Goals and Guardrails - Quality and Measurement

Inputs, Outputs, and Outcomes
Inputs are resources and activities invested in a particular program or strategy that are usually knowable at the beginning of a cycle and that are a measure of effort applied. In a school system, inputs are things like staff, books, programs, facilities, buses, and everything else we invest resources in prior to the start of school.

Outputs are the result of a particular set of inputs that is usually knowable in the midst of a cycle and that is a measure of the implementation of the program or strategy. In a school system, outputs are things like formative assessment scores, interim assessment scores, benchmark results, grades, quarterly financials, and everything else that measures our implementation of our inputs throughout the course of the school year.

Outcomes are the impact of the program or strategy that is usually knowable at the end of a cycle and that is a measure of the effect on the intended beneficiary. In a school system, outcomes are things like graduation rates, summative assessment scores, end of year audit results, end of year retention rates and everything else that measures the final result of all the inputs and outputs at the end of the school year.

All inputs, outputs, and outcomes in a school system come in one of two varieties: adult or student. As an example, adult outcomes are results at the end of the school year that are measures of what adults know or are able to do. Student outcomes are results at the end of the school year that are measures of what students know and are able to do. It is critical to remember that school systems only exist for one reason: to improve student outcomes. Not for adult outcomes, not for adult inputs, not for student outputs. School systems only exist to improve student outcomes. All other activities should be in service of this sole reason for existing.

Example
An example of a desired student outcome is having more students scoring 3 or better on their AP exams. That is a measure of what students know or are able to do by the end of the cycle; it's a student outcome. But outcomes don't simply occur. They are the result of inputs by staff that lead to outputs that then can lead to outcomes.

An input that might help lead to the aforementioned outcome is making more seats available in AP courses. This is an input in that it's the investing of resources at the beginning of the school year. Outputs that help lead to the desired outcome might be the increase in the number/percentage of students enrolled in or on track in the AP courses. These are outputs in that they are measures of the fidelity of implementation of the input and they are knowable throughout the school year.

Formative, Interim, & Summative Assessments
Closely related to the concepts of inputs, outputs, and outcomes are the concepts of formative, interim, and summative. Where inputs / outputs / outcomes can refer to any of the processes within an organization, formative / interim / summative focus specifically on methods of assessment and associated metrics. These terms are most commonly used in academic contexts.

Formative assessments are those given in the midst of an instructional cycle that measure student learning, narrowly tailored to specific student expectations. Generally created and/or administered on an ongoing basis by teachers or campuses for the purpose of refining instructional practice. Formative metrics are the data provided by formative assessments or other input evaluation tools and generally serve as data used to evaluate and refine the fidelity of implementation. Formative assessments and metrics are types of outputs but are likely not aligned to the summative assessment or metrics.

Interim assessments are those given in the midst of an instructional cycle that measure a cross-section of knowledge or skills. Generally administered up to two or three times per year by campuses or school systems for the purpose of predicting summative performance. Interim metrics are the data provided by interim assessments or other outcome predictive tools and generally serve as data used to predict summative results by tracking quantifiable progress at

critical and ongoing junctures during a cycle. Interim assessments and metrics are types of outputs and are explicitly aligned to the summative assessment or metrics.

Summative assessments are those given at the end of an instructional cycle that measure a cross-section of knowledge or skills over the course of an instructional cycle or school year. Generally administered at the end of a curricular unit, school year, or the transition to a new schooling experience by school systems or states for the purpose of evaluating school or school system effectiveness. Summative metrics are the data provided by summative assessments or other measurements of the final results in a school system. Summative metrics are used to track outcomes and to evaluate whether the school system/chief executive was successful. Summative metrics are generally a type of outcome -- and when measuring what students know or are able to do, are a type of student outcome.

Criterion-referenced & Norm-referenced

Another set of assessment-related concepts is the distinction between criterion-referenced and norm-referenced. These are two different ways of interpreting what the assigned rating will be after the raw scores are collected.

If assessment results are considered criterion-referenced, this means that answers given by the student are being compared to a set of fixed standards. A key element of criterion-referenced assessments is that every student whose raw score is above a certain threshold is assigned the same rating. For example, if a score of 9 right answers out of 10, 90%, is considered an 'A' rating, then everyone who scored 9/10 will earn an 'A'. Criterion-referenced assessment results are more ideal for identifying whether or not the students have a mastery of a predefined set of knowledge and/or skills.

If assessment results are considered norm-referenced, this means that answers given by the student are being compared to the answers provided by the other students. A key element of norm-referenced assessments is that it is possible to have an improved raw score but not have an improved assigned rating. For example, if a student improved their score by 5 points but all other students improved their score by 10 points, the student's assigned rating might actually go down. Norm-referenced assessment results are more ideal for identifying where students are in relation to each other.

Proficiency, Growth, & Comparison

Another set of distinctions critical to this work is the nature of expectation construction. Proficiency is about having expectations that are relative to a predefined competency level. Proficiency asks whether or not a student is at a particular level, usually expressed something like "on grade level" or "passing". An example of a proficiency goal might be, "the percentage of 3rd grade students demonstrating on-grade-level proficiency in reading will increase from 40% in June 2021 to 70% by June 2026".

Growth is about having expectations that are relative to a predefined starting point. Growth asks whether or not a student has demonstrated improvement from the starting point to an ending point, usually expressed something like "grew by more than a grade level" or "demonstrated adequate/expected growth". An example of a growth goal might be, "the percentage of 3rd grade students who began the year below grade level in reading and who grew by at least 1.25 grade levels will increase from 40% in June 2021 to 70% by June 2026".

Comparison is about having expectations that are relative to a control group. Comparison asks whether or not a student has demonstrated growth or proficiency relative to a different group of students, usually expressed as something like "closing the gap". An example of a comparison goal might be, "the percentage of 3rd grade students who were two grade levels behind in reading who closed the gap between themselves and the overall 3rd grade average will increase from 40% in June 2021 to 70% by June 2026".

Is Assessment Harmful To Students?

Categorically, emphatically, and unapologetically, no. Assessment is like any other tool — a screwdriver, a level, some pliers — in that it is both critically essential to effective construction (in this case constructing knowledge rather than buildings) and it is capable of being abused. Assessments used abusively certainly have the capacity,to be harmful. But anyone who suggests that all assessment is inherently harmful is profoundly misinformed and harbors beliefs that will

be harmful to children. Assessment plays several key roles in the learning process but two of the most vital are long term memory creation and instructional practice improvement.

Assessment: Long Term Memory Creation

Part of the neurological science of learning involves students migrating information out of short term working memory and into long term memory. One strategy for supporting this migration is assessment. When students are taught something on Monday of week 1 and then quizzed on it on Friday of week 1 (this is usually formative assessment), the process of migration begins. Then when the same material is assessed again at the end of the semester period (this is often an interim assessment), the process continues. When students are assessed on the same material again at the end of the year, requiring them to recall information learned months earlier (this is usually summative assessment), the associations with that learning are strengthened and reinforced in the brain's physical structures. In this way, appropriately designed and spaced assessment directly contributes to the learner's ability to master content.

Assessment: Instructional Practice Improvement

If there is any one place where the magic of education is most likely to occur, it's in the interactions between the learner and the educator. Study after study suggests that of all the factors school systems control, **quality of instruction has the largest impact on student performance**. As instructional quality continuously improves, so too are student outcomes more likely to improve. And **one of the most consistent paths for teachers to improve the quality of their instruction is to teach, assess what was taught, then reteach, then reassess**. This continuous improvement cycle pushes teachers to constantly evaluate what worked and what did not and make adjustments based on it. But this process is impossible to conduct without assessment data -- usually formative assessment data. Improving instructional practice is a vital key to improving student outcomes, but where there is no assessment, there will almost certainly be no improvement in instructional practice.

Is Standardized Assessment Harmful To Students?

Categorically, emphatically, and unapologetically, no. Standardized assessment is like any other tool — a screwdriver, a level, some pliers — in that it is both critically essential to effective construction (in this case constructing knowledge rather than buildings) and it is capable of being abused. Standardized assessments used abusively certainly have the capacity to be harmful. But anyone who suggests that all standardized assessment is inherently harmful is profoundly misinformed and harbors beliefs that will be harmful to children. Standardized assessment plays several key roles in the learning process but two of the most vital are performance transparency and system practice improvement.

Standardized Assessment: Performance Transparency

A commonly asked question is whether or not parents have the right to know how their child's school is performing compared to other schools. If the answer is "no", parents don't have the right to understand the relative performance of their child's school, then standardized assessment is less necessary. But if this is a right parents should have, then standardized assessment is essential. If the math teacher at one school gives students A's for the same work that the math teacher at another school gives her students B's, letter grades are no longer fair indicators of relative performance. One way of addressing unfairness in the system like this is to administer a common assessment at both schools, apply a common scoring rubric across both schools, and then routinely train teachers on the use of these assessments. In other words, to make it fair, you'd have to make the assessment standardized.

This transparency isn't just needed in K-12 performance. Many professions -- medicine, technology, engineering, construction, truck driving, and more -- rely on standardized assessment as part of their systems for determining who can enter the field as a means of promoting fairness and ensuring performance. This is used to protect all of us — ensure doctors in multiple hospitals all give the right medicine, engineers design bridges across the state that can all carry the right load, system administrators design servers across an organization that all protect our data, etc. When the transparency of performance across multiple sites matters, well-formed standardized assessment is an appropriate response.

Standardized Assessment: System Practice Improvement

Transparency isn't the only good reason to have standardized data. It's also immensely useful for helping complex systems identify outliers and make needed corrections. But if the data isn't apples to apples, it's harder to know what is noise and what is signal.

It's worth noting that standardized assessment can be formative, interim, or summative in nature and whichever of these is the case has a significant impact on how the data is appropriately used. It doesn't make sense to use formative data to predict performance on a summative assessment; it doesn't make sense to use summative data to try to inform weekly changes. Generally speaking. system-level practice improvement will rely on aggregated interim and summative data more so than formative data.

Is Standardized Assessment Biased?

Yes, because all assessments are biased, standardized or otherwise. Every assessment was created by someone who deployed aspects of their world view, their language, their understanding of the material into the assessment. Asking if standardized assessments are biased is the wrong question. What to ask instead: 1) is the bias that is present harmful to the collection of accurate data, and 2) what have we done to identify and address harmful bias? As an example, here's a question from a 3rd grade math assessment:

> If 300 crayons are added to 100 crayons, how many crayons are there?

Applying the wrong question actually creates more potential for harm. When asked, "is this question biased?" most people suggest it is not. That is asking the wrong question and that is the wrong answer to the wrong question. This question is clearly biased. The better question to ask is, "for whom is this question biased and is the bias that is present harmful to the collection of accurate data?" If children are being assessed who have no familiarity with crayons -- as can be the case with asylees or children experiencing similar traumas -- this is certainly a biased question. Does that bias harm the collection of accurate data? It depends on context. For the general population, probably not (the question doesn't require the test taker to know what a crayon is, only that it's an item being counted -- but only "probably not" because not knowing what a crayon is may slow a child down as they try to figure that out, which could have an impact on their overall performance). But if the assessment is only going to be administered to asylees, then yes, the question has a high enough likelihood of harm that I'd consider omitting it. Context matters.

Of equal importance is taking time to ensure that assessments have been vetted for bias. Each item in an assessment should be reviewed by multiple teams in an effort to identify any inappropriate items whether regarding grade level of the standards being assessed, grade level of the language in the questions, developmental appropriateness of the material, or bias.

It's worth noting that while all assessment has the capacity for harmful bias in its design, standardized assessments -- when well designed -- have a heightened capacity to protect against bias due to the more structured administration of the assessment -- a protection harder to achieve in non-standardized assessments.

What Are Inappropriate Uses Of Standardized Assessments?

The simple answer: using assessment data for what it wasn't designed to be used for. Academic assessments are not designed to tell you a child's worth, or their value. Anyone who asserts otherwise is using standardized assessment data in an inappropriate manner. And as mentioned, another example of misusing data is to use summative assessment data as if it's formative assessment data, or vice versa.

Does Constant Test Prepping Improve Performance On Standardized Assessments

No. This is a myth perpetuated by, not surprisingly, the test prep industry. What does support improved performance is a small amount of preparation regarding the item types that will be used in the assessment and the general design of the assessment. This familiarity helps a test taker be better positioned to demonstrate what they know on the assessment. That's a very good thing. But it only takes a few hours to do this. Endless test prepping that begins weeks and months before an assessment is time wasted and is a practice that is actively harmful to children. What improves performance on standardized assessments? When students are taught the assessed material to a depth of knowledge that allows them to use the material in new ways. There are no shortcuts.

Goals & Guardrails

The intention of goals is to reveal the community's vision for its students' outcomes. As such, goals are only about student outcomes -- what the community wants its students to know and be able to do. Ideal goals will be SMART (specific, measurable, attainable, results-focused, and time-bound), will describe what the community wants its students to know and/or be able to do, and will number between one and five (we generally recommend three). Goals describe what the school system is trying to accomplish.

Examples of goals include:

- The percentage of students who will enter kindergarten school-ready on a multidimensional assessment will increase from 21% on August 1, 2019 to 65% by August 1, 2024
- The percentage of graduates who are persisting in the second year of their post-secondary program will increase from W% on X to Y% by Z
- The percentage of free and reduced lunch-eligible students in kindergarten through 2nd grade who are reading/writing on or above grade level on the district's summative assessment will increase from W% on X to Y% by Z
- The percentage of students at underperforming schools who meet or exceed the state standard will increase from W% on X to Y% by Z
- The percentage of males of color who graduate with an associate's degree will increase from W% on X to Y% by Z
- Percentage of students from our PreK program who enter kindergarten ready will increase from 20% in August 2019 to 80% by August 2023.
- Percentage of graduates persisting in 2nd year post-secondary will increase from 25% on June 2019 to 50% by June 2023.
- Number of high performing campuses as measured by the SPF will increase from 2 in July 2019 to 5 by July 2023.
- The percentage of students at underperforming schools who meet or exceed the state standard will increase from 47% on June 2019 to 75% by June 2023.

Non-examples of goals include:

- Students will improve academically
- Percentage persisting in 2nd year post-secondary will grow by 5% each year
- The number of findings on the audit will decrease from 7 in June 2019 to 0 by June 2024
- Number of high performing campuses will increase to 5 by 2021
- Percentage of graduates having completed an associate's degree and/or been awarded an industry certification by graduation will grow from 10% to 30%
- The percentage of quality teachers retained each year will increase from 78% in May 2017 to 93% in May 2022

The community will also have other things it values beyond the vision. These other items relate to what the adults are doing to cause the goals to happen -- they are the inputs, not the outcomes. They are about the means, not the ends. We refer to the written version of these values as guardrails. Ideally a school board will adopt one to five such overarching statements (we generally recommend three). Guardrails describe how the school system will behave as it seeks to accomplish the goals.

Examples of guardrails include:

- The Superintendent will not allow underperforming campuses to have principals or teachers who rank in the bottom two quartiles of principal or teacher district-wide performance
- The Superintendent will not propose major decisions to the Board without first having engaged students, parents, community, and staff
- The Superintendent will not allow the number or percentage of students at underperforming campuses to remain the same or increase
- The Superintendent will not allow the inequitable treatment of students
- Do not allow the number or percentage of students in low performing campuses to increase or remain the same.
- Do not allow teacher/principal compensation or increases to be equal across a bell curve of teacher/principal performance

- Do not allow campuses to ineffectively implement PLCs
- Do not allow teachers or principals in the bottom two quartiles of performance to serve in low performing campuses.

Non-examples of guardrails include:

- Enroll fewer students in our low performing campuses.
- Close all low performing campuses.
- Keep parents happy.
- Do not fail to hire Mrs. Johnson.
- Do not do anything without prior approval from the board.
- Do not allow staff to ignore board member directives.

The more clearly and finitely defined the school board's adopted goals and guardrails are, the easier it is for the school board to ensure alignment between the school board's sayings and doings.

Interim Goals & Interim Guardrails

Interim goals and interim guardrails are measures of progress toward a defined goal or guardrail. By their nature, interim goals and interim guardrails are output metrics that are predictive of the goals / guardrails and that are influenceable by the chief executive. Interim goals and interim guardrails are defined by the chief executive, in consultation with the board, and are the data the chief executive uses to routinely report on performance to the board.

Questions To Ask When Evaluating Interim Metrics

Knowing whether or not interim metrics are sufficiently meaningful is always challenging. Here are a set of questions that staff, board members, and coaches should check off prior to their being accepted:

- What is the degree of correlation between the interim metric and the goal metric?
- Are the interim metrics leading indicators relative to the goal/guardrail rather than lagging?
- If all three of the interim metrics are accomplished, will that ensure the accomplishment of the goal?
- Do the interim metrics fully address the content of the goal/guardrail?
- Are each of the interim metrics updateable multiple times per year?
- Are each of the interim metrics outputs rather than inputs?
- Are there significant unintended consequences that need to be considered?
- Does each interim metric have a starting point, ending point, starting date (month and year), and an ending date (month and year)?

Goals / Guardrails / Summative Metrics

- Definition: Quantifiable outcome measurement of achieving/honoring your school system's vision or values
- Function: Used to track outcomes; whether the school system/chief executive was successful
- Guiding Question: What would honoring the vision and values of our community look like for this school system?
- Formula: The [measure] for [population/area] will [increase/decrease] from [starting point] on [starting month/year] to [ending point] by [ending month/year]
- Examples:
  - Number of high performing campuses as measured by the School Performance Framework will increase from W% on X [month/year] to Y% by Z [month/year]
  - Percentage of graduates persisting in their second year post-secondary without needing remedial courses will increase from W% on X [month/year] to Y% by Z [month/year]
  - Percentage of graduates having completed an associate's degree and/or been awarded an industry certification by graduation will grow from W% on X [month/year] to Y% by Z [month/year]
  - Percentage of students reading on grade level according to ABC instrument by the end of 3rd grade will increase from W% during X [month/year] to Y% during Z [month/year]

Strong Summative Metrics...	Developing Summative Metrics...	Weak Summative Metrics...
✔ Are **quantifiable** (SMART), and the data needed to measure can be reasonably collected. ✔ Clearly **align** with the vision and values of the community. ✔ Are developed with baseline data and **indicate an important improvement** in student outcomes. ✔ Include the **starting date** (month and year) and the **ending date** (month and year) by which the metric **will be achieved**. ✔ Include a **starting point** (baseline) and **ending point** (target) ✔ Will **challenge** and disrupt the organization and require adult behavior **change**. ✔ Were developed through **analysis** of student needs. ✔ Community members were **included** in the development process. ✔ Where feasible, rely on **external data** rather than data produced by the school system's staff. ✔ The data is collected using a robust system for ensuring **data integrity.**	Includes **vague** markers for success (e.g. "more," "better," "large amounts") and data cannot be collected within a reasonable time **Unclear** alignment with the vision and values of the community Not developed with baseline data and/or, if achieved, **will not have a significant impact on students** Can be accomplished by **only slightly tweaking** what the organization is currently doing Include a **date**, but it is **incorrect**, too broad, or **not aligned** with the vision and values of the community	o **Cannot** be measured. o **Do not indicate** overall success of the initiative, or are actually **formative metrics** or **interim metrics** o Are **not ambitious** or inspirational o Can be accomplished **without changing anything** the organization did last cycle o No **date** included

Interim Goal / Interim Guardrail / Interim Metrics

- Definition: Quantifiable output measurements of achievement that indicate progress towards the Summative Metric.
- Function: Used to track outputs; quantifiable progress at critical and ongoing junctures during a cycle
- Guiding Question: How will I measure whether or not I am on track to reach my Summative Metric?
- Formula: The [measure] for [population/area] will [increase/decrease] from [starting point] on [starting month/year] to [ending point] by [ending month/year]
- Examples:
  - The percentage of students on track in reading will increase from W% on X [month/year] date to Y% on Z date [month/year], as measured by district benchmark assessments
  - The percentage of 9th graders on track to graduate will grow from W% in X [month/year] to Y% [month/year] by Z as measured by the district's ABC on-track indicator

Strong Interim Metrics...	Developing Interim Metrics...	Weak Interim Metrics...
✔ Are **quantifiable** (SMART), and the data needed to measure can be **reasonably collected**. ✔ Lead **directly** to the **achievement** of the Summative Metric. ✔ Are **ambitious**, but can be completed **within the scope of the goal**, and will need to be **tracked on an ongoing basis** in order to ensure overall success. ✔ Refreshed data is available **multiple times per year** ✔ Include the **starting date** (month and year) and the **ending date** (month and year) by which the metric **will be achieved**. ✔ Include a **starting point** (baseline) and **ending point** (target) ✔ Will **challenge** and disrupt the organization and require adult behavior **change**. ✔ Are **predictive** of the Summative Metric ✔ Are **influenceable** by the chief executive ✔ Where feasible, rely on **external data** rather than data produced by the school system's staff.	Include **vague** markers for success (e.g. "more," "better," "large amounts") Do not account for **all aspects** necessary to goal/guardrail success OR are **not directly tied** to overall goal/guardrail success. Can be accomplished by **only slightly tweaking** what the organization is currently doing Indicate **trivial achievements** or do not necessarily relate to the overall goal/guardrail success.	○ **Cannot** be measured. ○ **Do not align** with nor lead directly to the Summative Metric. ○ Can be accomplished **without changing anything** the organization did last cycle ○ **Do not indicate ambitious achievements** or require little to no effort.

✔ The data is collected using a robust system for ensuring **data integrity.**		

Progress Monitoring

- Definition: Process of tracking progress regarding the extent to which the agreed upon results are likely to occur
- Practitioners: This should be conducted by staff responsible for managerial creation of results and their supervisors; this is both a managerial function (when it occurs between staff and their supervisors), and a governance function (when it occurs between the board and the chief executive)
- Frequency: Progress monitoring should be practiced no less than once per quarter (though it is most commonly a monthly practice) and no more than the frequency with which lead measure/interim measure data is refreshed.
- Function: Used to track interim metrics (outputs)
- Guiding Question: How will I know whether or not I am on track to reach my Summative Metric?

Blue	Results were completely delivered
Green	All interim metrics indicate that the results will be delivered
Yellow	Interim metrics indicate conflicting evidence regarding whether results will be delivered
Orange	Interim metrics indicate results are unlikely to be delivered without significant changes
Red	Interim metrics indicate results will not be delivered
Gray	Interim metrics are not available yet