What Do Grades Mean? Variation in Grading Criteria in American College and University Courses

Anastasiya A. Lipnevich, Thomas R. Guskey, Dana M. Murano & Jeffrey K. Smith

Published online: 27 Jul 2020

Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020) What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice.* DOI: 10.1080/0969594X.2020.1799190

ABSTRACT

This study examined differences in the criteria used by college and university instructors in the United States to assign course grades. Two hundred and fifty course syllabi (159 from universities and 91 from four-year colleges) developed by randomly selected instructors from five academic disciplines (education, mathematics, science, psychology, and English) were examined to determine the extent to which instructors employed different criteria in assigning course grades in introductory-level courses. Sources of variation in grade assignment included the use of product versus process criteria, the prevalence of using performance exams, and the framing criteria for grades. Differences between institution types and among academic disciplines were also investigated. Results revealed significant differences among the five academic disciplines in grading criteria and the use of examinations, with instructors in education and English relying more heavily on process criteria. A significant interaction between institution type and academic discipline in grading criteria was also identified. Theoretical, practical, and policy implications are discussed along with avenues for further research.

KEYWORDS: Grading, product, process, syllabi, criteria

Assigning fair, accurate, and meaningful grades in college and university courses presents a challenge for all instructors. Even those with significant training in pedagogy have rarely learned about effective grading methods or the advantages and shortcomings of various grading strategies. Should grades reflect achievement only, or is grading more effective when it incorporates multiple aspects of a student's performance, such as effort and study habits? These concerns exist at primary and secondary levels of education as well, and varying viewpoints abound (e.g. Lipnevich and Smith, 2009; Brookhart, 2011; Cross & Frary, 1999; Guskey & Link, 2019; Smith & Smith, 2019).

Studies conducted in Spain show that instructors are typically given great latitude in devising the grading system for a course (Pandero et al., 2019). For example, Panadero et al. (2019) stated that Spanish university regulations place the responsibility for assessment decisions at the departmental level, which, in turn, allows instructors to devise and implement assessments without many constraints. This is also the case in the US, where universities rarely offer recommendations regarding the types of assessments, weight of assessments in determining final grades, or how final course grades should be calculated. To our knowledge, however, few studies have systematically examined these differences in the U.S. and those have focused mainly on the secondary level (e.g. Gullikson, 1985; Guskey & Link, 2019; McMillan, 2001; Randall & Engelhard, 2010). Hence, our intent was to shed the light on grading practices and grading criteria in the United States, highlighting issues relevant in institutions of higher education across the world.

Assigning grades in college courses

In attempting to seek a balance among alternative approaches to grading, many instructors simply reflect upon what they experienced as students in order to establish grading procedures for their own courses, despite the questionable validity of such practices (Allen, 2005). From these experiences they choose policies and approaches that they believe are fair, reasonable, and educationally defensible (Boothroyd & McMorris, 1992).

In describing their grading procedures, university instructors generally state that they base grades on how well students have achieved the specified learning goals for a course. Most consider this to be the fairest and most equitable way to determine course grades (Dweck, 2000; Kovas, 1993). However, not all of the evidence instructors consider in assigning course grades reflects articulated learning goals. In determining course grades most instructors aggregate multiple sources of evidence, including scores from major exams and compositions, projects or reports, exhibits of student work, and laboratory assignments, along with class attendance or participation, punctuality in turning in assignments, and perceived effort (Hu, 2005). These diverse grading criteria can be grouped into three broad categories: *product, process*, and *progress* criteria (Guskey, 2006). Product criteria describe summative demonstrations of *what* students know and can do at the point of assessment. Process criteria reflect *how* students got to that point of achievement or behaviors that enabled their learning. Progress criteria demonstrate how much knowledge and skill students gained or improved throughout the course.

In addition to factors that contribute to the overall grade, instructors vary in how they frame their grading systems. Three main framing systems are most frequently used by college and university instructors: a 100-point system, a percentage system, and an open-point system (Smith & Smith, 2009). In the 100-point system, all assignments are assigned points, and the sum of perfect scores on all the assignments is 100. In the percentage system every assignment counts for a certain percentage of the final grade, and these are all scored on a 100% basis (e.g. a student may score up to 100 on an assignment that accounts for 25% of the final grade). Some instructors may use letter grades with this weighting system instead of numerical grades. Finally, the open point system allocates a specific number of points for each assignment. The number of points a student achieves is then divided by the total number of points possible to get a percentage score. A letter grade is then obtained by translating the numerical grade onto a grading scale. Some professors

use a 4- or 5-point scale for all their grading, avoiding the 100-point system altogether, but this approach is not common (Smith & Smith, 2009).

The current study explores variation in the elements of student performance and framing methods instructors in colleges and universities use to determine students' course grades. Additionally, we analyzed the use of exams, a form of product criteria, in order to determine whether there were differences among academic disciplines and institution types. Our purpose was to unpack some of components in what Cross and Frary (1999) amusingly refer to as 'hodgepodge' grades, or grades that are derived from diverse measures of student performance and behavior. We also hoped to identify trends that may exist among different academic disciplines and types of institutions. Describing elements that instructors consider in determining students' grades, and which overall approach they take to grading, should help us clarify just what grades mean.

The U.S. higher education context

Because the research presented here describes grading practices in colleges and universities in the United States, some context for that setting may be helpful for readers who do not have experience with U.S. higher education. To begin, universities are distinguished from colleges historically by whether they offer doctoral programs: universities do while colleges do not. Although this remains generally true today, several prestigious institutions that use the term 'college' yet have strong doctoral programs. For example, Dartmouth College and Boston College are actually both universities but use the traditional term 'college' in their names. Also, colleges tend to be smaller and focus more on teaching rather than research. In the U.S, the term 'college' almost never refers to a high school, as it often does internationally.

Furthermore, grading in American colleges and universities is typically considered to be part of the 'academic freedom' extended to faculty members, resulting in great variability in grading across courses. Although many institutions offer guidelines and recommended practices for grading, faculty members turn in their grades at the end of each academic term and, unless the distribution is highly unusual, it is accepted as final. In most instances, marks of 90–100 receive an A; 80–89 receive a B; and 70–79 receive a C. In some institutions, below 70 is a failing mark, whereas in others 60–69 is a D and below 60 is a failure. American readers will be surprised to learn that internationally, 50 is often a passing score, and 80–100 is often an A (Brookhart et al., 2016). Also, in the U.S., 'marks' and 'grades' are used interchangeably, whereas in an international context, 'marks' are typically numerical and 'grades' are letters.

We also note that in many countries, the term 'professor' only refers to those who hold the rank of full professor, whereas in the U.S., it refers to faculty who hold ranks of assistant, associate, or full professor. To avoid confusion, we use the term 'instructor' throughout to refer to anyone who teaches at the college/university level.

The nature of grades

Grades can be seen as a vehicle to track student progress and mastery of course material. Bailey and McTighe (1996) extend the purpose of grading to communicate information about student

achievement to stakeholders in addition to students, including parents, school administrators, postsecondary institutions, and potential employers. Nevertheless, there seems to be no clear consensus about the purpose of grading (Brookhart, 2011). This lack of consensus with regard to purpose makes it difficult to decide what evidence to use in determining students' grades (Brookhart & Nitko, 2008). Brookhart et al. (2016) report that grades typically are comprised of a wide range of cognitive indicators in addition to non-cognitive factors. Smith et al. (2001) questioned this approach from a measurement perspective and clarified the need for reform in grading policies. Different sources of evidence will vary in their appropriateness and validity depending on the identified purpose of a grade.

The purpose of grades also comes into question when we consider the rising rate of grade inflation over the past 70 years (Brookhart et al., 2016). On average today, A's represent 43% of all letter grades issued in higher education institutions, up dramatically from the 15% issued in 1960. Evidence also indicates more A's and B's are assigned in smaller colleges compared to larger universities (Rojstaczer & Healy, 2012).

Although grades are inherently part of academic culture, studies show that grades issued during the learning process may not always be effective in helping students improve their knowledge and skills (Klapp, 2015). Lipnevich & Smith (2009) found that college students who received descriptive feedback on an essay exam (i.e. comments on their work suggesting specific steps towards improvement) outperformed students who received evaluative feedback in the form of a grade (medium effect sizes). Moreover, students who received a letter grade in addition to descriptive comment did not improve nearly as much as their counterparts who received comments without a grade. This suggests that evaluative feedback, particularly in the form of letter grades, may not be effective in helping students improve the quality of their work.

In addition to the ambiguity of purpose, there is also a great deal of variation in how most college and university course grades are reported. The vast majority of grades are reported as single letter grades for each course (Brookhart, 2011). This requires instructors to combine all of the diverse sources of evidence they gather on students' performance into a single symbol (Brookhart, 1991, 2009; Cross & Frary, 1999). Even when instructors clarify their weighting strategies and employ computerized grading programs to ensure accuracy in their computations, the final grade can be a confusing amalgamation that is difficult to interpret and rarely presents a true picture of the complex nature of students' achievement or proficiency (Guskey, 2002; Sadler, 2010).

Amalgamated and differentiated grades

Amalgamated grades – the standard practice in most U.S. colleges and universities – reflect the aforementioned idea of the 'hodgepodge grade' (Cross & Frary, 1999). A student is evaluated on multiple written, oral, group assignments, participation, attendance, and exams, and then given one grade that is reflective of all components combined (Authors, 1996; Brookhart, 1991, 2009). Although various components of a student's performance, both achievement-based and behaviorally-based, are evaluated separately, all are aggregated into a final grade reflective of holistic performance (Cizek et al., 1996; O'Connor, 2009; Royal & Guskey, 2015). A grade of *A*, for example, may mean the student knew what was intended before instruction in the course

began (product); did not learn as well as expected but displayed exceptional effort (process); or simply made significant improvement (progress). Amalgamated grades combining various aspects of students' performance can be difficult for various stakeholders (including the student) to interpret and can result in grades losing clear or direct meaning.

An alternative approach is differentiated grading, where students are graded separately on different aspects of course performance. With this approach students receive individual grades for different aspects of evaluation such as academic performance, behavior, work ethic, etc. (see Guskey, 2011; Stiggins, 2008). This practice can be especially salient in fields such as medicine, nursing, music, and education, where performance is crucial to student success independent of their mastery of course content (Royal & Guskey, 2015; Webb et al., 2003). Although this practice is not a common practice in the United States, many K-12 schools in Canada have used this system many years (O'Connor, 2010). Ultimately, differentiated grades that assign independent grades to separate measures of student achievement, as opposed to grades that combine multiple aspects of student performance, can give a more meaningful and accurate account of student performance in various areas (Guskey, 2002).

In the next section, we examine three broad categories of criteria used for grading: product, process, and progress.

Product criteria

Product criteria relate to *what* students know and are able to do at a specific point in time (Guskey, 2006). University instructors who use product criteria normally base grades on final examination scores, final products (reports or projects), overall assessments, and other culminating demonstrations of learning. Product criteria can best be understood as summarizing students' achievement or mastery of course academic goals. They are the most frequently used grading criteria in post-secondary institutions (Milton et al., 1986).

Grades based on product criteria align with summative evaluation of student achievement (Guskey, 2011; O'Connor, 2009, 2010). Summative assessment was initially differentiated from formative assessment in education by Bloom (1968, 1971)), who based his concepts on distinctions in program evaluation originated by Scriven (1967). Summative assessments are intended to objectively measure what students have learned at a given point in time (Elwood & Murphy, 2015; Taras, 2005) and are the primary form of product criteria.

Exams are the most prevalent examples of product criteria in college and university courses. Most scholars advocate for objective measurement of student skills on achievement-based exams, but recognize other external factors can influence students' performance. Test anxiety, for example, has been studied intensively by educators and psychologists worldwide for the past 50 years, resulting in a multitude of theoretical approaches, causal mechanisms, and proposed interventions to help students cope (Zeidner, 2007). High levels of test anxiety have been repeatedly shown to relate to decreased performance on exams (e.g. Cassady & Johnson, 2002; Seipp, 1991). Alternatively, exam performance can also be influenced by testwiseness, the ability to do well on exams by being attuned to the characteristics and subtleties of the exam format (Fagley, 1987; Sarnacki, 1979; Smith, 1982).

Exams are not the only source of product criteria, however. Alternative approaches to product grading exist, and are oftentimes preferred by students. In one teacher education program, for example, 82 percent of students who selected an oral presentation as a summative assessment reported that they preferred this type of summative assessment in comparison to a traditional written exam or essay (Turner et al., 2013). Qualitative data from this study revealed, however, that some students experienced a wide range of stress and negative emotions pertaining to the oral presentation.

Process criteria

Process criteria are used by instructors who believe that grades should reflect not only students' final achievement but also *how* they got there. Instructors who consider students' effort or work habits when assigning grades are using process criteria. So are those who count ongoing classroom formative assessments and concept checks, punctuality in turning in assignments, class participation, or attendance. Process criteria can be crucial in a holistic assessment of student performance. Medical education departments have changed assessment practices in order to incorporate measures of student characteristics that contribute to successful performance. Process criteria are essential in evaluating potential candidates in these fields (Guskey, 2011; Usherwood et al., 1995; Webb et al., 2003). Supervisors assessing capstone projects of engineering students in undergraduate programs have also advocated for the evaluation of process in addition to product criteria (Lawson et al., 2015).

In general, process criteria can be separated into three broad categories (Guskey & Link, 2019). The first is learning enablers, which include formative assessments, homework, and class participation. They refer to ongoing indicators of student engagement with the course. The second set includes social and emotional characteristics. For example, ethics, compassion, perseverance, professionalism, enthusiasm, etc. The final category is compliance, which is indexed by students' turning in assignments on time, punctuality, and professional behavior in the classroom (e.g. not texting or engaging in course-irrelevant conversations). By assessing process, instructors are able to capture aspects of student performance not necessarily directly included in product criteria.

Progress criteria

Progress criteria are based on how much students gain from their learning experiences. Other names for progress criteria include 'learning gain,' 'improvement scoring,' 'value-added learning,' and 'educational growth.' Typically, progress criteria involve the difference between a measure of where the student was at the beginning of a course, and where they are at the end. This might be assessed through overall summative pre- and post-measures given at the beginning and end of a course, or a tally of the number of learning objectives achieved by students during the course (see, e.g. Guskey, 2001; Wiggins, 1996).

Grade assignment

Most instructors at the undergraduate level employ an amalgamated grading system, as opposed to assigning differentiated grades to students for different aspects of performance. These

amalgamated grades typically comprise a combination of product, process, and progress criteria (Guskey, 2006). Although instructors defend this practice on the basis of fairness (Tippin et al., 2012), it significantly confounds the meaning of the grade (see McMillan, 2001). Given these circumstances surrounding grade assignment, the interpretation of grades becomes questionable, especially in attempts to compare grades that come from different institution types or academic subjects.

In addition to the different criteria that constitute the actual grades, university faculty also present their grade formulation to students in different ways. In other words, in addition to *what* constitutes grades, college and university instructors differ in *how* they calculate grades. As described above, Smith and Smith (2009) identified three common ways in which professors frame course grading systems to their students: a 100-point system, a percentage system, and an open point system. These grade framing approaches are depicted below in Table 1.

Table 1. Grade framing approaches

Table 1. Grade framing approaches (Smith & Smith, 2009).

Course Component	**100-Point System**	**Percentage System**	**Open Point System**
Participation	5	5%	25
Homework	5	5%	25
Quizzes	10	10%	50
Midterm Exam	20	20%	100
Final Exam	25	25%	125
Term Paper	25	25%	125
Course Presentation	10	10%	50
Total	100	100%	500 (then total is divided by 5)

Although different approaches to grade framing result in mathematically identical course grades, Smith and Smith (2009) found that grade framing approaches had different effects on student perception of the course. The researchers randomly assigned undergraduate students to receive a syllabus describing the percentage system, 100-point system, or open-point system. In each condition, the grading approach consisted of identical assignments that were assigned the same ultimate weight in determining the overall course grade. After viewing the course assignments framed according to each condition, students were asked to complete a survey about how they would react to various assignments in the course. All survey items were on a 5-point (strongly agree to strongly agree) Likert scale and assessed the following constructs: motivation, anxiety, confidence, effort, demonstration, self-efficacy, usefulness, and preference. After completing the survey, students also gave qualitative feedback about their reactions to each of the course assignments, under their respective framing condition.

Results showed that student motivation, effort, and confidence in completing assessment assignments were significantly lower in the 100-point framing condition, compared to the open-point system. The difference was theoretically attributed to Tversky and Kahneman (1981) work,

which suggests that the 100-point system may indicate a 'loss-situation' to students, while the open-point system reflects a 'gain orientation', in which students are earning course points, as opposed to losing them. Smith and Smith (2009) work therefore suggests that although various grade framing approaches are identical mathematically, they may not be perceived identically in the minds of students. We were therefore interested in the frequency with which the various approaches to grade framing are realized in practice.

Current study

Because of variation in grading criteria across different universities and colleges, it is difficult to objectively interpret the meaning of the grades students receive. An 'A' in one course may be very different from an 'A' in another course, both within and among departments and institutions, with a variety of underlying factors both contributing and weighing differently upon the amalgamated grade. The purpose of this study was to explore the variation among college and university faculty members in their use of product, process, and progress criteria in assigning course grades in introductory level courses. Specifically, we were interested in the percentage of course grades that are determined by performance on exams, a specific form of product criteria. Additionally, we sought to determine if there were differences in approaches to grade framing. For each source of variation, we were interested in any significant differences among academic disciplines and between different types of institutions. In exploring these differences, we aimed to gain a better understanding of the trends in various constituents and calculation methods of overall course grades.

Method

Syllabi

A syllabus is an official document in which course instructors present descriptions of course content, along with expectations, responsibilities, assignments, and criteria for evaluation (Panadero et al., 2019; Stanny et al., 2015). The syllabus is viewed as a contract between an instructor and the student. All syllabi are expected to define assessment approaches and include information about types of assessment (e.g. examinations, essays), and weights of individual assessments in determining the final course grade, etc.

Syllabi analysis offers a window into the instructional and evaluative practices that instructors employ. It has been used in prior studies to explore numerous research questions ranging from the alignment of syllabi with learning outcomes to the evaluation approaches used in courses (Bers et al., 2000; Cashwell & Young, 2004; Panadero et al., 2019; Rathbun et al., 2017). In the U.S. instructors have significant latitude in designing their courses, but must adhere to specific university policies in designing their course syllabi. In most universities, departing from the policies delineated in the syllabus is considered a contractual violation and represents a legal issue. Hence, syllabi represent a valuable source for identifying what grading practices may look like within and across disciplines.

Materials and procedure

To determine the differences in grading criteria used by college and university instructors in various academic disciplines, course syllabi were gathered from randomly selected college and university websites. From each institution's website, one introductory level course at the undergraduate level was randomly selected in each of five academic disciplines: English, mathematics, science, psychology, and education. Students in most U.S. colleges and universities are required to take courses in English, mathematics, science, and psychology during their first two years of studies, with these four departments representing the largest and rather distinct academic domains (NCES, 2020). We also selected education because instructors in departments of education may be expected to have deeper understanding of pedagogy and educational assessment.

If the syllabi of the selected course did not include a detailed description of the criteria by which course grades would be assigned, another introductory course within that department and institution was chosen. If institutions did not have one syllabus per academic domain available, we contacted department chairs and requested representative syllabi in a specific domain. Our final sample included 50 syllabi from each of the five academic disciplines. One hundred and fifty-nine (64%) of the selected syllabi came from large, comprehensive universities and 91 (36%) were from smaller, four-year colleges. These proportions approximate the numbers of students attending such institutions overall. All institutions were in the United States, and institutions awarding doctorate degrees were considered 'universities', whereas all others were considered 'colleges. [1] '

Each course syllabus was read to determine the specific sources of evidence that would be considered in determining the course grade. These sources of evidence were then independently coded by all authors and two graduate research assistants as reflecting product, process, or progress criteria. Assignments coded as *product* included exams, midterms, finals, papers, presentations, portfolios, essays, article reviews, book reviews, lesson plans, and literature summaries. Constituents of *process* criteria included homework, attendance, reflection papers, journal entries, discussion board posts, participation, classroom quizzes, take-home assignments, in-class assignments, peer evaluations, classroom observations, and email correspondence. *Progress* criteria entailed grading based on learning-gain scores typically based on course pre-test, post-test comparisons.

Syllabi were then coded for grade framing approach, categorized by the 100-point system, percentage system, and open point system. Inter-rater reliability in coding was.95. The few instances of disagreement were discussed and consensus reached. For each course we recorded the number of different sources of evidence employed in each of the three categories of criteria (i.e. product, process, and progress), and the percentages assigned to each category in determining course grades. Percentage scores were transformed using a natural log transformation in order to account for the values of 100% in the distribution, and then treated as continuous variables in analyses. Grade framing was coded for each syllabus and analyzed as a categorical variable.

Results

Variation in grading criteria

We began our analysis by looking into some of the nuanced differences between uses of product, process, and progress criteria. No syllabi in the sample courses showed evidence related to progress criteria considered in determining students' course grades. In some instances, we observed that process criteria related to attendance or class participation were considered a direct portion of a student's grade. In several cases, as much as 20% of the course grade was based on regular class attendance. More often, however, instead of being included directly as a portion of the grade, process criteria served as a source of grade reduction. Forty-two per cent of syllabi analyzed included criteria upon which grade deductions would be made, and these stemmed exclusively from process criteria. In several instances, for example, a second unexcused absence resulted in a 5% reduction in the course grade. Similarly, in many courses, turning in an assignment a day late resulted in a 10% reduction in the assignment grade. An additional process criterion cited as a possible deduction was any disrespectful or unprofessional behavior in a class setting. Instructors in English classes specified deductions most frequently, with 84% of the English syllabi reviewed reflecting deductions based on process criteria. Science, mathematics, psychology, and education instructors incorporated deductions into their grading criteria less often, with just 32%, 22%, 36%, 40% of course syllabi containing explicit mention of deductions. Professors at small colleges specified deductions more frequently than professors at larger universities (58% versus 34%). In all cases across institution types and academic departments, grade deductions were based on exclusively process criteria.

Following our descriptive review of progress, process, and product criteria on syllabi, we sought to examine if there were differences in process and product criteria percentages by academic subject area and academic institution type. All mean percentages by academic subject and institution type are reported in Table 2.

Table 2. Means and standard deviations of process, product, and exam percentages by academic subject area and institution type.

Table 2. Means and standard deviations of process, product, and exam percentages by academic subject area and institution type.

	n	***M*(*SD*) Process**	***M*(*SD*) Product**	***M*(*SD*) Exams**
Overall	250	.27(.19)	.72(.19)	.47(.33)
Academic subject	250			
English	50	.26 (.16)	.74 (.15)	.16 (.22)
Mathematics	50	.25 (.15)	.75 (.15)	.61 (.22)
Science	50	.31 (.17)	.69 (.17)	.61 (.22)
Psychology	50	.21 (.23)	.78 (.24)	.63 (.32)
Education	50	.34 (.21)	.64 (.23)	.25 (.21)
Institution Type	250			
College	91	.26 (.19)	.73 (.19)	.47 (.29)
University	159	.28 (.19)	.71 (.20)	.48 (.34)

Variation in Use of process criteria by academic subject and institution type

A paired samples t-test was done to determine if the amount of product versus process criteria significantly differed overall. Across all 250 syllabi, product criteria were used significantly more than process criteria in determining students' course grades ($t_{(249)} = 18.32$, $p < .01$, $d = 2.30$).

We sought to determine if differences exist by academic subject area and institution type in the use of process criteria. A one-way ANOVA with raw percentage data revealed a significant difference among academic subjects in the use of process criteria ($F_{(4, 245)} = 3.80$, $p = .01$). Due to the nature of data reflecting percentages and their distributions, a log transformation was applied to the raw percentages and the analysis repeated (Atkinson, 1985). This approach also indicated significant differences among academic disciplines in use of process criteria ($F_{(4, 245)} = 6.24$, $p < .01$). We therefore report inferential statistics with the log transformation throughout the results section to properly account for the distribution of percentages in the raw data.

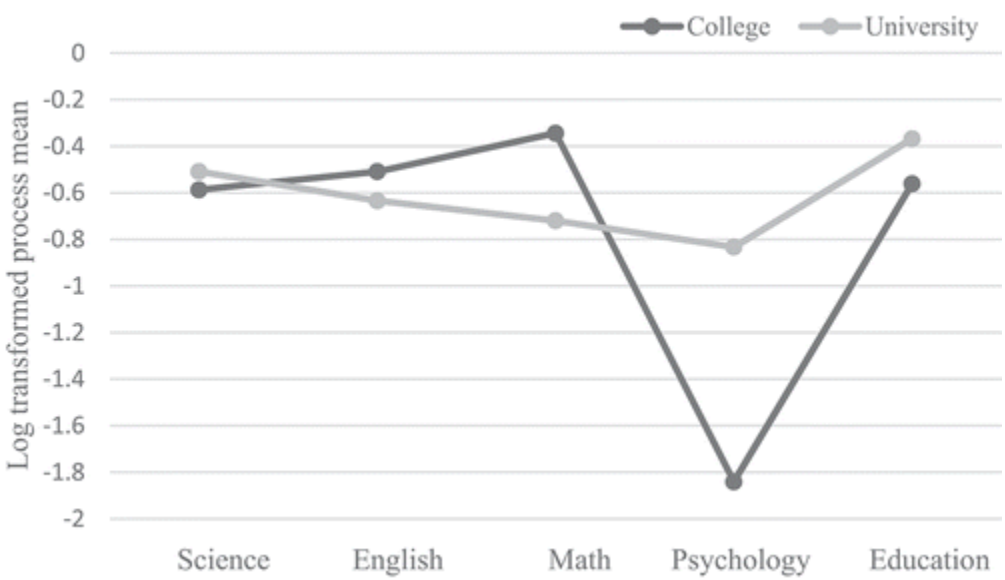
Tukey post hoc comparisons of the transformed data revealed significant differences between academic subject areas. We found that psychology courses use significantly less process criteria than science, English, mathematics, and education courses (all *p*s < .01), indicating that psychology courses used the least process criteria when compared to other domains.

Although differences appeared among subject domains in the use of process and product and process criteria, there did not appear to be differences in criteria use between colleges and universities. University syllabi had a mean of 28.2% use of process criteria, whereas courses taught in four-year colleges used a mean of 26.1% process criteria ($t = 1.23$, $p = .22$). When

considering institution type alone, both types of institutions use similar portions of product and process criteria in determining grades.

Last, we tested a possible interaction between academic subject and institution type in the percentage of process criteria used in students' grade calculations. The overall model was significant ($F_{(9, 240)} = 5.32$, $p < .01$, $\eta p^2 = .17$), with a significant main effect of academic subject ($F_{(4, 240)} = 9.27$, $p < .01$, $\eta p^2 = .13$), and a significant interaction effect of academic subject by institution type ($F_{(4, 240)} = 4.80$, $p < .01$, $\eta p^2 = .07$). Comparisons of results across different types of institutions revealed several interesting patterns, including a discipline by institution interaction, depicted by Figure 1. We see that psychology courses in colleges use less process criteria than those in universities ($d = -.85$, 95% CI = (−.24 – −1.45), and that mathematics courses in colleges use more process criteria than those in universities ($d = .64$, 95% CI = (.04–1.2).

Figure 1. Interaction of academic subject and institution type on use of process criteria.

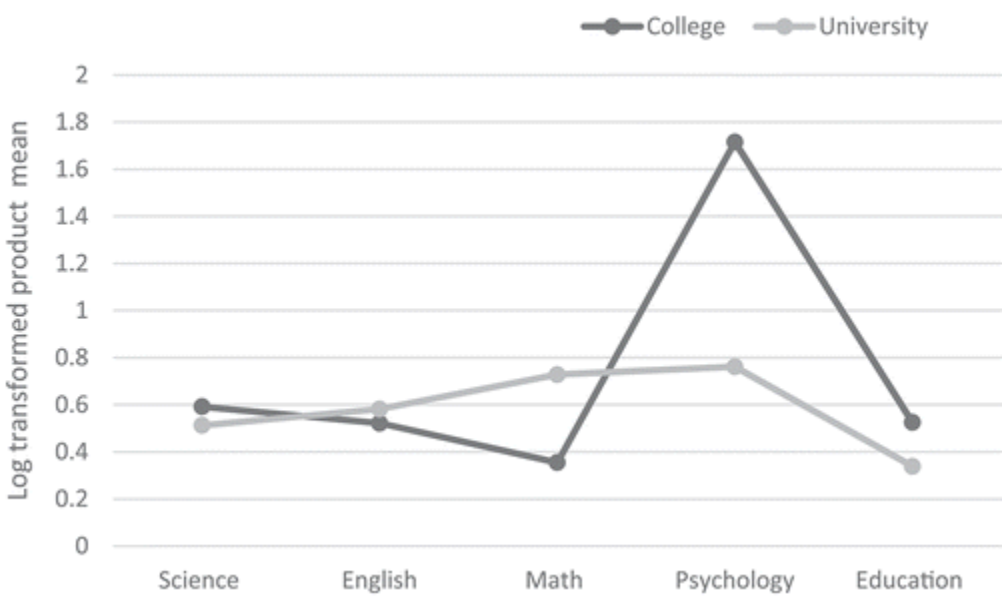


Variation in use of product criteria by academic subject and institution type

We repeated the same analyses to examine differences in the use of product criteria. A one-way ANOVA revealed significant differences among academic domains in the percentage of product criteria used in grade calculation ($F_{(4, 245)} = 4.04$, $p < .001$). Applying the same logit transformation to appropriately handle the distribution of percentages, the interpretations of this finding remain constant ($F_{(4, 245)} = 5.26$, $p < .01$). Tukey post-hoc analyses showed that psychology used significantly more product criteria than each of the remaining four domains, with each pairwise between-subject comparison reaching statistical significance (all ps < .05). Education differed significantly from mathematics, the second most frequent user of product criteria ($p = .05$). Overall, we saw that mathematics and psychology courses used product criteria most frequently across subject domains.

Although there was no significant difference by institution type on percentage of product criteria used ($t_{(248)} = -1.32$, $p = .19$), we tested the interaction between institution type and academic domain to determine if the effect of academic domain remained constant across institution type. The overall model was significant ($F_{(9, 240)} = 4.72$, $p < .01$, $\eta p^2 = .15$), with a significant main effect of academic subject ($F_{(4, 240)} = 7.83$, $p < .01$, $\eta p^2 = .12$), and a significant interaction effect of academic subject by institution type ($F_{(4, 240)} = 4.48$, $p < .01$, $\eta p^2 = .07$). Analogous to the findings for use of process criteria, we see that psychology courses in university settings use significantly more product criteria than psychology courses in university settings ($d = .83$, 95% CI = (.22–1.44). We also found that mathematics courses in universities use significantly more product criteria than mathematics courses in colleges ($d = .64$, 95% CI = (.04–1.23). Figure 2 shows this interaction effect.

Figure 2. Interaction of academic subject and institution type on use of product criteria.



Variation in exam usage by academic subject and institution type

We were also interested in learning more about different types of product criteria used in grading. Exams were the most frequently used form of product criteria, with 78.8% of the syllabi in the sample incorporating exams as part of a student's overall grade. We examined differences in subject areas and institution type in how heavily exams were weighed in students' grades. Table 1 shows average percentages of students' grades derived from exams by academic subject and institution type.

We fit a one-way ANOVA model to determine if exam percentage differed by academic subject area. The use of exams as grading criteria differed significantly both when considering raw

percentage data ($F_{(4, 245)} = 55.89$, $p < .01$) and after applying a log transformation to correct for the distribution of percentage data ($F_{(4, 245)} = 55.00$, $p < .01$). We used the transformed data for the remainder of the analyses. Tukey post hoc comparisons revealed significant pairwise differences (all *ps* < .05) between English and science, mathematics, and psychology courses, showing that English courses used significantly fewer exams than the other three academic domains. Pairwise comparisons between education and science, mathematics, and psychology were also all statistically significant ($p < .05$), showing that education courses also relied less on exams than other content areas. There were no significant differences between English and education courses. These findings show that education and English courses use fewer exams as product criteria in comparison to mathematics, science, and psychology.

Although there was no significant difference by institution type on percentage of product criteria used ($t_{(248)} = -.108$, $p = .91$), we tested the interaction between institution type and academic domain to determine if the effect of academic domain remained constant across institution type on exam percentage. The overall model was significant ($F_{(9, 240)} = 26.18$, $p < .01$, $\eta p^2 = .50$), with a significant main effect of academic subject ($F_{(4, 240)} = 48.56$, $p < .01$, $\eta p^2 = .45$), and a significant interaction effect of academic subject by institution type ($F_{(4, 240)} = 2.60$, $p = .04$, $\eta p^2 = .04$). Figure 3 shows the interaction between academic subjects and institution type on exam percentage.

Figure 3. Interaction of academic subject and institution type on use of exams.

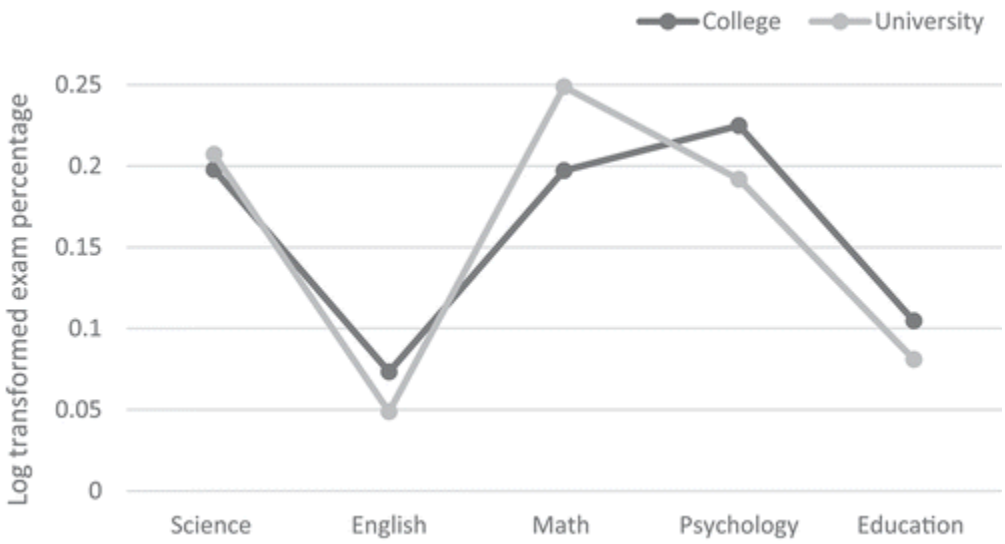
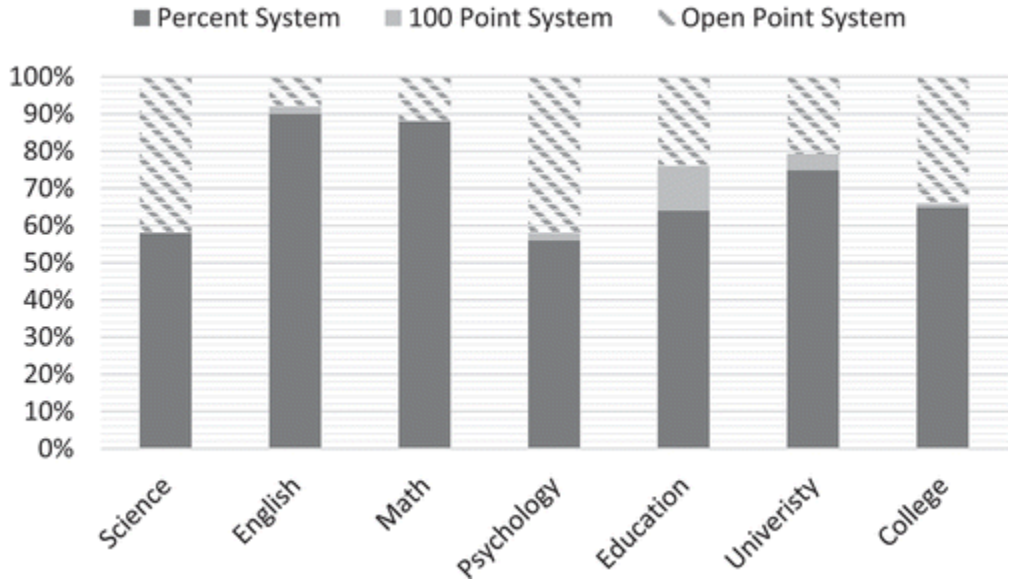


Figure 4. Grade framing approaches by academic subject and institution type.



Variation in grade framing

Finally, we examined differences among different types of grade framing. The percentage system was used most frequently (71.2% of syllabi), followed by the open-point system (25.6% of syllabi), with only 3.2% of syllabi using the 100-point approach. There were significant differences between college and university use of grade framing ($\chi 2 = 6.794$, $p = .033$), with larger universities showing more use of the percentage system than smaller colleges. There was also a significant difference in framing system use among the five academic domains ($\chi 2 = 43.643$, $p < .01$). We see that few syllabi used the 100-point framing approach, which can be deemed a positive finding based on decreased student motivation, effort, and confidence when 100-point grade framing approaches are used (Smith & Smith, 2009). See Figure 4 for grade framing approaches.

Discussion

The current study aimed at investigating grading practices of instructors who teach introductory level courses in colleges and universities across different academic departments. Overall, and consistent with the trends in the literature, we found that product criteria were used more frequently than process criteria when assigning grades. Although most instructors combine both product and process criteria, the ratios at which they combine them vary greatly. There is also a great deal of variance in how highly weighted exams are in course grades across academic departments. In some instances, only product criteria were considered in determining course grades while others weighed process criteria more heavily. Instructors also vary in whether they apply deductions, or whether grades were calculated using solely additive methods. Lastly, college and university instructors used different framing approaches in their syllabi, with most

syllabi using the percentage-system, but others using open-point approaches and a few employing the 100-point framing approach.

In disciplines like psychology and mathematics, product criteria were used significantly more than process criteria in grading students. Psychology instructors, for example, tend to base the grades they assign almost exclusively on product criteria (exams and quizzes) and rarely considered process criteria (attendance, class participation, or ongoing online posts). On the other hand, instructors in domains such as English and education used more process criteria, and, hence, relied less heavily on exams in the computation of course grades.

Domains in which product criteria are used more frequently compared to process criteria present opportunities for future research. Subjects that integrate the most process criteria, such as English and education, seem to include assignments that are focused on writing, and, possibly, instructors do not consider final written assignments to be exams. Disciplines like mathematics and psychology focus more on problem-solving and design principles, concepts that may be harder to assess using process criteria; exams may be the most accessible and reliable means of measuring whether or not students can solve specific problems. Alternatively, it may be that instructors believe these 'non-academic' factors should have less influence on grades that are supposed to reflect academic achievement. That is, it may be more a difference in perceptions of the purpose of the grade rather than ease or difficulty in measurement (Allen, 2005; Guskey, 2006; Smith & Smith, 2009). A further exploration of how to assess mathematics-centric domains using more process criteria could be beneficial to professors in those fields. Interestingly, process criteria were used as frequent bases for course reduction. Tardiness with assignments and homework submissions, attendance, and other indicators of process criteria were often counted to deduct points. This presents additional support to the idea that amalgamated grades are difficult to interpret and higher education may benefit from differentiating assessment criteria. After all, a stellar mastery of the course content (as indicated by product criteria) may be hidden behind poor organizational habits and inefficient time management (indexed through process criteria) rendering interpretations of grades highly problematic (Guskey & Link, 2019).

An additional salient finding is the absence of progress criteria in the calculation of any grades in the sampled courses. Several plausible explanations exist for this, the foremost being the difficulty in measuring student progress, and furthermore, an indistinct interpretation between counting grades received earlier on in the semester at face value, or as interpreting differences between lower grades and higher grades earned later on in a course as scores on product-criteria. Calculating learning-change scores for each student is difficult and time consuming, and a likely reason why faculty are unlikely to engage in such a process. Further, and most importantly, at the post-secondary level, interest appears to be not on what students gain, but rather on certifying competence. That is, the main question appears to be: Did students master the specific competencies or reached learning goals specified for the course? Ceiling effects, where students who are already scoring in the maximum range at the beginning of a course, also present an obstacle to measuring progress. The use of technology in evaluating student-learning progress, such as through interactive tablets or formative in-class data collection, yields a promising avenue for the integration of progress criteria into grade calculations (Office of Educational Technology; US Department of Education, 2016). It may also simply be the case that faculty

believe that grades should solely reflect where a student stands at the end of instruction as opposed to considering where the student started.

Although informative, this study is not without limitations. To begin, we considered introductory courses only. Because these courses often have large enrolments, especially at larger universities, the efficiency of grading can become a factor. In psychology, mathematics, and to a degree, science, the use of objective exams makes assignment of grades much more efficient. In English, where writing development may be a key objective, and in education, where modelling good instructional practice may be a concern, the use of such exams may be less attractive. As a result of the selection procedure and screening syllabi for inclusion based on academic subjects, our group sizes for college and university courses were uneven. Subsequent studies could use stratified random sampling at each level of both variables from a sampling frame of open-access university syllabi only to create a sample that is balanced in all variables. The sample of syllabi also reflected courses from four-year institutions, at a minimum. Future studies could examine differences that may exist between four-year colleges and universities and two-year community colleges as well. As mentioned in the introduction, the syllabi from this paper also came solely from American universities, and we note that grading practices differ internationally (see, e.g. Brookhart et al., 2016). In an increasingly global and internationally interdependent world, further studies could examine between different institutions across different countries.

An exploratory study such as this also gives rise to a multitude of further studies and lines of research. Many of these could be experimental in nature, with different combinations of product and process criteria being experimentally manipulated in course construction, and subsequently testing student achievement, student affect, and the distribution of grades as several plausible outcome variables. Experimental studies could also be conducted to develop and test different types of process criteria in domains, which rely less on process criteria, such as psychology and mathematics. These proposed studies could follow the procedure of Smith and Smith (2009), in which students were given a syllabus containing various criteria, based on condition, and asked to respond to cognitive and affective items to measure their reaction to the course assignments.

Practical implications and conclusion

This study examined the variation that occurs within college and university grading practices. It also raises a discussion of best practices. Considering what we know, should instructors be using product criteria solely in evaluating their students? We see from past evidence that product criteria, such as written and oral exams, can lead to negative emotions in students. We also know that product criteria align with summative assessment, a practice that educators are increasingly moving away from on the landscape of education reform. On the other hand, if process criteria are used, should an A in a calculus course be a reflection of punctuality in addition to calculus ability?

Of particular interest is the greatest use of process criteria in the field of education. Process criteria, including the use of formative assessment, has been shown to be advantageous in evaluating student outcomes and in establishing assessment for learning (Black & Wiliam, 1998; Shute, 2008). It is interesting, however, to see that these process criteria, which are often considered in education to be solely for formative purposes, are included as components of

summative assessment (grading). Although the use of process criteria can be seen as advantageous on many levels, discrepancies still remain as to whether these components should be evaluated as separate grades, or as part of an overall amalgamated grade.

Overall, the variation within college and university grading policies demonstrates that instructors are evaluating their students on both product and process criteria, a deviation from arguments that grades should be based solely on achievement (Cross & Frary, 1999). Components such as punctuality, attendance, turning in assignments on time, and working interactively in group settings are considered by many instructors as important aspects of students' grades, framed as process criteria. We clearly see that instructors value skills other than those reflecting academic achievement exclusively.

From this study, we recognize the vast amount of differentiation that exists in grading policies at colleges and universities across America. Resolving challenges to the validity of college and university course grades and resolving disputes about grade inflation and other related grading issues will require a clear understanding of the criteria instructors use in assigning grades. The results of this investigation show significant differences in grading criteria and grade framing exist among different academic disciplines and different institutions. Recognizing these differences, understanding their antecedents, working to resolve them, and contextualizing them within the scope of today's educational landscape will be a crucial step in meaningful grading reform in college and university classrooms.

An important step to bringing clearer meaning to grades and greater consistency among instructors their grading policies and practices would be to reach consensus on the purpose of grades. Faculty members should come together to debate but then decide what a course grade represents. This does not imply that college or university governing boards should dictate to instructors how they should grade. That would be a challengeable infringement on academic freedom. But as education professionals, we should be able to agree on what a course grade represents and what purpose it serves. With the purpose clear, questions about the appropriateness of specific grading policies and practices will be much easier to address.

Suppose, for example, that after discussion and debate, a college or university faculty, or even the faculty of a single academic department, reaches consensus on the purpose of a course grade as, 'to accurately and meaningfully describe students' attainment of course learning goals.' Although simple and direct, such a purpose statement has direct implications for how course grades can be determined. Specifically, it implies that course grades will be based on product criteria only, and that other evidence related to process criteria (e.g. class attendance) will not be used in determining students' course grades. Instructors would certainly want to stress the importance of regular class attendance and the correlation between regular class attendance and course grades. But because evidence on attendance is not a measure of 'students' attainment of course learning goals,' it would not be considered in determining course grades.

Consistency of purpose would do much to increase both the reliability and the validity of college and university course grades. Although instructors might still vary widely in the types of evidence on student learning they gather and the procedures they use in combining that evidence

in determining students' course grades, the meaning of the grade would be clearer to all stakeholders.

In sum, the goal of this exploratory study was to describe the criteria that university instructors in the US use to assign course grades. We discussed various assessment types that are used in calculating grades, and explored differences across four-year colleges and universities, as well as among five distinct academic domains. It is our hope that this study will inspire educators and policy makers to open a conversation about the nature and meaning of grades and to take action towards systematizing grading practices, both in the US and beyond.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. The data upon which the findings of this study are based are available on request from the corresponding author.

References

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House , 78*(5), 218–223. https://doi.org/10.3200/TCHS.78.5.218-223 [Taylor & Francis Online], [Google Scholar]

Atkinson, A. C. (1985). *Plots, transformations, and regression: An introduction to graphical methods of diagnostic regression analysis*. Oxford University Press. [Google Scholar]

Bailey, J., & McTighe, J. (1996). Reporting achievement at the secondary level: What and how? In T. R. Guskey (Ed.), *Communicating student learning: ASCD yearbook 1996* (pp. 119–140). ASCD. [Google Scholar]

Bers, T. H., Davis, B. D., & Taylor, B. (2000). The use of syllabi in assessments: Unobtrusive indicators and tools for faculty development. *Assessment Update*, *12*(3), 4–7. [Google Scholar]

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7–74. [Taylor & Francis Online], [Google Scholar]

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, *1*(2), 1–12. (ERIC Document Reproduction No. ED053419). [Google Scholar]

Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47–63). Holt, Rinehart & Winston. [Google Scholar]

Boothroyd, R. A., & McMorris, R. F. (1992). *What do teachers know about testing and how did they find out?* [Paper presentation]. Annual meeting of the National Council on Measurement in Education, San Francisco, CA. [Google Scholar]

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, *10*(1), 35–36. https://doi.org/10.1111/j.1745-3992.1991.tb00182.x [Crossref], [Google Scholar]

Brookhart, S. M. (2009). *Grading*. Pearson Education. [Google Scholar]

Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership*, *69*(3), 10–14. [Web of Science ®], [Google Scholar]

Brookhart, S. M. , Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, *86*(4), 803–848. https://doi.org/10.3102/0034654316672069 [Crossref], [Web of Science ®], [Google Scholar]

Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Pearson Education. [Google Scholar]

Cashwell, C. S., & Young, J. S. (2004). Spirituality in counselor training: A content analysis of syllabi from introductory spirituality courses. *Counseling and Values*, *48*(2), 96–109. https://doi.org/10.1002/j.2161-007X.2004.tb00237.x [Crossref], [Google Scholar]

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety an academic performance. *Contemporary Educational Psychology*, *27*(2), 270–295. https://doi.org/10.1006/ceps.2001.1094 [Crossref], [Web of Science ®], [Google Scholar]

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment* , 3(2), 159–179. https://doi.org/10.1207/s15326977ea0302_3 [Taylor & Francis Online], [Google Scholar]

Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, *12*(1), 53–72. https://doi.org/10.1207/s15324818ame1201_4 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development* . Psychology Press. [Google Scholar]

Elwood, J., & Murphy, P. (2015). Assessment systems as cultural scripts: A sociocultural theoretical lens on assessment practice and products. *Assessment in Education: Principles, Policy, & Practice*, *22*(2), 182–192. https://doi.org/10.1080/0969594X.2015.1021568 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Fagley, N. S. (1987). Positional response bias in multiple-choice tests of learning: Its relation to testwiseness and guessing strategy. *Journal of Educational Psychology*, *79*(1), 95–97. https://doi.org/10.1037/0022-0663.79.1.95 [Crossref], [Web of Science ®], [Google Scholar]

Gullikson, A. R. (1985). Student evaluation techniques and their relationship to grade and curriculum. *Journal of Educational Research*, *79*(2), 96–100. https://doi.org/10.1080/00220671.1985.10885657 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Guskey, T. R. (1996). Reporting on student learning: Lessons from the past – Prescriptions for the future. In T. R. Guskey (Ed.), Communicating Student Learning. 1996 Yearbook of the Association for Supervision and Curriculum Development (pp. 13-24). Alexandria, VA: Association for Supervision and Curriculum Development. [Google Scholar]

Guskey, T. R. (2001). Helping standards make the grade. *Educational Leadership*, *59*(1), 20-27. [Google Scholar]

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and teaching*, *8*(3),381–391. doi:10.1080/135406002100000512 [Taylor & Francis Online], [Google Scholar]

Guskey, T. R. (2006). Making high school grades meaningful. *Phi Delta Kappan*, *87*(9), 670–675. [Crossref], [Web of Science ®], [Google Scholar]

Guskey, T. R. (2011). Five obstacles to grading reform. *Educational Leadership*, *69*(3),16–21. [Web of Science ®], [Google Scholar]

Guskey, T. R , & Link, L. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, *26*(3), 303–320. https://doi.org/10.1080/0969594X.2018.1555515 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Hu, S. (Ed.). (2005). Beyond grade inflation: Grading problems in higher education. *ASHE Higher Education Report*, *30*(6), 1–99. https://doi.org/10.1002/aehe.3006 [Google Scholar]

Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy, & Practice*, *22*(3), 303–323. https://doi.org/10.1080/0969594X.2014.988121 [Taylor & Francis Online], [Google Scholar]

Kovas, M. A. (1993). Make your grading motivating: Keys to performance-based evaluation. *Quill and Scroll*, *68*(1), 10–11. [Google Scholar]

Lawson, J., Rasul, M. G., Howard, P., Martin, F., Hadgraft, R., & Jarman, R. (2015). Getting it right: The case for supervisors assessing process in capstone projects. *International Journal of Engineering Education*, *31*(6), 1810–1818. [Web of Science ®], [Google Scholar]

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, *15*(4),319–339. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20–32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x [Crossref], [Google Scholar]

Milton, O., Pollio, H. R., & Eison, J. A. (1986). *Making sense of college grades: Why the grading system does not work and what can be done about it*. Jossey-Bass. [Google Scholar]

O'Connor, K. (2009). *The last frontier: Tackling the grading dilemma*. Solution Tree Press. [Google Scholar]

O'Connor, K. (2010). Grades: When, why, what impact, and how? *Education Canada*, *50*(2), 38–41. [Google Scholar]

Office of Educational Technology; US Department of Education. (2016). *Future ready learning: Reimagining the role of technology in education*. 2016 National Educational Technology Plan. [Google Scholar]

Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: Examination tradition with diversity by faculty. *Assessment & Evaluation in Higher Education*, *44*(3), 379–397. https://doi.org/10.1080/02602938.2018.1512553 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, *26*(7), 1372–1380. https://doi.org/10.1016/j.tate.2010.03.008 [Crossref], [Web of Science ®], [Google Scholar]

Rathbun, G. A., Leatherman, J., & Jensen, R. (2017). Evaluating the impact of an academic teacher development program: Practical realities of an evidence-based study. *Assessment & Evaluation in Higher Education*, *42*(4), 548–563. https://doi.org/10.1080/02602938.2016.1161004 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Rojstaczer, S., & Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading, 1940–2009. *Teachers College Record*, *114*(7), 1–23. [Web of Science ®], [Google Scholar]

Royal, K. D., & Guskey, T. R. (2015). A case for differentiated grades. *Medical Science Education*, *25*(3), 323–325. https://doi.org/10.1007/s40670-015-0127-5 [Crossref], [Google Scholar]

Sadler, D. R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education*, *35*(6), 727–743.

https://doi.org/10.1080/02602930902977756 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Sarnacki, R. (1979). An examination of test-wiseness in the cognitive test domain. *Review of Educational Research*, *49*(2), 252–279. https://doi.org/10.3102/00346543049002252 [Crossref], [Web of Science ®], [Google Scholar]

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.), *Perspectives on curriculum evaluation* (pp. 1967). Chicago, Rand McNally and Co. [Google Scholar]

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety, Stress, and Coping*, *4*(1), 27–41. [Google Scholar]

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795 [Crossref], [Web of Science ®], [Google Scholar]

Smith, J. K. (1982). Converging on correct answers: A peculiarity of multiple-choice items. *Journal of Educational Measurement*, *19*(3), 211–220. https://doi.org/10.1111/j.1745-3984.1982.tb00129.x [Crossref], [Web of Science ®], [Google Scholar]

Smith, J. K., & Smith, L. F. (2019). Grading in higher education. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading* (pp. 195–213). Association for Supervision and Curriculum Development. [Google Scholar]

Smith, J. K., & Smith, L. F. (2009). The impact of framing effect on student preferences for university grading systems. *Studies in Educational Evaluation*, 35(4), 160–167. https://doi.org/10.1016/j.stueduc.2009.11.001 [Crossref], [Google Scholar]

Smith, J. K., Smith, L. F., & DeLisi, R. (2001). *Natural classroom assessment: Designing seamless instruction & assessment*. Corwin Press. [Google Scholar]

Stanny, C., Gonzalez, M., & McGowan, B. (2015). Assessing the culture of teaching and learning through a syllabus review. *Assessment & Evaluation in Higher Education*, *40*(7), 898–913. https://doi.org/10.1080/02602938.2014.956684 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Stiggins, R. J. (2008). Report cards: Assessments for learning. In *Student-involved assessment for learning* (5th ed., pp. 267–310). Prentice. [Google Scholar]

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466–478. https://doi.org/10.1111/j.1467-8527.2005.00307.x [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Tippin, G. K., Lafrenier, K. D., & Page, S. (2012). Student perception of academic grading: Personality, academic orientation, and effort. *Active Learning in Higher Education*, *13*(1),

51–61. https://doi.org/10.1177/1469787411429187 [Crossref], [Web of Science ®], [Google Scholar]

Turner, K., Roberts, L., Heal, C., & Wright, L. (2013). Oral presentation as a form of summative assessment in a master's level PGCE module: The student perspective. *Assessment & Evaluation in Higher Education*, *38*(6), 662–673. https://doi.org/10.1080/02602938.2012.680016 [Taylor & Francis Online], [Web of Science ®], [Google Scholar]

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683 [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Usherwood, T., Challis, M., Joesbury, H., & Hannay, D. (1995). Competence-based summative assessment of a student-directed course: Involvement of key stakeholders. *Medical Education*, *29*(2), 144–149. https://doi.org/10.1111/j.1365-2923.1995.tb02818.x [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Webb, C., Endacott, R., Gray, M. A., Jasper, M. A., McMullan, M., & Scholes, J. (2003). Evaluating portfolio assessment systems: What are the appropriate criteria? *Nurse Education Today*, *23*(8), 600–609. https://doi.org/10.1016/S0260-6917(03)00098-4 [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Wiggins, G. (1996). Honesty and fairness: Toward better grading and reporting. In T. R. Guskey (Ed.), *Communicating student learning: 1996 yearbook of the association for supervision and curriculum development* (pp. 141–176). Association for Supervision and Curriculum Development. [Google Scholar]

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts and findings. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 165). Elsevier. [Crossref], [Google Scholar]

Author information

Anastasiya A. Lipnevich, Ph.D., is Professor of Educational Psychology at Queens College and the Graduate Center, City University of New York. Her research interests include instructional feedback, formative assessment, attitudes towards mathematics, alternative ways of cognitive and non-cognitive assessment, and the role of psychosocial characteristics in individuals' academic and life achievement.

Thomas R. Guskey, Ph.D., is Professor Emeritus in the College of Education, University of Kentucky. His research focuses on professional learning, education reform, teacher change, assessment and grading. His most recent books include: *Get Set, Go! Creating Successful Grading and Reporting Systems* (2020, Solution Tree) and *What We Know About Grading: What Works, What Doesn't, and What's Next?* (with S. Brookhart, 2019, Association for Supervision and Curriculum Development).

Dana M. Murano, Ph.D., is a Research Scientist in ACT's Learning Division. She completed her PhD in Educational Psychology with a specialization in Learning, Development, and Instruction at the City University of New York. Her research focuses primarily on the development and assessment of social and emotional skills in students.

Jeffrey K. Smith, Ph.D., is Professor and Dean of the College of Education at the University of Otago in New Zealand. From 1976 to 2005, he was a Professor at Rutgers University. He has also served as Head of the Office of Research and Evaluation at The Metropolitan Museum of Art. He received his A.B. degree from Princeton University and his Ph.D. from the University of Chicago.