A B or not a B? A Proposed Framework for Discussing Grade Aggregation in Standards-Based Assessment

Dr. James Blake Hylton, Ohio Northern University

Dr. Hylton is an Assistant Professor of Mechanical Engineering at Ohio Northern University. He previously completed his graduate studies in Mechanical Engineering at Purdue University, where he conducted research in both the School of Mechanical Engineering and the School of Engineering Education. Prior to Purdue, he completed his undergraduate work at the University of Tulsa, also in Mechanical Engineering. He currently teaches first-year engineering courses as well as various courses in Mechanical Engineering, primarily in the mechanics area. His pedagogical research areas include standards-based assessment and curriculum design, the later currently focused on incorporating entrepreneurial thinking into the engineering curriculum.

Mr. Matthew Walker, Ohio Northern University

Matthew Walker is a sophomore Computer Engineering student at Ohio Northern University. He is also minoring in Applied Mathematics. He is the President of the student chapter of the American Society for Engineering Education and has a passion for teaching. His previous classroom experiences include a year of being in the education college and working in the field with high school math classes that range from Algebra II to Calculus BC

A B or not a B?
A Proposed Framework for Discussing Grade Aggregation in Standards-Based Assessment

Introduction

While grading and assessment has long been a discussion point among educational researchers, there has been a recent resurgence in concern over the efficacy of traditional point-based grading systems. Many of these discussions center around the meaning and value of the proverbial 'point'. In a traditional grading scheme, in which students complete a series of assignments, each given a point value and weighting, what is the conceptual value of a point? On what grounds does the accumulation of some threshold number of points constitute mastery of the topic at hand? Is such a numerical marker valuable to a learner in reflecting on their progress and accumulated knowledge? The broad answer to such questions is that points are largely arbitrary, varying wildly in meaning across institutions, courses, or even across assignments.

Trends in pedagogy have shifted strongly in the direction of more experiential, authentic learning activities such as project-based and active learning. As the nature of the classroom activity has changed, important questions have been raised about the efficacy of traditional grading schemes. Separation has been observed between course objectives and assessment practices, and the ability of the traditional grading schemes to effectively assess student competency and achievement has been called into question by Sadler [1], among others. Guskey notes five key obstacles to grading reform. He notes that grading has long been viewed as a means of differentiation between students rather than a tool by which to assess a student's competency and communicate that assessment to relevant stakeholders. [2] Recently, however, educational researchers and practitioners have begun to question the efficacy of such a perspective.

Standards-, criteria-, or objectives-based course design has emerged as a possible path forward for the grading reform efforts. The central premise of such systems is the alignment of course assessments with clearly delineated course outcomes. [3, 4] Their history is traceable as far as 1949 in Tyler's *Basic Principles of Curriculum and Instruction* [5] and are now popularized in such models as Wiggins and McTighe's *Understanding by Design* [6]. The shift in application from course design to assessment first emerged in K12 during the 1990s before gaining traction more recently in higher education. [7] The rapid popularization of the approach has sparked a parallel proliferation of terminology and methods, with Sadler [1] providing an effective summary of the topic and Ziegenfuss [8] adding a review of the underlying design philosophies. Heywood examines a related method, criterion referenced grading, and identifies that successful implementation of such schemes require a well-defined list of objectives, clear and well-developed rubrics and guidelines for assessment, open communication with students on expectations and assessment methods, and a centering of course content around the established standards and corresponding assessment methodology. [9] A well-stated and more recent review by Muñoz [10] concludes that such methods are the clear next step in educational assessment.

Applications of standards-based assessment are wide ranging in context and approach. Parker, assessing student writing across the curriculum, used a semester-spanning aggregation method

based on the number of assignments demonstrating the desired level and number of competencies. Also notable was the implementation of a standards-based approach without the use of rubrics, instead using a checklist-based assessment. [11] Siniawski, conducting a mechanics of materials course, applies a unique plus, check, check minus scoring approach, ultimately translating rubric scores into traditional numeric values for aggregation. [4] Post utilized standards-based assessment in a fluid mechanics course and corresponding lab and later in a thermodynamics course, giving students multiple opportunities to pass each objective and assessing each instance on a pure binary scale. [12, 13] Hylton examined the use of standards-based grading in a first-year engineering course, with emphasis on its applications to traditional written exams, while Marbouti used a standards-based grading system as the basis for a predictive model for first-year engineering student success. [14, 15]

The full impact of a standards-based system on student learning are the subject of ongoing discussion and research. Scriffiny provides seven reasons for standards-based grading, including arguments that it gives greater meaning to grades, provides greater control for the instructor, and aids teachers in adjusting instruction due to greater transparency into student understanding. [16] Sadler [2] notes a number of advantages of such a scheme, including the value of personalized, meaningful feedback and transparency in the grading process. Carberry [3] began to quantify the benefits, observing the impacts on cognitive and affective behaviors and noting increases in self-efficacy and sophistication of students' epistemological beliefs. Attwood [17] builds on this finding, exploring the impacts at both large public institutions and smaller private colleges and reporting significant boosts in student self-efficacy. Heywood [9] observed these motivational effects to be consistent across student performance levels. Gusky goes so far as to that, if properly constructed with sufficient detail, a standards-based assessment approach "facilitates teaching and learning better than almost any other grading method." [18]

Purpose of Work

Though standards-based grading methods have grown in popularity, little guidance has yet been published with respect best practices. Carberry et al summarize the deployment in several higher education settings, attempting to distill best practices and impacts, but are limited by a relatively shallow literature pool. [19] Suggestions about how to develop rubrics are widespread (e.g. [20]), but the literature is sparse with respect to combining standards-based scores across assignments or formulating an overarching course grade in a standards-based system.

The primary purpose of this work is to propose a methodology-based classification scheme through which to frame future discussion around standards-based grading score aggregation. A series of exemplars of the grade aggregation methods encompassed by the classification scheme are provided. The exemplars were generated by applying various schemes to a set of hypothetical student profiles for a first-year engineering course.

The secondary purpose is to use those exemplars to provide a cursory examination of how various schemes may impact students at various levels of achievement. While this secondary exploration is far from a robust exploration of grading scheme impacts, it is intended to aid users in making decisions about what scheme may be appropriate for their course and teaching style.

Classification Framework

Discussion of standards-based grading necessarily involves some examination of how individual assessment activities have ultimately been aggregated into an overall course grade. Such discussion is challenging, however, when practitioners have no common framework through which to describe their aggregation scheme within the broader literature context. To this end, four key dimensions were identified to categorize standards-based assessment methods – data type, aggregation approach, weighting scheme, and aggregation scope. These dimensions are based on the various methods presented in the literature cited above and the authors' own experiences in applying standards-based grading. The framework is depicted in Figure 1.
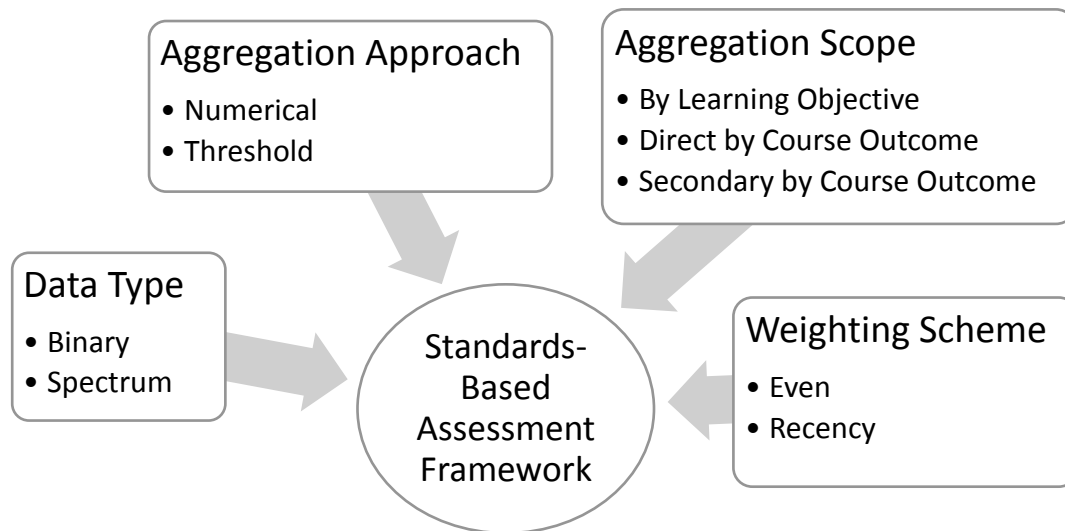

Figure 1: Visual representation of classification framework and associated dimensions

Dimension 1: Data Type
Data Type refers to the form of the underlying assessment data and how scores are generated within a given assignment. This dimension includes two variations – binary and spectrum. Binary type methods use a pass/fail assessment for each individual rubric row. For rubric rows which are not already binary (such as those included in Figure 2 below) a threshold should be identified above which a score is considered to be passing. For example, any score ≥50% of the maximum possible point value may be considered to have passed. If multiple rubric rows mapped to the same learning objective appear on a single assignment, they may be treated separately or combined using a similar threshold condition applied to the set. For example, using a 50% pass threshold, if three rubric rows on a single assignment rubric mapped to the same objective and at least two of them met the pass threshold, then the objective would be considered passed for that assignment. Spectrum type methods fully leverage the richness of a multi-level rubric by retaining the relative score of intermediate levels. For example, a student scoring as "Underachieved" on objective 5.4 in the rubric below (Figure 2) would be recorded with a score of 2.5/6 (42%), rather than a zero as would be earned under a binary type method.

Description of Presentation Topic			
5.4: Demonstrate curiosity about the world and about your profession.			
No Evidence [0 pts]	**Underachieved [2.5 pts]**	**Partially Achieved [5 pts]**	**Fully Achieved [6 pts]**
Memo missing.	Well described topic but no sources, or very poor description with sources.	Vague or confusing topic description, or missing one source.	Memo clearly explains the topic to be presented and includes 2 sources.
Engineering and Broader Impacts			
5.1: Demonstrate an awareness of engineering and non-engineering disciplines and their role within society.			
No Evidence [0 pts]	**Underachieved [1.5 pts]**	**Partially Achieved [2.5 pts]**	**Fully Achieved [3 pts]**
Memo missing.	Engineering AND societal impact missing.	Engineering OR societal impact missing.	Memo addresses the engineering AND societal impact of topic.

Figure 2: Example rubrics with mapped objectives and fully defined performance standards

Dimension 2: Aggregation Approach

Aggregation Approach describes how scores across various assignments are combined into an overall course grade. This dimension includes two varieties – numerical aggregation and threshold aggregation. In both approaches, a proficiency percentage is calculated for each learning objective or, depending on the scope classification, course outcome. This proficiency percentage is calculated by combining the constituent assignment scores for the learning objective or course outcome. For example, in a binary type method a student who passed a given objective on five of the six assignments in which that objective appeared has a proficiency of $(1+1+1+1+1+0)/6 = 83.33\%$. Using a numerical aggregation approach, the proficiency percentages of all objectives (or outcomes) are averaged in a traditional sense to obtain the course grade. For example, in a course with three course outcomes, a student receiving proficiency percentages of 55%, 83.33%, and 73% would receive an overall course grade of 70.4%. Using a threshold aggregation approach, each proficiency percentage is converted to a binary pass/fail using a pass threshold. This percentage can be adjusted per the expectations of the instructor, but 60% was used to generate the exemplars discussed later in this work. For example, the student mentioned above would receive a FAIL, PASS, PASS for the three course outcomes and an overall course grade of $(0+1+1)/3 = 67\%$.

Dimension 3: Weighting Scheme

The weighting scheme dimension captures how individual assignments and rubric scores are valued relative to one another during calculation of proficiency percentages, including even weighting and recency weighting. Using even weighting, all assignments and objectives are weighted equally (effectively unweighted). Recency weighting devalues older assignments when the same objective has been reassessed more recently. In this study, the exemplars were created using a sliding scale of 100%, 80%, 60%, 40%, and 20% applied to the most recent through fifth most recent occurrence of a given objective, respectively, with any earlier occurrences being weighted by 0%. Other weighting schemes could conceivably be applied, such as weighting objectives or outcomes by their respective frequency of appearance or perceived importance by the instructor, but only even and recency weighting are included in this discussion.

Dimension 4: Aggregation Scope

Aggregation Scope refers to the level at which the aggregation occurs. In a flat course design, in which there is only one level of course outcome or learning objective, this attribute has only one variation – averages are conducted on whatever outcomes/objectives are available. In a two-tiered course design, a set of high level course outcomes each contain a set of related learning objectives. For example, Figure 3 depicts a course with three course outcomes, each containing several related learning objectives. Each learning objective is in turn assessed some number of times over the course of the term, such as in the rubric shown in Figure 2. Under such a two-tiered system, three scoping variations are possible – aggregation by learning objective, direct aggregation by course outcome, and secondary aggregation by course outcome. Aggregation by learning objective combines all assessment scores mapped to a particular learning objective then combines those learning objective averages to obtain the overall course grade, effectively ignoring the higher-level outcome structure. Direct aggregation by course outcome combines all assessment scores mapped to a particular course outcome, ignoring the learning objective bins, then combines those course outcome scores to obtain an overall course grade. Secondary aggregation by course outcome combines assessment scores by learning objective, then combines learning objective scores by course outcome, and then combines those course outcome scores into an overall course grade. Aggregation in each case is conducted according to the approach and weighting dimensions as previously discussed. Student scores are shown in Figure 2 as calculated using each of the three different scoping varieties discussed above, using an even weighted numerical aggregation approach based on spectrum type data.

The impact of the scoping dimension ties to the effective weighting of an individual assessment activity. Consider the impact of the zero in objective 3.2 for the example in Figure 3. In the learning objective average approach, the zero is averaged equally with nine other learning objectives and the impact is minimal. In the direct average course outcome approach, that zero is combined with the four scores in objective 3.1 – effectively weighted as 20% of the CO3 score and 7% (20% of 1/3) of the overall grade. In the secondary course outcome approach, that same zero is instead 50% of the CO3 score and 16% (50% of 1/3) of the overall course grade.

Course Outcome	Learning Objective	Frequency of Appearance	Learning Objective Average	Direct Average CO Score	Secondary Average CO Score
1	1.1	4	91.67	86.04	90.60
	1.2	1	100.00		
	1.3	14	83.22		
	1.4	2	87.50		
2	2.1	5	78.33	65.33	72.61
	2.2	5	26.00		
	2.3	3	86.11		
	2.4	2	100.00		
3	3.1	4	75.00	60.00	37.50
	3.2	1	0.00		
	Course Grade		72.78	70.46	66.90

Figure 3: Calculation examples for variations on scoping dimension

Development of Exemplars

To provide a more concrete understanding of the framework, a total of nineteen different aggregation methods were selected from among the various combinations of framework attributes and applied to hypothetical student data to form a set of exemplars.

As a baseline for comparison, Method 1 represents a traditional summative assessment approach rather than a truly standards-based method. Student scores were aggregated by assignment then summed to provide an overall course grade. Their grade was then based on their earned percentage of total possible points. A traditional 90/80/70 sliding scale was used to assign a corresponding letter grade.

Methods 2-19 explore the various permutations of the key assessment dimensions as outlined above. Figure 4 below details each of the methods, as characterized by the proposed dimensions.

To provide an authentic context, the course structure, assignments, and rubrics for the exemplars are taken from a first-year engineering course at a small, private university. The course provides a broad introduction to engineering topics, including design, analysis, and communication, captured across six course outcomes. There are also more specific learning objectives, written to be deployed at the assignment level, which are each mapped back to the broader course outcomes. This two-tiered structure allows a high degree of specificity at the rubric level while keeping the bigger picture broader and uncluttered by an overabundance of objectives.

Recent efforts towards implementing a standards-based grading system have led to the development of learning-objective mapped rubrics across all course assignments. Figure 2 above provides an example of part of one such rubric. The rubrics translate each performance level to a numerical point value, which may be used as is in certain aggregation schemes and ignored or reweighted in others.

Rubric-level scores for all assignments were generated for three hypothetical exemplar students – a high performer (Student A), a mid-level performer (Student B), and a low-performer (Student C). While hypothetical, scores were based loosely on observed trends among real students enrolled in the course. Figure 4 provides the complete case study results for each of the methods applied. Complete calculation tables may be made available for those seeking a more detailed look at the underlying mathematics of each approach but are not possible to format succinctly for inclusion in this publication.

Method	Type	Approach	Weighting	Scope	Student A		Student B		Student C	
1	--	--	--	--	93.84	A	77.7	C	54.99	F
2	Binary	Numerical	Even	By learning objective	93.72	A	70.94	C	37.82	F
3	Binary	Threshold	Even	By learning objective	95.45	A	72.73	C	22.73	F
4	Binary	Numerical	Even	Direct by course outcome	95.1	A	67.47	D	35.51	F
5	Binary	Threshold	Even	Direct by course outcome	100	A	83.33	B	0	F
6	Binary	Numerical	Even	Secondary by course outcome	92.7	A	67.12	D	40.54	F
7	Binary	Threshold	Even	Secondary by course outcome	83.33	B	66.67	D	33.33	F
8	Spectrum	Numerical	Even	By learning objective	94.96	A	74.57	C	47.26	F
9	Spectrum	Threshold	Even	By learning objective	100	A	68.18	D	31.82	F
10	Spectrum	Numerical	Even	Direct by course outcome	94.96	A	69.85	C	44.06	F
11	Spectrum	Threshold	Even	Direct by course outcome	100	A	100	A	16.67	F
12	Spectrum	Numerical	Even	Secondary by course outcome	95.84	A	71.46	C	43.96	F
13	Spectrum	Threshold	Even	Secondary by course outcome	100	A	83.33	B	50	F
14	Spectrum	Numerical	Recency	By learning objective	95.1	A	74.17	C	47.78	F
15	Spectrum	Threshold	Recency	By learning objective	100	A	68.18	D	36.36	F
16	Spectrum	Numerical	Recency	Direct by course outcome	96.53	A	75.19	C	58.23	F
17	Spectrum	Threshold	Recency	Direct by course outcome	100	A	66.67	D	66.67	D
18	Spectrum	Numerical	Recency	Secondary by course outcome	95.72	A	70.64	C	44.68	F
19	Spectrum	Threshold	Recency	Secondary by course outcome	100	A	83.33	B	50	F

Figure 4: Summary of methods applied

Comparisons and Insights

A full investigation of the impacts of various dimensions is left for a series of latter works, but comparison of exemplar results can provide some high-level observations for prospective practitioners to consider as well as some best-practices for standards-based grading made apparent through this exploratory exercise. Broadly speaking, the impact of a standards-based grading scheme (as compared to a traditional grading scheme) is seen to have little impact on the overall course grades of the high performing and low performing exemplars. Though further study is needed, this suggests that those students who do very well are likely to succeed and those who do very poorly are likely to do so regardless of grading system. Only under one method (Method 7) did the high performing exemplar drop to a B, due to receiving a secondary outcome average of 59.52 on one of the six outcomes – within what the author would consider "rounding range" of the 60% threshold used. Digging deeper, this in turn was the result of failing one of three related learning objectives under that outcome, with the low score coming in an area assessed only once. This speaks to the importance of ensuring that all objectives (or outcomes, if using direct aggregation by outcome) provide multiple opportunities for students to demonstrate proficiency. This also speaks to a secondary value of considering a standards-based approach in that it reveals gaps in the curriculum that may otherwise go unnoticed.

Turning the attention to the low performing student, only in one instance (Method 17) did they raise their grade from F to D. This method applied direct aggregation by course outcome using a recency weighted threshold approach. The elevated grade is largely an artifact of the underlying course design, rather than the grading scheme, in that the course is back weighted with team project work. When rubrics are grouped by outcome and then weighted towards the end of the semester, team deliverables dominate the calculation of course grade and the student can ride on the coattails of their teammates. This provides two pieces of insight – 1) the importance of extracting individual contributions to team assignments; and 2) the need for courses being assessed by a recency weighted scheme to ensure that assessment on an individual basis is possible throughout the term. It is also worth observing that three of the four outcomes passed were scored in the 60% range, meaning that they only just made the threshold. This highlights the fact that threshold aggregation methods do not allow differentiation between an outcome score very close to the threshold and a score well beyond it. In courses which cover an abundance of learning objectives (or course outcomes, depending on the method scope), this last point is somewhat mitigated by shear number of scores included in the overall grade. In courses with a smaller number of outcomes, however, barely passing even a small number of them may result in an inflated grade profile.

Additional insights may be derived from the middle performing student. Across the nineteen methods tested, this student received one A, three B's, nine C's, and six D's. Percentage scores range from 66.7% to 100%, with an average of 74.3%. This student was designed intentionally to model one who receives generally high scores but struggles on a few key learning objectives. Middle scoring students are those who are most susceptible to minor variations in assessment approach and grading leniency, banking on partial credit to see them through the course. As

such, the impact of variations in standards-based grading method has a tremendous impact on the final grade of this exemplar.

Examining the high scoring case, Method 11, scores were directly aggregated by course outcome, somewhat masking the low scoring individual objectives, and then evaluated by a threshold approach. Of the six outcomes, two met the threshold by less than 1% and two more by less than 6%. This becomes an extreme case of the previous observation regarding students who barely meet the threshold appearing stronger than their underlying competency may be. Careful consideration is needed as to whether this is appropriate and what success in a course truly means. Practitioners who are concerned about mitigating cases such as this may want to consider an additional rider placed on the threshold aggregation approach, such as a limitation on how many outcomes can fall within 10% of the threshold in order to receive a particular letter grade, or perhaps should consider a numerical aggregation approach instead (e.g. Method 10, which assigned this student a 69.85%). Similar but lessened trends are also observed with Methods 13 and 19, which used secondary aggregation by outcome and produced a similar masking effect.

Turning to observations about general methods, rather than specific student cases, the binary type methods, as compared to spectrum type methods of similar scope, weighting, and approach, produce a minimal impact on the high performing student, some reduction for the mid performing student, and a significant reduction for the low performing student. This impact is exacerbated for objectives and outcomes with few assessment opportunities and for cases using a smaller number of objectives or outcomes in the aggregation. Conversely, the negative impact of the binary system is mitigated when assessment opportunities are many and a large number of objectives are included in the aggregation. The simplicity of the approach cannot be understated, but neither can student perceptions that such an approach is harsh and unforgiving (e.g. [12]). Proper distribution of assessment opportunities and open dialogue with students may mitigate the negative perceptions and potential negative impacts on student grades. Ramifications for at-risk populations, including underserved minority students, requires additional study and consideration before adopting such a grading scheme.

Recency weighting schemes are significantly more complex than other methods presented and the full impacts of various permutations of this method require additional study. The impacts of this weighting approach appear to be highly dependent on the underlying course design and likely on the profile of the student in question. There is also some question as to what impact such an approach would have on student behaviors if it were deployed during the semester and advertised to students. Significantly more study is required to fully understand the impact of this approach and the answer may just as likely remain "it depends".

Conclusions and Future Directions

A classification framework for standards-based assessment methods has been presented, encompassing four key dimensions – type, approach, weighting, and scope. Examination of nineteen variations, as applied to three hypothetical student records, has been provided as exemplars. General insights and observations about standards-based grading are provided. In summary, those insights and possible directions for future work include:

1) It is hypothesized that the highest and lowest performing students are not likely to be impacted by any change in assessment scheme, but middle performing students may see significant shifts in final course grade depending on the scheme applied and how it interacts with the underlying course design and student profile. More robust study is needed to confirm this hypothesis.
2) Regardless of method, it is critical that multiple assessment opportunities be provided for all learning objectives (or outcomes, depending on scope). This becomes increasingly important for binary type methods and methods which use a threshold aggregation approach based on a small number of objectives or outcomes.
3) The masking effect of team-based assignments on a student being carried by stronger teammates should be carefully considered, especially if there are few individual assessments of a particular objective or if a recency weighted scheme is applied to a course back-heavy with team assessments.
4) Threshold based aggregation methods reduce the ability to differentiate between barely proficient from highly proficient and, in some extreme cases, may result in highly inflated course grades. Careful course and assessment design is critical when deploying such methods. Novice practitioners may feel more comfortable using a numerical aggregation approach.
5) Implementation of a binary type assessment scheme can significantly reduce the instructor assessment burden and simplify bookkeeping but requires careful course design to ensure sufficient opportunities for students to demonstrate proficiency on each learning objective. Transparency and continuous dialogue is also critical to mitigating student perceptions of unfairness.
6) Impacts of assessment schemes which may be viewed as overly penalizing, such as those using a binary type or threshold aggregation approach, must be carefully studied with respect to their impact on at-risk students and underserved minority populations.
7) The effects on student mindset and behavior during live deployment of certain assessment schemes, such as those using a recency weighting or a threshold aggregation approach, begs additional study. Preliminary work suggests that standards-based grading in general produces an increase in student self-efficacy and motivation [17], but the underlying driver of this effect has not yet been attributed to any particular element of the assessment approach.
8) Novice practitioners may feel more comfortable using an even-weighted numerical aggregation approach, at least during early implementations of standards-based grading, until underlying structural issues can be resolved (such as an underassessed objective) and the faculty member better understands how to deploy and navigate the standards-based approach.

References

1. Sadler, D. R. "Interpretations of criteria-based assessment and grading in higher education," *Assessment & Evaluation in Higher Education, 30*(2), pp. 175-194, 2005
2. Guskey, T. R. "Five obstacles to grading reform," *Educational Leadership*, 17–21, November 2011
3. A.R. Carberry, M. T. Siniawski, and J. D. N. Dionisio, "Standards-based grading: Preliminary studies to quantify changes in affective and cognitive student behaviors," In Proc. Frontiers Edu. Conf. 2012.

4. M. T. Siniawski, A. R. Carberry, and J. D. N. Dionisio, "Standards-based grading: An Alternative to score-based assessment," In Proc. Amer. Soc. Eng. Edu. Pacific-Southwest Regional Conf., 2012.
5. R. W. Tyler, *Basic principles of curriculum and instruction*. Chicago: Univ. of Chicago Press, 1949.
6. G. P. Wiggins and J. McTighe, *Understanding by design*, Assoc. Supervision & Curr. Dev., 2005.
7. Reeves, D. 2003. Making standards work: how to implement standards-based assessments in the classroom, school, and district, 3rd ed., Advanced Learning Press, Englewood, CO
8. D. H. Ziegenfuss, "A phenomenographic analysis of course design in the academy," *J. Ethnographic & Qualitative Res.,* vol. 2, pp. 70-79, 2007.
9. J. Heywood, "The evolution of a criterion referenced system of grading for engineering science coursework," In Proc. Frontiers in Edu. Conf., pp. 1–6 & 22–25, Oct, 2014.
10. B. M. A. Muñoz, & T. R. Guskey, "Standards-based grading and reporting will improve education," Kappan, pp. 64-68, Apr 2015.
11. P. J. Parker, B. Bocher, & A. Polebitski, "Assessing Student Writing Competencies in Environmental Engineering Courses," In Proc. 121st Am. Soc. Eng. Edu. Annu. Conf. & Expo., pp. 24.205.1–12, 2014.
12. S. L. Post, "Standards-Based Grading in a Fluid Mechanics Course," In Proc. Am. Soc. Eng. Edu. Annu. Conf. & Expo, pp. 24.1099.110, June 2014.
13. S. L. Post, "Standards-Based Grading in a Thermodynamics Course," vol. 7, no. 1, pp. 173–182, 2017
14. J. B. Hylton & H. Diefes-Dux, "A Standards-based Assessment Strategy for Written Exams," 2016 ASEE Annu. Conf. Expo., 2016.
15. Marbouti, Farshid, "A standards-based grading model to predict students' success in a first-year engineering course," 2016. Open Access Dissertations, 675. https://docs.lib.purdue.edu/open_access_dissertations/675
16. P. L. Scriffiny, "Seven Reasons for Standards-Based Grading," *Educational Leadership*, 66(2), pp. 70-74, October 2008
17. S. A. Atwood, M. T. Siniawski, & A. R. Carberry, "Using Standards-based Grading to Effectively Assess Project-based Design Courses," In Proc. 121st Am. Soc. Eng. Edu. Annu. Conf. & Expo., pp. 24.1345.1–10, 2014.
18. T. R. Guskey, "Helping Standards Make the Grade," Edu. Leadership, pp. 20–27, Sept, 2001.
19. A.R. Carberry, M.T. Siniawski, S.A. Atwood, & H. Diefes-Dux, "Best Practices for Using Standards-based Grading in Engineering Courses," In Proc. 123rd Am. Soc. Eng. Edu. Annu. Conf. & Expo., June 2016
20. Reddy, Y. M., & Andrade, H. "A review of rubric use in higher education," *Assessment & evaluation in higher education*, 35(4), pp 435-448, 2010