ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #40776

Grading: The (Mis)use of Mathematics in Measuring Student Learning and its Disproportionate Impact on Equity and Inclusion

Prof. Sharona Krinsky, California State University, Los Angeles

Sharona Krinsky is an instructor and course coordinator in the Mathematics department at California State University, Los Angeles and the co-PI of the NSF funded project "Commitment to Learning Instilled by a Mastery-Based Undergraduate Program (CLIMB-UP). She works with faculty on redesigning courses to utilize the principles of mastery-based grading in order to enhance student success and enable increased equity, inclusion, and access to careers in STEM fields for students from historically underrepresented groups. Sharona is a founding organizer of "The Grading Conference", an annual two-day online conference focused on reforming grading as we know it across STEM fields throughout higher education, now entering its fifth year. She coordinates a large general education Quantitative Reasoning with Statistics course for over 1,400 students per year as well as teaches a wide range of mathematics courses including Calculus and Linear Algebra.

Mr. Robert Christopher Bosley, California State University, Los Angeles
Dr. Dina Verdin, Arizona State University, Polytechnic Campus

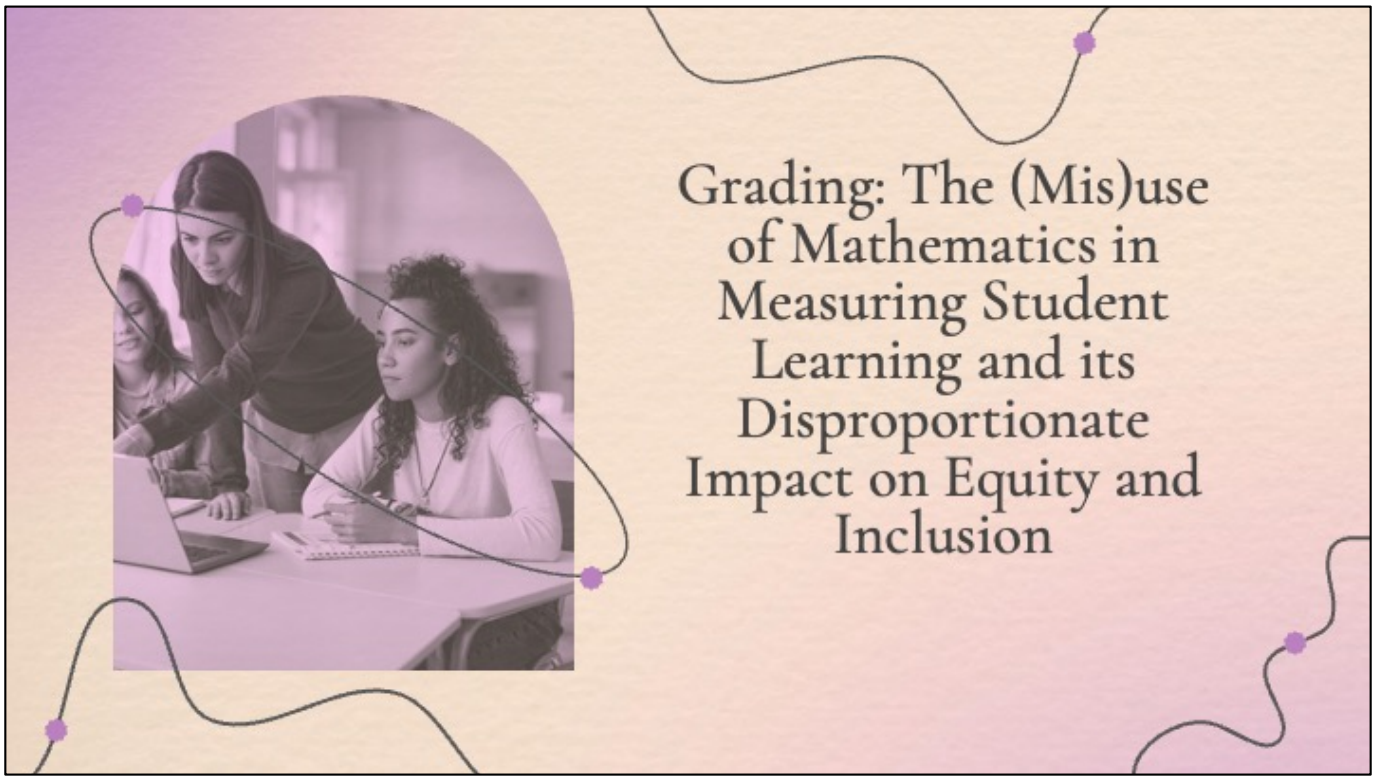
Dina Verdín, PhD is an Assistant Professor of Engineering in the Ira A. Fulton Schools of Engineering at Arizona State University. She graduated from San José State University with a BS in Industrial Systems Engineering and from Purdue University with an MS in Industrial Engineering and PhD in Engineering Education. Dina is a 2016 recipient of the National Science Foundation's Graduate Research Fellowship and an Honorable Mention for the Ford Foundation Fellowship Program. Her research interest focuses on changing the deficit base perspective of first-generation college students by providing asset-based approaches to understanding this population. Dina is interested in understanding how first-generation college students author their identities as engineers and negotiate their multiple identities in the current culture of engineering. Dina has won several awards including the 2022-2023 Outstanding Research Publication Award by the American Educational Research Association (AERA) Division I, 2018 ASEE/IEEE Frontiers in Education Conference Best Diversity Paper Award, 2019 College of Engineering Outstanding Graduate Student Research Award and the Alliance for Graduate Education and the Professoriate (AGEP) Distinguished Scholar Award. Dina's dissertation proposal was selected as part of the top 3 in the 2018 American Educational Research Association (AERA) Division D In-Progress Research Gala. Dina was a 2016 recipient of the National Science Foundation's Graduate Research Fellowship and an Honorable Mention for the Ford Foundation Fellowship Program.

Eva Schiorring, STEMEVAL

Eva Schiorring has almost two decades of experience in research and evaluation and special knowledge about STEM education in community colleges and four-year institutions. She presently serves as the external evaluator for four NSF-funded projects. The

Dr. Emily L. Allen, California State University, Los Angeles

Emily L. Allen, Ph.D., is Dean of the College of Engineering, Computer Science, and Technology at California State University, Los Angeles. She believes in a collaborative, student-centered approach to research, education, academic administration and lea



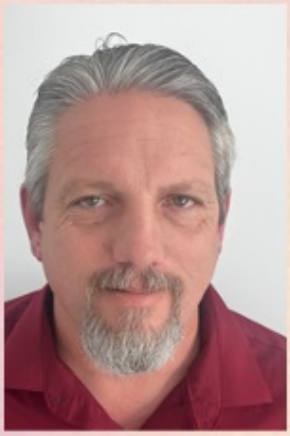
In this presentation, we will examine the beliefs that are unconsciously embedded in our current grading systems, the mathematical implications of many of our current choices within traditional grading systems, the historical development of our traditional grading systems and what the implications are of those choices on efforts to promote Equity and Inclusion.

We report on the results of a 3-year NSF IUSE funded project supporting the redesign of sophomore level Engineering courses to utilize alternative grading methods such as mastery-based grading, looking at both faculty experiences with the redesign process and student experiences of taking the redesigned courses.

WHO WE ARE

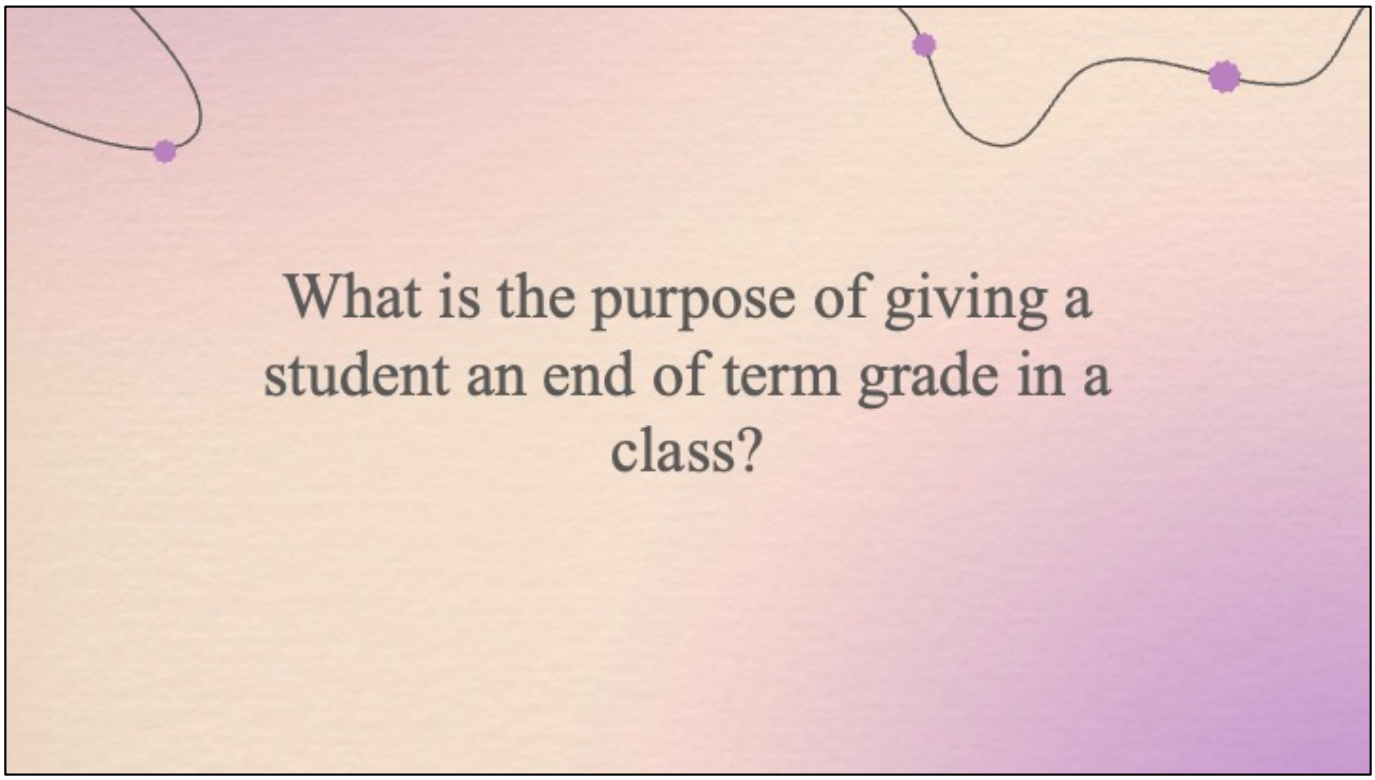
Sharona Krinsky
California State University, Los Angeles, Math

Co-PI and Faculty Development Facilitator on the NSF IUSE Funded CLIMB-UP Grant



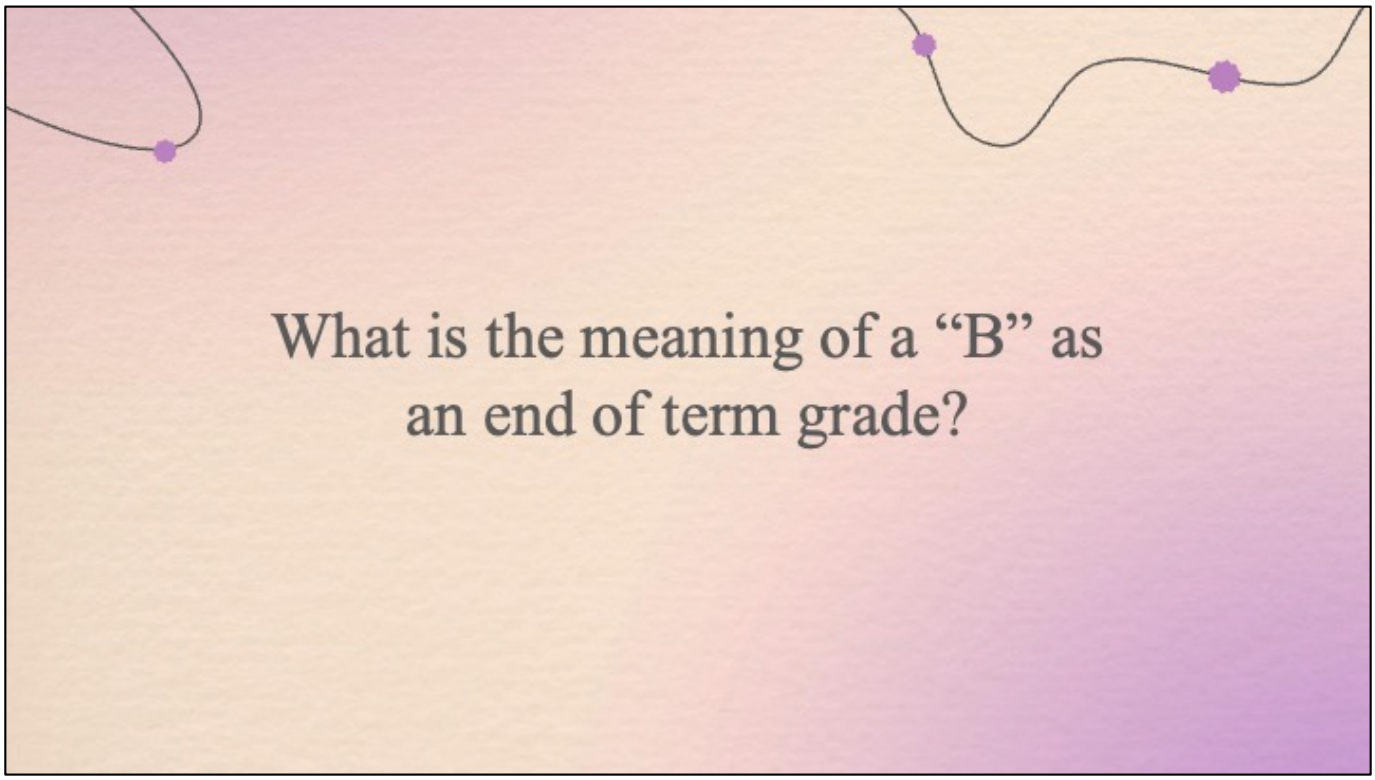
Robert Bosley
California State University, Los Angeles, Math

Faculty Development Facilitator on the NSF IUSE Funded CLIMB-UP Grant



What is the purpose of giving a student an end of term grade in a class? In asking this question of hundreds of educators over the last six years, the top three answers given have been:

1. Communicate the amount and/or quality of what a student has learned.
2. Communicate to the student whether or not they have learned enough to be successful in a subsequent class.
3. Rank students for use in future opportunities such as admission for advanced degrees, professional schools, jobs, scholarships, etc.

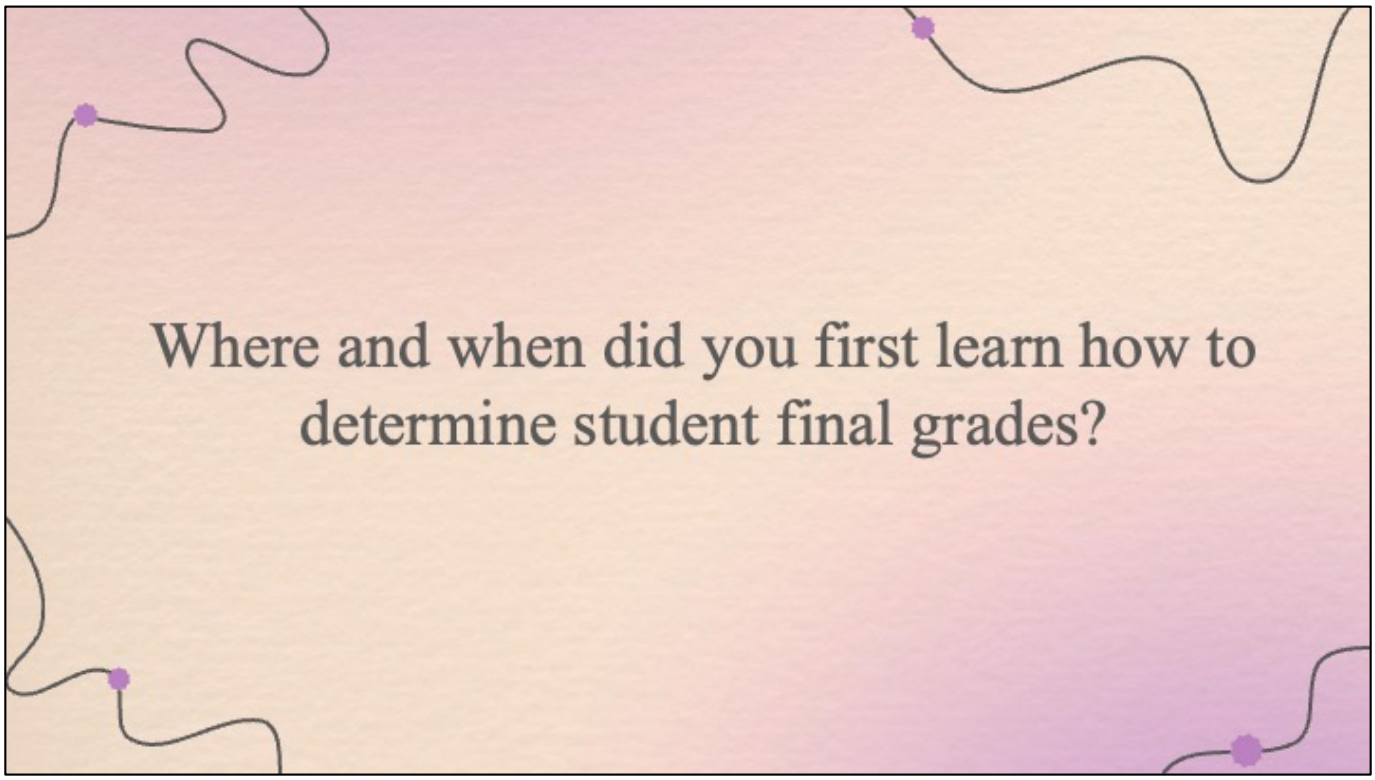


In addition to the question asked a student gets a “B” in a class – what do we want it to mean?

Top three answers from previous faculty asked:

- Student is likely to be successful in a subsequent course.
- Student knows most of the material but not all of it or not perfectly.
- Depends on who gave the B (common answer)

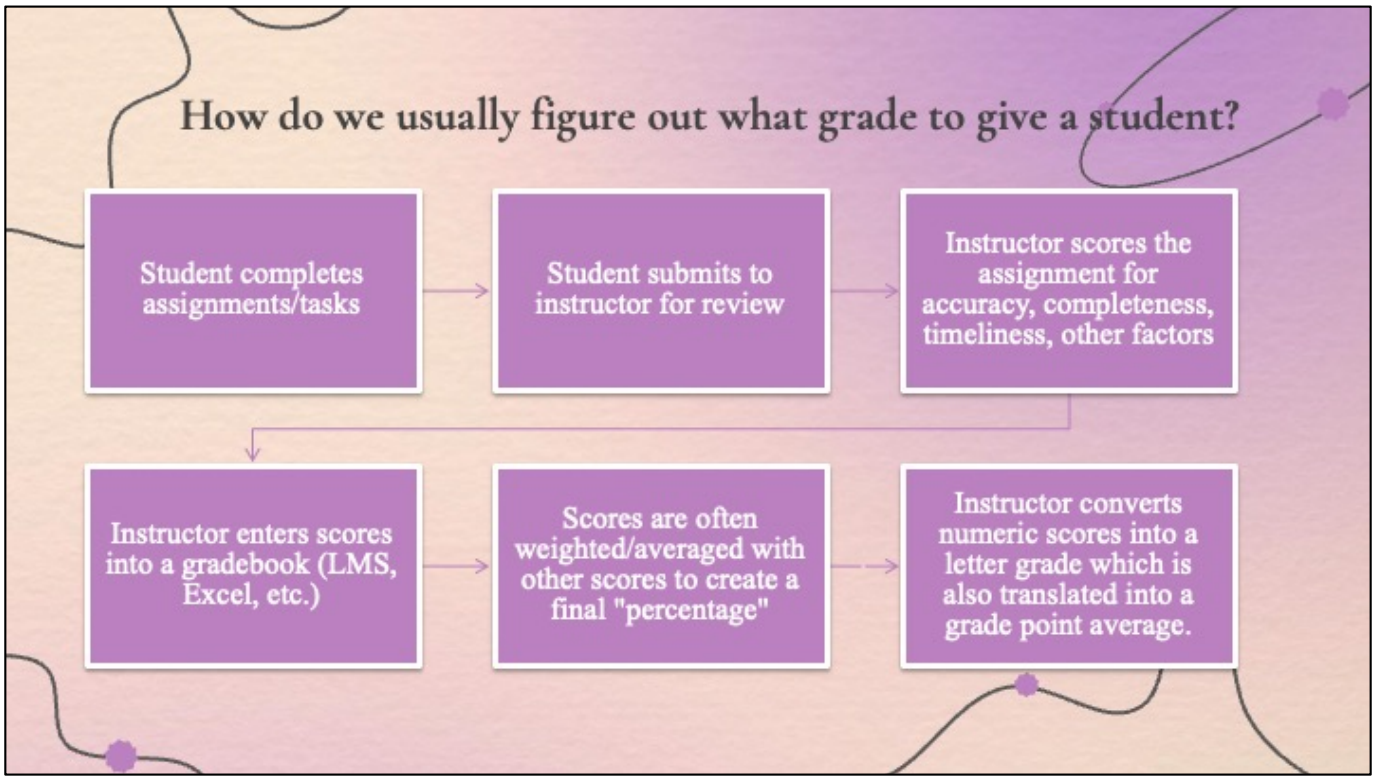
The next question we ask – does it actually mean what we want it to mean? Almost unanimously faculty say “No”.



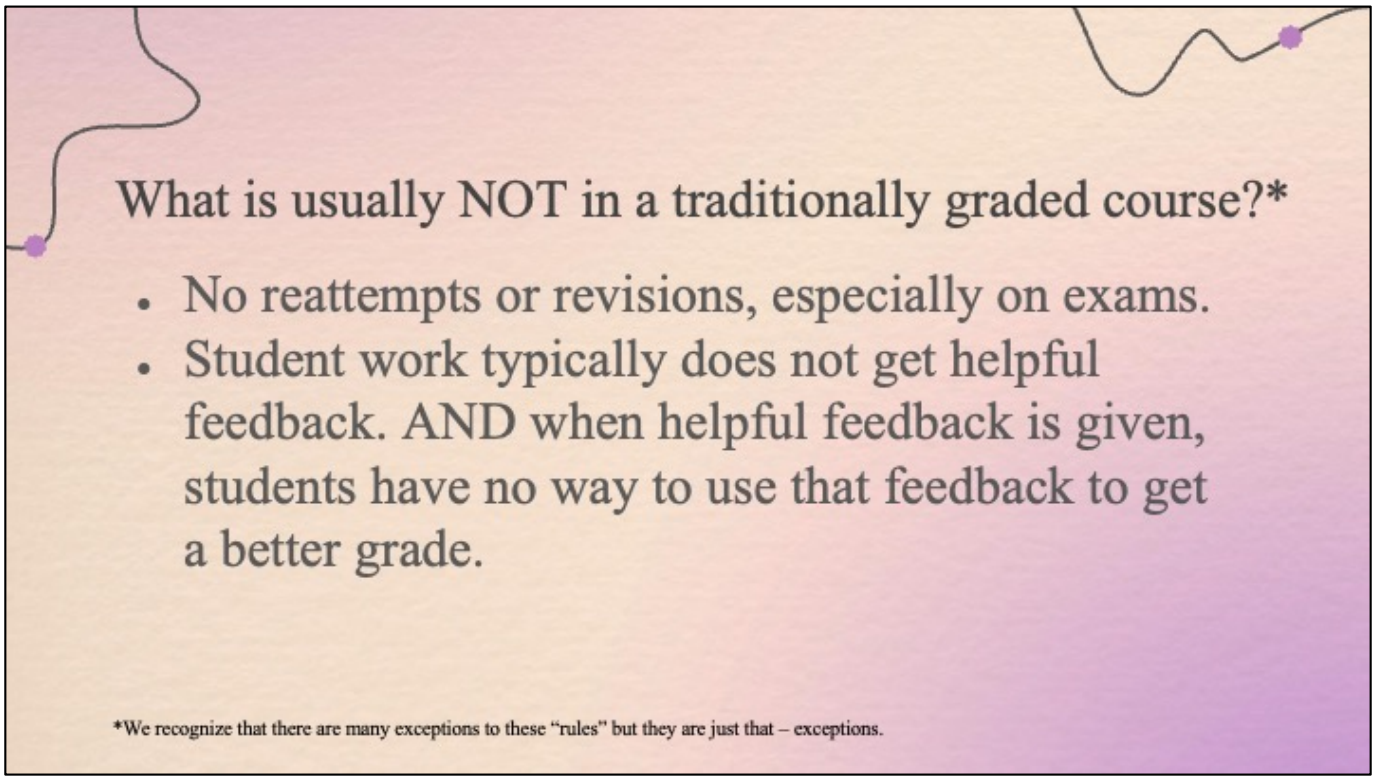
- Participants in the workshops are asked to answer the question about when and how they learned how to determine final grades. The possible answer choices are:

    1. I had a specific course which focused on research-based strategies to determine student final grades during my credential program.
    2. I learned from a mentor teacher.
    3. I learned by collaborating with other teachers after I started teaching.
    4. The way I grade is how I was graded as a student.
    5. I learned from the functionality of a digital gradebook.

- Over 80% of respondents answer either #4 or #5, either how they were graded or based on the options available in a digital gradebook. Very rarely has anyone in the workshop had a specific course about grading during a faculty training or credentialing program.

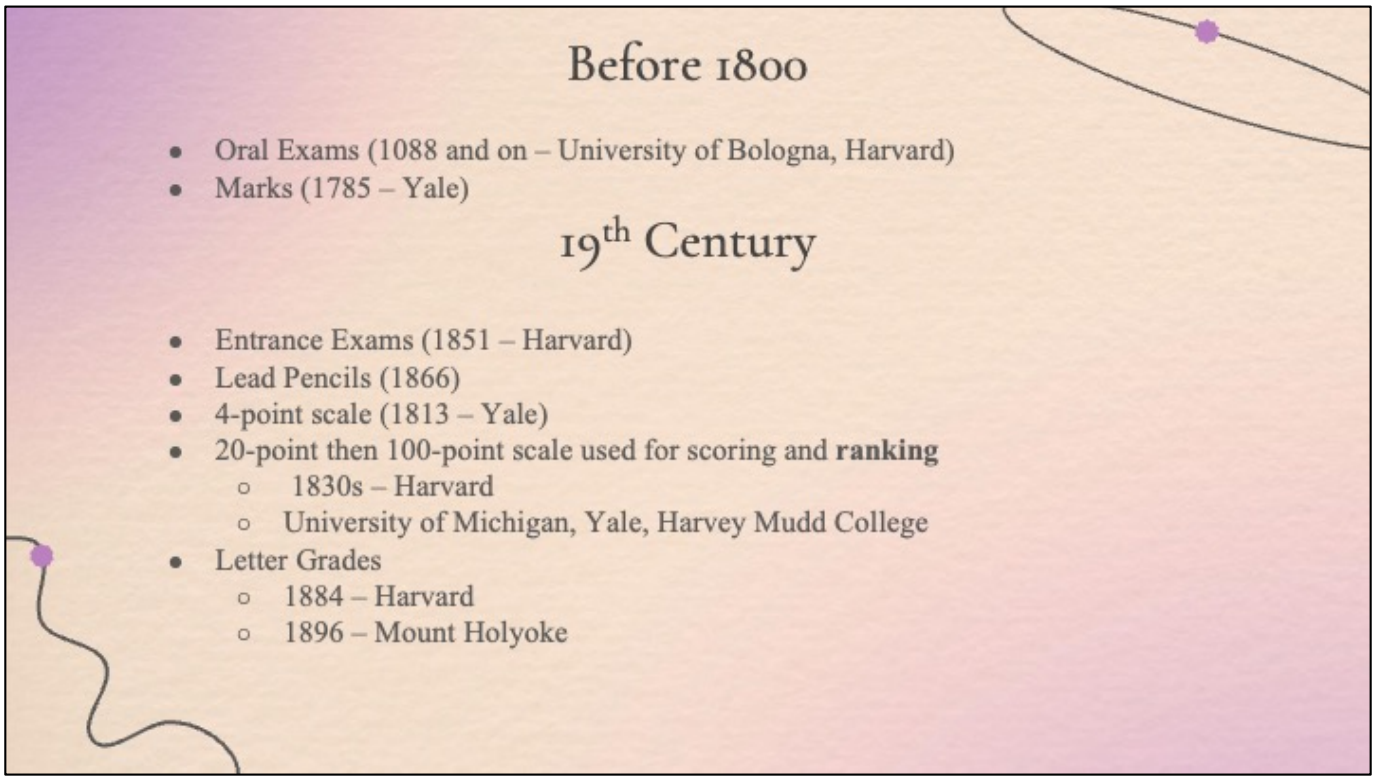


- Typical sequence of course activities that results in scores or grades in a gradebook, especially in higher education STEM courses.



- There is extensive research demonstrating that retrieval practice, the practice of retrieving knowledge from memory, reflecting on that knowledge, and interleaving practice of that knowledge with other topics is effective at making knowledge stick. Known as "the testing effect", research has shown that a number of successful retrievals are required to give a lasting effect.
    - Brown, P.C., Roediger, H. L., & McDaniel, M. A. (2014). *Make It Stick: The Science of Successful Learning.* Cambridge, MA: The Belknap Press of Harvard University Press
    - Carey, B. (2014). *How We Learn.* New York: Random House
    - Rawson K. A., Dunlosky J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? Journal of Experimental Psychology: General, 140(3), 283–302.
- The impact of grades vs feedback on student work has also been documented as well as the components of what effective feedback looks like.
    - Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. Journal of Educational Psychology, 78(3), 210–216. https://doi.org/10.1037/0022-0663.78.3.210
    - Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). Evaluation to improve learning. New York, NY: McGraw-Hill.


A Brief History of Grading in the US



Source Material:

- Nilson, L. & Stanny, C. J. (2015). *Specifications grading : restoring rigor, motivating students, and saving faculty time*. Stylus Publishing.
- Inoue, A. B. (2019). *Labor-based grading contracts : building equity and inclusion in the compassionate writing classroom*. The Wac Clearinghouse.
- Blum, S. & Kohn., A. (2020). *Ungrading : why rating students undermines learning (and what to do instead)*. West Virginia University Press.

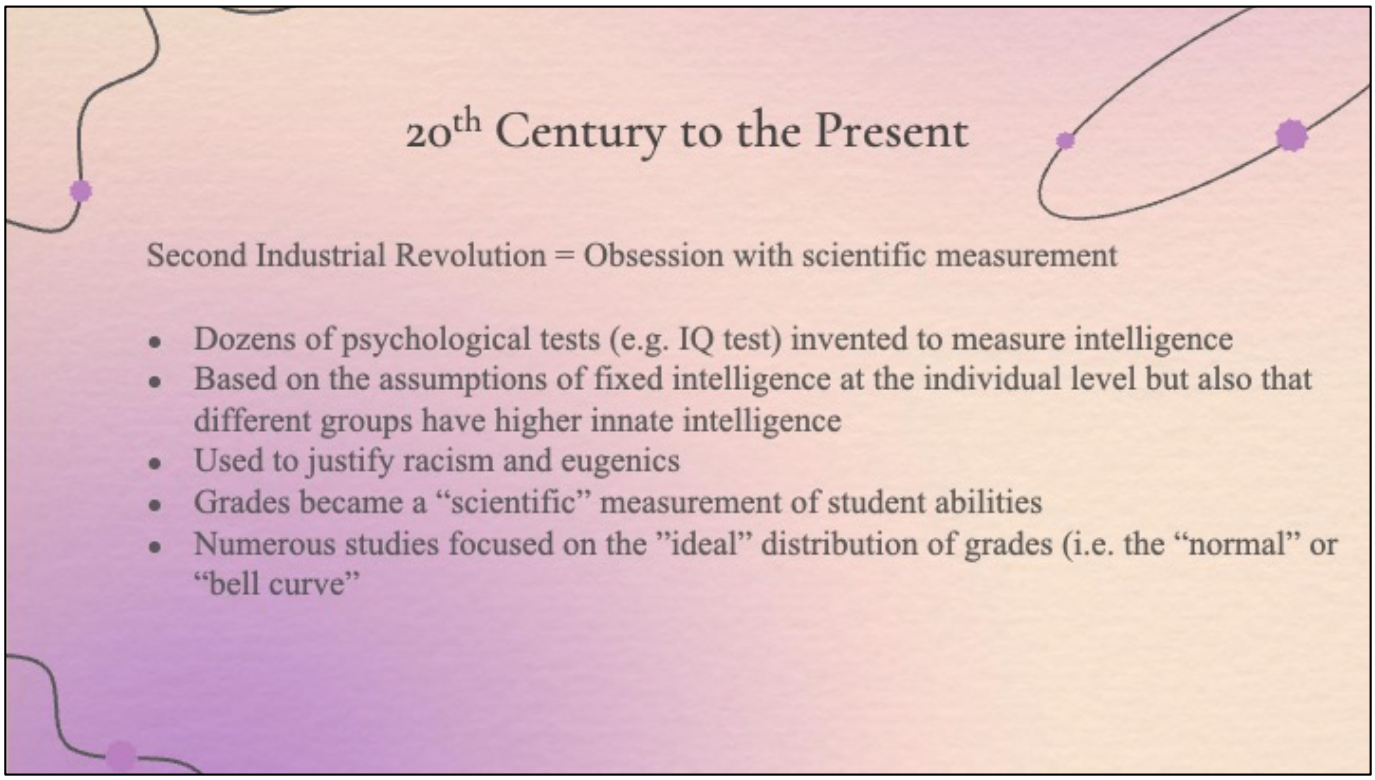
Before 1800

- University of Bologna (founded in 1088) one public oral exam at the end of their education.
- Harvard (founded 1636) public oral examinations involving translation of the Old and New Testaments into Latin. In 1650 Harvard increased the frequency of examinations to once per year.
- Earliest record is at Yale in 1785. Students received one of three descriptive adjectives: *Optimi, Inferiores (Boni) and Pejores* (the best, inferior but good, the worst)

19th century

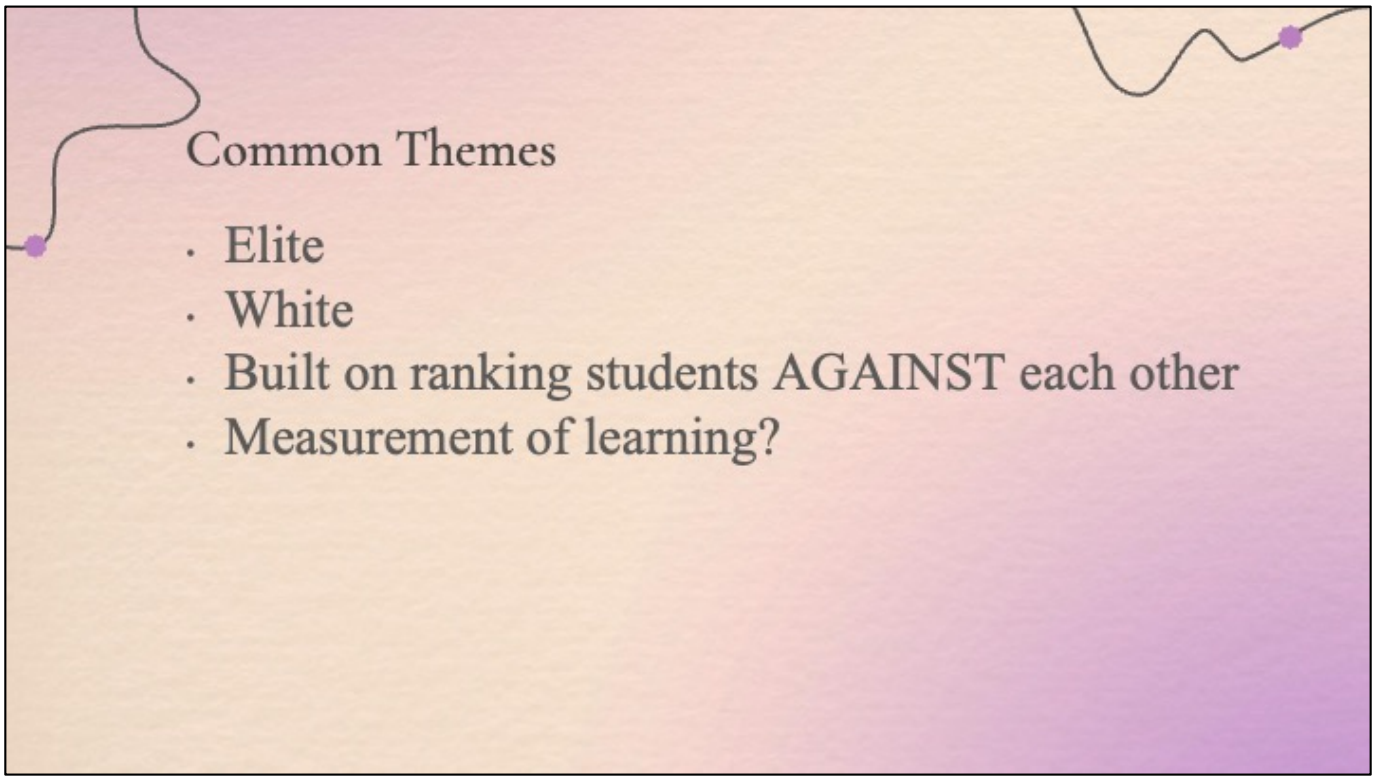
- Beginning to see a shift from oral exams to written exams. (Lead pencils began mass production in 1866)
- Harvard began using written entrance exams in 1851 Instructors began to get more autonomy in evaluating student exams.
- Marking begins:
  - Yale in 1813 – 4-point scale and recording the average results in a ledger
  - Harvard in the 1830s – shift from descriptive adjectives to a 20-point scale used for exams in rhetoric and physics then to a 9-point scale, then to a 4-point scale, then a 100-point scale.
- 100-point scale was used to rank students into six separate divisions:
  - Division 1: 90 or above, Division 2: 75 to 89, Etc.
- Also seeing these fluctuations in scoring and ranking systems at the University of Michigan, Yale and other elite institutions including Harvey Mudd College.
  - 1884 – Harvard institutes letter grades

- 1896 – Mount Holyoke College introduces A, B, C, D and E (then changed from E to F)



Source Material:

- Nilson, L. & Stanny, C. J. (2015). *Specifications grading : restoring rigor, motivating students, and saving faculty time*. Stylus Publishing.
- Inoue, A. B. (2019). *Labor-based grading contracts : building equity and inclusion in the compassionate writing classroom*. The Wac Clearinghouse.
- Blum, S. & Kohn., A. (2020). *Ungrading : why rating students undermines learning (and what to do instead)*. West Virginia University Press.
- Reynolds, Cecil R., et al. *"The problem of bias in psychological assessment."* Mastering Modern Psychological Testing, 2021, pp. 573–613, https://doi.org/10.1007/978-3-030-59455-8_15.
- Eyler, Joshua. *How Humans Learn: The Science and Stories behind Effective College Teaching*. West Virginia University Press, 2021.

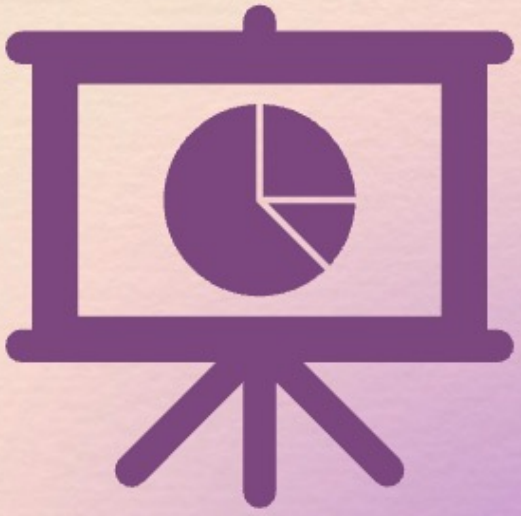


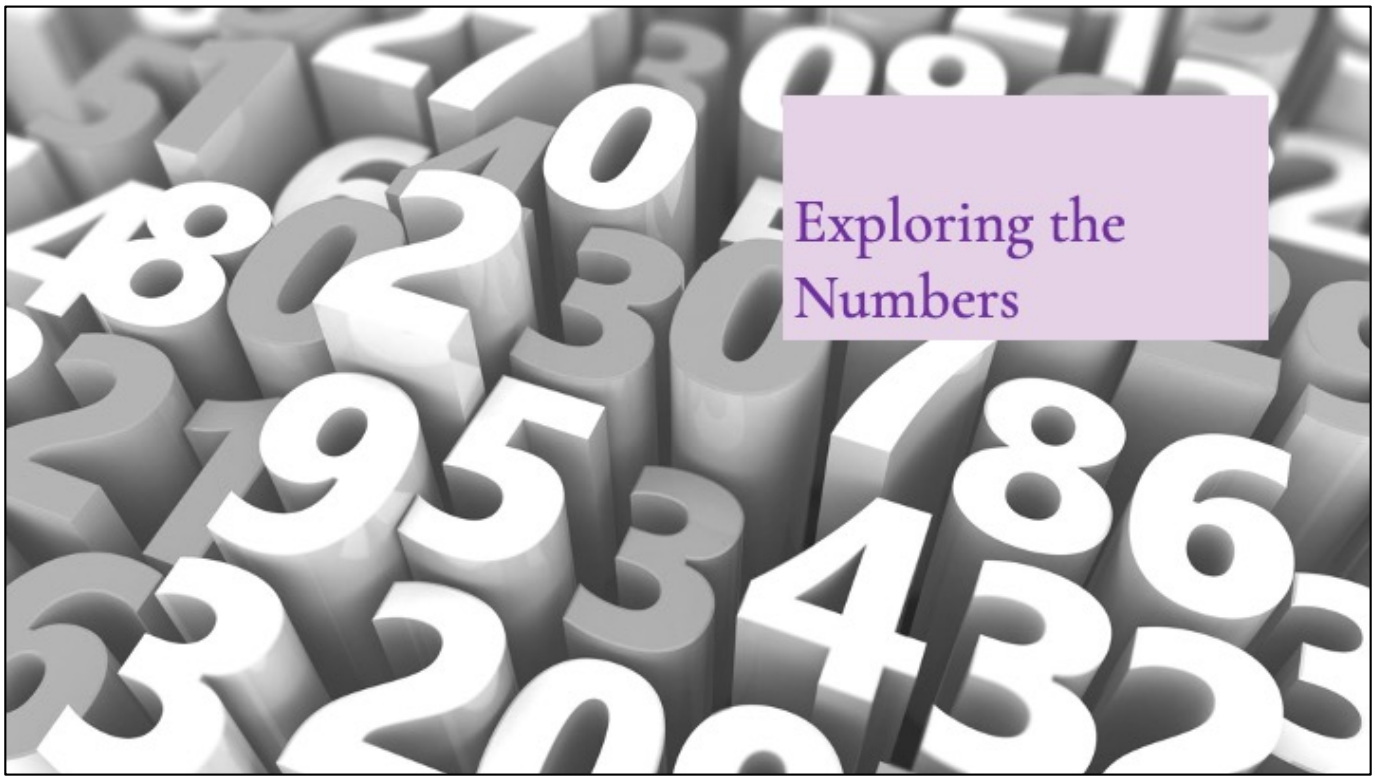
- Schinske, Jeffrey, and Kimberly Tanner. *"Teaching more by grading less (or differently)."* CBE—Life Sciences Education, vol. 13, no. 2, 2014, pp. 159–166, https://doi.org/10.1187/cbe.cbe-14-03-0054.

Does the traditional grading system produce a final grade that meets the purpose of a final grade?

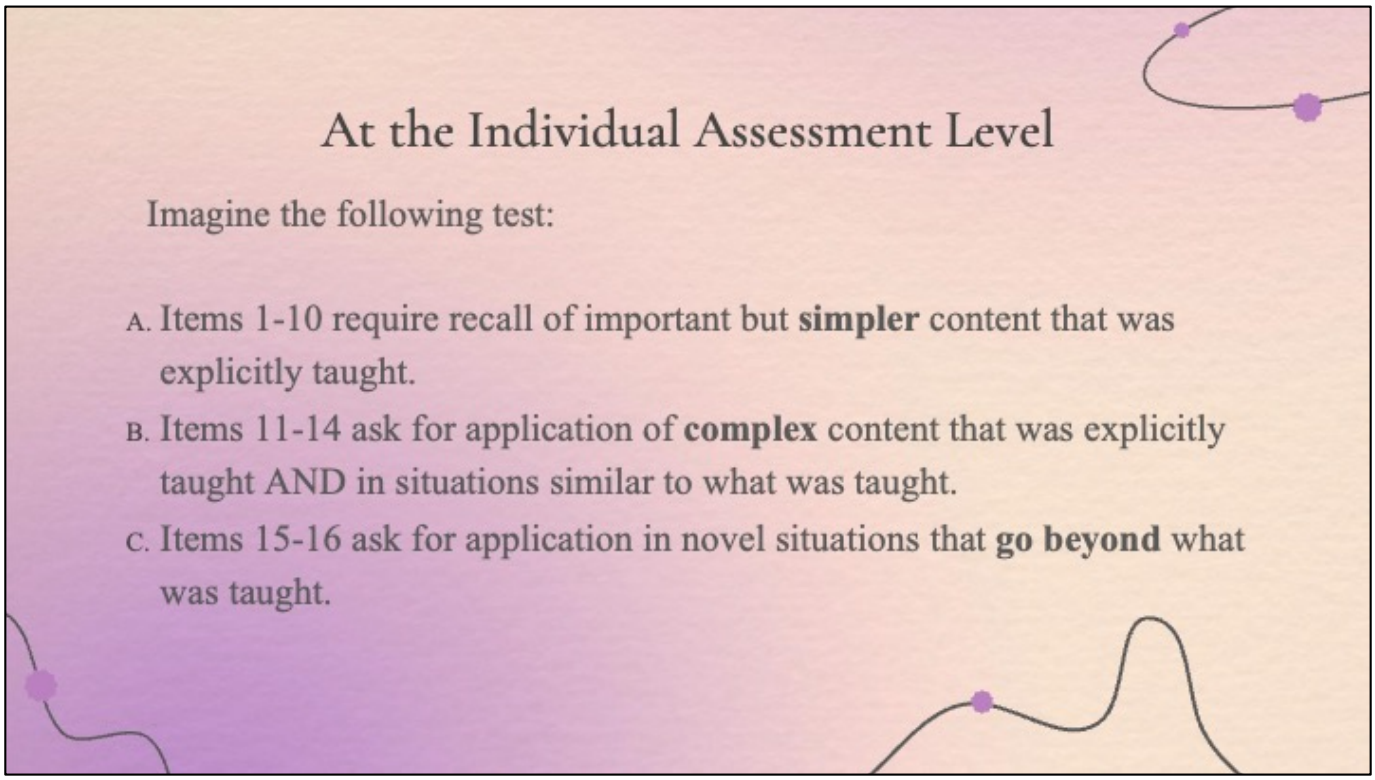
Let's look at these aspects:

1. How we grade a single assignment or assessment.
2. The impact of grade categories.
3. The use of averages.

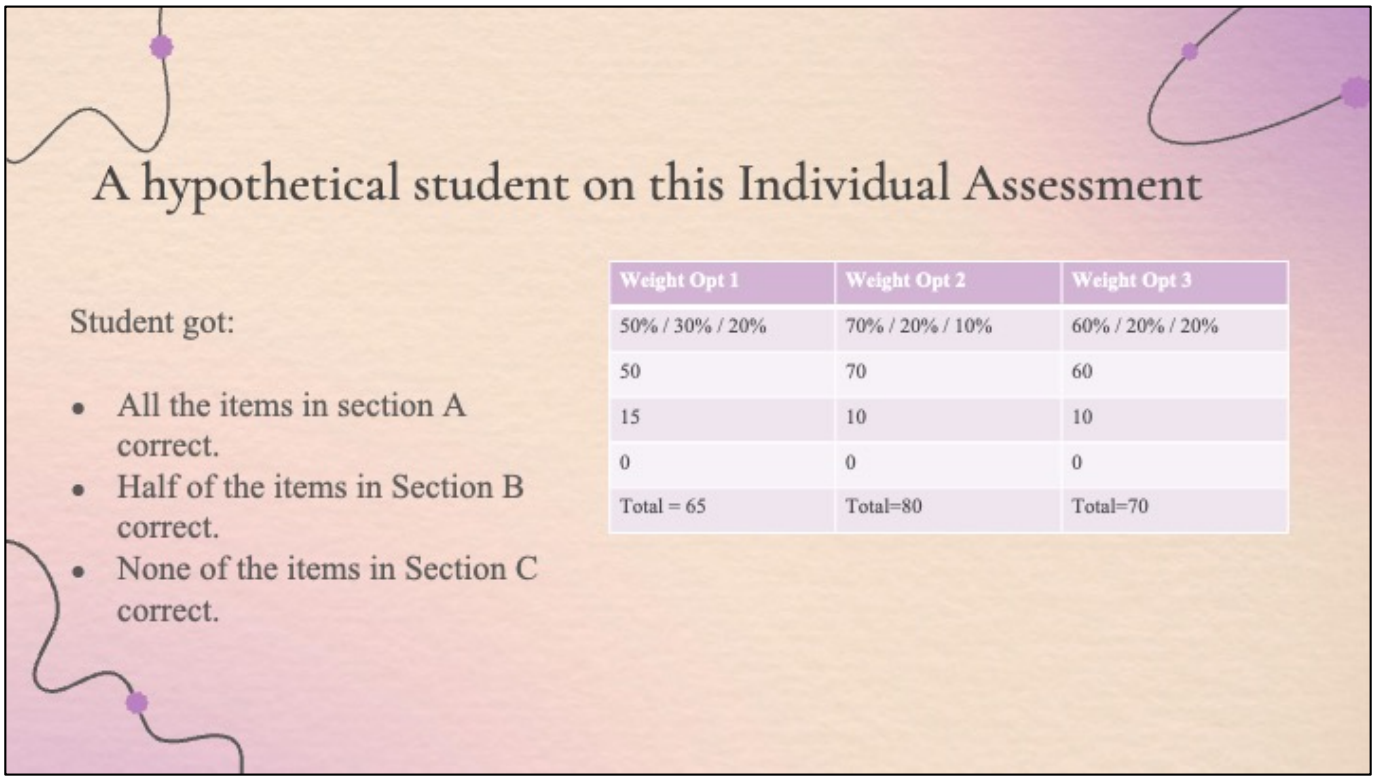




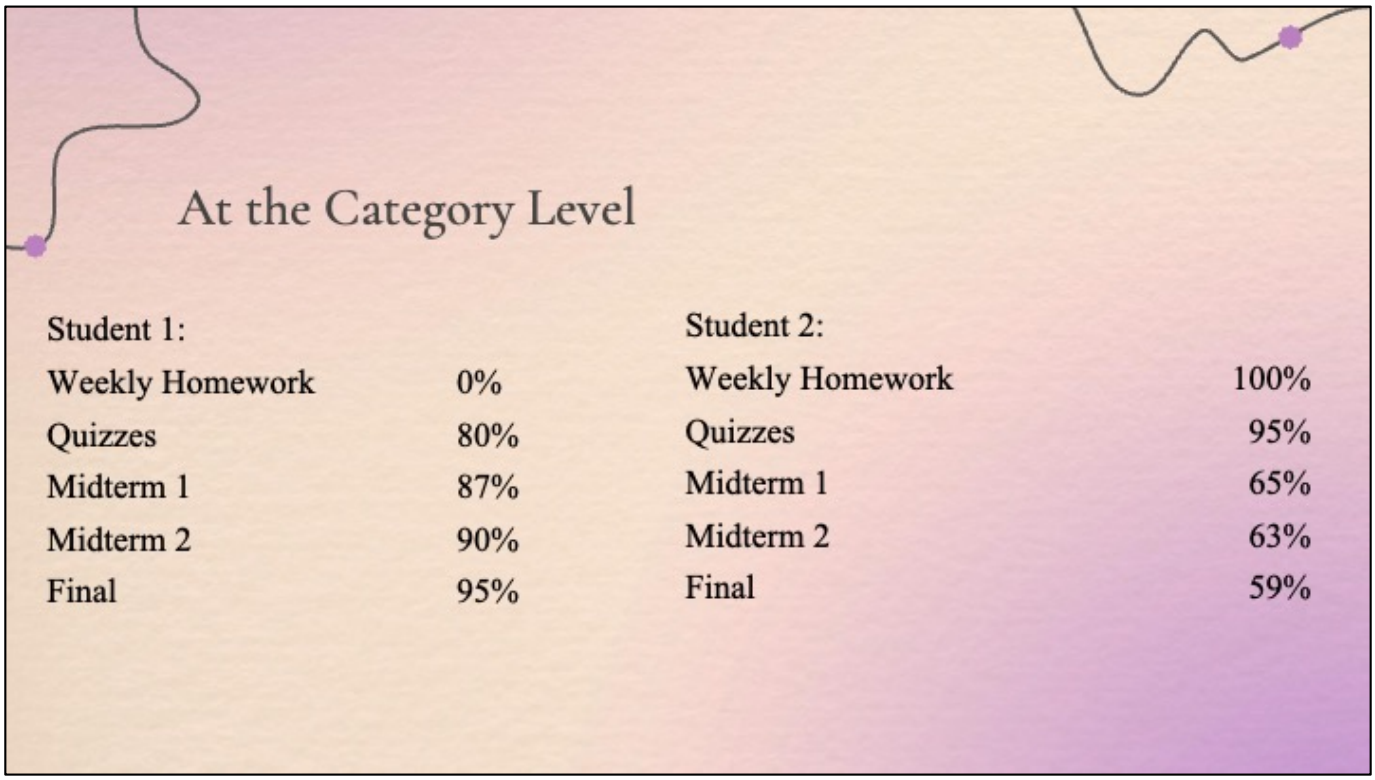
- Section header



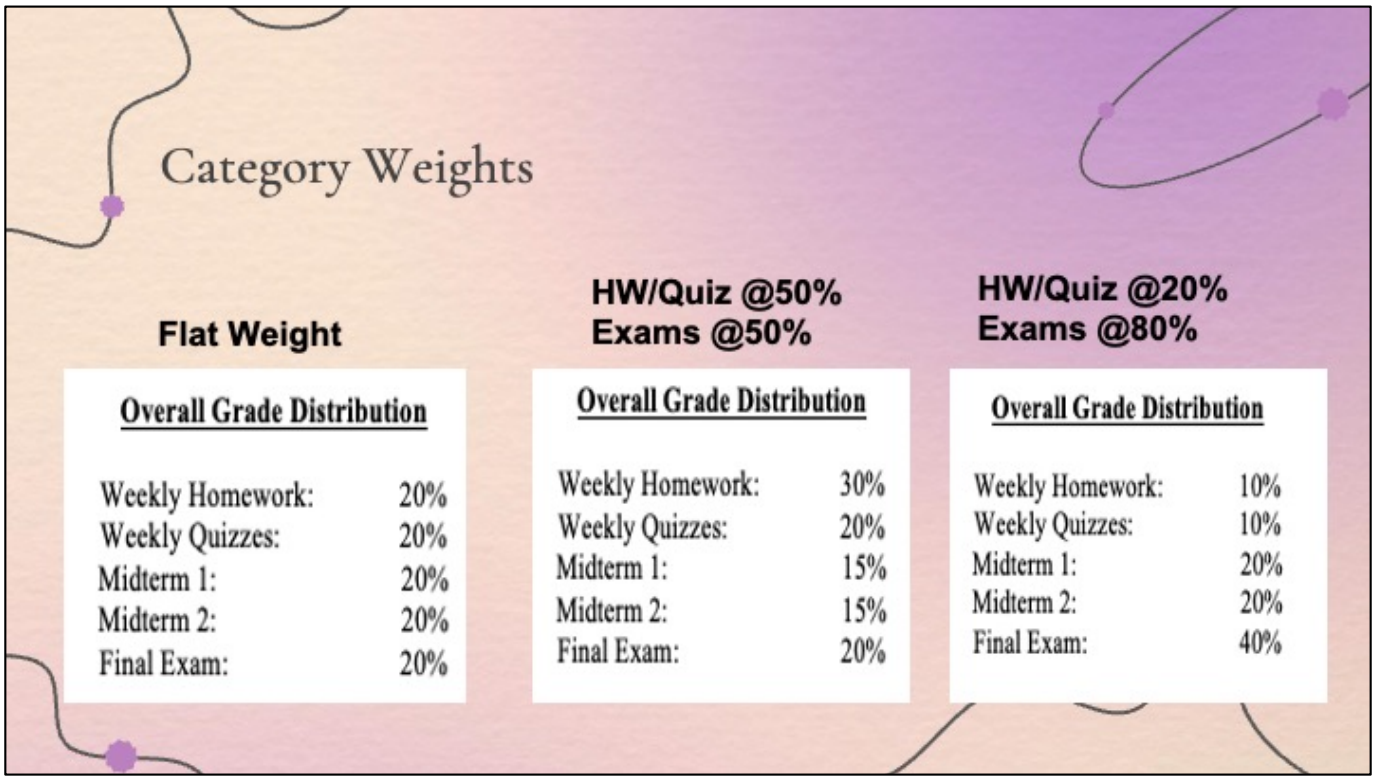
- Interactive Exercise for the participant assign a number of "points" for each of A through C, total should be 100 – you decide how many points to give each "group".
- In a typical workshop of 30 or more participants, a few main "methods":
    1. Flat – all three categories get the same weight
    2. Explicitly taught simple gets more weight
    3. Explicitly taught complex gets more weight



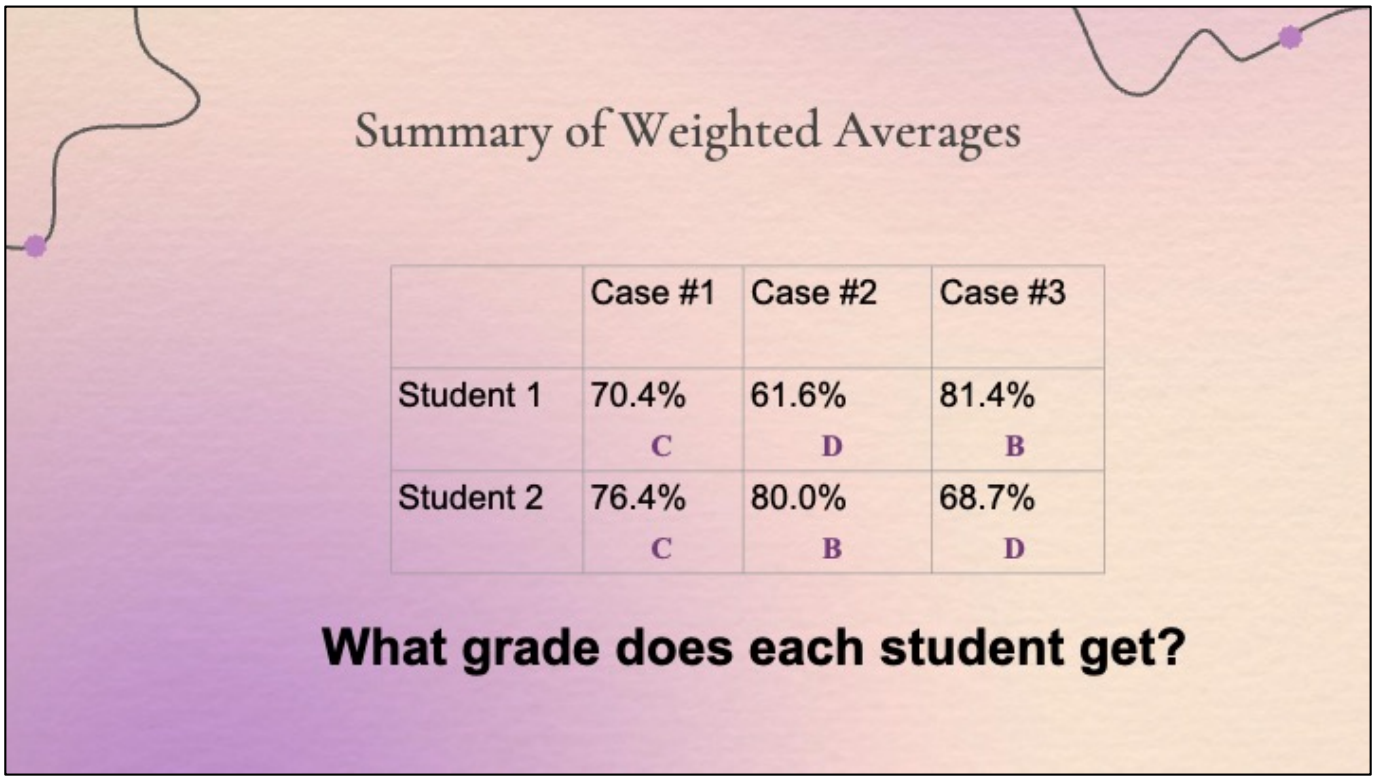
- Give the student a total score based on what you decided
- Then here are some common ones (might not be exactly what you gave)
- Assuming the "usual" 100 point scale – anything from a D to a B
    - Workshop participants will typically have between 20 and 90 on a 100-point scale
- Every person in the room will likely have a different answer/ component or explanation. Students have to deal with all of this across 3-6 classes per term



- pParticipants make a note or take a moment to narratively describe student 1 vs student 2

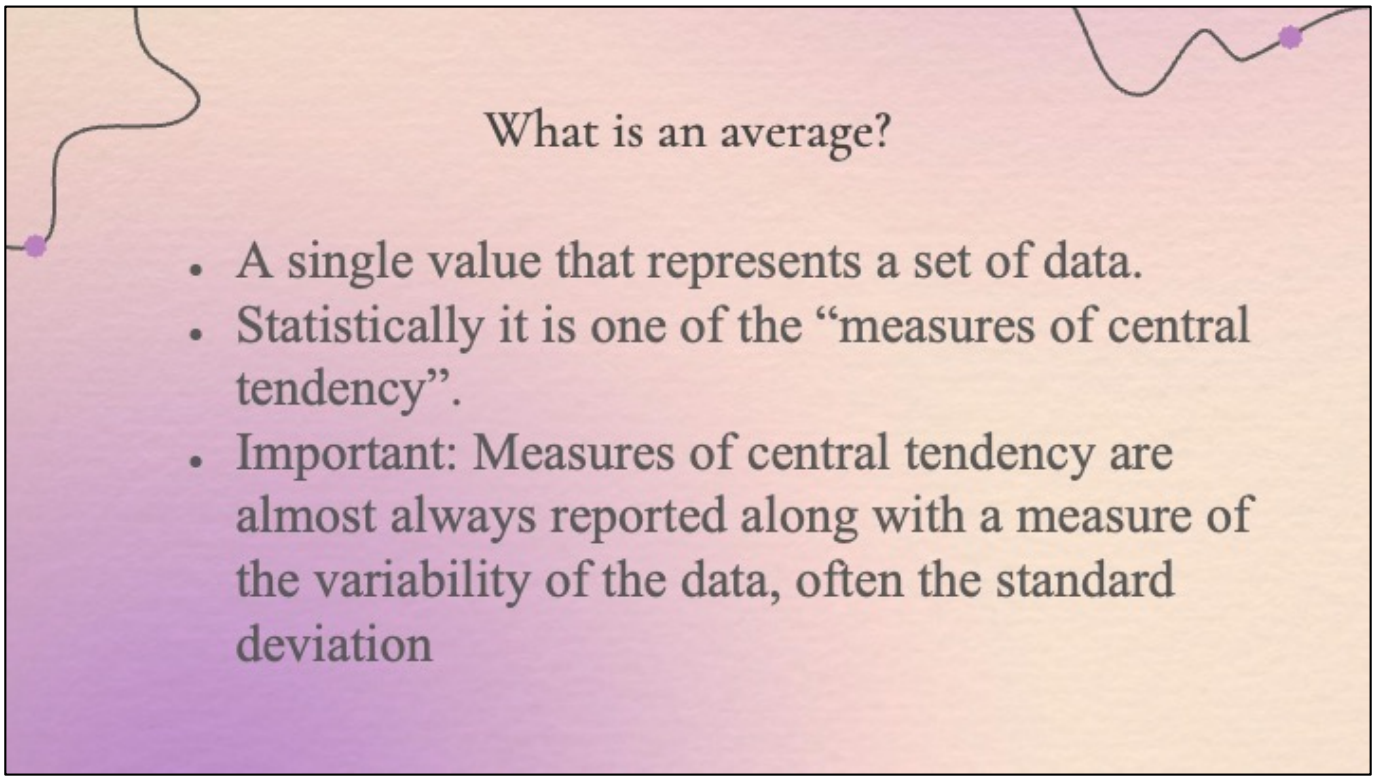


- Three common weight breakdowns between categories. Typical representations in STEM related fields, looking at Homework and Quizzes vs “Exams”.
- Note – there are infinitely many variations on this topic.



Using the "usual" 100-point scale - the same work gets wildly varying final grades based on the hypothetical weighting presented.

This is presuming 90-100 A, 80-89.99 B, 70-70.99 C, 60-69.99 D, <60 F



- The way we calculate averages in grading we usually mean the *weighted arithmetic mean.*
- A central tendency is a central or typical value for a probability distribution. Also, the tendency of quantitative data to cluster around some central value.

The Use of Averages

Which Student Would You Choose To Pack Your Parachute?

	1st Try	2nd Try	3rd Try	4th Try	5th Try	6th Try	7th Try
Student A	95	40	80	55	90	50	80
Student B	40	55	50	80	80	90	95
Student C	95	80	90	80	50	55	40
Mastery	60	60	60	60	60	60	60

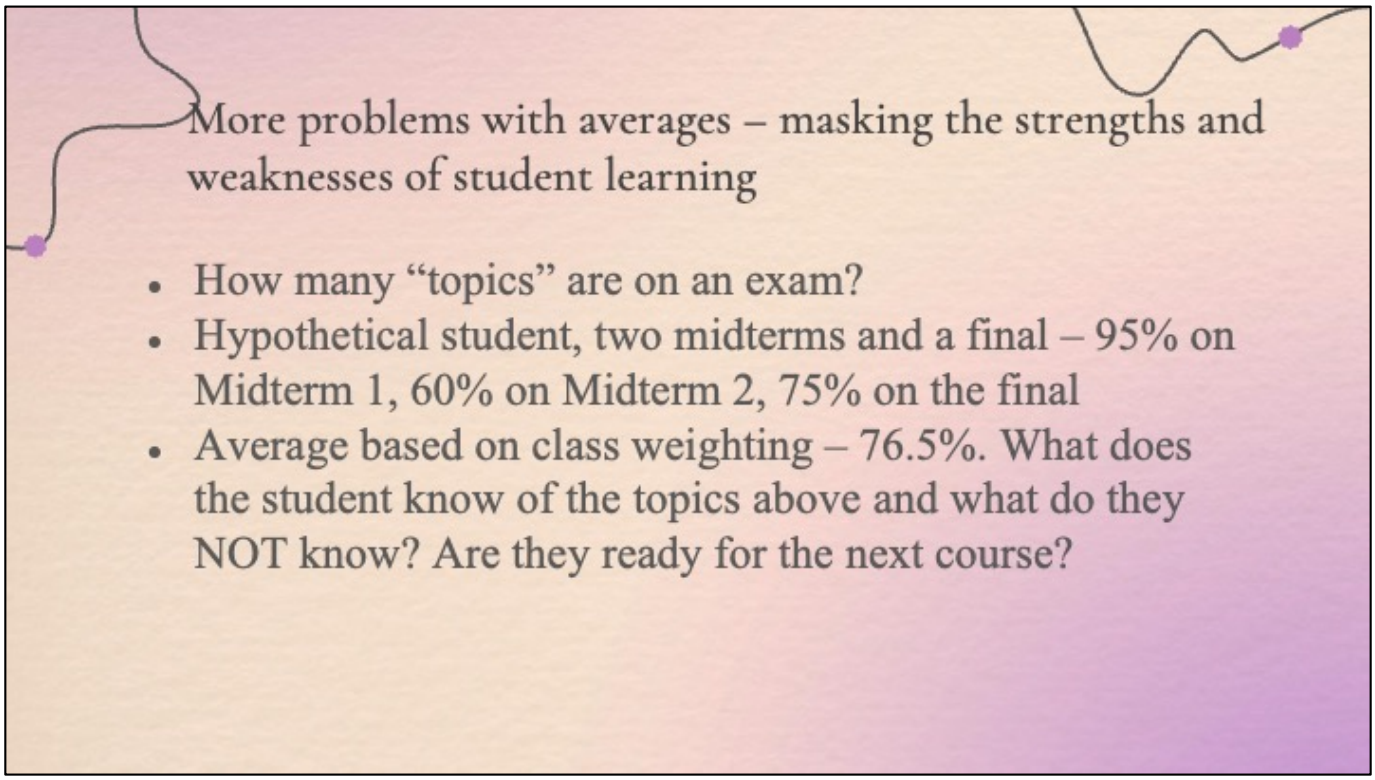
Adapted from *How to Grade for Learning, 3rd Edition* (O'Connor, 2009)

Student A: Inconsistent

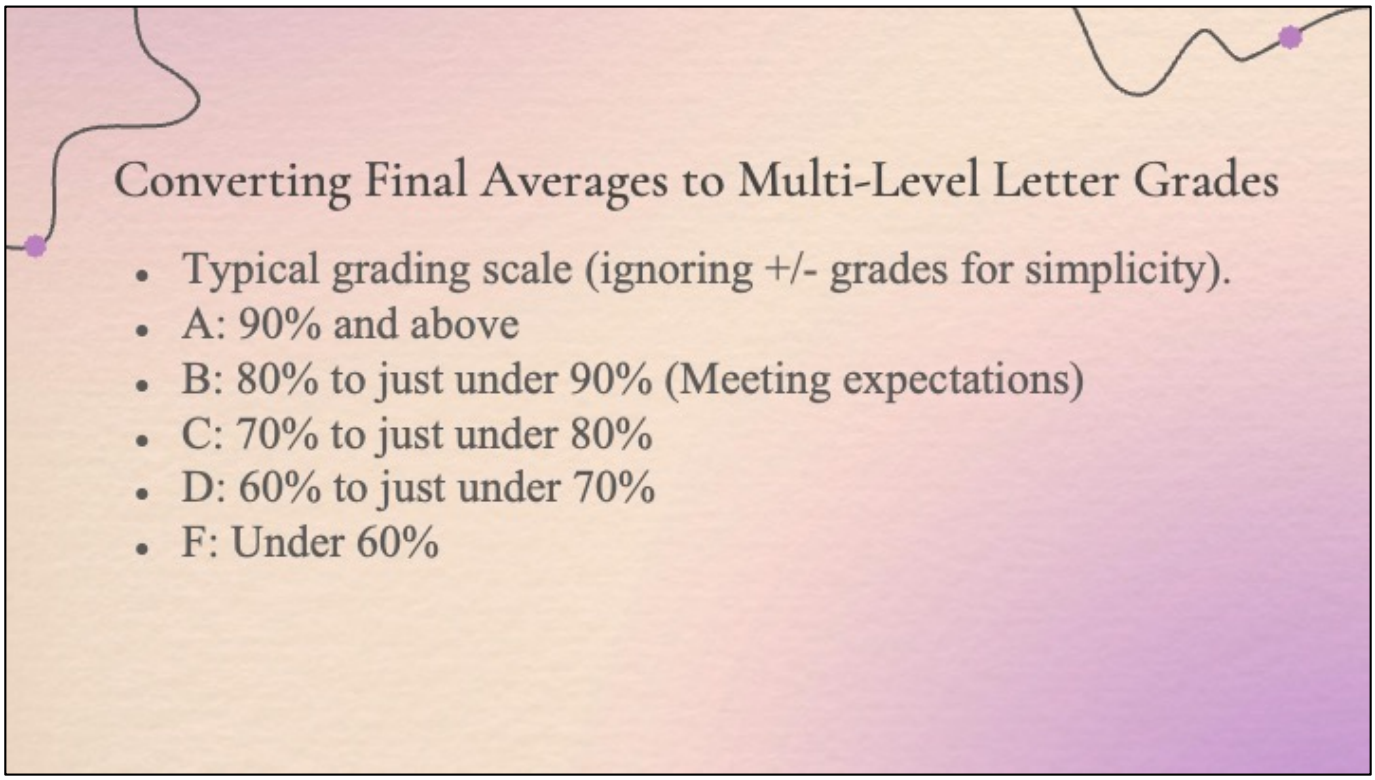
Student B: Consistent Growth

Student C: Consistent Deterioration

- How averages hide the story – all three students average out to 70 mathematically despite having distinct, and potential important, differences in their learning and abilities at the end of the course.



- For example, Calculus 2 has three major areas (typically)
  - Advanced Integration Techniques
  - Sequences and Series
  - Motion in Space
- How much of each of these topics does a student know if they have a "76.5%' in the course at a specific time?



- Failing grades occupy three times the area of the 100-point scale and the range dedicated to meeting standards (60% for failing, 20% for "passing).

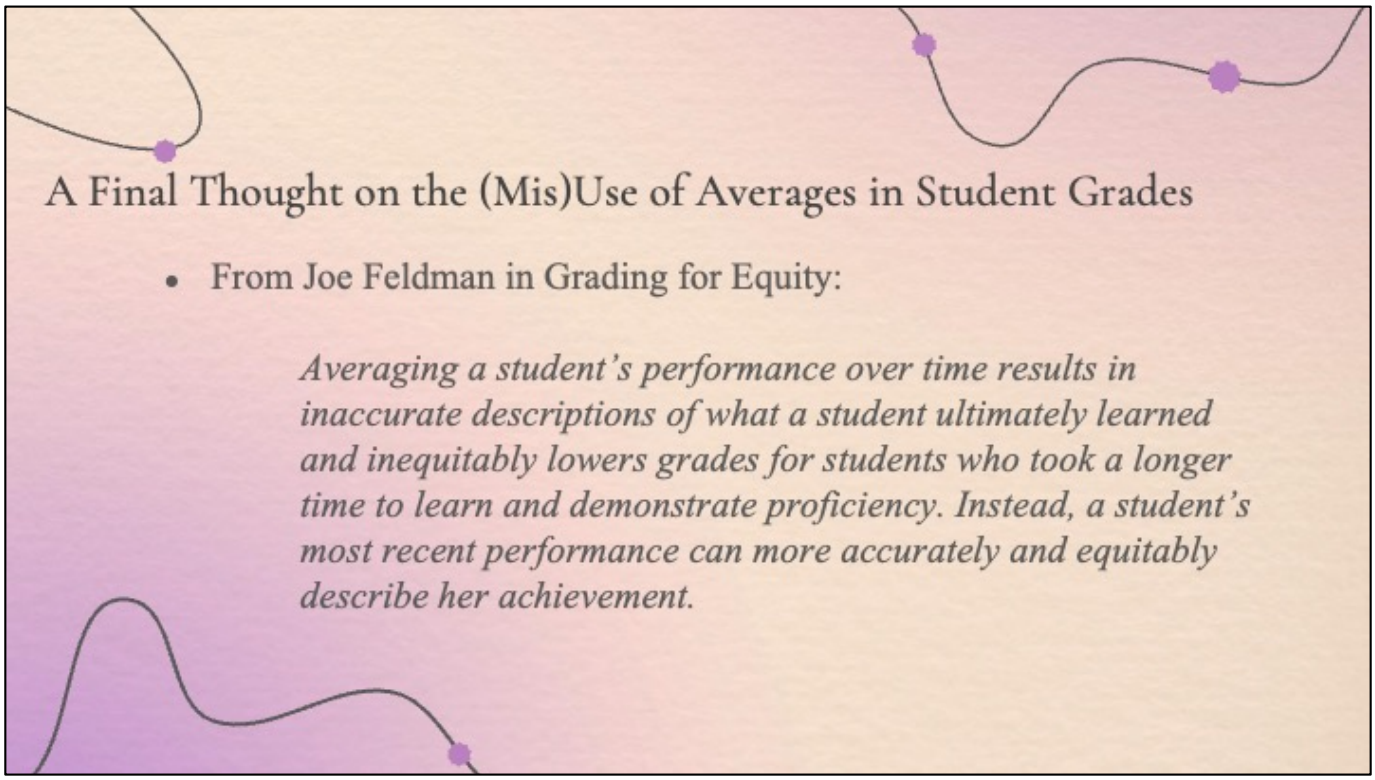
- https://eric.ed.gov/?id=EJ1022559 – The Arguments and Data in Favor of Minimum Grading

    - "In our retrospective study, we found no evidence of grade inflation or social promotion of those students receiving minimum grades in a large urban High School using a school-wide "macro" minimum grading system over a seven-year period. We also found most of the benefits posited by minimum grading theory, including students who received minimum grades doing significantly better on state exams than would be predicted by the overall GPA's with the opposite results being true for the other students in this High School."

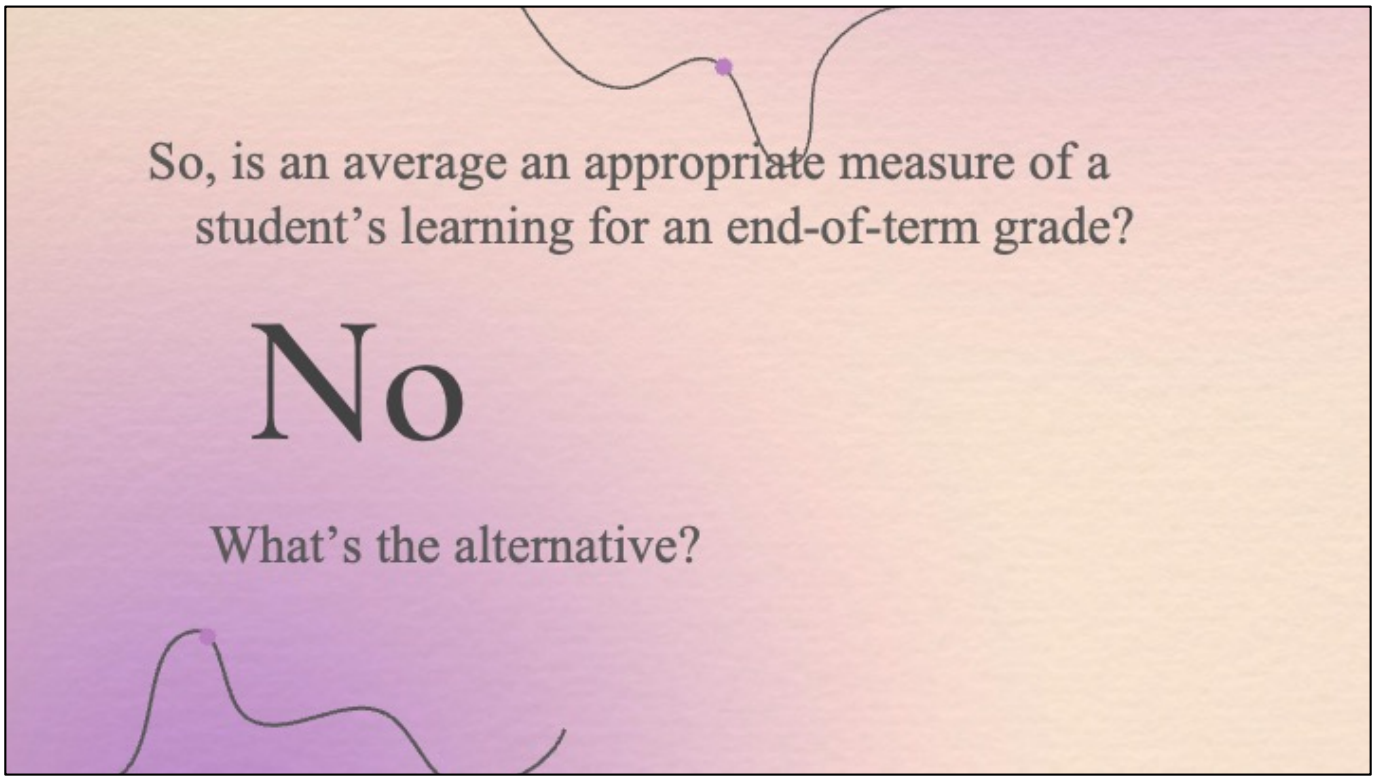
- https://www.mwera.org/MWER/volumes/v25/issue4/v25n4-Carifio-Carey-POINT-COUNTERPOINT-SECTION.pdf

    - "Grading schemes instituted at a time when only a few advantaged students were either expected or allowed to advance to the higher levels of learning cannot adequately serve the increasingly diverse population of students found in today's schools"

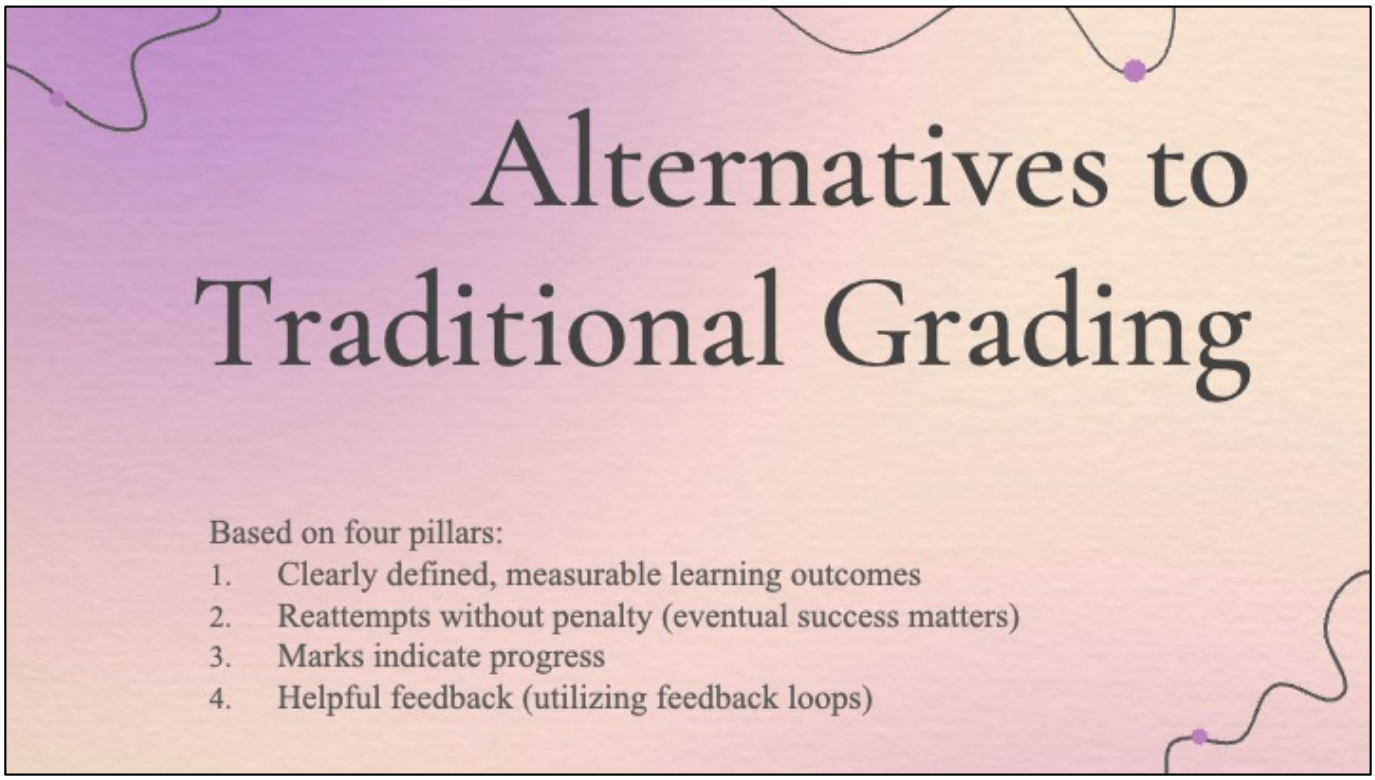
- Note – the entire country of New Zealand uses a 50% minimum in schools.



- Feldman, Joe. *Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms*. Corwin, 2018.



- We have shown that the math is inconsistent and often misrepresentative. it measures a lot of things but is relatively useless as a measure of and communication tool of a student's actual learning.
- So is there an option that is NOT traditional points-and-percentages grading that does a better job of measuring and communicating student learning?



Clearly defined, measurable learning outcomes: Three key components:

1. The learning outcomes can be externally measured. I.e. a student has to take an action (describe, define, solve, write, explain, etc,) that can be measured against a clear set of expectations.
2. The learning outcome is something a student does. I.e. I (the student) can solve a problem, explain a technique, write a thesis statement, identify the plot, etc.
3. The learning outcome should have the right amount of detail to guide students as to what "success" looks like without being prescriptive.

Reattempts without penalty: Students have multiple opportunities to demonstrate success on a learning outcome without earlier, unsuccessful attempts counting in their final grade.

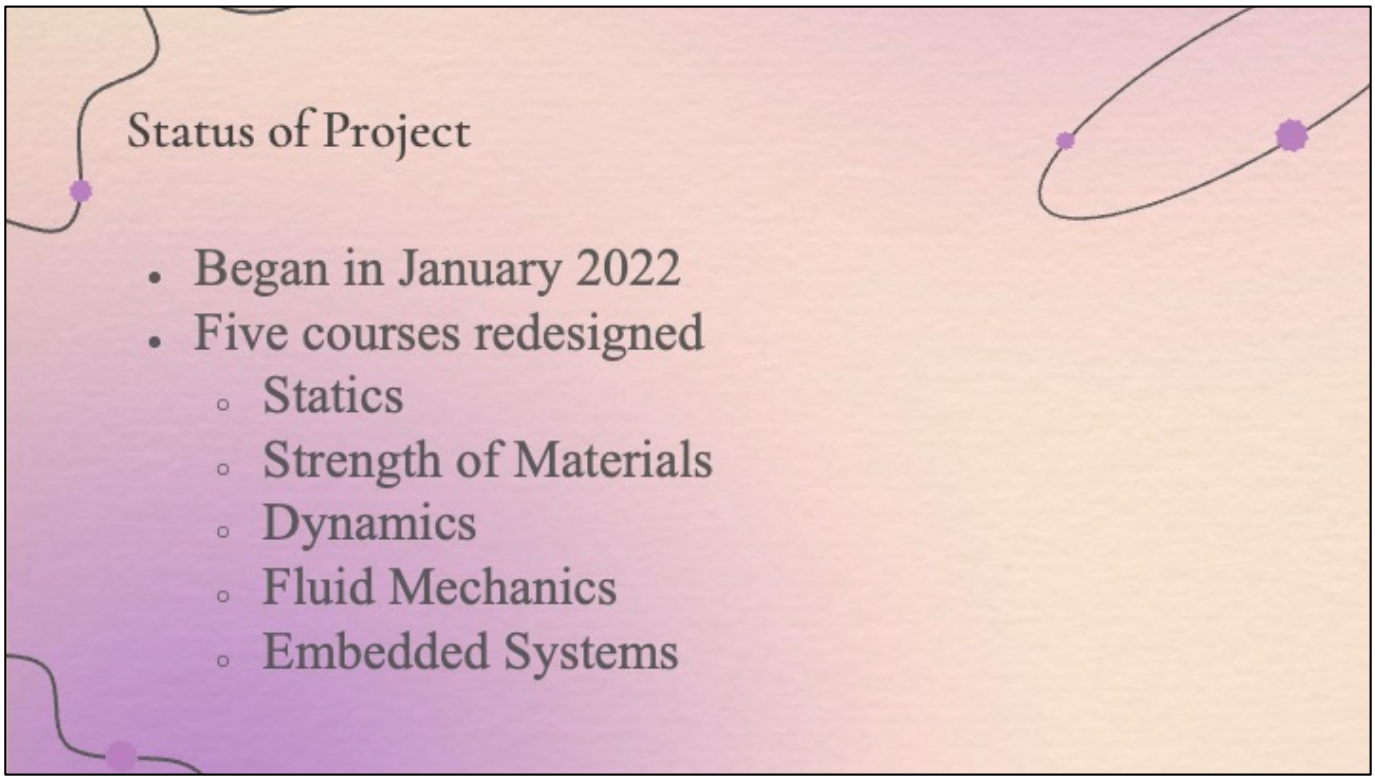
Marks indicate progress: Instructor feedback is focused on success on the learning outcomes and what a student still needs to demonstrate (if applicable).

Helpful Feedback: Feedback is goal oriented (based on the learning outcomes), actionable (something a student needs to DO to move forward), timely (enough time to review feedback and learn from it before attempting it again), and specific (identifying specific things not at a sufficient level for success and scaffolding options to improve.

- Clark, D., & Talbert, R., et al. *Grading for Growth: A Guide to Alternative Grading Practices That Promote Authentic Learning and Student Engagement in Higher Education.* Routledge, 2023.
- Nilson, L. & Stanny, C.j. . *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*. Stylus Publishing, 2015.

- Blum, S., et al. *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead)*. West Virginia University Press, 2020.
- Stommel, Jesse. *Undoing the Grade: Why We Grade, and How to Stop.* Hybrid Pedagogy, Inc., 2023.

- Currently in year 3.



- Statics, Strength of Materials, Dynamics, Embedded Systems, and Fluid Mechanics have been redesigned at two institutions:
    - California State University Los Angeles
    - Pasadena City College (feeder community college to Cal State LA)

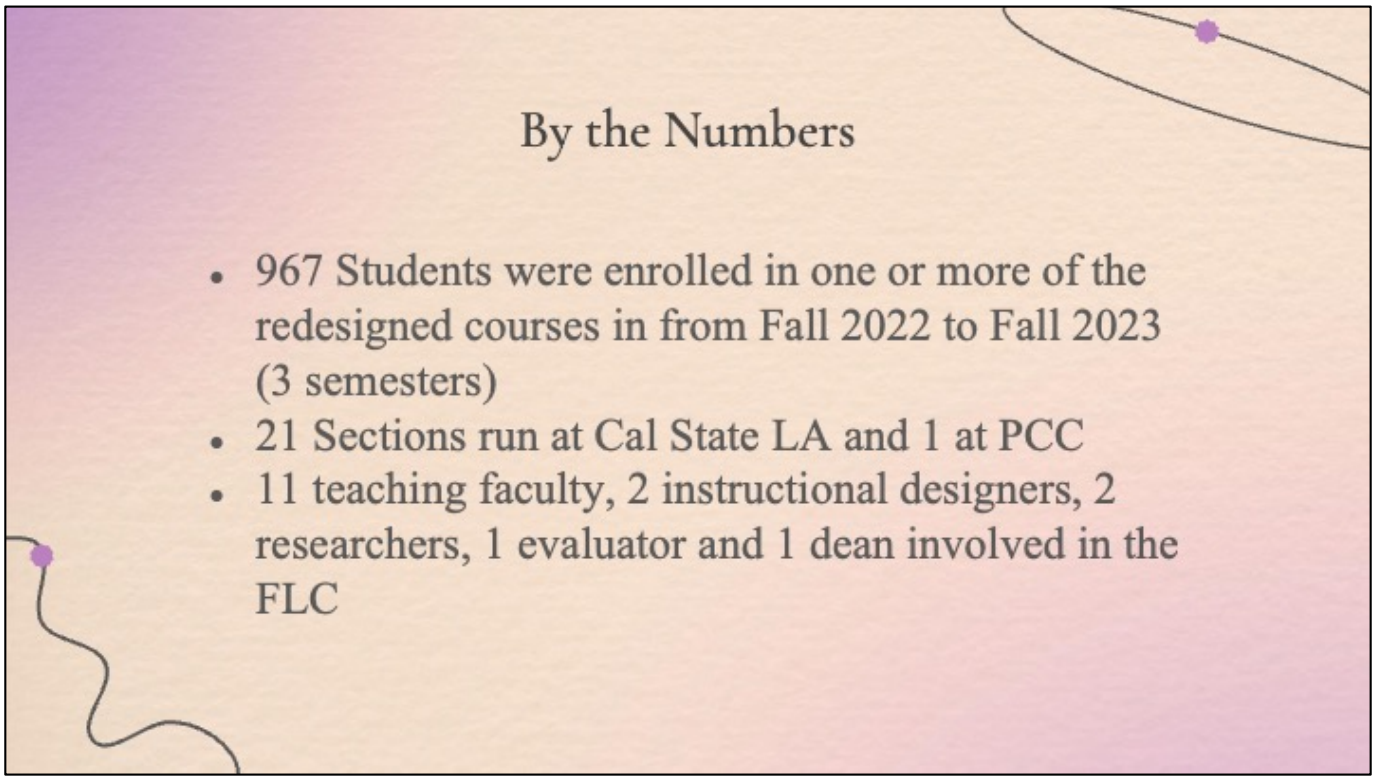
- 9 original designing instructors, 2 new implementing instructors (on-boarding after the redesign)
- Redesigned courses have just completed their third semester of implementation
- Utilizes instructional student assistants to support student learning and support instructor feedback and grading

- Initial faculty training for designing instructors: 4-day intensive zoom-based, synchronous training (approximately 6 hours per day)
- Followed by a nine-month, team-based redesign process. 2-3 instructors per course team
- Course teams met bi-weekly to:
    - Develop learning outcomes
    - Decide on common grading architecture choices including:
        - What does "success" on a learning outcome look like?
        - How will evidence of learning be gathered?
        - How much evidence is needed?
        - How will a final grade be determined?

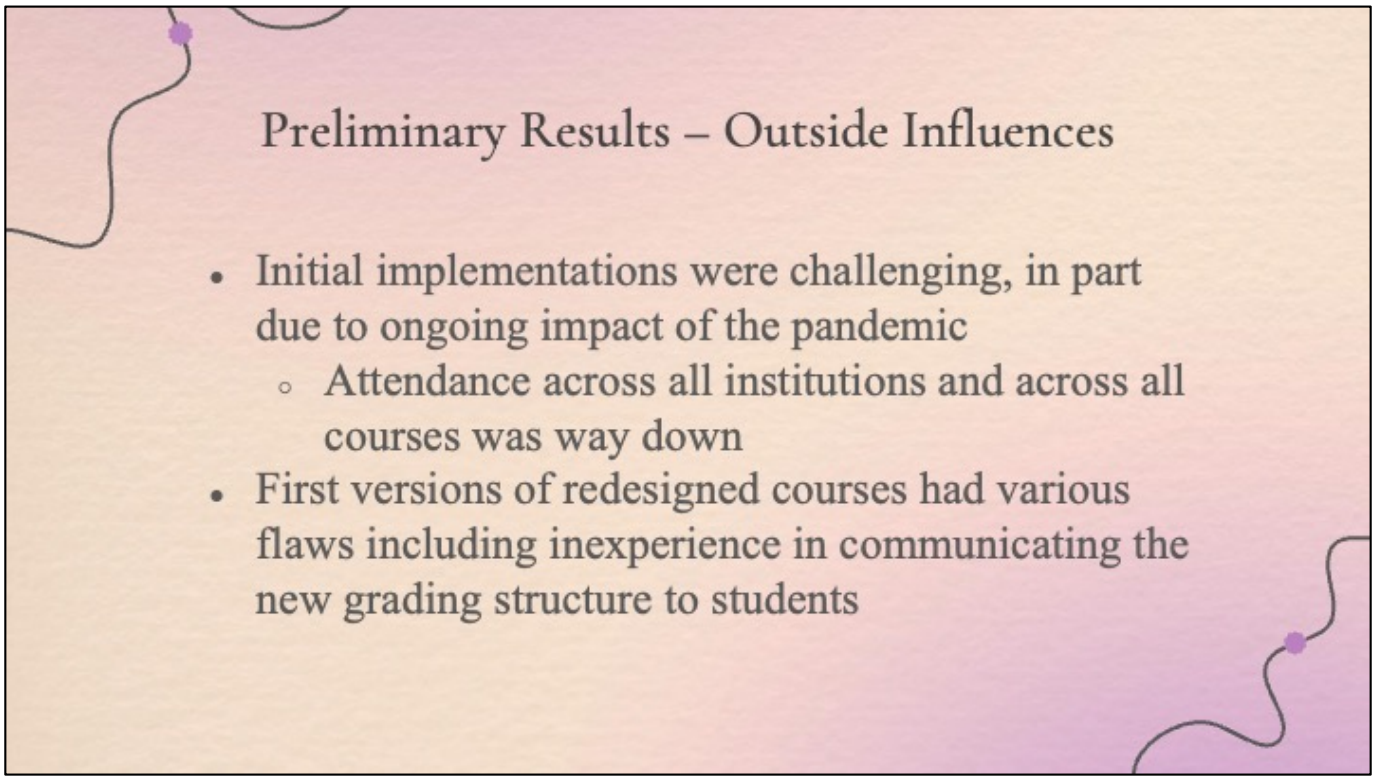
- Adopting faculty training: 2-day intensive, zoom-based. Included designing faculty AND adopting faculty

- Partner student research project with Dr. Dina Verdin, Arizona State University, to explore the impact of "mastery-based grading" on the growth mindset and Engineering

identity of students in the redesigned courses.



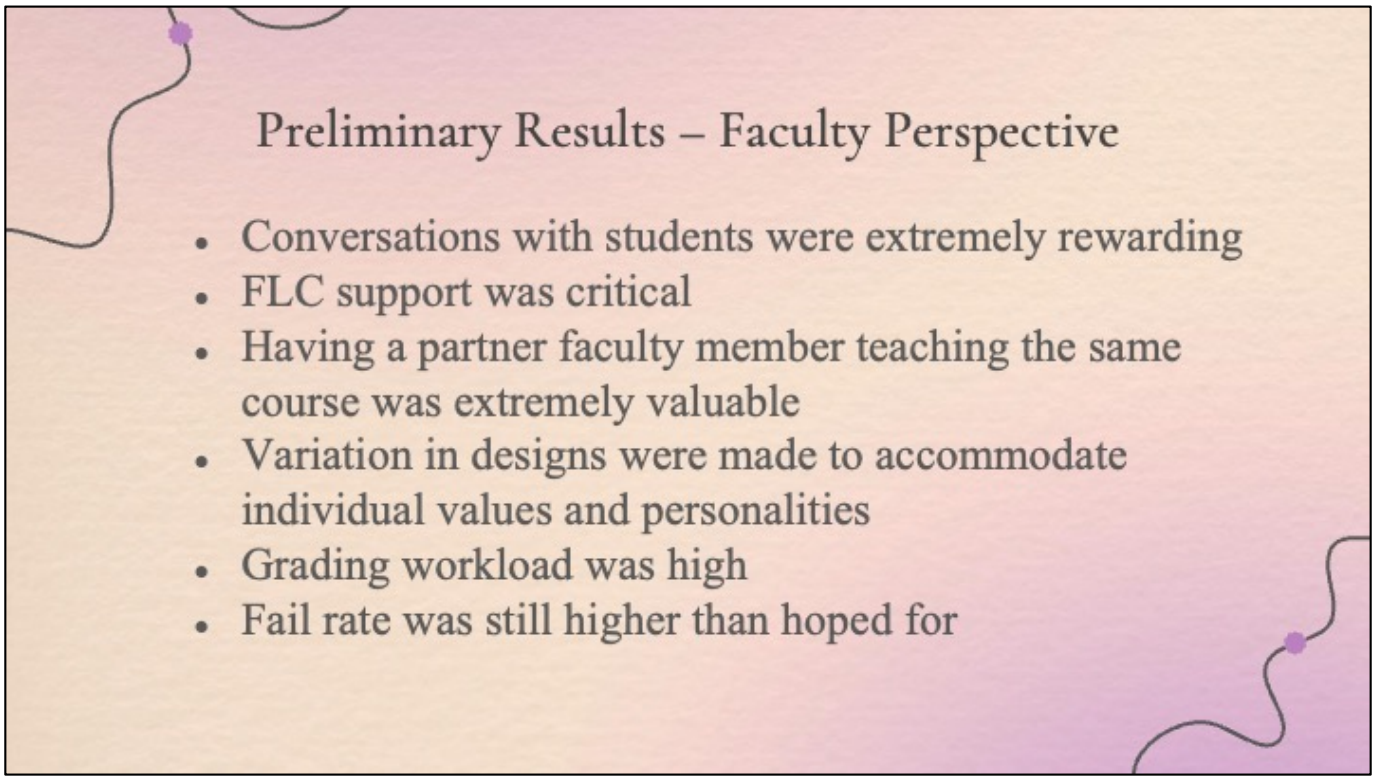
In Fall 2022 and Spring 2023, most students enrolled in one of the topics were in a redesigned course. In Fall 2023, students were split between redesigned versions of the courses and non-redesigned versions (taught by instructors who did not participate in the grant and therefore had not redesigned their courses or adopted a course that had been redesigned.



A core component of the CLIMB grant was the creation and support of a faculty learning community (FLC) for the faculty redesigning the courses and then teaching the redesigned courses. The FLC met approximately monthly during the redesign process with individual course teams meeting more frequently. Once the redesigned courses launched, the FLC tried various meeting frequencies incuding monthly and biweekly. By Fall 2023, the FLC meetings were held more sporadically, with more focus on individual ongoing instructor support by the faculty facilitators.

Interviews of all members of the faculty learning community were conducted twice during year 2 of the project. Seven faculty were interviewed at the end of the Fall 2022 semester and nine faculty members (seven returning and two new) were interviewed at the end of the Spring 2023 semester. Key findings included:

- Faculty and students struggled to re-engage after COVID and a prolonged absence from campus and face-to-face instruction.
- Most team members reported low and attendance and flagging engagement, a pattern that was not limited to the redesigned courses but rather was reported consistently across Cal State LA.
- Attempts at communicating the new grading system that gave the false impression that students could “wait” to engage with the feedback loops. Many students skipped early attempts at demonstrating learning of the learning outcomes, thereby not receiving critical feedback.



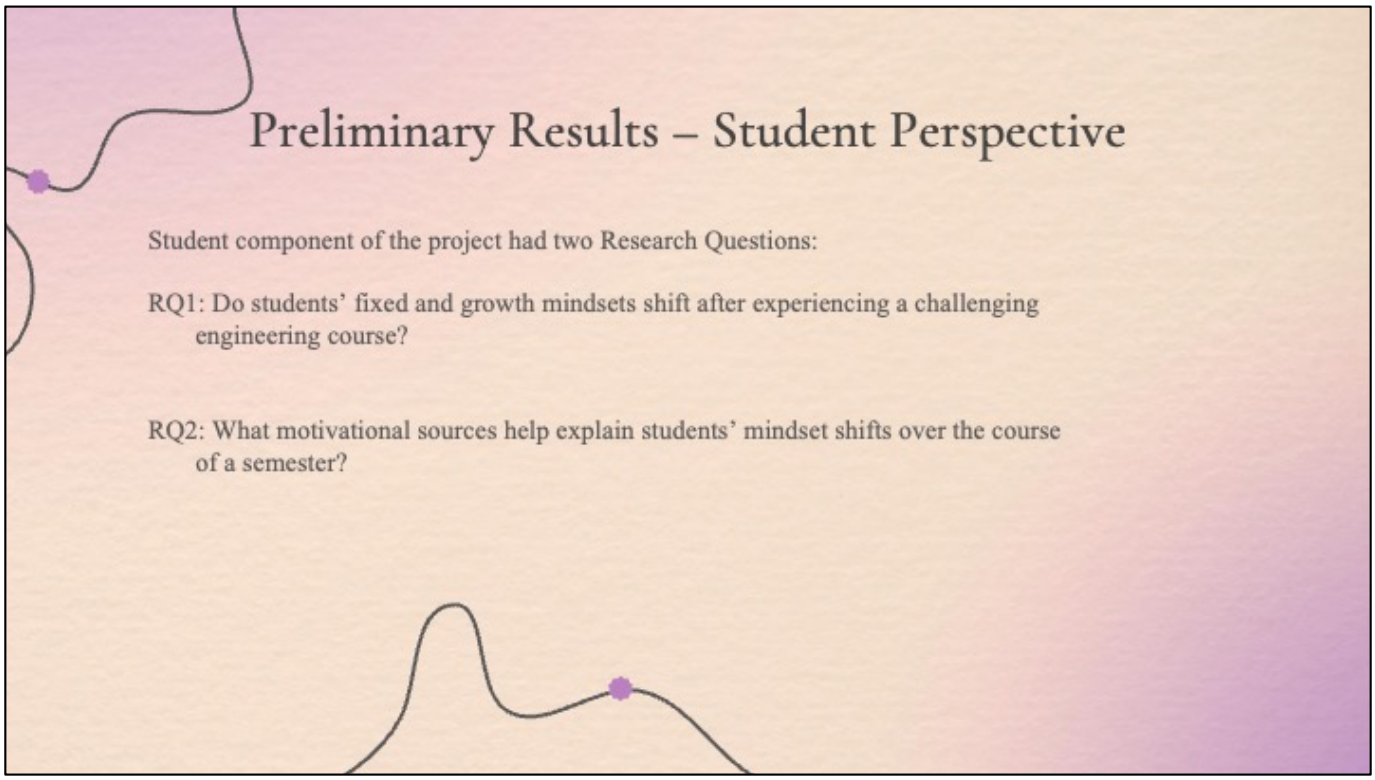
Key findings from interviews continued:

Some key initial errors in implementation included:

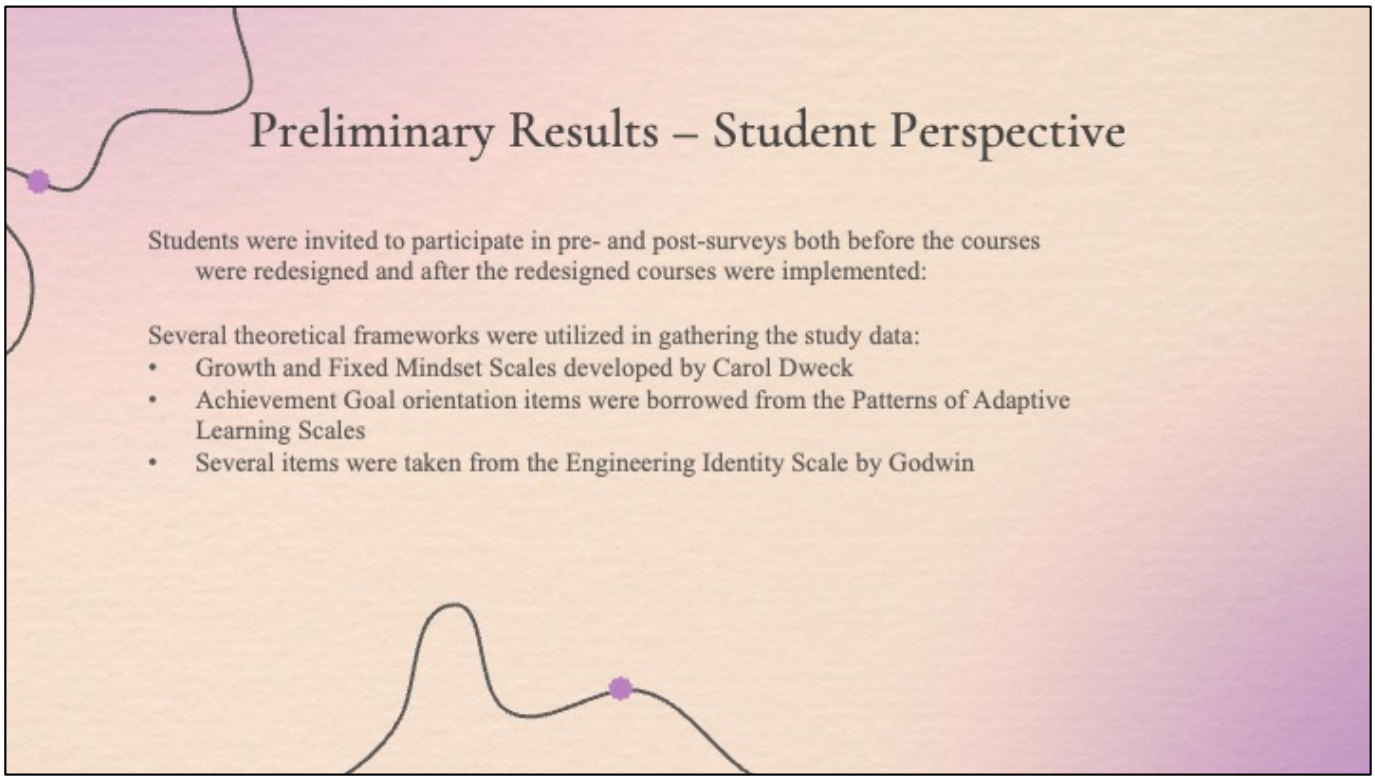
- Workload varied dramatically depending on implementation and the number of students in a course.
- Adding too many additional attempts to demonstrate mastery of the learning outcomes, resulting in faculty overload.
- Too many/too few learning outcomes. Faculty struggled with finding the right balance between smaller, more focused learning outcomes, which increased the number of learning outcomes used in the course, and broader, more wide-ranging learning outcomes, which reduced the number of outcomes but increased the difficulty for students knowing what to learn in order to be successful on the learning outcomes.
- Calendaring and workload became an issue for both faculty and students.

Key positives from the interviews:

- Starting with a framework already developed got the semester off to a good start.
- Use of instructional student assistants for instructional support to students and grading feedback was critical and extremely valuable.
- Coaching from experiences facilitators and opportunities to work with a team was appreciated.
- Despite concerns and challenges, faculty indicated that they would ”not go back” to traditional grading due to increased student engagement and learning by students.



- This is the data from our own project – shows connection between mastery grading, mastery approach using achievement goal theory and growth mindset. Based on pre- and post-survey data.
- Data from Year 1 of the grant available at: https://peer.asee.org/examining-engineering-students-shift-in-mindsets-over-the-course-of-a-semester-a-longitudinal-study

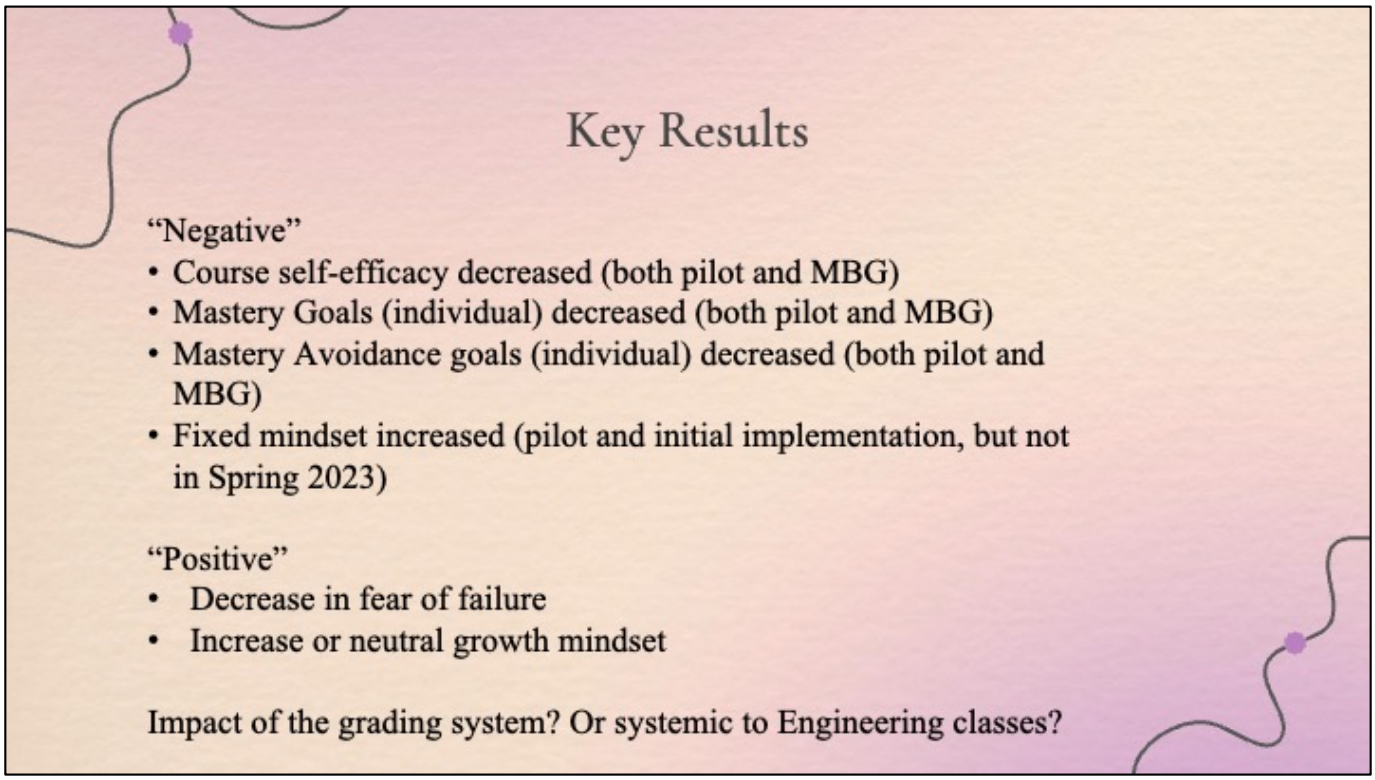


Sources:

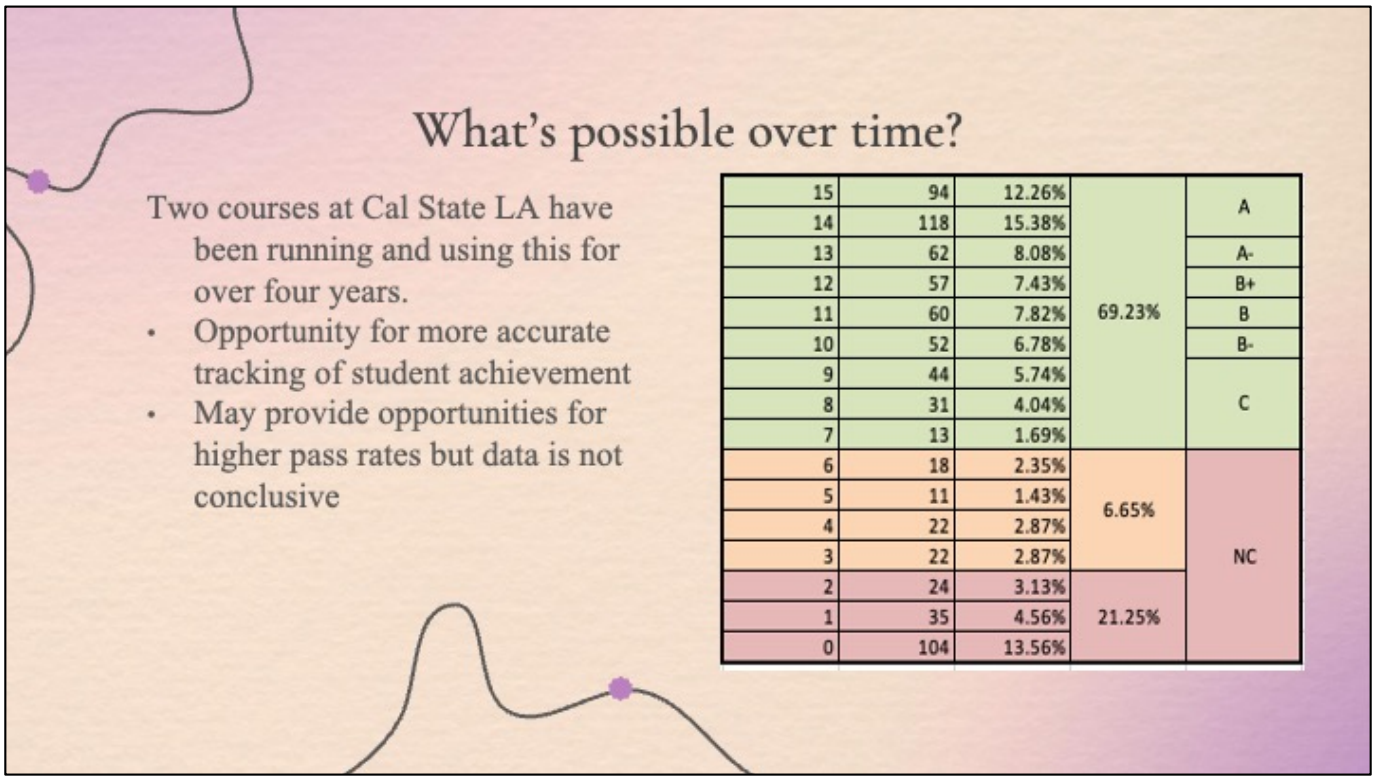
- C. S. Dweck, *Mindset: The New Psychology of Success*. Penguin Random House, 2006.
- C. S. Dweck, *Mindset: Changing the way you think to fulfil your potential.* Robinson,
- 2017. [Online]. Available: https://www.ptonline.com/articles/how-to-get-better-mfiresults
- C. Midgley et al., *"Manual for the Patterns of Adaptive Learning Scales."* pp. 734–763, 2000.
- A. Godwin, *"The development of a measure of engineering identity,"* ASEE Annu. Conf. Expo. Conf. Proc., vol. 2016-June, 2016, doi: 10.18260/p.26122.

Previous publications from this grant.

- https://peer.asee.org/examining-engineering-students-shift-in-mindsets-over-the-course-of-a-semester-a-longitudinal-study
- https://asu.elsevierpure.com/en/publications/mastery-learning-in-undergraduate-engineering-courses-a-systemati



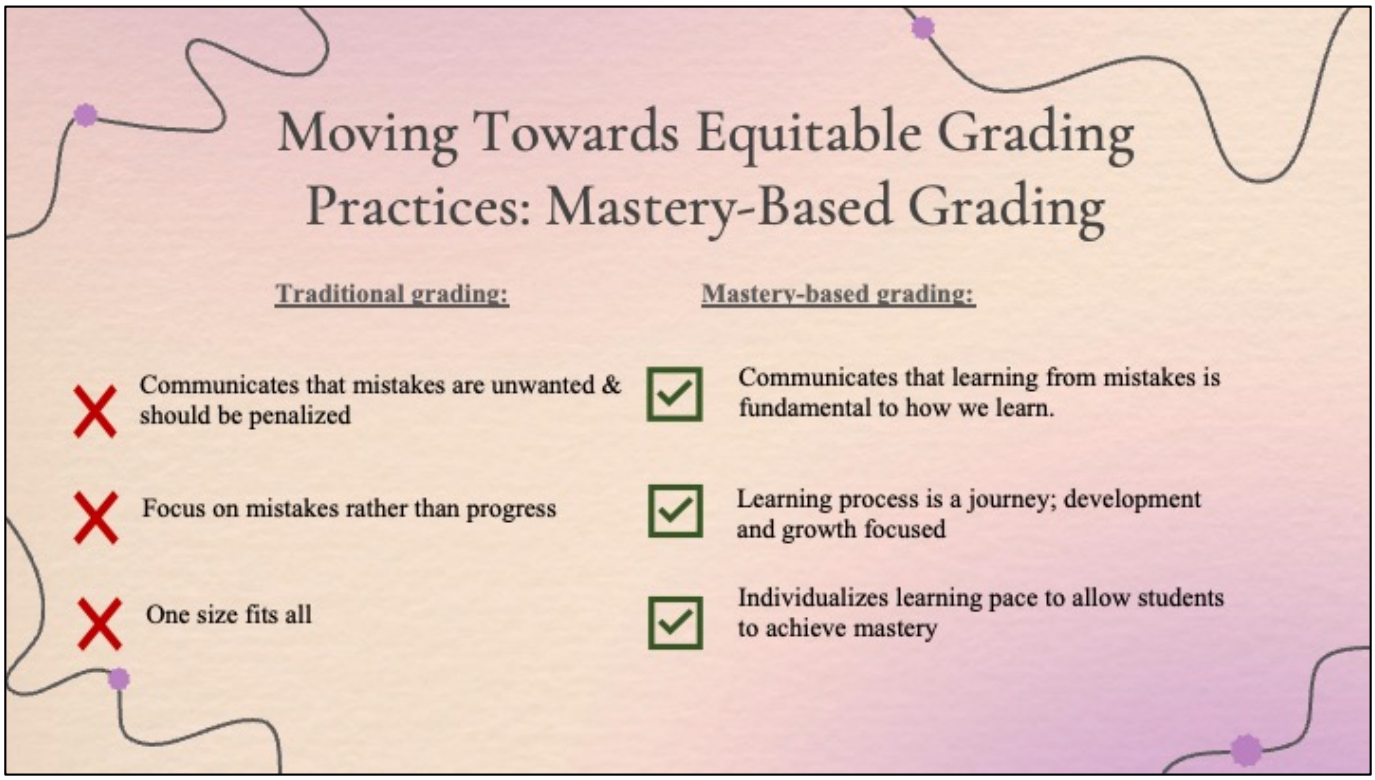
- Results given are all statistically significant using a variety of statistical tests, t-test, multi-level modeling. Level 1 variables included classroom mastery goal structure, classroom performance goal structure, classroom performance avoidance goal structure, external recognition. Level 2 variables included Final Course Grade and Course type.
- Difficulties:
    - Getting sufficient post-course surveys from students has been challenging
    - Variations in implementation prevent sufficient sample sizes for subdivision



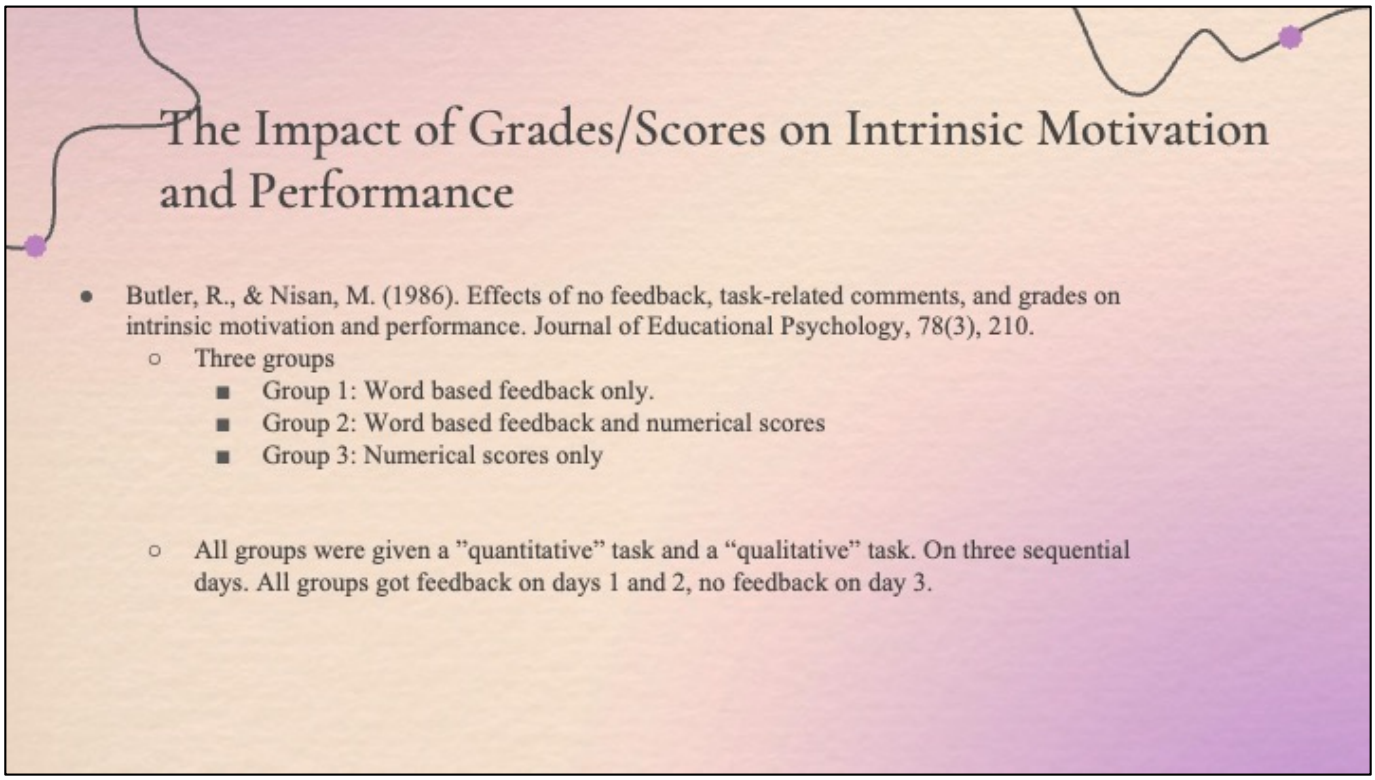
- Data shown is Fall 2023 – 767students across 32 sections of a coordinated Quantitative Reasoning with Statistics course.

- Each grade level is associated with a specific number of learning outcomes completed. Data is available also showing WHICH outcomes were completed by each student.

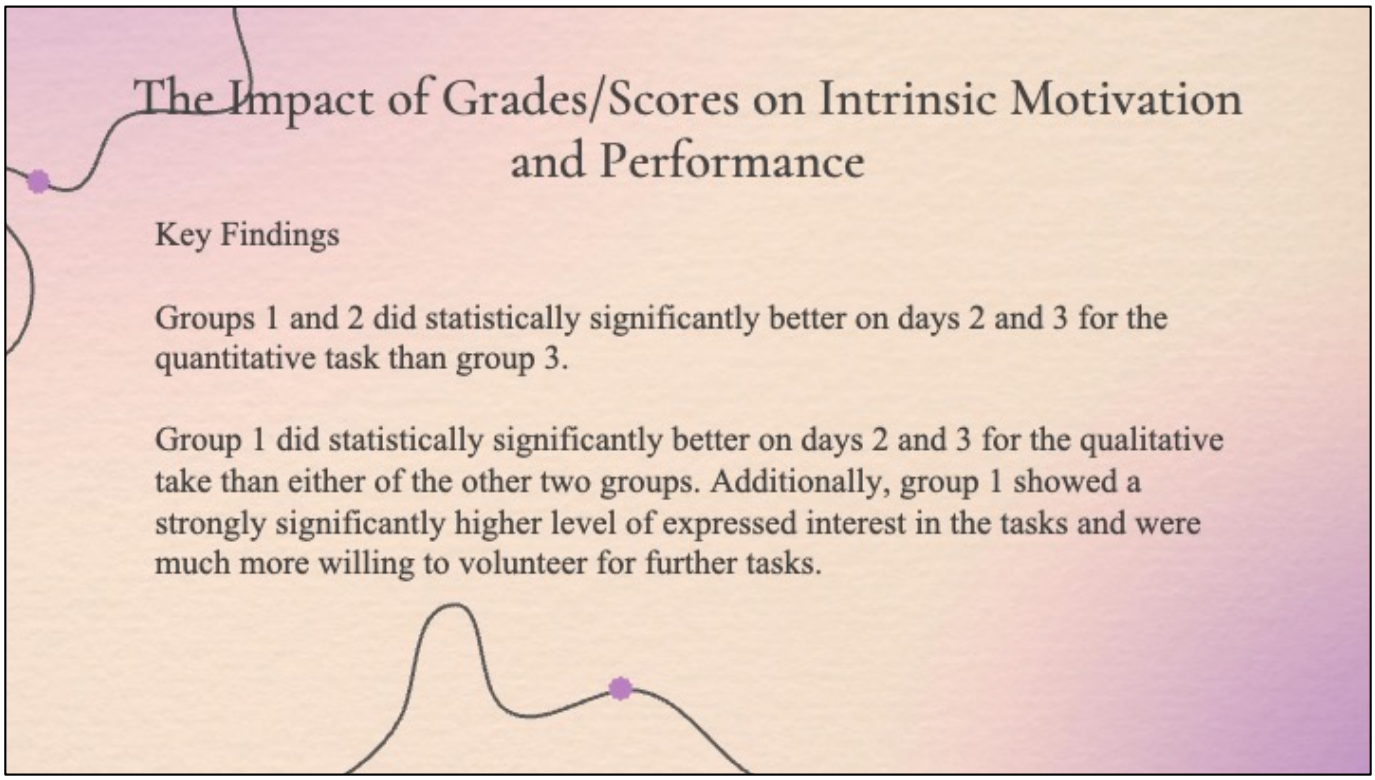
- Since Fall 2018 over 8,600 students have taken this course, with over 70% passing it, the vast majority with a B or better,



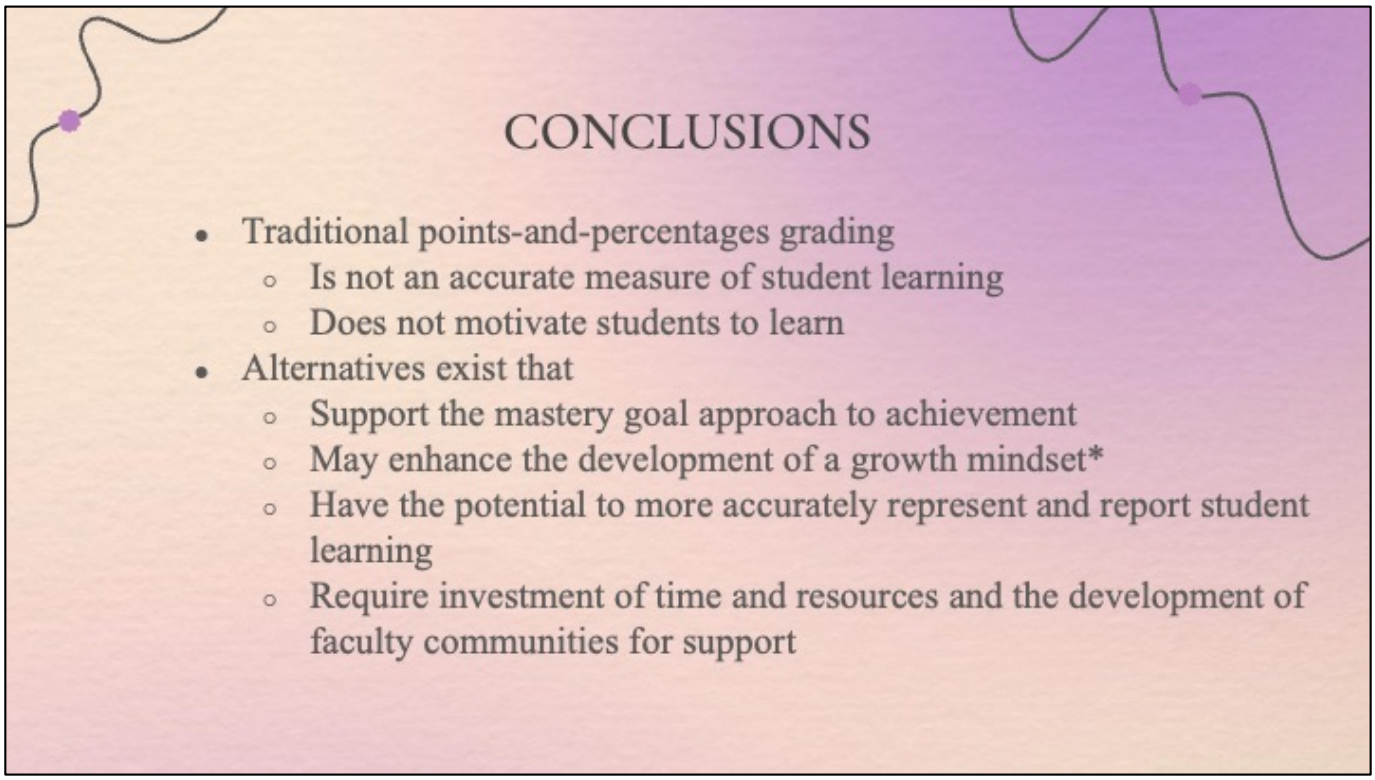
- Alternative Grading Practices are not necessarily automatically more equitable. Any grading system can be poorly implemented and be used to harm students. However, intentional use of alternative grading practices opens up the opportunity for more equitable practices such as those listed.



- More details about the impact of feedback alone versus feedback with numerical scores.
- https://psycnet.apa.org/doiLanding?doi=10.1037%2F0022-0663.78.3.210
- Other publications include Black, 2013, Butler 2008, Kanfer and Ackerman 1989, Kluger and DeNisi 1996, Black, Harrison, Lee, Marshall and William 2004



- https://psycnet.apa.org/doiLanding?doi=10.1037%2F0022-0663.78.3.210



* Key area of current research

Thank you – Please contact us!!

Sharona Krinsky
@southbaysharona on Twitter, Bluesky, Instagram

skrinsk@calstatela.edu
Sharona.Krinsky@thegradingconference.com

Robert Bosley
@Bosley74Bosley on Twitter

rbosley@calstatela.edu
Robert.Bosley@thegradingconference.com