

Multiple Measures: Establishing and Assessing Credibility

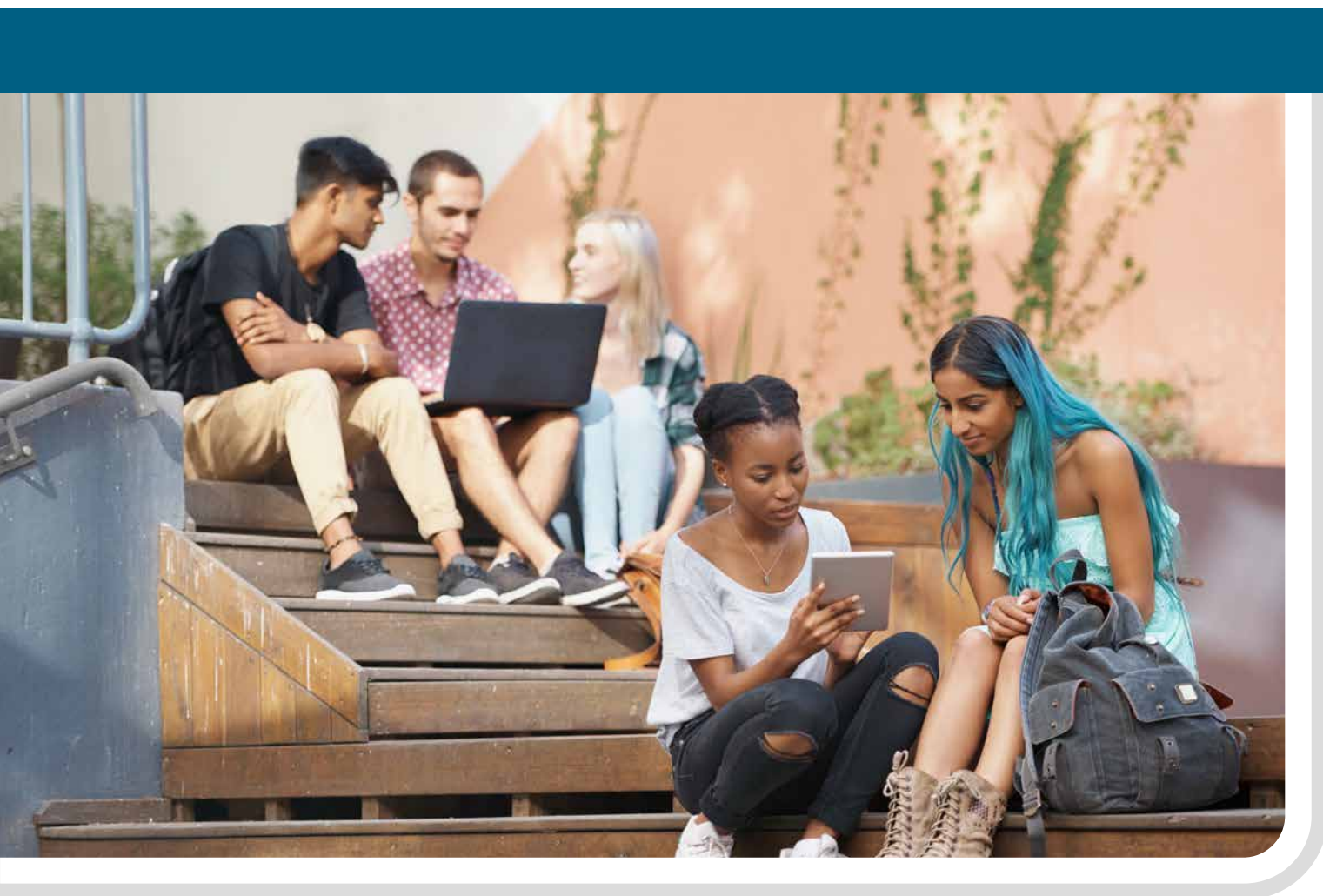
©Copyright 2023. National Association of Charter School Authorizers, Chicago, Illinois, U.S.A.
All rights reserved under both international and Pan American copyright conventions.

nacsa
NATIONAL ASSOCIATION OF
CHARTER SCHOOL AUTHORIZERS

PURPOSE OF THIS RESOURCE

This resource is for authorizers and schools who have already thought about what specific student outcomes they seek to achieve, beyond those typically measured by formal accountability. Through guided conversations, development of a logic model, or other similar processes you (authorizer, school, or authorizer/school teams) have begun to hone in on specific outcomes. Now you're considering how to best measure progress towards those outcomes. Ensuring that your measurement tool has **credibility** with internal and external stakeholders will go a long way towards successful data collection and evaluation. This resource can help identify measurement tools that will be credible, facilitating effective program implementation and evaluation of outcomes.

If your authorizing office or school is just beginning to consider Multiple Measures, start with this [Multiple Measures Readiness Assessment](#).



WHAT DO WE MEAN BY CREDIBILITY?

Credibility refers to how much stakeholders (authorizing staff and board, community members, families, students, school leaders, teachers, school board, etc.) believe in a tool as an effective way to measure something. When we think about the credibility of sound data collection tools, we often talk about [validity and reliability](#).

→ VALIDITY

Validity is the extent to which the tool measures the concept you want it to measure. For example, does your 20-question survey on parent engagement measure how engaged parents are or does it measure whether parents are willing to complete a 20-question survey? Similarly, the DIBELS is a way to measure reading behaviors and includes some comprehension questions. It allows teachers to see the reading strategies students are using, and potential problems with decoding. The test does not, however, provide in-depth information about a student's comprehension strategies across a range of texts.

→ RELIABILITY

Reliability refers to how well your tool will give you the same expected result when used multiple times and with a variety of users. We often also use the word consistency because we look for a consistent result from the tool over time. For example, if Mr. Jones uses a writing rubric with his class fall semester, he should be able to see some progress made by Malachi, but not wildly different when using the rubric this week as he will next week.

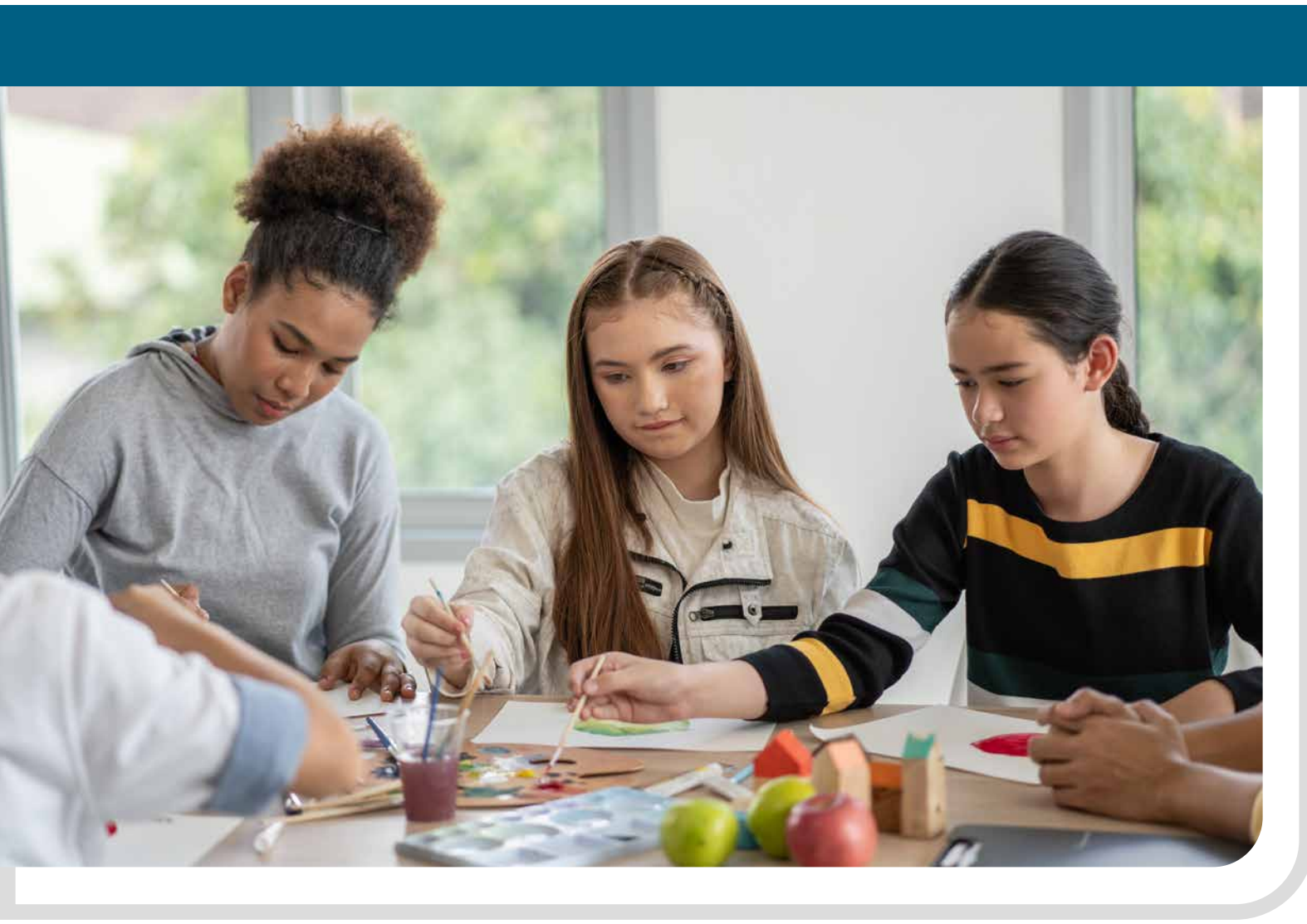
Credibility, however, often requires more than simply determining validity and reliability. This is particularly important to remember since some of the tools you choose to measure student outcomes may not have a long enough history of use to have been tested yet for validity and reliability. (That's the challenge with innovation: there is no clear evidence at the outset that what you're doing will work!)

Testing the credibility of your tool will require you to defend it to your stakeholders. One way to do so is to have a clear logic model or other documented theory of action that outlines the rationale for using a particular tool to measure a particular outcome. Otherwise, you may have a valid and reliable measure that none of your stakeholders believe in or care about. For example, Arkansas used the PARCC assessment, which was both valid and reliable, but quickly changed to the ACT Aspire because PARCC lacked credibility due to the debates around Common Core and its assessments.

FORMAL ACCOUNTABILITY AND SCHOOL IMPROVEMENT

When choosing a measure, be clear on its purpose. Any measure of student learning, achievement, wellness, etc. should **drive school improvement**, regardless of whether the measure is used for formal accountability. If a measure is part of formal accountability (i.e., included in the contractual performance expectations), establishing the credibility of the assessment tool is that much more important, and an authorizer will likely have a higher bar for rigor, validity, and reliability.

If both the school and authorizer are at earlier stages of implementing Multiple Measures, some measures could be geared only towards school improvement until both the school and authorizer are comfortable adding them into the formal accountability structure. The school and authorizer must consider their tolerance for uncertainty as well as for innovation—which could result in failure. Some authorizers may not accept measures that haven't been tested for validity and reliability; others will trust a tool created specifically for a school's identified outcome, with a logic model or something similar to defend its use.



ESTABLISHING THE CREDIBILITY OF MEASUREMENT TOOLS

Not all measurement tools that a school or authorizer chooses to evaluate student outcomes will meet the criteria of having credibility. Validity and reliability are still typically expected, though not necessary, for a tool to have credibility.

There are several criteria that will impact a tool's perceived credibility by your stakeholders, even if you have the ability to test for validity and reliability. Here are some things to consider:



HAVE YOU IDENTIFIED YOUR STAKEHOLDERS?

Authorizers, teachers, parents, school leadership, and board members should all be engaged as you determine the credibility of your measurement tool. Having these stakeholders engaged early in the process of identifying desired outcomes and potential ways to measure them will make it easier to get buy-in around the credibility of a tool.



ARE TEACHERS ENGAGED?

As many have learned through standardized testing, when teachers don't believe a data collection tool is credible, collecting good data can be difficult. But when teachers see a direct tie between student outcomes, measures, data collection tools, and school improvement, they are more likely to believe in them. Schools should bring teachers into the decision making early and often to avoid challenges later, when teachers are essential to collecting the necessary data.



DOES YOUR COMMUNITY BELIEVE IN IT?

The purpose of measuring mission-related or school-specific outcomes is to tell a more comprehensive story to relevant stakeholders—including parents and the community—about the impact a school is having on students. Schools and authorizers should consider ways to engage parents, students, and community in defining what excellence means to them, as you develop and/or prior to rolling out your measurement tool, so that you know whether they will perceive the data you collect as credible.

USING THIS RESOURCE

The next section provides ways that Multiple Measures might be used to measure progress towards various school goals. The following table is not comprehensive. It illustrates potential measures across several student outcome areas where additional metrics and measurement tools could supplement more traditional accountability measures. In each table row, explore how the example attempts to quantify and measure the potential metric. Then click on a specific measurement tool to learn more about its strengths, challenges, and how to consider its credibility.

SAMPLE MEASURES AND TOOLS

STUDENT OUTCOME AREAS	DESCRIPTION	SAMPLE MEASURE	SAMPLE METRIC	POTENTIAL MEASUREMENT TOOL(S)	REAL WORLD EXAMPLE
Social-Emotional Wellness	Social-Emotional Wellness (or learning) refers to the knowledge, skills, and attitudes for developing healthy identities, managing emotions, achieving personal and collective goals, feeling and showing empathy for others, establishing and maintaining supportive relationships, and making responsible and caring decisions. ¹	Belonging and Connectedness	Percentage of students who feel personally connected to the school community.	<ul style="list-style-type: none"> • Survey • Attendance 	To create a culture of high achievement, a middle school believes that belonging and connectedness to the school community is important. As such, they have developed advisory programs to support students and provide relevant professional development (PD) to teachers. The school wants to determine the effectiveness of these programs and PD and is considering using a semi-annual survey to understand student perceptions of their connectedness to the school community, including how much they personally feel they belong and/or feel accepted by various stakeholder groups. In addition, the school believes high levels of connectedness will be reflected in attendance rates.
		Emotional and Behavior Management	Percentage of students who can demonstrate the ability to effectively manage their emotions.	<ul style="list-style-type: none"> • Rubrics • Discipline Data 	A school with a mission focused on emotional intelligence and social wellness implements weekly Circle programs in advisory. Students are taught skills and given opportunities to effectively communicate about and manage emotions. Students and advisors use a rubric to evaluate student engagement in Circles as well as their ability to effectively manage their emotions. The school is being very intentional to ensure their rubric reflects cultural differences.
Post-Secondary Readiness	Post-Secondary Readiness refers broadly to preparing students transitioning from high school for multiple pathways after graduation, or specifically, being ready for college. ²	Credit Attainment	Percentage of students earning enough credits each year to be on track for five- or six-year graduation.	<ul style="list-style-type: none"> • Transcripts/Credit Attainment • G.P.A. • Attendance 	A high school designed specifically to support academic re-entry may have nontraditional academic goals for students, and uses attendance, transcripts, and credit attainment as tools to track students' progress towards engagement and obtaining a high school diploma.
		College Persistence	Percentage of graduated seniors who persisted through a certain number of years of college.	<ul style="list-style-type: none"> • Survey • National Clearinghouse Data • State Post-Secondary Tracking Systems 	A college-prep-focused high school reviews National Student Clearinghouse data to track college entrance and persistence rates. For data that may not be available, the school surveys graduated students about their college enrollment status for a certain number of years to determine post-secondary persistence rates.
Career and Workforce Readiness/21st Century Skills	Career and Workforce Readiness refers to the concrete knowledge, skills, and dispositions that students must demonstrate mastery of to be prepared for post-secondary success. Broadly, these skills include academic knowledge, problem solving/critical thinking, communication/collaboration, resilience, and others. However, they also include specific workforce readiness measures related to technical and vocational training. ³	CTE Credentialing	Percentage of students who earn an industry credential through an approved CTE pathway.	<ul style="list-style-type: none"> • Credit Attainment • CTE Credential 	A school with a CTE/vocational focus has goals for the percentage of students that graduate with an industry certification or specialized CTE certificate to enter the workforce.
		Navigating the Employment Market	Percentage of students demonstrating satisfactory performance on workplace readiness portfolio.	<ul style="list-style-type: none"> • Rubrics 	A high school whose mission is to prepare students for multiple pathways after graduation requires seniors to create a portfolio that documents their search for a hypothetical job of interest and includes a professional resume and a job-specific cover letter. This portfolio will be evaluated using a rubric the school is developing, leveraging examples in use in the field. The measure will initially be used only for school improvement. The school and authorizer will consider its inclusion in formal accountability in the next contract term, after two years of data has been gathered.
Life/Adult Skills	Life and/or "Adult" skills encompass the behavioral, cognitive, or interpersonal skills and abilities that enable students to effectively navigate the opportunities, demands, and challenges of life. ⁴	Financial Literacy	Percentage of students demonstrating mastery in concepts of personal financial literacy.	<ul style="list-style-type: none"> • Standardized Assessments • Credit Attainment 	A K-12 school, whose focus includes preparing students for the real world, embeds instruction in financial literacy through the elementary and middle school years and requires a class on Personal Financial Literacy in high school. The course curriculum is based on the standards and frameworks published by the Consumer Financial Protection Bureau, Council for Economic Education, and JumpStart Coalition. Course assessments measure student mastery on various topics of financial literacy, and ultimately, the school has a goal for the percentage of students earning a "C" or better for the course and scoring 80% or higher on a related test of financial literacy. Since the school is using an established standardized assessment aligned to the course curriculum, the measure will be included in the school's formal accountability framework.
		Conflict Resolution	Percentage of students who effectively manage conflicts with peers using proven conflict management strategies.	<ul style="list-style-type: none"> • Rubrics • Discipline Data 	An elementary school implements a behavioral intervention curriculum that teaches specific conflict resolution techniques and skills. Students are then assessed on their ability to use those skills effectively via a rubric, and discipline data is monitored to assess the impact on incident frequency.
Personal Development	Personal Development Goals refer to measurable annual goals consistent with a student's needs and abilities, as identified by the student's present levels of performance. Typically, these goals are found in a student's Individualized Education Program (IEP), Personal Development Plan (PDP), 504 Plan, etc.	IEP Goal Attainment	Percentage of students with disabilities who meet a certain percentage of their annual IEP goals.	<ul style="list-style-type: none"> • Rubrics • Discipline Data • Grades • Credit Attainment 	A school expects each student with an IEP to meet a certain percentage of their annual goals as outlined in their IEP. The varied nature of these individualized goals means the measurement tools can vary greatly, depending on the goal. The school and authorizer will consider whether this should be included in the formal accountability framework after the initial year of implementation.
Academic Learning	Academic Learning goals surround the core academic knowledge, skills, and concepts students are expected to master. Many of these competencies are assessed through established measures, however, some specific mission-aligned academic goals may require additional measures.	Critical Thinking/ Problem Solving	Percentage of students who can effectively apply problem solving skills to real world problems.	<ul style="list-style-type: none"> • Rubrics 	A school whose mission is centered on public policy and civic engagement requires students to submit a policy memo that addresses and proposes a potential solution to a problem in the school community. Teachers use a school-developed rubric, adapted from well-established national models, to assess students' ability to diagnose a problem and support an evidence-based solution through critical thinking and problem solving. The school and authorizer agreed to pilot this rubric for one year before considering whether to include the measure in formal accountability.
		Proficiency in a Language Other Than English	Percentage of students demonstrating proficiency in Chinese (Mandarin).	<ul style="list-style-type: none"> • Credit Attainment • Standardized Assessments • Grades 	A dual-language immersion school expects each student to attain proficiency in Chinese. The school uses a mix of standardized assessments, grades, and course credit attainment to assess each student's mastery of the Chinese language.

1. Adapted from [CASEL](#)
2. [Definition from WestEd](#)
3. Adapted from the [American Institutes for Research](#)
4. Adapted from [Berkeley Well-Being Institute](#)



Surveys are an important method of collecting data on individuals' perceptions and feelings about their school and themselves.

OUTCOMES TO MEASURE

In general, surveys are used to measure feelings and perceptions that an outsider would not be able to know without asking a student, teacher, parent, community member, or school leader. Social-emotional skills, belonging, school culture, teacher satisfaction, family satisfaction, relationships, and individual goals can all be measured well with surveys. Sometimes surveys can collect data on how much/how often something happens, but these questions rely on an individual's memories and can have skewed results. Surveys are not ideal for collecting data on the actual quality of a school or program or its components because surveys can only capture what someone reports rather than what actually happened.



STRENGTHS

There are many ready-made survey tools to help measure outcomes such as social-emotional skills, belonging, school culture, teacher satisfaction, family satisfaction, etc. These ready-made tools likely report their validity and/or reliability so you can have confidence that you will get accurate results when administering them. Schools or school systems can create their own surveys to measure an outcome for which there is no existing measurement tool, but they should plan to conduct due diligence to determine whether the tool is valid and reliable. Surveys can be done quickly and gain insight from a diverse group of people without a great deal of effort.

CHALLENGES

Without testing of a newly developed survey instrument, it is difficult to gain credibility with a broad stakeholder group. Because there are many existing tools used by schools around the country, most authorizers and others interested in school accountability would expect a rationale for using a self-developed survey instead, as well as some type of testing to have faith in that tool as a credible measure. Additionally, most surveys now are given online. These surveys often have low response rates (less than 20%) unless proper follow up and incentives are planned for prior to administering the survey. Low response rates will bias your sample because only certain types of people are likely to respond. To have an unbiased, representative sample, the target response rate should be about 80% or higher across all relevant groups.

OTHER CONSIDERATIONS

There is a temptation to create your own survey to perfectly suit your context. Unless you are studying the implementation of a specific program, it is best to use existing valid and reliable tools widely used for the outcomes you are measuring.

Plan prior to the start of a school year when and how often you want to give a survey. Plan when and where you will make the survey available and how you will follow up with those who do not respond. Give dedicated time and space to take the survey to ensure a high response rate.

Surveys are not necessarily ideal for the youngest students (3rd grade or younger) who cannot read and/or easily express their emotions or opinions. If you want similar skills or perceptions measured for younger students, consider observations, survey input from their families and/or teachers, or interviews and focus groups.

LITERATURE

[Considerations for Use of Assessments of Social-Emotional Competence](#)
[EdInstruments from Annenberg Institute for School Reform at Brown University](#)

TOOLS FOR IMPLEMENTATION

[Sample size calculator](#)

[← Back to Sample Measures](#)



Rubrics are usually composed of a set of competencies. Proficiency for each competency is explicitly defined in the rubric, along with what exceeding expectations, as well as not meeting expectations, would look like.

OUTCOMES TO MEASURE

Rubrics are helpful for measuring growth or mastery of a broad range of competencies that are nuanced and not well measured by traditional assessments, such as playing a musical instrument, giving a speech, or writing. Rubrics are fundamentally tools for describing and assessing student performance. For example, rubrics are particularly useful for assessing writing, since it is a complex task that contains multiple competencies, such as sentence structure, use of evidence, use of persuasive devices, or creativity. In a rubric, each one of these competencies includes clear descriptions of performance for each level or score, typically one through four.

STRENGTHS

Rubrics are especially effective for helping teachers explicitly set expectations for what they are looking for from students. In the grading of assignments using a rubric, there is a high degree of transparency because teachers are able to indicate which competencies students mastered or are working to master. If more than one indicator is used to define and measure a competency, teachers are able to provide highly nuanced feedback through the rubric that helps students see what they need to do to meet expectations. Rubrics are particularly helpful for skills-based tasks such as writing, projects, artwork, presentations and communication, or even behavior.

CHALLENGES

Without testing of a newly developed rubric instrument, it is difficult to gain credibility with a broad stakeholder group. While rubrics are great for measuring many skills and competencies for both students and teachers, they require a substantial amount of time devoted to calibrating scoring across raters through substantial team grading of assignments. Reliability for a rubric comes from creating consensus, or agreement, which is achieved if raters assign the same score, or consistency, which provides a measure of correlation between the scores of raters. If consensus or consistency is not achieved, stakeholders will not trust the results from your rubric. In this case, credibility is highly dependent on the rubric's reliability. Whether an existing and widely used rubric or one developed for a unique school model is used, obtaining a high degree of reliability within a team of raters is still necessary. Authorizers and schools both must pay attention to this.

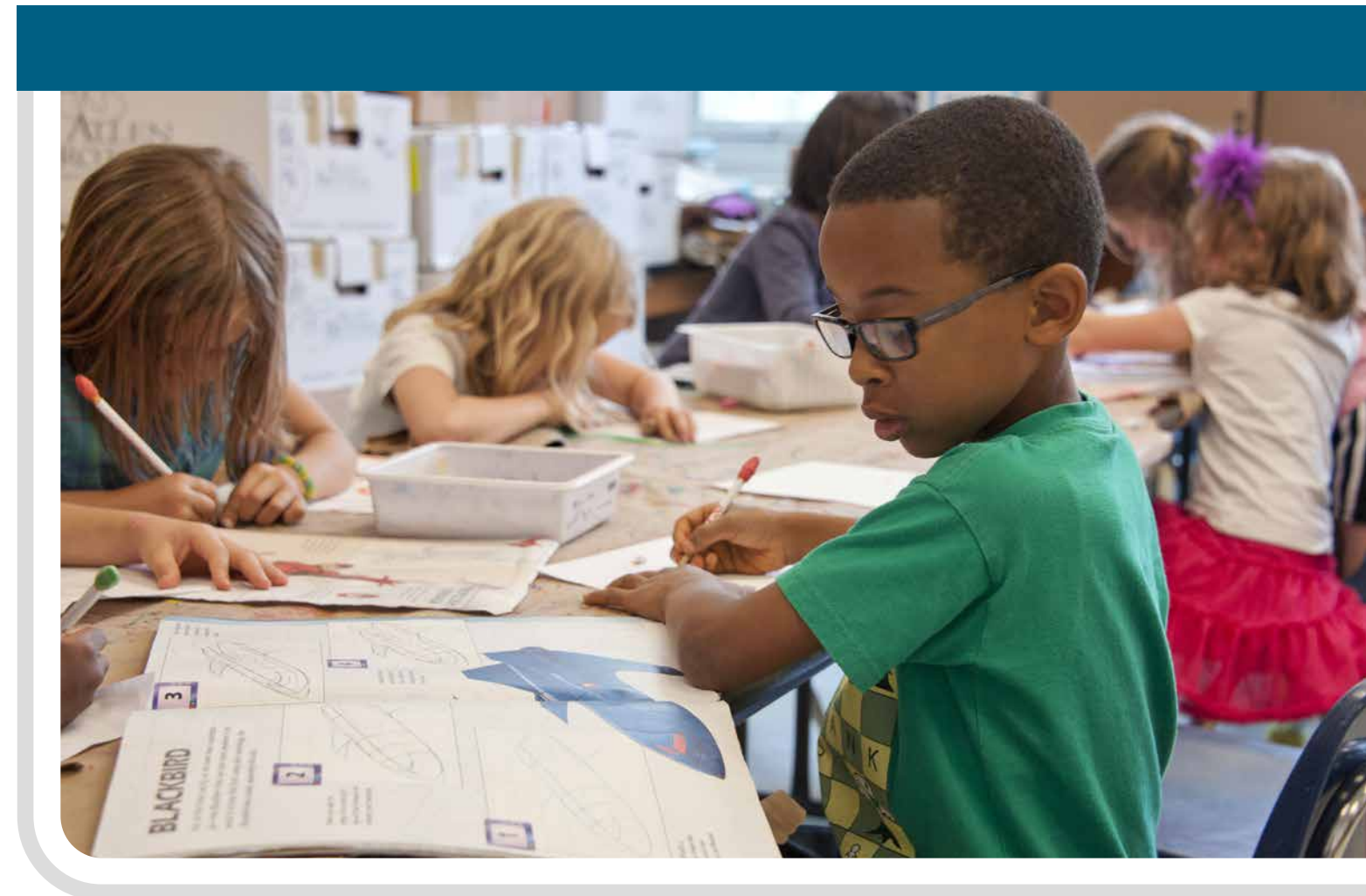
OTHER CONSIDERATIONS

Existing rubrics have the benefit of a higher degree of validity because there is evidence that a sample of individuals or teams have trusted the rubric as a tool to measure a given competency. If a school model is unique enough, however, the local stakeholders may not consider an existing rubric to be valid for their context. For example, schools may have different definitions for post-secondary readiness. If one school defines post-secondary readiness as including life skills while another includes public speaking, they would not be able to share the same rubric. As long as the rubric is directly tied to the school's definition of post-secondary readiness, it will be considered valid by the school's stakeholders. Authorizers may need to determine how much leeway they allow schools in determining the individual competencies they will accept as valid for accountability purposes.

REFERENCES

- [Calibration Protocol for Grading Student Work](#)
- [Modules on How to Create a Rubric](#)
- [Inter-Rater Agreement or Reliability](#)

[← Back to Sample Measures](#)





Standardized Assessments are usually norm- or criterion-referenced tests used with a sufficiently large number of students across many environmental and schooling contexts.

OUTCOMES TO MEASURE

There are standardized assessments to evaluate student learning in almost any content area taught in schools, from second languages to biology; from financial literacy to health. Norm-referenced standardized assessments provide student achievement results as compared to similar peers across the country. Criterion-referenced standardized assessments score students on how much specific content they know, regardless of what other students may know. Standardized assessments can also be summative or formative. Summative assessments are used to understand what a student has learned at the end of an instructional period—chapter, unit, course, or year; formative assessments are used to monitor student learning to provide ongoing feedback that can be used by instructors to improve their teaching and by students to improve their learning. Each assessment company or product will typically specify which type of assessment it is, including its primary intended purpose.

STRENGTHS

Standardized assessments are most often used because they have strong validity and reliability. They are considered objective measures of standards, concepts, or content across a wide variety of content areas. Statewide assessments are certainly the most often-used assessments in performance frameworks for school evaluation, and they provide meaningful information. Beyond these statewide tests, there are other standardized assessments that can also be used in a variety of learning areas, to provide valuable information about student performance and school quality. These assessments are relatively easily administered to any age group. (For example, while statewide assessments typically do not begin until 3rd grade, other standardized assessments can be used with younger students.) In theory, results are available quickly (though this is not typically the case with statewide assessments). Comparability is another strength of most standardized assessments, though this depends on the size of the sample used for norming or the criterion used as the content of the assessment.

CHALLENGES

Many states and schools administered standardized assessments prior to being required to by the No Child Left Behind (NCLB) legislation in 2001. However, since then, statewide assessments in schools have maintained an outsized influence as the primary measure of school quality at the federal, state, and local levels. As a result, many feel that standardized tests, particularly statewide assessments, are overemphasized for accountability. Further, because results from statewide assessments lag—often as long as 4-6 months, their usefulness to teaching and learning is sometimes questioned (although some of those tests are designed to provide information on school quality, not to provide real-time data for instructional practices). The high stakes related to some standardized assessments, including statewide assessments or SATs, ACTs, AP exams etc. can also create an environment of stress and anxiety in some students and teachers. Finally, in some states, debates about the credibility of specific assessments have created inconsistent test administration leading to more than four different statewide assessments being used over a period of six years. These challenges can impact perceptions about all types of standardized tests thus impacting credibility.

OTHER CONSIDERATIONS

State or local policies likely require all public schools to administer a specific set of standardized assessments for certain grade levels, at certain times during the year. These assessments can often take multiple days to complete; however, not all standardized assessments are long. Schools and authorizers should thoughtfully consider what standardized assessments to administer, in addition to statewide assessments, for which learning outcomes and when.

The assessment delivery method and format can greatly impact how quickly results are available as well as the type of learning best measured. Computer-based assessments typically can provide very quick results, whereas pencil and paper tests likely take longer. Furthermore, multiple choice tests are easy to grade and more useful for certain concrete learning areas, whereas open-ended response questions likely take more time to evaluate and are suited to more complex skill assessment.

Putting too much weight on any given assessment has the potential to create undue pressure and/or anxiety for students and teachers. Schools should consider the way assessments, and the preparation for them, are communicated to students, to ensure that they are trying their best, so results are accurate, without causing undue burden.

REFERENCES

[Assessment Design Toolkit](#)

[A Policy Maker's Guide to Assessments](#)

[Testing Our Schools: A Guide for Parents](#)

[← Back to Sample Measures](#)



CTE/Industry Credentials are certifications students can earn through coursework, apprenticeship hours, and/or assessments that employers and post-secondary institutions accept as demonstration of workforce readiness skills.

OUTCOMES TO MEASURE

Career and Technical Education (CTE) has become increasingly recognized as an avenue toward post-secondary success for high school students. Federal and state governments have created new workforce readiness opportunities through Perkins Grants and changes to graduation requirements. Beyond creating CTE pathways, there has been a push for students to earn industry-recognized credentials or certifications that employers can use to assess a student's readiness for a job. Students can meet these requirements through industry assessments like ACT WorkKeys or the Armed Services Vocational Aptitude Battery (ASVAB). But it is increasingly popular to accept industry certifications such as Adobe Certified Associate Photoshop, Autodesk Inventor, OSHA 10, or the National Health Science Certification.



STRENGTHS

Measuring industry credentials is relatively easy because a student either passes the requirements and receives the credential or does not. Depending on how the credential is administered, the school may receive this information directly. In other cases, schools have to work with students or outside organizations to obtain the information about students who earn industry credentials. These credentials are seen as attractive because they are supposed to be externally valued by employers, giving them credibility with a number of stakeholders. More importantly, industry credentials acknowledge that students may be on different post-secondary pathways. This provides an opportunity for students to spend time working toward the pathway which interests them the most.

CHALLENGES

The evidence base linking industry certifications to higher employment rates in related fields, higher pay, or higher levels of college enrollment and persistence for students is mixed and can vary with the type of certification a student has achieved (e.g., cosmetology, health science, technology, agriculture, etc). For schools that focus on CTE and early career readiness, industry credentials may be highly credible because stakeholders are choosing the school specifically for the purpose of earning them; however, this may not be the case in other types of schools.

OTHER CONSIDERATIONS

The overall credibility of industry certifications is strengthened by how they connect to employment opportunities. Schools can develop relationships with local employers who require the credentials made available to students. Without these direct connections between credentials and employers, students are less likely to find a job that actually uses the credential they earn. For example, students earning a National Green Infrastructure Certification should have ways to connect to potential employers who require that certification. Course offerings and their outcomes must be connected to post-secondary opportunities, or they cannot actually hold value for students or their communities.

REFERENCES

[Littleton Public Schools Approved Certifications](#)

[DCPS CTE Industry-Recognized Certifications](#)

[How Industry Based Certifications in High School Relate to Other Outcomes](#)

[How Attaining Industry-Recognized Credentials in High School Shapes Education and Employment Outcomes](#)

[Work Readiness Certification and Industry Credential](#)

[← Back to Sample Measures](#)



IEP goal attainment measures whether a student in special education has successfully met the goals defined in their Individualized Education Programs (IEPs).

OUTCOMES TO MEASURE

Students with disabilities have legally mandated Individualized Education Programs (IEPs) that define their present levels of performance, areas of academic and non-academic need, goals to address their needs, and the supports and services needed to ensure students meet their goals. IEP goals are drafted by special educators and typically aligned with grade level standards. Student IEP teams, composed of, at minimum, a school administrator, parent/guardian, general education teacher, special education teacher, and the student (as appropriate), agree upon student IEP goals and how they will be measured. Because IEPs are designed specifically for each student, they are rarely the same across students. Furthermore, IEP goals are not always academic, and once students reach at least 16 years old, IEPs must contain post-secondary transition goals. IEP goal attainment can thus be defined in a variety of ways. An overall average percent of IEP goals met would calculate the average percent of IEP goals all students in a special education program meet. This average percent could just include academic goals. A school could also measure the percent of students who met all their IEP goals.

STRENGTHS

IEP goals address the diverse needs of students with disabilities in a school. For schools that specialize in educating students with disabilities or continuously enroll a large percentage of students with disabilities, IEP goal attainment can be an effective way to measure the performance of this student group, and particularly students with higher needs, given that statewide assessments, even when modified may not align to students' individualized programs.

CHALLENGES

IEP goals and how we measure students' progress towards meeting them is highly subjective and variable. There can be large differences in IEP goal quality across special educators at the same school, based on experience and training. Compliance in special education does not often focus on IEP goal quality or ensuring that progress on goals is measured with quality. There are potential legal repercussions for schools if students do not meet IEP goals. This creates an incentive for schools to create IEP goals that are assured to be met. Given these challenges, this measure should only be used in special settings and circumstances.

OTHER CONSIDERATIONS

Authorizers and schools should carefully consider if they want to include these types of measures, as IEP goals are designed to be supportive, formative assessments of student progress. Inclusion of these goals in an accountability system may create unintended impacts on the students who need the most support. Furthermore, if IEP goal attainment is included as a measure, school performance in this area should be viewed in relation to disaggregated performance on statewide assessments and other measures for students with disabilities.

REFERENCES

[Teacher-reported IEP Goal Data Collection Methods](#)

[IEP quality and Academic Achievement](#)

[Measuring and Reporting Progress Toward Mastery of Annual Goals](#)

[← Back to Sample Measures](#)





Credit attainment is a count of the number of credits a student has earned in order to graduate from high school.

OUTCOMES TO MEASURE

High school students earn credits that help them meet graduation requirements. States typically measure credit accumulation for graduation in Carnegie Units (CUs), with each course being worth one CU. Each state has different graduation requirements, but Local Education Agencies (LEAs) can also set more rigorous graduation standards than their state's. Credits for graduation also have to be in specific subject areas. For example, some states require art credit or four years of all core subjects. When measuring credit attainment as an outcome, it is important to take into account these variations in requirements. Instead of actually measuring the number of credits students have earned, it will be more accurate to measure the percent of students who have met all credit requirements or each type of credit requirement. For instance, 95% of students met all of their math credit requirements (4 CUs) by the end of 12th grade.

STRENGTHS

High school credits are easily collected without worries about fidelity. This data lives in all student information systems and can be easily analyzed. For schools that specialize in serving students who are at risk of not graduating, credit attainment is a particularly credible measure that aligns with the real goals of students, families, and schools.

CHALLENGES

Graduation course requirements vary significantly across states, and, for charter schools, across LEAs. If there is a desire to compare this outcome across schools, this outcome may not be ideal. Additionally, earning credit does not necessarily mean students are learning, increasing their skills, or receiving a high-quality education. Using credit attainment for accountability can create additional incentives for schools to give students grades right above the cutoff for getting credit. For example, schools that assign credit if students earn at least a D in a class may end up giving students a D instead of an F, regardless of student success. In addition, different students take different courses to obtain their credit. Some students take AP or Honors courses, others take college preparatory courses, and some students with disabilities may take modified courses. Without specificity or care, aggregating all these types of courses will mask the access and rigor of the classes students actually take to get credit.



OTHER CONSIDERATIONS

Some students may be on alternative graduation pathways with different requirements than their peers, based on state policy or Individualized Education Program (IEP) team decisions. These differences, if overlooked, may skew a school's data. Credit attainment is also not a useful measure outside of the high school context, as elementary and middle school students do not accrue credits the same way high school students do.

We know that completion of specific courses, like 9th grade Algebra, or being on track to graduate based on local graduation requirements, are strong indicators of on-time high school graduation. Schools should analyze their own requirements to understand which credits a student must have at the end of each school year to identify who would be on- or off-track to graduate. This approach to measuring credit attainment may be far more valuable for a typical high school than simple credit accumulation.

Early college high schools or schools that prioritize dual enrollment may consider focusing on college credit attainment rather than high school credit attainment. For example, a school and authorizer may set a target for the percentage of students across all identified student groups who achieve at least one college credit upon graduation from high school. This could potentially be through dual enrollment courses or successful scores on AP or IB exams.

REFERENCES

[Credit Structure of American Schools](#)
[On-track Indicators to Predict Graduation](#)

[← Back to Sample Measures](#)



Grade Point Average is the average of grades earned in courses based on A=4, B=3, C=2, D=1, and F=0 points, weighted by the number of credits each course is worth.

OUTCOMES TO MEASURE

Grade Point Average (GPA) is commonly calculated starting in high school for students as a continuous average of a student's grades over their high school years. Different institutions calculate GPA slightly differently based on whether they give different numerical values to plus or minus grades as well as additional GPA points for Advanced Placement (AP), International Baccalaureate (IB), dual enrollment (college-level), or honors courses. GPA can also be measured on a semesterly or annual basis as a way to measure trends as a school grows, or to measure the health of a school's program at a point in time. Cumulative GPA has gained particular importance as university admissions have become more competitive. Some universities now have an average GPA of accepted students that is over a 4.0 as a result of the many advanced courses students take, some of which give students a 5.0 instead of a 4.0 for an A. A school's grading policies can have implications for GPA as some people are increasingly skeptical about the impact that grade inflation at some schools can have on GPA. Nevertheless, GPA continues to be an exceptionally good indicator of college success, often as good, if not better than, entrance exams like the SAT or ACT that were designed to be indicators of future college success.

STRENGTHS

GPA is a relatively easy measure to regularly calculate for any high school. Most students quickly learn how to calculate their own GPAs and use it as a motivator for improving over time. Because of its wide use by students, schools, and colleges, GPA is widely accepted by almost all stakeholders as a credible measure of student success. GPA has the additional benefit of being very good at predicting longer-term success, particularly for students with less than a 3.0 GPA at the end of high school. Students with low GPAs in high school likely have not learned important academic and non-academic skills that help individuals be successful after high school.

CHALLENGES

GPA, potentially, contains considerable bias, based on the courses made available to students as well as local grading policies, if trying to compare across schools. Students who do not have access to advanced courses or dual enrollment opportunities are at a disadvantage because they cannot earn an extra grade point to contribute to their average. If assessing one school over time, however, this may be a good way of showing increased access to these types of classes at a given

school. With the importance of GPA, grade inflation has been an increasing concern. Some schools see it as a testament to their rigor when few students earn As in courses; other schools may see it as their goal that all students get As. These differences can potentially have considerable consequences for the comparison of GPA across schools.



OTHER CONSIDERATIONS

GPA is not typically used in meaningful ways in middle school when students are not getting course credit toward graduation. However, middle schools that do use GPA may be able to use it as a measure of student success. Some schools use alternative grading methods that may not assign grades to students who achieve below a certain standard, in an effort to use a revision process to improve mastery of content. For example, a school may only assign grades if they are above a C and consider all other courses still in-progress until the student revises work to meet the standard of a C grade. This type of practice can skew cumulative GPA data for particular schools if in-progress courses are not considered in the calculation of cumulative GPA. Credit/No Credit classes can similarly cause GPAs at a school to skew slightly, compared to other schools.

REFERENCES

[Predictiveness of High School GPA on College Success](#)
[Predicting Long-Term College Success Using ACT and GPA](#)
[High School GPA Beats Out ACT in Predicting College Readiness](#)
[What Matters Most for College Completion](#)

[← Back to Sample Measures](#)



Attendance is a count (or percent) of days or class periods in which a student was identified as present in, or absent from, school.



OUTCOMES TO MEASURE

Attendance as an outcome measure can be calculated in a variety of ways. The most common may be the overall attendance rate of all students, as well as disaggregated attendance rate for identified student groups. It has become increasingly important to measure the percentage of students who are chronically absent, typically defined as missing at least 10% of school days in a year. Attendance itself is not typically considered an outcome because it is necessary for other outcomes to be possible. Increasingly, attendance is seen as a proxy for engagement and a necessary measure to assess how well schools are meeting students' needs. Reducing chronic absenteeism is often a key outcome of interest to policymakers, as students who are chronically absent are most likely to be from under-resourced communities.

STRENGTHS

Attendance measures are particularly attractive because the data is already collected with fidelity, as a result of its relationship with state requirements and school funding. In addition, using attendance data is generally easy, because all schools have a set number of instructional days, making it easy to create an average attendance rate.

CHALLENGES

Since there is some variation in how attendance can be defined, it is important to create a clear definition and to collect the data based on that definition, with fidelity. Attendance can be defined as a student being present at any point during a school day, being present from morning bell to dismissal bell, or being in attendance for the majority of the day. Additionally, with the prevalence of online or virtual learning during the height of the pandemic and since, definitions of attendance are evolving. If schools will be compared to one another using attendance measures, each school should use the same definition. Even within schools, leadership must ensure that staff have a shared application of attendance definitions.

OTHER CONSIDERATIONS

While attendance is highly correlated with student achievement, research has consistently found that teachers who impact student attendance are not typically the same teachers who impact achievement. This may mean that schools need to enact different interventions, programs, and/or supports to improve student achievement versus student attendance. Additionally, authorizers and schools must determine what are acceptable attendance measures and targets, if those targets are the same for all schools, and/or if schools are able to individualize attendance targets for students based on students' engagement/attendance history (particularly for alternative high school programs).

REFERENCES

[Relationship between Teachers and Student Attendance](#)

[How Measurement and Modeling of Attendance Matter](#)

[Reducing Chronic Absenteeism under the Every Student Succeeds Act](#)

[← Back to Sample Measures](#)