WHITE PAPERS

JumpRope

Calculation Methods

Take an in-depth look at the seven calculation methods available in our gradebook, learn how they work, and discover the advantages and challenges of each.

Contents

Calculation Methods

One of the trickiest aspects of mastery-based grading is the way that scores are calculated in the system. There are pros and cons to each approach, so it's a matter of finding the proper balance between the philosophical and the practical. JumpRope supports seven different strategies/algorithms, and it's up to the teacher, school, or district to choose.

CALCULATING SCORES FOR STANDARDS

The most important numbers in JumpRope's gradebook are the standard-level calculations per student. These are based on one or more assessments, and answer that most-important question: Given all of the evidence, what level of mastery has this student demonstrated?

This white paper looks at the seven calculation methods available in our gradebook and discusses how they work on a technical level, as well as the advantages and challenges associated with each one.

1. Most Recent

This calculation type carries the most recent assessment score achieved, based on the date assigned to the assessment (rather than the date the score was entered). In other words, as a teacher gathers additional data from classroom activities and assessments, the proficiency level of the standard constantly updates to reflect the most recent—and only the most recent—evidence available.

Advantages with the Most Recent calculation

- It is very simple to understand how it works. This goes for teachers as well as students and parents. Don't underestimate this advantage—some of the other options are powerful but extremely complex mathematically or conceptually.
- It does a good job of representing "current mastery" assuming that the most recent assessment is at least as (if not more) rigorous than the last.
- It's easy to control. If you feel that a student needs another opportunity to show mastery (because the most recent data point is not accurately representing their work), simply put in another score for the student and the mastery on the standard will update right away.
- You can see it grow over time. As you add more evidence, student scores will grow substantially over time, versus other calculation types that will see less "movement" in their scores as new scores come in.

Challenges with the Most Recent calculation

- Bad day syndrome. If a student has a bad day on the day of the test, none of the evidence leading up to the test tempers this.

  POSSIBLE SOLUTION Provide additional opportunities to show mastery for students that didn't do well. Let them re-take the test or some version of it, for example.

- The multiple modalities problem. In many cases, you may wish to provide opportunities for students to show mastery in different ways, such as written vs. spoken. The Most Recent calculation type will take the most recent score (and last alphabetically, actually, in the case that two assessments have the same date).

POSSIBLE SOLUTION If you create your assessment as "Final Paper and Performance" instead of having two different assessments, you can essentially do the "combining the modalities" step offline and enter the combined score into JumpRope for that target. Doing so loses some data specificity, but it will represent mastery on the standard more accurately.

ANOTHER POSSIBLE SOLUTION One way to approach this problem is to write the standards (if you're able to create your own) in JumpRope in such a way that they are specific to a certain modality. In other words, you can split the standard into two standards and score them separately if the modality is very important to represent.

- Students "gaming the system." Once the student figures out that only the most recent assessment counts, many teachers immediately worry about losing the motivational aspect of the grade and fear that students will intentionally sandbag early work (or simply be lazy).

COMMENT 1 Consider the student in this scenario and what it means about the student, the standard, and the assessment(s) that a student is able to "slack off" early and still show mastery. Might this be evidence more of a student's work habits than their academic ability?

COMMENT 2 Just about every grading system ever is "gamed" by kids once they figure it out. So while this may be "easier to game" in some senses, it is not a problem that is unique to the Most Recent calculation type.

COMMENT 3 In our experience, it's rare for students to substantially change their existing behavior based on this new calculation type. Students who are engaged in their learning rarely ever suddenly stop turning things in just because only the most recent score represents their mastery.

2. Power Law

This calculation type is based on research on cognitive development. It is a time-based average, and automatically adjusts assessment weights to give higher weight to the more recent assessments. The result is a mastery score on the standard that represents the entire body of evidence but considers the trend and more recent evidence instead of simply averaging the scores. The algorithm for this method was developed by a research team examining thousands of pieces of data on student cognitive development.

The Power Law is perhaps best understood by example:

	ASSESSMENT 1	ASSESSMENT 2	ASSESSMENT 3	ASSESSMENT 4	POWER LAW SCORE
STUDENT 1	1.00	2.00	3.00	4.00	4.00
STUDENT 2	1.00	3.00	2.00	4.00	3.66
STUDENT 3	2.00	4.00	1.00	3.00	2.16
STUDENT 4	4.00	3.00	2.00	1.00	1.28

Fig. 1 Power Law examples

Advantages of the Power Law calculation

- It responds to growth and does not penalize students for low scores early in the learning process. At the same time, it is not as "harsh" as the Most Recent calculation and incorporates consistency and trends as opposed to just the most recent score.
- It's backed by research. JumpRope didn't make up Power Law—rather, we implement an algorithm that is based on research into the cognitive aspects of learning. For much more detail, check out Marzano's book: Transforming Classroom Grading.
- It is generally intuitive. If we were to ask you to look across all student scores for each standard every time you added an assessment and come up with a number to represent their "current" level of mastery, most teachers would generally agree with the Power Law. It arrives at the same/similar outcome with a lot less work on your part. This doesn't apply to every teacher in every situation, but JumpRope has spoken with many teachers at different grade levels and content areas and to a large extent there is intuitive agreement.

Challenges of the Power Law calculation

- It is difficult to explain (for example, to parents and students). Basically, it's an enormously-complex algorithm and it's exceedingly difficult to reproduce the mathematics behind it to figure out exactly where the number came from.

  SOLUTION 1 Use the chart above to explain the Power Law intuitively. We do not recommend trying to use the algorithm in most cases as it tends to cause added confusion. In our experience, if you are able to have the conversation objectively (e.g., without the emotion surrounding a student's specific scores), most students and parents will agree that student 1 and student 4 "deserve" different overall scores. Once you've made that point, you explain that if the scores were averaged then every student would have a 2.5, so you use an equation that accounts for their growth.

  SOLUTION 2 We strongly recommend explaining it to students and parents before they first see their "grades." If you find yourself explaining it for the first time with the student's own grade in front of them, the conversation is often too emotionally charged to make logical arguments.

- There are some quirks. Certain scenarios can lead to some counterintuitive scores emerging from the Power Law. A classic example of this is the situation where a student scores a 3, 3, 3, followed by a 4 would get a 3.6 in the chart above; this is actually a lower score than a student that receives a 1, 2, 3, followed by a 4. Why? The Power Law has less evidence of growth and thus statistically responds with a value that represents the "lack of confidence" in the student's ability to acquire the knowledge or skill.

  SOLUTION 1 In some cases, it may make sense to override this score. Most of JumpRope's gradebook setups allow you to override student scores (depending on the settings at your school). Just make sure that this is the exception and not the rule.

  SOLUTION 2 If early assessments are formative, consider representing that in JumpRope's gradebook. In other words, did that first score truly mean that the student "met the standard" vs. "did what they were supposed to do?" The Power Law gives us the incredible opportunity to shift our feedback towards standards, and this is actually an example of how giving students lower scores on formative assessments (assuming they truly are at a lower level of mastery) actually *boosts* the student's final score on the standard instead of punishing them—which makes sense in the context of mastery-based grading.

- It does not account for teacher-assigned assessment weights. This is the most difficult bit for many teachers: You don't have control over the weight of each assignment towards the final grade. In essence, you can think of the Power Law as automatically assigning weights to the assessments based on the order in which they were given (that's a bit of an oversimplification, but it captures the essence of it). The reality is that the Power Law was simply not designed to work with teacher weights and for us to support weights would mean modifying it and losing the "research-based" aspect. The Decaying Average calculation is our attempt to build in the construct of growth-based calculations while also accounting for weights, but it does not work with the Power Law.

  SOLUTION 1 Assess often. Effectively, the number of times you assess something becomes your new technique to "weigh" things. Consider, for example, scoring each draft of a paper separately or a test re-take as an additional assessment in JumpRope's gradebook instead of editing the existing score. These strategies work well with the Power Law and effectively make the paper and the test (in these examples) "worth more" towards the overall standard mastery.

SOLUTION 2 Use the special weight of 0.01. If you give an assessment a weight of 0.01, it has a special property with regards to the Power Law: It will "count" like a normal assessment until you score at least one assessment with a weight of 1 or higher, at which time it will still show up on reports but will be completely thrown out of the Power Law calculation. It's a trick that we built in to let you score formative assessments (weight = 0.01) during the learning process and then to later give a summative assessment (weight >= 1) that "trumps" the formative. It will still spit out a score early on, but the summative assessments will be the only scores that count in the end (and it works with multiple weight >=1 assessments). Be sure to note that an assessment with a weight of 0 (zero) will not count at all in the Power Law either. These are the only two exceptions to the weight being ignored.

SOLUTION 3 Give up some control and let it do its thing. We have found that attempts to exert hyper-specific control over student overall grades where Power Law is being used can be an exercise in frustration. It works best when the focus is shifted from the overall grade toward a focus on student mastery of specific skills. It's also useful to remember that all grading systems have quirks and weaknesses.

There are a couple of special cases to be aware of

- If you have fewer than three scores on an assessment, the Power Law doesn't work mathematically. JumpRope's gradebook just shows the most recent score if you have two scores for a target, and shows the only score if you've only entered one.
- You need to avoid having two assessments on the same date. If both occur on the same date, only one will be factored into the score because the gradebook cannot determine the order in which they were given. Consider simply fudging the system to move the date on one of the assessments and/or pre-combining the scores (think entering an overall "presentation and paper" score instead of an individual presentation and individual paper score with the same due date).

3. Weighted Average

This calculation type takes an average of all assessments for each student on each learning target. If teachers assign a weight to an assessment, this will incorporate the appropriate weight when calculating the average. This method is the easiest to understand, since it is close to traditional grading systems.

Advantages of the Weighted Average calculation

- It's easy to understand because it matches well with other gradebooks' default setting. Traditional grading has been in use for hundreds of years and most grading policies involve averaging scores on assignments together with certain types of assignments counting more than others towards a class grade.
- It allows teachers to indicate the relative weight of assessments directly, without the algorithm "taking over" and impacting the overall score.

Challenges with the Weighted Average calculation

- It does not represent the learning path or process well. In order to get the highest grade on a standard, a student must receive a 4 on every single assessment along the way. Real-world learning patterns—and standards-based philosophy—support the concept that a student often

grows towards mastery over time without being penalized for not knowing it from the start. If the student has already mastered the concept, why are you teaching it?

- In part due to the challenge above, the weighted average calculation tends to encourage teachers to score students based on whether they "did what they were supposed to do" rather than measuring them against the criteria of the standard. In other words, teachers will feel pressure to give students a score of 4 on early, easier, scaffolded assignments even though the student hasn't demonstrated mastery of the standard at full rigor. This corrupts the mastery-based data and makes it mathematically impossible to represent a student's growth over time.

4. Max Value

This simple calculation type takes the highest score ever achieved by each student on each learning target. In this way, the highest level of mastery ever demonstrated is what is considered the final level of mastery on a standard.

Advantages of the Max Value calculation

- It is very easy to understand.
- When teachers truly measure against the criteria of the standard, this can be a robust method to determine overall mastery per standard. It works best when early assignments don't even offer the opportunity to receive a 4 if they are scaffolded. Only on robust summative assessments, after many previous formative assessments, would a teacher offer a student the opportunity to earn a 4.

Challenges with the Max Value calculation

- It doesn't do a great job of representing consistency if a standard demands multiple demonstrations of mastery before mastery can be achieved.
- To work well, it often involves a complete re-design of classroom assessments and grading.
- Student, parents, and teachers need to accept that scores will very often be lower than we've become accustomed to during the learning process, and it is up to the teacher to communicate the "expected" level of mastery and thus answer a student who asks "am I falling behind with this 2, or am I on pace to eventually master it?"

5. Decaying Average

A decaying average, in effect, allows newer assessments to automatically "count more" in the overall score. Similar to the Power Law, as a student improves the more recent scores are contributing more to the overall rating of the standard.

JumpRope offers two decaying average calculations. The first one, referred to as the Exponential Decaying Average assigns progressively-decreasing weights to older assessments. Working backwards, each assessment is worth 66.667% of the teacher-assigned weight, compounded exponentially. Teacher weights still apply.

This is a calculation method that we designed internally (unlike the Power Law which is based on Marzano's work) as an alternative to the Power Law with fewer of its downsides but which still honors the concept of "more recent evidence is more important when determining a student's level of mastery."

Advantages of the Exponential Decaying Average calculation

- This strikes a good balance between other methods, splitting the difference between the Weighted Average and Power Law methods.
- Teachers can still indicate relative weight of assessments while it also considers more recent evidence more heavily than earlier evidence to reward growth.

- The growth trend is represented in the calculation without the complexity of the algorithm of the Power Law. It's easier to understand mathematically.

Challenges with the Exponential Decaying Average calculation

- Much like the weighted average, it's technically impossible to get the highest score on a standard if you ever receive a lower score on any assignment (though with the decaying average, it's more likely that a student scoring high on more recent assignments will "round up" to a 4).
- Despite being simpler than the power law, it's still an exponential equation that you'll need to be prepared to explain. Also, you'll probably still need a calculator to reproduce it.

The second decaying average, referred to as the Rolling Decaying Average, calculates a score for the standard by taking 25% of the current mastery score and adding it to 75% of the new assignment score. (Other values that partition current mastery score and the newest score are available as custom calculations.)

Advantages of the Rolling Decaying Average calculation

- It hits a good balance between other methods, splitting the difference between the Weighted Average and Power Law methods.
- The growth trend is represented in the calculation without the complexity of the algorithm of the Power Law or the Exponential Decaying Average.
- Because it is easier to understand mathematically, teachers find it easier to communicate to students and parents.

Challenges with the Rolling Decaying Average calculation

- Much like the weighted average, it's technically impossible to get the highest score on a standard if you ever receive a lower score on any assignment (though with the decaying average, it's more likely that a student scoring high on more recent assignments will "round up" to a 4).

6. Teacher Judgment (or Final Rating)

Technically not a calculation method at all, the Final Rating system allows teachers to "eyeball" the formative evidence to come up with an overall score for each student and standard. Essentially, it allows a teacher to collect evidence throughout the learning experience and reserve "judgment" and calculation until a learning experience has completed (perhaps at the end of a unit or course). The only difference between this and the Most Recent calculation is that Final Rating will not produce any score at the standard level until a special Final Rating score is entered into the system, whereas the Most Recent method continuously recalculates the standard score as evidence is added incrementally.

Advantages of the Final Rating method

- The educator gets to use professional judgment based on the available data in JumpRope along with any other objective or subjective data sources.
- There's no math to explain to others - it falls on the teacher to justify the score anecdotally and with the formative assessment scores in the system.

Challenges with the Final Rating method

- There's no math to explain to do the work for you - it falls on the teacher to justify the score anecdotally and with the formative assessment scores in the system.
- It's a lot of work for the teacher at the end of the learning experience.
- Mastery reports are not meaningful during the learning process, as JumpRope's gradebook will report a "U" for un-scored until the rating is entered.
- It opens the door for subjectivity.
- Teachers may forget to enter or update the rating as students go through the learning experience, and this can lead to "stale" data in reports.

7. Custom Calculations

In addition to the calculations detailed above, JumpRope's gradebook supports the design of custom calculations that can follow more complex logic to determine the overall standard score per student. For example, a school may decide to only consider the "top three" scores within a standard to determine the overall score, or may decide to limit the maximum score a student can receive if they have any evidence marked as "missing." The sky is the limit, but JumpRope charges to develop custom calculations and you should first carefully consider the pros and cons.

Advantages of the Custom Calculation method

- You can precisely represent your grading policy (and all of it's quirks!).
- You can set up JumpRope's gradebook to mirror legacy and home-grown systems.

Challenges with the Custom Calculation method

- It's not as easy as you think to precisely define this. What about the outliers and edge cases? What if you didn't consider certain content area's assessment patterns? If you've never used this "in production," it's very likely that it will evolve over time, and you're at the software company's mercy in terms of cost, time frame, and viability of desired changes.
- You may not have any research or external organization to help justify or explain your methodology, which means more of the burden will fall on you to explain, justify, and operationalize the calculation.

Conclusion

It's important for our partner schools and districts to discuss what calculation method to use, but JumpRope always cautions against pursuing the holy grail/pot of gold when it comes to the calculation.

We have seen schools and districts change their mind several times only to realize that they're chasing their own tail. Every situation is different, and computers aren't going to be able to always match a teacher's intuition about a student. In the meantime, the most effective solution we've seen is to provide some standardization/structure for the sake of clear communication and to help educators understand the chosen system well enough to know when and how to handle exceptions.

It's also worth noting that the conversation around calculations in a standards-based gradebook doesn't end at the standard level. While calculating higher-level aggregate scores such as course grades tends to be less complex than representing mastery on an individual standard over time, there are still many things to consider:

- Should each standard be treated equally within a course?
- Are academic standards and habits of work standards both part of the course grade?
- When a standard is addressed in more than one course experience (with more than one teacher), is it valid to combine those data points for an overall standard score?
- Should course or graduation requirements be based on an average level of proficiency across multiple standards, or on meeting all standards to a certain level?
- How should schools handle things like eligibility (sports, extracurriculars, etc.) and recognition (honor roll, awards, etc.) that have traditionally been addressed by grades?

For more information about professional development and consulting contact us at info@jumpro.pe.