7-27-2020

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

Anastasiya A. Lipnevich
Queens College

Thomas R. Guskey
University of Kentucky, GUSKEY@UKY.EDU

Dana M. Murano
ACT, Inc.

Jeffrey K. Smith
University of Otago, New Zealand

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

Digital Object Identifier (DOI)
https://doi.org/10.1080/0969594X.2020.1799190

Notes/Citation Information
Published in *Assessment in Education*, v. 27, issue 5.

What Does an "A" Mean?

Variation in Grading Criteria in College and University Courses

Anastasiya A. Lipnevich

Queens College and the Graduate Center, City University of New York

Thomas R. Guskey

University of Kentucky

Dana M. Murano

The Graduate Center, City University of New York

Jeffrey K. Smith

University of Otago

Correspondence concerning this article should be addressed to Anastasiya A. Lipnevich, The Graduate Center, City University of New York, Department of Educational Psychology, 3rd Floor, 365 5th Ave, New York, NY 10016 email: a.lipnevich@gmail.com

Lipnevich, A. A., Guskey, T. R., Murano, D. M., Kim, T. H., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice*, 27(5), 480-500.

Abstract

This study examined differences in the criteria used by college and university instructors 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, Math, Science, Psychology, and English) were examined to determine the extent to which they employ different criteria in assigning grades in introductory level courses. Sources of variation in grade assignment included the use of product and process criteria as outlined by Guskey (2006), the prevalence of using exam performance to determine course grades, and the framing criteria for grades (Smith & Smith, 2009). Differences between institution types and among academic disciplines were also investigated. Results revealed significant differences among the five academic disciplines in grading criteria and exam prevalence. 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*

What Does an "A" Mean?

Variation in Grading Criteria in College and University Courses

Assigning fair, accurate, and meaningful grades in college and university courses presents a challenge for all professors and 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 grading 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., Brookhart, 2011; Cross & Frary, 1999; Guskey, 2001; Lipnevich & Smith, 2009; Smith & Smith, 2019).

In attempting to seek a balance among these alternatives, many instructors simply reflect on 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, professors and instructors generally state that they base grades on how well students achieve the specified learning goals for a course. Most instructors, as well as most students, consider this to be a fair and equitable way to determine course grades (Dweck, 2000; Kovas, 1993). But not all of the evidence that instructors consider in assigning course grades reflect specified learning goals. That is, in assigning course grades, they typically aggregate multiple sources of evidence. These sources may include scores from major exams and compositions, projects or reports, exhibits of student work, laboratory

assignments, class attendance or participation, punctuality in turning in assignments, and perceived effort (Hu, 2005). Studies also show that instructors vary greatly in the procedures they use to combine or summarize this evidence in assigning course grades (Cizek, Fitzgerald, & Rachor, 1996). The criteria professors typically use 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. And progress criteria demonstrate how much knowledge and skill students gained throughout the course of their learning endeavors.

In addition to factors that contribute to the overall grade, faculty members 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). Finally, the open point system allocates a maximum 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.

Variations are possible in any of these approaches. In some cases, conditions are set around how graded assignments contribute to the final grade. For example, students might be required to achieve a minimum passing score on the midterm and final exam to pass a course. In addition, even when the included assignments add up to the same letter grades, students'

reactions and expectations often differ depending on the grade framing system used. This further contributes to the confusion and non-uniformity of grade assignments by college and university instructors.

The current study explores variation in the elements of student performance and framing methods professors and instructors in colleges and universities use to determine students' course grades. Additionally, we analyzed the prevalence of exams, a form of product criteria, in order to determine whether there were differences between academic disciplines and institution types. In doing so, we attempted to unpack some of components in what Cross and Frary (1999) amusingly refer to as "hodgepodge" grades, or grades that stem 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. Understanding what elements professors and instructors consider in determining students' grades, and which overall approach they take to grading, should reveal the deeper meaning of grades that lies beneath the seemingly superficial nature of single letter grades.

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 on purpose makes decisions about what evidence to use in determining students' grades difficult to make (Brookhart & Nitko, 2008). Brookhart et al. (2016) reported that grades typically comprised a range of non-cognitive factors in addition to the standard set of cognitive indicators.

Smith, Smith, and DeLisi (2001) questioned this 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 alarming 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). With this saturation of high grades in American institutions of higher education, grades continue to lose meaning and identifying their purpose becomes more elusive.

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. Lipnevich and Smith (2009) found, for example, that when receiving feedback on writing assignments, students who received descriptive feedback (i.e. comments on their work suggesting specific steps toward 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 as effective in helping students improve the quality of their work. Furthermore, students who received letter grades in primary school years through 6th grade achieved fewer and completed fewer levels of higher education, compared to students who received descriptive evaluation throughout their primary school years (Klapp, 2015). Studies with students at all levels of schooling generally agree that letter grades

alone may not be most beneficial to student growth, especially when unaccompanied by guidance from a teacher.

In addition to the apparent 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 professors and 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 the weighting strategies, they use to combine these elements and employ computerized grading programs to ensure accuracy in their computations, the final grade can be a confusing amalgamation that is impossible to interpret and rarely presents a true picture of students' achievement or proficiency (Guskey, 2002; Sadler, 2010).

Amalgamated and Differentiated Grades

Amalgamated grades - the standard practice in most US colleges and universities - reflects 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 (Brookhart, 1993; Guskey, 1996). 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 (Royal & Guskey, 2015; O'Connor, 2009a). A grade of *A*, for example, may mean the student knew what was intended before instruction in the course began (product); or, did not learn as well as expected but displayed exceptional effort (process); or, simply made significant improvement (progress). Amalgamated grades combining various performance

aspects into a single letter can become difficult to interpret and can result in grades losing meaning.

An alternative approach is differentiated grading, where a student is graded separately on different aspects of course performance. With this approach students receive individual, independent grades for different aspects of evaluation (i.e. performance, behavioral, work ethic; etc. Guskey, 2011; Stiggins, 2008). A practice such as this can be especially salient in fields such as medicine and nursing, where performance is crucial to student success independent of mastery of course content (Royal & Guskey, 2015; Webb, Endacott, Gray, Jasper, McMullan, & Scholes, 2003). Although this is not a common practice in the United States, many schools in Canada have already adopted this system (O'Connor, 2010). Ultimately, differentiated grades that assign independent grades to separate measures of student achievement, as opposed to grades that reflect combined multiple aspects of student performance, can give a more meaningful and accurate account of student performance in various areas (Guskey, 2015).

Differentiated Grades: Product criteria

Product criteria relate to *what* students know and are able to do at a specific point in time (Guskey 2006). Professors and 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 have a history of being the most predominantly used vehicle of assessment in post-secondary institutions (Milton, Pollio, & Eiuson, 1986).

Grades based on product criteria align with summative evaluation of student achievement (Guskey, 2011; O'Connor, 2009a). Summative assessment, as initially differentiated from

formative assessment in education by Bloom (1968), based on distinctions in program evaluation originated by Scriven (1967), can be seen as a judgment made on accumulated evidence at a given point in time (Taras, 2005). Summative assessments are intended to objectively measure what a student has learned and can do independently of their teachers and other learners at the point of administration (Elwood & Murphy, 2015). Summative assessment can be understood as restrictive forms of product criteria, often in the form of exams, that do not necessarily create a reciprocal relationship between assessment and learning (Black & William, 1998).

Exams are the most prevalent examples of product criteria in college and university courses. Some scholars advocate for objective measurement and grading of student skills on achievement-based exams (e.g., O'Connor, 2010a), but there are many implications of exams that must be taken into consideration. Test anxiety 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). On the other side of this coin is 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 grades do exist, and are oftentimes preferred by students. Eighty-two percent of students who made an oral presentation as a summative assessment in a teacher education program reported that they preferred this type of summative assessment in comparison to a traditional written exam or essay (Turner, Roberts, Heal, & Wright, 2013). Qualitative data from the study revealed, however, that some students also experienced a wide range of stress and negative

emotions pertaining to the oral presentation. Thus, summative assessments, while useful and popular at the tertiary level, have problems associated with anxiety and testwiseness.

Differentiated Grades: Process criteria

Process criteria are used by professors and 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 quizzes 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 (Usherwood, Challis, Joesbury, & Hannay, 1995), are essential in evaluating potential candidates in these fields (Webb et al., 2003; Guskey, 2015). Supervisors assessing capstone projects of engineering students in undergraduate programs have also advocated for the evaluation of process – in addition to product criteria (Lawson, Rasul, Howard, Martin, & Hadgraft, 2015).

In general, process criteria can be separated into three broad categories (Guskey, 2019). The first one is learning enablers, which include formative assessments, homework, class participation and 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.

Differentiated Grades: 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." Some educators draw distinctions between "progress," which they measure backward from a final performance standard or goal, and "growth," which is measured forward from the place a student begins on a learning continuum (Wiggins, 1996). When achievement is judged using well-defined learning standards that include graduated levels of performance, however, progress and growth criteria can be considered synonymous. Progress criteria are highly individualized among students (Guskey, 2001), and can therefore be difficult to measure and track throughout a semester in calculating a grade.

Grade Assignment

Although most professors at the undergraduate level adopt an amalgamated grading system, as opposed to assigning differentiated grades to students for each aspect of performance, these amalgamated grades often comprise product, process, and progress criteria (Guskey, 2006). When grades are amalgamated, we do not know precisely what the grade means, or what constituents underlie its formation. Some professors also vary their grading criteria from student to student, taking into account individual circumstances. Although professors and instructors defend this practice on the basis of fairness (Tippin, Lafrenier, & Page, 2012), it significantly confounds the meaning of any grade. Given these circumstances surrounding grade assignment,

the interpretation of grades becomes questionable, especially in attempts to compare one grade to another that come from different institution types and academic subjects.

In addition to 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, professors and 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 (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 overall 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 set of

questions about how they would react to various assignments in the course. All 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 questionnaire items, 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's work (1981), 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's (2009) work therefore suggests that although various grade framing approaches are identical mathematically, they may not be perceived identically in the eyes of students. We were therefore interested in grade framing as an additional factor of variation in grading.

Current Study

With variation of grading criteria across different universities and colleges, it becomes difficult to objectively interpret the grades students receive. An "A" in one course may be very different from an "A" in another course, with a plethora 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 grade 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 as Units of Analysis

A syllabus is an official document in which course instructors present descriptions of the course content, along with expectations, responsibilities, assignments, and criteria for evaluation (Stanny, Gonzalez, and McGowan, 2015). Syllabus is viewed as a contract between an instructor and a student, and it is expected that all syllabi would define assessment approaches and include information about types of assessment (e.g. examinations, essays), weights of individual assessments in the final course grade, etc. Syllabus analysis offers a window into the instructional and evaluative practices that instructors employ, and has been used in prior studies to explore a range of research questions ranging from the alignment of syllabi with learning outcomes to the evaluation approaches in Spain, among others (Bers, Davis, & Taylor 2000; Panadero, Fraile, Fernández Ruiz, Castilla-Estévez & Ruiz, 2018; Cashwell & Young 2004; Rathbun, Leatherman, and Jensen, 2017). In the US 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 professors and 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. 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 selected. If universities 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 of the selected syllabi came from large, comprehensive universities and 91 were from smaller, four-year colleges. These proportions approximate the numbers of students attending such institutions overall. All institutions were in the United States. Institutions awarding Ph.D. 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 tests, in-class assignments, peer evaluations, classroom observations, and email correspondence.

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

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 by looking into some of the nuanced differences between uses of product and process criteria. None of the sample courses showed evidence related to progress criteria considered in determining students' course grades. In some instances, process criteria related to attendance or class participation were considered a direct portion of the 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 percent 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. Professors and instructors in English classes specified deductions most frequently, with 84% of the English syllabi reviewed reflecting deductions based on process criteria. Science, math, psychology, and Education instructors incorporated deductions into their grading criteria less often, with 32%, 22%, 36%, 40% of course syllabi containing explicit mention of deductions. Professors at small colleges specified deductions more than professors at larger universities (58% versus 34%). In all cases across institution types and academic departments, grade deductions were based on process criteria.

Figure 1 depicts percentages of product and process criteria in assigning course grades. Instructors in education departments use more process criteria than professors in the other four domains. However, all academic domains used product criteria more frequently than process criteria in determining students' course grades.

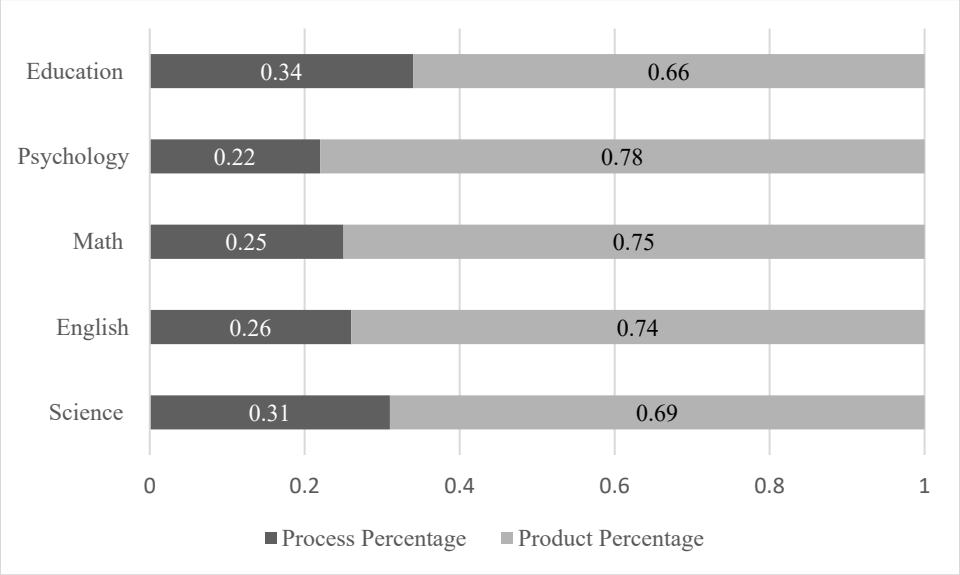


Figure 1. Product and Process Grading Criteria by Academic Subject

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 ($t = .84$, $p = .40$). We see that there is much greater variation among academic subjects in grading criteria than there is between colleges and universities. When aggregated by institution type, both types of institutions use predominantly product criteria in determining grades.

A one-way ANOVA done with raw percentage data revealed a significant difference among academic subjects and use of process criteria ($F_{(4, 249)} = 3.80$, $p = .005$). 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 ($F_{(4, 249)} = 6.24$, $p < .001$). Tukey post-hoc comparisons of the transformed data revealed significant differences between English and psychology ($p = .004$), with English using significantly more process criteria than psychology, and education and Psychology ($p < .001$), showing that both English and education classes used significantly more process criteria than psychology classes. Psychology differed significantly from science and math as well ($p = .001$, $p < .001$), indicating that psychology courses used the least process criteria when compared to other domains. Ultimately, English and education instructors included a broader range of sources of evidence and also were more likely to include process criteria (e.g. class attendance and participation), either as a portion of the grade or as a source of grade reduction.

Similar analyses were used to examine differences in the use of product criteria. Significant differences among academic domains in the percentage of product criteria were revealed ($F_{(4, 249)} = 4.04$, $p = .003$). Applying the same logit transformation to appropriately handle the distribution of percentages, the interpretations of this finding remain constant ($F_{(4, 249)} = 5.26$, $p < .001$). Psychology used significantly more product criteria than each of the remaining four domains, with each pairwise between-subject comparison reaching statistical significance. Education differed significantly from math, the second most frequent user of product criteria ($p = .02$). Overall, math and psychology courses used product criteria most frequently across subject domains.

Although there were no main effects of institution type on grading criteria used ($F_{(1, 249)} = .711$, $p = .40$), 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 ($F_{(9, 240)} = 5.32$, $p < .001$), with a significant main effect of academic subject ($F_{(4, 240)} = 9.27$, $p < .001$), and a significant interaction effect of academic subject by institution type ($F_{(4, 240)} = 4.80$, $p = .001$). Comparisons of results across different types of institutions revealed several interesting patterns, including a discipline by institution interaction, depicted by Figure 2.

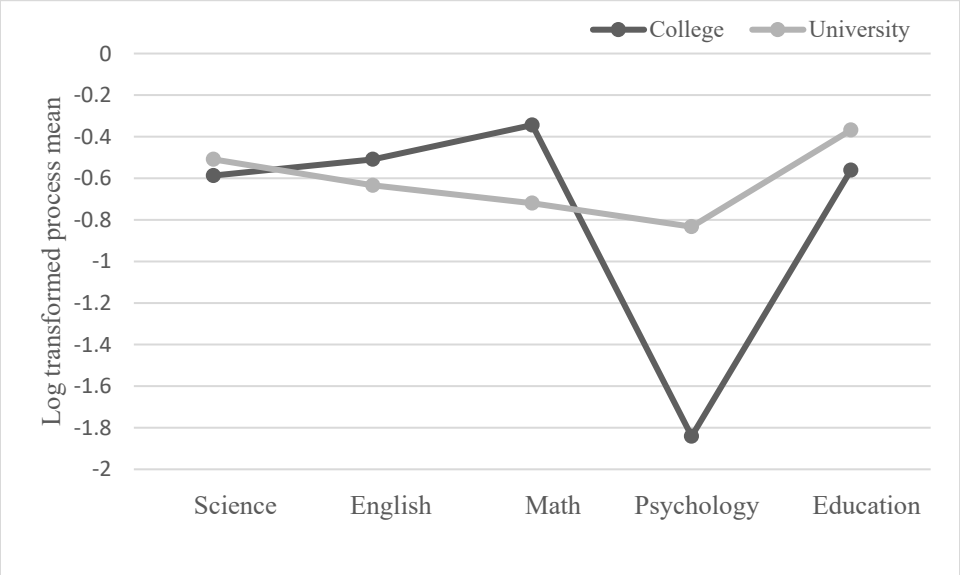


Figure 2. Means of Transformed Process Criteria by Subject and Institution Type

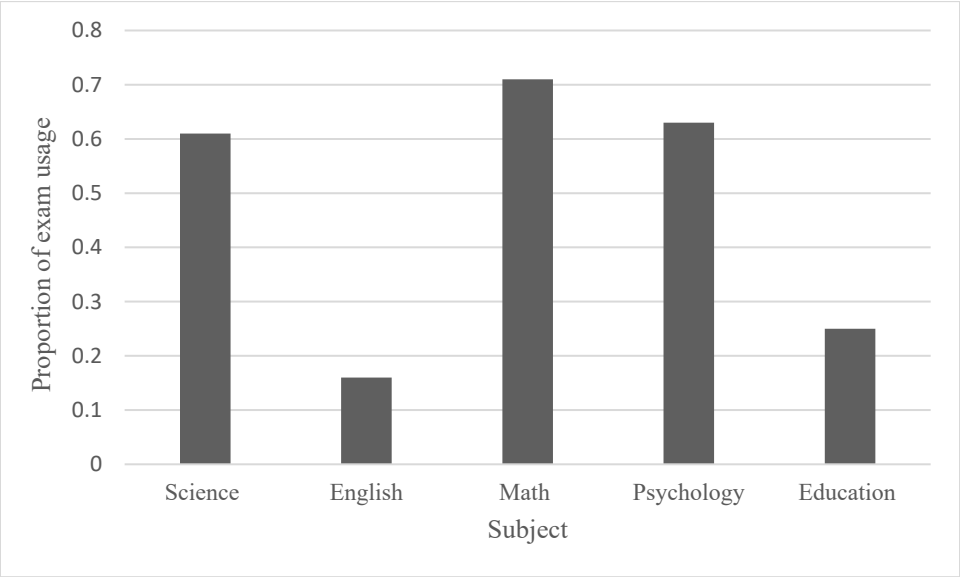
Although differences can be seen in math and education, the large difference to be seen is that professors in psychology departments in colleges are far less likely to use process criteria than those in large institutions. Of course, it can also be seen that those professors are far less likely to use process criteria than their colleagues in the other four disciplines examined. Overall, it seems that process criteria are used similarly in terms of frequency across departments and types of institutions, with the singular exception of college psychology departments.

Variation in Exam Prevalence

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 weighed on students' grades. Figure 3 shows differences among academic domains in how much of students' grades came from performance on exams. We see that Math, Science, and Psychology courses weighted exams

heavily in grade calculation, whereas English and Education courses placed less weight on exams.

Figure 3. *Percentage of Grades Constituted by Exams by Academic Subject*



The use of exams as grading criteria differed significantly both when considering raw percentage data ($F_{(4, 249)} = 9.42$, $p < .001$) and after applying a log transformation to correct for the distribution of percentage data [2]($F_{(4, 249)} = 55.00$, $p < .001$). We used the transformed data for the remainder of the analyses. Tukey post-hoc comparisons revealed significant pairwise differences ($p < .05$) between English and science, math, and psychology, confirming that English used significantly fewer exams as product criteria than the other three academic domains. Pairwise comparisons between education and science, math, and psychology were also all statistically significant ($p < .05$). These findings confirm that education and English use fewer

[2] We used the transformed data in our analyses but show figures in percentages, whenever appropriate, for ease of interpretation.

exams as product criteria in comparison to math, science, and psychology. It appears that the domains more likely to use product criteria also weigh exams heavily in the calculation of student grades.

Variation in Grade Framing

Last, 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 ($\chi 2 = 43.643$, $p < .001$). 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).

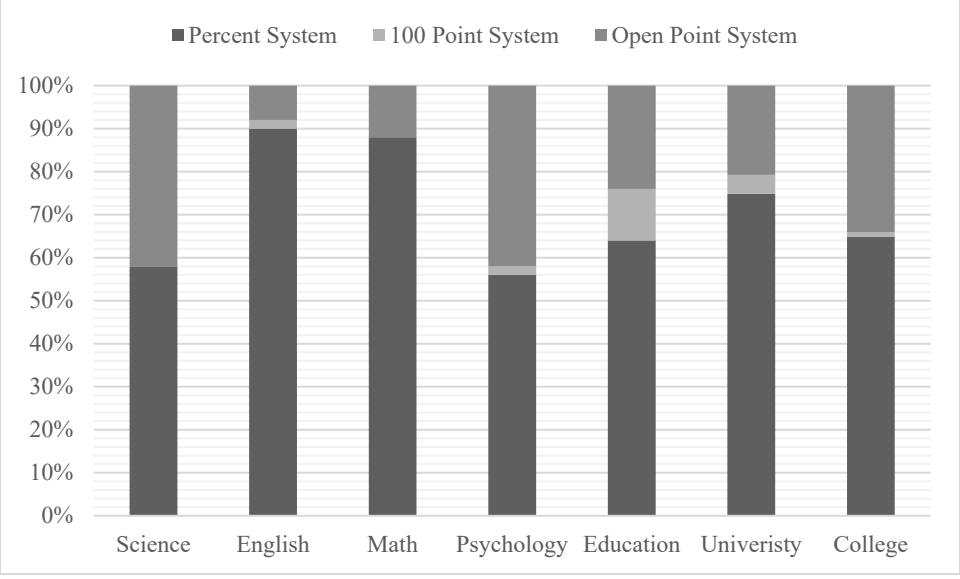


Figure 4. Grade Framing Approaches by Academic Subject and Institution Type

Discussion

The current study aimed at investigating grading practices of professors and instructors who teach introductory level courses in colleges and universities across different academic departments. Overall, and consistently with the trends in the literature, we found that product criteria were used more frequently than process criteria when assigning grades. Although most professors 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 calculating course grades, and in others grades were based more heavily on process criteria. Professors also vary in whether they apply deductions, or whether grades were calculated using solely additive methods. Lastly, professors 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 math, 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, professors 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 over process criteria present opportunities for future research. Subjects that integrate the most process criteria, such as English and education, seem most conducive to assignments that are writing-centric. Disciplines like math 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 professors and 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. A further exploration of how to assess mathematics-centric domains using more process criteria could be beneficial to professors in those fields.

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. 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, 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 factoring in where the student started.

While informative, this study is not without limitations. To begin, we considered introductory courses only. Because these courses often have large enrollments, 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. The syllabi from this paper also solely came from American universities. 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 emotion, 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 have demonstrated less modes of process criteria, such as psychology and math. 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.

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 professors 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 professors are evaluating their students on both process 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 professors as important aspects of students' grades, framed as process criteria. We clearly see that professors value skills other than those reflecting academic achievement exclusively.

Simultaneously, the landscape of K-12 education is undergoing rapid changes as the ESSA (Every Student Succeeds Act) standards are implemented. These standards will require student assessment to be broad and inclusive of additional indictors of "success", such as student personal growth and engagement, in addition to traditional achievement scores on standardized tests (US Department of Education, 2016). Certainly, the notion of using differentiated grades to reflect different aspects of student performance seems feasibly integrated into this new initiative. Whether colleges and universities follow suit, however, is to be determined. Considering the vast majority of institutions that already incorporate process and product criteria into an amalgamated grade, switching to a differentiated system of grading, where students could receive separate grades to reflect different aspects of achievement, does not at all seem to be an implausible next step. ESSA could very well be the catalyst that shifts grading standards from amalgamated to differentiated, in which product, process, and progress criteria would all be recognized as equally important, yet independent, grades in and of themselves.

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 professors and 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, 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.

References

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House, 78*(5), 218-223.

Atkinson AC. (1985). *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford, UK: Oxford University Press.

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). Alexandra, VA: ASCD.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability, 21,* 5-31.

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

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, *10*(1), 35-36.

Brookhart, S. M. (2009). *Grading*. Upper Saddle River, NJ: Pearson Education.

Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership*, *69*(3), 10-14.

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).

Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, NJ: Pearson Education.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety an academic performance. *Contemporary Educational Psychology, 27*(2), 270-295.

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.

Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, *12*(1), 53-72.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. New York, NY: Psychology Press.

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.

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.

Hattie, J., & Timperley, H. (2007). The powder of feedback. *Review of Educational Research. 77(1),* 81-112.

Guskey, T. R., (2019). *Get set, go! Implementing successful reforms in grading and reporting*. Bloomington, IL: Solution Tree.

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.

Guskey, T. R. (2001). Helping standards make the grade. *Educational Leadership, September 2010,* 20-27.

Guskey, T. R. (2002). Computerized gradebooks and the myth of objectivity. *Phi Delta Kappan*, *83*(10), 775-780.

Guskey, T. R. (2006). Making high school grades meaningful. *Phi Delta Kappan*, *87*(9), 670-675.

Guskey, T. R. (2011). Five obstacles to grading reform. *Educational Leadership, 69*(3), 16-21.

Hu, S. (Ed.) (2005). Beyond grade inflation: Grading problems in higher education. *ASHE Higher Education Report*, *30*(6), 1-99.

Jones, T . (2009). Investing in assessment: The role of formative assessment in teaching psychological research methods. *Psychological Teaching Review, 15*(2), 74-81.

Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy, & Practice, 22*(3), 303-323.

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

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.

Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2013). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science,* 1-21.

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied, 15(4),* 319-333.

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.* San Francisco: Jossey-Bass.

O'Connor, K. (2010). Grades: When, why, what impact, and how? *Education Canada, Spring 2010,* 38-41.

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

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.

Royal, K. D., & Guskey, T. R. (2015). A case for differentiated grades. *Medical Science Education, 25,* 323-325.

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

Sarnacki, R. (1979). An examination of test-wiseness in the cognitive test domain: *Review of Educational Research, 49*(2), 252-279.

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

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Smith, J. K. (1982). Converging on correct answers: a peculiarity of multiple choice items. *Journal of Educational Measurement, 19*(3), 211-220.

Smith, J. K., & Smith, L. F. (2009). The impact of framing effect on student preferences for university grading systems. *Studies in Educational Evaluation, 35,* 160-167.

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). Alexandria, VA: Association for Supervision and Curriculum Development.

Stiggins, R. J. (2008). Report cards: Assessments for learning. In *Student-involved assessment for learning* (5th ed., pp. 267-210). Upper Saddle River, NJ: Prentice.

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.

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.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453-458.

United States Department of Education (2016). *Transitioning to the Every Student Succeeds Act (ESSA): Frequently asked questions.*

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,* 144-149.

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,* 600-609.

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). Alexandria, VA: Association for Supervision and Curriculum Development.

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

Appendix A: Coding Criteria

We are coding the tasks into two main categories
1. Product criteria (relate to *what* students know and are able to do at a specific point in time)
2. Process criteria (reflect students' effort or work habits when assigning grades).

Here are examples of tasks that belong in each of these categories:

Product	**Process**
Final/Midterm exam	Homework
Paper	Attendance
Presentation	Reflections

Portfolio	Journal
Essay	Discussion board
Article review	Participation
Book review	Quizzes
Lesson plan	Home tests
Performance	In-class assignments
Summary of literature	Email correspondence

So, here is an example:

	Product	Process	N/A	Total	Subject	College	Deductions
Molecules and Cells	300 (3 exams)	200 (attendance, homework, blackboard quizzes)		500	1 (Science)	1 (university)	0 (=none)
College Writing	50 (portfolio final)	50 (attendance, portfolio process)		100	2 (English)	1 (university)	1 (=yes)
					3 (Math)		
					4 (Psychology)		
					5 (Education)		

You derive this by adding up separate tasks. So, with Molecules and Cells (first syllabus for universities), you add 3 exams, worth 100 points each. Exams are *Product*.
Then, you add 30 in quizzes, 100 in homework, and 70 in attendance, and you get 200 for *Process*. For deductions, just see if there is anything (code as 1 if Yes, and 0 if none).

Here are some odd ones that I found:
Lab, Fieldtrips, and Service Learning Project will be coded as **Process**
Peer evaluation as **Process**
Field observation as **Process**
Personal ownership of a book as **N/A**