clear, foster higher-order cognitive development and creativity, assess authentically, have high interrater agreement, and be simple. Nilson, *Specifications Grading*, 9–13.
40. *Ibid.*, 12.
41. *Ibid.*
42. For more on arbitrariness in grading and how to avoid it, see Loftis, "Beyond Information Recall."
43. Ambrose et al., *How Learning Works*, 74–79. A useful application of their findings to philosophy is Green, "How to Motivate Students."
44. Ambrose et al., *How Learning Works*, 76–77.
45. One traditional system, the menu approach, maximizes the connection between student choice and motivation to such a degree that students may be more motivated within the menu approach than in a (somewhat poor) specifications system. For more on the menu approach, see Mills, "Letting Students Choose."
46. Nilson, *Specifications Grading*, 11–12.
47. *Ibid.*, 71.

Bibliography

- Ambrose, Susan A., Michael W. Bridges, Michele DiPietro, Marsha C. Lovett, Marie K. Norman, and Richard E. Mayer. *How Learning Works: Seven Research-Based Principles for Smart Teaching*. San Francisco: Jossey-Bass, 2010.
- Armocost, Robert L., and Julia Pet-Armocost. "Using Mastery-Based Grading to Facilitate Learning," in *33rd Annual Frontiers in Education, Westminster, Colo., November 5–8, 2003*, T3A-20, IEEE. <https://doi.org/10.1109/FIE.2003.1263320>

- Behun, William. "Specifications Grading: A Useful Two-Year College Alternative," *The APA Newsletter on Philosophy in Two-Year Colleges* 17(1) (Fall 2017): 3–4.
- Blackstone, Bethany, and Elizabeth Oldmixon. "Specifications Grading in Political Science," *Journal of Political Science Education* 15(2) (2019): 191–205.
<https://doi.org/10.1080/15512169.2018.1447948>
- Bonner, Matthew W. "Grading Rigor in Counselor Education: A Specifications Grading Framework," *Educational Research Quarterly* 39(4) (June 2016): 21–42.
- Cahill, Ann J., and Stephen Bloch-Schulman. "Argumentation Step-by-Step: Learning Critical Thinking through Deliberate Practices," *Teaching Philosophy* 35(1) (March 2012): 41–62. <https://doi.org/10.5840/teachphil20123514>
- Close, Daryl. "Fair Grades," *Teaching Philosophy* 32(4) (December 2009): 361–98.
<https://doi.org/10.5840/teachphil200932439>
- Elkins, Donna M. "Grading to Learn: An Analysis of the Importance and Application of Specifications Grading in a Communication Course," *Kentucky Journal of Communication* 35(2) (Fall 2016): 26–48.
- Green, Paul. "How to Motivate Students: A Primer for Learner-Centered Teachers," *AAPT Studies in Pedagogy* 1 (2015): 47–60. <https://doi.org/10.5840/aaptstudies20159184>
- Immerwahr, John. "The Case for Motivational Grading," *Teaching Philosophy* 34(4) (December 2011): 335–46. <https://doi.org/10.5840/teachphil201134446>
- Jones, Jason B. "Experimenting with Specifications Grading," *The Chronicle of Higher Education Blogs: ProfHacker* (blog), March 23, 2016. <http://www.chronicle.com/blogs/profhacker/experimenting-with-specifications-grading/61912>.
- Loftis, J. Robert. "Beyond Information Recall: Sophisticated Multiple Choice Questions in Philosophy," *AAPT Studies in Pedagogy* 5 (2019): 89–122.
<https://doi.org/10.5840/aaptstudies2019121144>
- Marcus, Russell. "Scaffolding for Fine Philosophical Skills," *AAPT Studies in Pedagogy* 5 (2019): 34–67. <https://doi.org/10.5840/aaptstudies201912642>
- Martin, Langdon J. "Introducing Components of Specifications Grading to a General Chemistry I Course," in *Enhancing Retention in Introductory Chemistry Courses: Teaching Practices and Assessments*, ed. Supaporn Kradtap Harwell and Tanyo Gupta, 105–19. Washington, D.C.: American Chemical Society, 2019.
<https://doi.org/10.1021/bk-2019-1330.ch007>
- McCrickerd, Jennifer. "Review of *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*," *Teaching Philosophy* 43(1) (March 2020): 114–17.
<https://doi.org/10.5840/teachphil20204317>
- McCrickerd, Jennifer. "What Can Be Fairly Factored into Final Grades?," *Teaching Philosophy* 35(3) (September 2012): 275–91. <https://doi.org/10.5840/teachphil201235329>
- Mills, Andrew. "Letting Students Choose: Investigating the Menu Approach to Graded Work," *AAPT Studies in Pedagogy* 5 (2019): 68–88.
<https://doi.org/10.5840/aaptstudies2019121043>
- Mittell, Jason. "Rethinking Grading: An In-Progress Experiment," *Just TV* (blog), February 17, 2016. <https://justtv.wordpress.com/2016/02/16/rethinking-grading-an-in-progress-experiment/>.
- Mulnix, Jennifer Wilson, and Alida Liberman. "Philosophers Folding Origami: Illustrating Essential Strategies for Learner-Centered Teaching," *Teaching Philosophy* 40(4) (December 2017): 437–462. <https://doi.org/10.5840/teachphil201812579>
- Nilson, Linda. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*. Reprint edition. Sterling, Va.: Stylus Publishing, 2014.

- Rapaport, William. "A Triage Theory of Grading: The Good, the Bad, and the Middling," *Teaching Philosophy* 34(4) (December 2011): 347–72. <https://doi.org/10.5840/teachphil201134447>
- Roberson, Christian. "Techniques for Using Specifications Grading in Computer Science," *Journal of Computing Sciences in College* 33(6) (June 2018): 192–93.
- Rylaarsdam, Robin Pals, and Cheryl Heinz. "Specifications Based Grading: Nearly Pointless Education that Points Students to Key Concepts," *The FASEB Journal* 30(1) Supplement (April 1, 2016): 662.13-662.13.
- Scott, Rebecca. "Syllabus Showcase: Rebecca Scott, Introduction to Philosophy," *Blog of the APA* (blog), October 23, 2019. <https://blog.apaonline.org/2019/10/23/syllabus-showcase-rebecca-scott-introduction-to-philosophy/>.
- Talbert, Robert. "If You Don't Succeed, Try Again: Timed Tests Using Specs Grading," *The Chronicle of Higher Education: Casting Out Nines* (blog), February 26, 2015. <http://www.chronicle.com/blognetwork/castingoutnines/2015/02/16/if-you-dont-succeed-try-again-timed-tests-using-specs-grading/>.
- Talbert, Robert. "Observations and Changes for Specs Grading, 8 Weeks In," *The Chronicle of Higher Education: Casting Out Nines* (blog), March 6, 2015. <http://www.chronicle.com/blognetwork/castingoutnines/2015/03/06/observations-and-changes-for-specs-grading-8-weeks-in/>.
- Talbert, Robert. "Putting Specs Grading to Work," *The Chronicle of Higher Education: Casting Out Nines* (blog), January 29, 2015. <http://www.chronicle.com/blognetwork/castingoutnines/2015/01/19/putting-specs-grading-to-work/>.
- Talbert, Robert. "Specifications Grading: We May Have a Winner," *Robert Talbert, PhD* (blog), April 28, 2017. <http://rtalbert.org/specs-grading-iteration-winner/>.

Sarah E. Vitale is Assistant Professor of Philosophy at Ball State University. Her research focuses on Marx and post-Marxism, especially on the notions of production, labor, and human nature, as well as contemporary feminist theory and the scholarship of teaching and learning. She is co-editor of the Radical Philosophy Review, the journal of the Radical Philosophy Association, as well as of The Weariness of Democracy (Palgrave Macmillan, 2019). Her recent publications include "Post-Marxist Political Ontology and the Foreclosure of Radical Newness (Philosophy Today 64:3) and "Beyond Homo Laborans: Marx's Dialectical Account of Human Essence" (Social Theory and Practice 46:3). At Ball State, she teaches Introduction to Philosophy as well as classes on contemporary philosophy, especially nineteenth- and twentieth-century continental philosophy and critical theory. sevital@bsu.edu

David W. Concepción is Professor of Philosophy at Ball State University. His research on teaching addresses inclusive pedagogy, feminist pedagogy, and course design. He has received the Lenssen Prize for research on teaching in philosophy, six university-level awards for teaching, the American Association of Philosophy Teachers (AAPT) Award for Outstanding Leadership in Advancing the Teaching of Philosophy, numerous grants, and the American Philosophical Association (APA) Prize for Excellence and Innovation. He is a past president of the AAPT and is currently chair of the APA Committee on Teaching Philosophy. dwconcepcion@bsu.edu