MACHON SIACH at SAR High School מכון שיח

Honoring the memory of Belda K. Lindenbaum z"l

Objectivity and Subjectivity in Grading

by Dr. Victor Parkinson

Introduction:

In the summer of 2018, I attended an Advanced Placement Summer Institute on teaching AP Physics C at St. Johnsbury Academy in Vermont. The workshop was taught by Gardner Friedlander, who has taught AP Physics for multiple decades, written AP Physics questions and graded AP Physics exams.

The process of grading (or "reading", as it is called) in excess of fifty thousand AP Physics C exams, historically, involves approximately one thousand physics teachers gathering in person for one week to become readers of the exam using a rubric.

We thus arrive at the first challenge of grading this exam: consistency between readers. Readers undertake a one-day calibration and training process on a set of pre-scored answers until everyone at the table agrees, down to each individual point, on what score these answers receive. Once the table begins grading, each reader's grading is checked by the table leader until the leader is satisfied that the reader has properly internalized how to grade with that rubric.

Once a student's exam has had all of its multiple choice and free response questions graded, the points earned are added up to form a composite score out of 90, after which a student's score of 1 through 5 is determined. Here is a typical set of cut scores for AP Physics C Mechanics:

Exam Score	Cut Score
5	55-90
4	43-54
3	32-42
2	21-31
1	0-21

This process is designed with the express intention that a student who scores a 5 on the AP Physics C exam would earn an A in an equivalent introductory college physics course, a student with a 4 would earn a B, and so on. This calibration of cut scores to college course achievement levels is conducted every year and the cut scores adjusted accordingly.

This entire machinery of writing, calibrating, administering, and reading the AP Physics C exam is the single most objective grading process I have yet encountered. The extent of this machinery is necessary to achieve some measure of objectivity in grading. It is therefore not surprising that a single teacher, independently writing assessments, calibrating and grading them, and using a rubric they wrote to grade students they personally know, would be far less objective.

The solution to this challenge is not to attempt to create a system where each grading decision made by a teacher is, on its own, objective. Adopting the AP Physics C exam approach of creating consensus between a group of teachers and a pre-written rubric is, during the academic year, not feasible. Instead, one solution is acknowledging that subjective decisions form the basis of grading and creating consensus between the student and the teacher. This philosophy leads us to skill-based grading.

Section One: Relevant Educational Research

Axioms

In the grand tradition of Euclid, I begin with five personal axioms that represent my teaching philosophy.

Axiom #1. All communicated information should be as up-to-date and as accurate as possible.

Axiom #2. Grading is communication. Grades on a transcript communicate to students, to parents, and to institutions with which a student shares the transcript.

Axiom #3. The best way to learn physics is by repeatedly doing it so students form a sort of muscle memory for how to apply physics to analyze situations and solve problems.

Axiom #4. The best way to assess a student's ability level on a physics skill is to provide them with a completely new task that requires that skill and to record their thought process as they work through the task. The closest practical implementations of this ideal are 1-on-1 oral exams and individual written exams wherein students show their thought process.

Axiom #5. All students are capable of complete mastery of high school physics at their assigned course level within one school year.

Theorems

Next, I will use these axioms, in combination with each other and with relevant educational research, to draw conclusions.

Theorem #1: The bell curve has no place in grading students.
This theorem follows straightforwardly from Axiom 5 and from the definition of the bell curve, which is a mathematical function that describes the random distribution of a variable with a mean value. Given that all students are capable of mastering high school physics, the distribution of their grades should not be *random*. It might be that the distribution of their ability levels in physics is random at the very beginning of their first course on the subject, but the very purpose of instructing them in physics is to ensure that all students reach mastery.

Corollary #1: There's no such thing as too many A grades.
This statement would provoke violent disagreement from some college professors. Covington, (1992); Johnson, Johnson & Tauer, (1979); and Johnson, Skon, & Johnson (1980) all show that a sense of competition is detrimental to student learning.

Theorem #2: The 100-point grading scale is biased towards failure.
This theorem hardly needs the support of education research. At SAR, a grade of 0 to 64 is considered an F, which is 64% of the 100-point scale. Additionally, the school culture is such that grades of D are borderline unacceptable, thus yielding a supermajority 69% of the 100-point grading scale devoted to recording student failure. Not only is this grossly unnecessary in a school that routinely sends every graduate to college, but it sends a message that the school is very focused on student failure (Axiom 2).

Theorem #3: The 100-point grading scale promotes false objectivity and inconsistency. Let's begin with two blind grading experiments done first in 1913 and again in 2011 (Starch, 1913; Brimi, 2011), which provided English teachers with a piece of student writing and math teachers with a student geometry paper. In the Starch experiment, grades varied from 64% to 98% on the English paper, and 28% to 95% on the geometry paper. Brimi's experiment trained English teachers on a grading approach and still found a variance of 50% to 96%.

Corollary #2: A grading scale with fewer increments promotes consistency between teachers, and reduces errors in grading.
Compare the 100-point scale with a smaller scale: 4, 3, 2, 1, 0. There are only five levels each student might receive as a grade, and only four boundaries between grade levels. With broader levels and fewer boundaries, a teacher is much less likely to mistakenly classify a student's work (Guskey, 2015, pp 27-29; Feldman; 2019, pp 80-81). In a follow-up study to his work in 1913, Starch again asked English teachers to grade an essay, this time requiring them to use a 5-point grading scale, which achieved statistically significant reductions in grading variations between teachers (Sims, 1933; Starch, 1915).

Theorem #4: For a cumulative subject, averaging numerical scores over an entire semester or school year results in an inaccurate grade.
Axiom 1 leads the way here. If a student struggles with the skills that are taught in September, in a cumulative subject such as physics, the student organically has multiple opportunities to practice those skills again. Indeed, in physics, mastering the skills of later topics often *requires* analytical skills that are more than sufficient to handle the early material. It is a common success story in physics classes to see a student gradually improve over time, and such a student upon returning to the initial exercises is able to complete them with relative ease. There is no good reason to mathematically reduce this student's grade by including what is now out-of-date and inaccurate information.

To be clear, there are courses and subjects where the material is not highly cumulative. In such cases, unlike physics, it may be that once something is assessed early in the course, it is not assessed again later, and thus that score should indeed be part of the student's class grade.

Theorem #5: Any use of zero grades reduces student motivation.
Axiom 2 tells us that all grades are communication, even the ones that are not assigned to students. Zero is, literally, nothing. Teachers sometimes believe that a zero works in the same way an EMT would use a defibrillator – to shock the system back into action. Educational research does not support this conclusion. Guskey (2004) reports that "no studies support the use of zeros or low grades as effective punishments. Instead of prompting greater effort, zeros and the low grades they yield more often cause students to withdraw from learning."

Corollary #3: The combined use of averaging and a score of zero is a particular hindrance to student learning.
What grade would you give a student who earned the following scores on successive assessments, scored out of 100: 0, 78, 81, 82, 79, 83? The mean is 67%. What if there were ten more scores all hovering around 80? At some point, it becomes obvious to the teacher that this is a "B student", or maybe a "low B student," not a "D student." Use of a zero with averaging of scores immediately results in inaccurate information about a student's understanding (Axiom 1). It can also be argued that the zero itself is inaccurate information about a student's understanding – what student knows literally nothing?

Theorem #5: Individualized comments from the teacher promote learning more than a numerical score or a letter grade.
Numerous experiments with groups of students divided into grade-only, grade-and-feedback, and feedback-only subgroups have been conducted (Page, 1958; Stewart & White, 1976; Elawar & Corno, 1985, to name a few), all showing that providing students with feedback has a positive effect on their learning. It is also important to recognize that feedback must be done properly in order to achieve the best version of this effect (Bloom, Hastings, & Madaus, 1981, Guskey, 2019). Given Axiom 2, it is logical to enhance the communication aspect of grades beyond numerical values and single letters.

We now move to theorems concerning the use of skill-based grading. It is also frequently called standards-based grading, so henceforth I shall refer to it as SBG.

Theorem #6: SBG is aligned to the psychology of learning physics.

Psychology defines "intrinsic motivation" as internal desires to complete an action, such as personal satisfaction. Conversely, "extrinsic motivation" is the desire to complete an action in order to receive some form of external reward (Spielman, Jenkins, Lovett, 2020). SBG's core structure of rating a student's mastery level skill by skill speaks to a student's intrinsic motivation. We are saying to the student, "Your reward for practicing this skill is that you mastered this skill;" thus, the enjoyment of the practice-then-mastery was the motivation. Further research shows better academic behaviors and habits with a goal-oriented grading structure and poor academic behaviors with an ability-oriented grading structure (Kaplan & Maehr, 1999; Kaplan & Midgely, 1999).

Corollary #4: Psychological research shows that providing extrinsic rewards can diminish intrinsic motivation.

A meta-analysis of 128 studies shows significant empirical support for this result (Deci, Koestner, & Ryan, 1999). If the goal is to increase the intrinsic motivation of students to learn physics, then eliminating the extrinsic rewards of high numerical scores is a good first step.

Theorem #7: SBG is an effective teaching tool.

SBG would not be of much use if it did not generate the results we seek according to Axiom 5: that every student can obtain mastery. Fortunately, SBG has been tested in many classrooms, and evidence shows that it is at least as effective as traditional methods (Bloom, 1984; Schoen et. al., 2003; Wambugu & Changeiywo, 2008; Adeyemo & Babajide, 2014; Tietyen, 2017).

Theorem #8: Students should self-evaluate, with guidance, as part of the learning process.
Multiple large studies and meta-analyses of studies in education (Brown & Harris, 2013; Brown, Andrade, Chen, 2015; Panadero, Jonsson, & Botella, 2017; Andrade, 2019) show the two parts of this theorem: that students who engage in any form of self-evaluation experience a benefit to their learning, and that students who are guided in self-evaluation experience a much larger benefit. Examples of guidance in self-evaluation include teacher modeling, the use of rubrics, and transparent connections to recent tasks. The last two points are themselves fundamental features of SBG. Additionally, self-evaluation is consistent with Axiom 2 because students communicate with themselves about their "grade" and Axiom 3 because practicing the skills is enhanced by self-reflection.

Theorem #9: SBG and physics pair like peanut butter and jelly.

You'll forgive me for waxing poetic, but it is especially easy to implement Axiom 3 and Axiom 4 in SBG. Axiom 3 is all about learning physics through repeated practice, and SBG's core process involves repeated practice until mastery. Axiom 4 states that the individual exam is the best way to assess a student's understanding, and the practice of reassessment, central to the structure of SBG, provides for multiple exam environments in which to assess students.

Section Two: Implementation

In the fall semester of 2021, the following implementation of SBG was made for Physics 12.31.

0. What's Not Changing?

The textbook, the curriculum of topics, the exams, the labs, and the encyclopedia assignments. Most in-class lessons remained the same although some in-class time was devoted to student self-evaluation of learned skills.

1. Checklists of Skills

On the first day of the course, we completed a hands-on exercise in scientific modeling that introduced students to the kinds of thinking I expected from them.

Thereafter, I presented a skills checklist spreadsheet that can be constructed with sortable columns and color-coding, so that later on, students could see at a glance what they needed to spend more time working on or what skills gave them the most trouble. Each student's spreadsheet was split into tabs either by month or by topic:

Skill Name	Skill Description	Skill Type	Relevant Book Chapter	Current Understanding Score	Current Effort Score

As we learned new skills, I added skill names, descriptions, and types to the students' spreadsheets, along with relevant chapters in the book. Approximately once a week, students self-evaluated their mastery on recently learned skills and their investment in completing the associated work according to the rubrics below. This took some class time; however, students increased their self-evaluation speed as the semester progressed.

Effort Rubric Understanding Rubric		
Maximum effort	I work until the task is complete, and push myself to exceed the requirements.	I could teach the class!
Full effort	I work until the task is complete.	I understand it, but sometimes make minor mistakes.
More effort	I give it some effort and ask questions to clarify my understanding.	I mostly know it, but I have noticeable gaps in my understanding.
Some effort	I give it some effort, but stop when it becomes tough.	I'm beginning to understand it.
Little to no effort	I give it very little effort.	I have no idea, or I haven't done the work.

On each lab report, encyclopedia entry, and exam, the feedback given to students took the form of how well they demonstrated mastery of the skills required to do that particular piece of work. No numerical scores were provided on assignments or exams to privilege feedback above numbers

Each exam marked the end of a "unit." This checklist of skills simultaneously served multiple purposes: a living document that recorded the student's understanding, a road map of where the course was going and how topics fit together, and a review sheet for the exam.

2. Reassessment

At the end of the semester, there were reassessment days equal in number to the exams that occurred. On each day, all students reassessed skills they did not fully master. In the rare event that a student arrived at the end of the semester having already mastered all skills, more in-depth work was be given to the student. This activity is mandatory for all students to be consistent with the philosophy that all students should push themselves to deeper learning.

2a) How does reassessment fit into the grading system?

Reassessment work only has an upside, meaning that it did not lower a students' score. Consider the following hypothetical case:

Student X showed mastery of skills 1-7 on their first at-home assignment and proficiency on skills 8-12 on the second at-home assignment. Exam 1 tests skills 1-12. On Exam 1, Student X showed mastery of skills 1-5, proficiency on skills 6 and 7, and mastery of skills 8- 12. It turns out that at the time of the first at-home assignment, Student X didn't yet fully understand skills 6 and 7 but received significant help from a classmate. In Student X's skill checklist, I record mastery of skills 1-5, proficiency on skills 6 and 7, and mastery of skills 8-12.

On reassessment day #1, corresponding to Exam 1, Student X showed mastery of skills 6 and 7, and proficiency on skills 8 through 12. Student X reviewed and practiced skills 6 and 7, mastering them in preparation for reassessment, but did not practice 8 through 12. Student X is recorded as having mastered all 12 Exam 1 skills as part of determining their semester grade.

The philosophy at work here is that once a student shows mastery of a skill, the skill checklist records that they *have shown* mastery even if they do not do it perfectly in a later instance.

2b) How much work is reassessment for the teacher?

Before reassessment, I solved all the questions myself, of course, to create answer keys; make sure the questions are clear, self-consistent, and physically valid; and knew which skills each question addressed. Therefore, evaluating each student's work on a particular question was a near-instantaneous process.

It is important to acknowledge that preparing this material presents a significant workload to the teacher. However, it is mitigated by the efficiency of using a checklist of skills to decide a students' semester letter grade. For most students, this entire process will be a small adjustment to a grade that is largely already decided – I need only grade those skills that make the difference on the checklist. The process is further mitigated by the upside-only nature of the reassessment.

Work that was obviously poor quality did not need to be examined in detail.

3. Determining Semester Quality of Work Grades, Part One: Letter and Minus Grades.

To determine a student's letter grade at the end of the semester, each student's skill checklist was compared against a Semester Checklist containing broader skill sets. Each Semester Checklist had criteria necessary for the student to achieve an A, A-, B, B-, etc.

The following skill sets were covered during the fall semester:
1. Experimental analysis
2. Motion Graphs
3. Dot Diagrams
4. Projectile Motion Vectors
5. Free Body Diagrams
6. Circular Motion, Solar System, and Gravity

Here is one example of a piece of the grade checklist.

To receive a B, a student must show proficiency on the following skill sets.

1. **Experimental Analysis.** In a lab environment, show mastery of Linear Equation Graphing, Error Analysis: Random vs Systematic, and proficiency on 5 of the following 6 skills:
    - Error Analysis: Rulers
    - Error Analysis: Motion Sensors
    - Error Analysis: Dot Timers
    - Error Analysis: Trajectory Video
    - Error Analysis: Force Sensors
    - Error Analysis: Stopwatches

2. **Motion Graphs**. In an exam environment, show mastery of the following 2 skills:
    - Kinematics: Constant a Graphing
    - Kinematics: Position Graph Drawing

Show proficiency on the following 2 skills:
- Kinematics: Velocity Graph Drawing
- Kinematics: Acceleration Graph Drawing

Additionally, show proficiency on 2 of the following 3 skills:
- Kinematics: Position Graph Reading
- Kinematics: Velocity Graph Reading
- Kinematics: Acceleration Graph Reading

3. **Dot Diagrams.** In an exam environment, show mastery of the following 4 skills: •
    1-D Dot Diagrams: Basic Cases
    - 2-D Dot Diagrams: Basic Cases
    - 1-D Dot Diagrams: Descriptions
    - 2-D Dot Diagrams: Descriptions

Additionally, show proficiency on the following 3 skills:
- 1-D Dot Diagrams: Turnaround
- 1-D Dot Diagrams: Motion Graphs

- 2-D Dot Diagrams: Comparison

4. Determining Semester Quality of Work Grades, Part Two: Plus Grades

Historically, I have used an extra credit system called "Good Questions". The nature of this work, and the name of the system, were inspired by this quote from particle physicist I.I. Rabi:

"My mother made me a scientist without ever intending to. Every other Jewish mother in Brooklyn would ask her child after school: So? Did you learn anything today? But not my mother. "Izzy," she would say, "did you ask a good question today?" That difference — asking good questions — made me become a scientist." – I.I. Rabi, Parents Magazine (Sept 1993)

The work goes like this: students write and submit an entry consisting of some out-of-classroom context, a question about the physics occurring in that context, and their best attempt at identifying relevant principles of physics. Students are not required to answer their own questions, but they are encouraged to comment on other students' questions. The Good Questions Extra Credit system awarded students points 0-5 for each entry, and these points were added to numerical exam scores as a means of improving their average exam score, which in turn would improve their letter grade.

In the new grading system, there are no numerical exam scores, and there is no averaging to determine a grade. Any optional work that influences a quality of work grade will be completed more often by students who are more motivated, and who have environments and resources that enable them to spend additional time (Feldman, 2019, pp 112-115). The quality of work grade should not be a measure of motivation or resources.

Instead, Good Questions entries moved students to the plus version of the letter grade, e.g. B to B+. To go from a D to a D+ required two entries, a C to a C+ required four, a B to a B+ required six, and an A to an A+ required eight. A student could not move from a B+ to an A- with Good Questions entries, because an A- has other, specific criteria (see above) that must be fulfilled. Similarly, a student who has only met the B- criteria cannot use Good Questions entries to move to a B grade.

5. Investment in Learning Grade

All students began the year with an investment in learning grade of 80 on a scale from 60 to 100. For each student, that grade was adjusted up and down according to the following:

- -1 for each tardy, -2 for each absence
- +2 for each on time assignment, -2 for each late assignment
- -4 for each missing assignment
- +1 for each on time pre-class video
- +2 for each good week of in-class participation (This means answering questions when asked and being engaged in classwork. The intention is to not award high grades just for being cheerful or extroverted.)
- -10 for any violation of Academic Integrity, unless it is the student's first violation at SAR, and they have completed an Academic Integrity Repair Plan, in which case no change.

Notably, redemption is built directly into this system in the same way that reassessment is built directly into the quality of work system. A student who turned in the first five assignments late moved to an investment score of 70. If that student then turned in the remaining five assignments on time, this canceled out the prior lates, returning the student to an 80. If, additionally, the student had five good weeks of in-class participation and completed three pre-class videos on time, the student received a grade of 93 for investment in learning.
Note that the implied causality of the above example could be reversed for the same resulting grade. In this way, the system as outlined measures how much of the course the student was invested in, without regard for when the investment occurred. Trends in the student's investment were, of course, described in the anecdotal.

6. The Final Assessment

Physics is a cumulative subject, and as such a final assessment is necessary both to obtain a "last possible" check on a student's understanding and to assess the student on integration of the concepts and skills covered throughout the year. Yet each of these two purposes is a substantial undertaking and must therefore be handled separately. Thus, the final assessment consisted of two parts: another sequence of reassessment days that covered the entire year and a timed exam that assessed the students on integration of concepts and skills.

The final reassessment days followed the spring semester reassessment days after at least a week to allow for in-class review. Unlike the semester reassessment days, the students need the freedom to show their improvement on a much wider range of skills. Therefore, instead of using a question bank, students were required to write and solve their own questions. Given that students were allowed to prepare and use a reference sheet, measures should be taken to avoid students copying a question and solution from an external source and submitting it as their own. A simple solution exists: Students were required to incorporate randomly generated object(s), setting(s), graph shape(s), and/or numerical value(s) that were provided to them at the start of the class period. In this way, even a student who copied a question must rewrite it to match the given prompts.

A student's final quality of work grade for the year was determined by a checklist that allowed the student to show improvement on any skill set covered in either the fall or the spring semester as well as their ability level in integrating concepts and skills. A student's final investment in learning grade for the year was determined by a weighted average of their fall and spring semester investment in learning grades, with a higher weight applied to the higher grade.

References

Adayemo, S.A., Babajide, V.F.T. (2014). "Effects of Mastery Learning Approach on Students' Achievement in Physics." *International Journal of Scientific & Engineering Research*. Vol 5, Issue 2, February 2014, pp910-920.

Andrade H.L. (2019) A Critical Review of Research on Student Self-Assessment. *Front. Educ.* 4:87. doi: 10.3389/feduc.2019.00087

Bloom, B.S., Madaus, G.F., & Hastings, J.T. (1981). *Evaluation to improve learning.* New York, NY: McGraw-Hill.

Bloom, B.S., (1984). "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring". *Educational Researcher,* Vol 13, 6 (Jun-Jul 1984), pp 4-16.

Brimi, Hunter M. (2011) "Reliability of Grading High School Work in English," *Practical Assessment, Research, and Evaluation*: Vol. 16 , Article 17.
DOI: https://doi.org/10.7275/j531-fz38

Brown, G. T., and Harris, L. R. (2013). "Student self-assessment," in *Sage Handbook of Research on Classroom Assessment*, ed J. H. McMillan (Los Angeles, CA: Sage), 367–393. doi: 10.4135/9781452218649.n21

Brown, G., Andrade, H., and Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assess. Educ.* 22, 444–457. doi: 10.1080/0969594X.2014.996523

Covington, M. V. (1992). *Making the grade: A self-worth perspective on motivation and school reform.* Cambridge University Press. https://doi.org/10.1017/CBO9781139173582

Craig T. A. (2011). *Effects of standards-based report cards on student learning*. (Doctoral dissertation). Retrieved from https://repository.library.northeastern.edu/files/neu:1127

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology, 77*(2), 162–173. https://doi.org/10.1037/0022-0663.77.2.162

Feldman, Joe. (2019). *Grading for Equity.* Corwin, a SAGE Company, Thousand Oaks, California. ISBN 978-1-5063-9157-1.

Guskey, Thomas R. (2015). *On your mark: challenging the conventions of grading and reporting.* Solution Tree Press, Bloomington IN. ISBN 978-1-935542-77-3.

Guskey, T.R. (2019, Oct. 28). Grades versus comments: Research on student feedback. *Phi Delta*

Kappan, 101 (3), 42-47.

Johnson D.W., Skon L, Johnson R.T. Effects of Cooperative, Competitive, and Individualistic Conditions on Children's Problem-solving Performance. *American Educational Research Journal*. 1980;17(1):83-93. doi:10.3102/00028312017001083.

Johnson, R. T., Johnson, D. W., & Tauer, M. (1979). The effects of cooperative, competitive, and individualistic goal structures on students' attitudes and achievement. *The Journal of Psychology: Interdisciplinary and Applied, 102*(2), 191–198. https://doi.org/10.1080/00223980.1979.9923487

Kaplan, A., & Maehr, M. L. (1999). Achievement goals and student well-being. *Contemporary Educational Psychology, 24*(4), 330–358. https://doi.org/10.1006/ceps.1999.0993

Kaplan, A., & Midgley, C. (1999). The relationship between perceptions of the classroom goal structure and early adolescents' affect in school: The mediating role of coping strategies. *Learning and Individual Differences, 11*(2), 187–212. https://doi.org/10.1016/S1041-6080(00)80005-9

Page, E.B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *Journal of Educational Psychology, 49* (2), 173-181. Starch, D. (1913). "Reliability and distribution of grades." *Science*, 38(983), pp 630-636.

Panadero, E., Jonsson, A., and Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analyses. *Educ. Res. Rev.* 22, 74–98. doi: 10.1016/j.edurev.2017.08.004

Schoen, H. L., Cebulla, K. J., Finn, K. F., & Fi, C. (2003). Teacher variables that relate to student achievement when using a standards-based curriculum. *Journal for Research in Mathematics Education, 34*(3), 228–259. https://doi.org/10.2307/30034779

Spielman, R., Jenkins, W., Lovett, M. *Psychology 2e*. OpenStax, Houston, TX. April 22, 2020. ISBN-13: 978-1-975076-45-0.

Starch, D. (1915) "Can the variability of marks be reduced?" *School and Society,* 2(33), pp 242-243.

Stewart, L. G., & White, M. A. (1976). Teacher comments, letter grades, and student performance: What do we really know? *Journal of Educational Psychology, 68*(4), 488–500. https://doi.org/10.1037/0022-0663.68.4.488
Tietyen, E.D. (2017). "Standards-based grading a high school physics classroom." Master's Thesis, Montana State University. https://scholarworks.montana.edu/xmlui/handle/1/13699 , Accessed July 2021.

Wambugu, P. & Changeiywo, J. (2008). Effects of Mastery Learning Approach on Secondary School Students' Physics Achievement. Eurasia Journal of Mathematics, Science and Technology Education. 4. 10.12973/ejmste/75352.