NYS 21st CCLC
Findings from Year 2 Annual Evaluation Reports (2018-19)

Submitted by Measurement Incorporated
July 2020

Table of Contents

Table of Contents		
PART I.	**Introduction**	2
PART II.	**Notable Strengths of AERs**	3
A.	*Point of Service Observations*	
B.	*Conclusions and Recommendations*	
B. (1)	*Most Common Themes Among Conclusions & Recommendations*	
B. (2)	*Quality of Recommendations*	
PART III.	**Notable Opportunities for Improving AERs**	7
A.	*Usefulness of Performance Indicators*	
B.	*Appropriateness of Assessment/Measurement Tools and Alignment with Performance Indicators*	
C.	*Alignment Between Performance Indicators, Supporting Activities and Goals/Objectives*	
D.	*Reporting of Status of PIs*	
E.	*Communications Between Evaluators and Stakeholders*	
E.(1)	*Reported Communications*	
E.(2)	*Implications*	
PART IV.	**Summary and Recommendations**	19
Appendix	**: Year 2 AER Review Template**	22

I. Introduction

Measurement Incorporated (MI) is contracted with the New York State Education Department (NYSED) for the statewide evaluation of New York's 21st Century Community Learning Centers (21CCLC) program. As required under this contract, MI's evaluation team conducts a comprehensive review each year of a selection of sub-grantees' Annual Evaluation Reports (AERs). This review is implemented each year for a stratified random sample of reports, representing approximately 20 percent of sub-grantees. A structured AER Review Template (see Appendix A) is used to ensure consistency of approach across projects as well as among reviewers.

The Year 2 reports (2018-2019) contain a panoply of essential information, including data collection and analysis methodologies and reports on the success with which they were implemented; study findings from both qualitative and quantitative sources, including the degree to which the programs' local objectives were met; and conclusions and recommendations for ongoing program improvement. These data are invaluable to program managers "at all levels of the system,"[1] from sub-grantees and their teams of service providers, to regional Resource Centers and District managers, to the state program coordinator and management team, and to the federal funding director at the U.S. Education Department.

Because of the importance of these documents, stakeholders at all levels are encouraged to take advantage of the findings from this review process to ensure the highest possible quality of local evaluation and program implementation. While this summary provides an overview of the findings from a selected sample of AERs, the recommendations and principles are broadly applicable and relevant to all program staff, local evaluators, and key stakeholders at the local level. Awareness of these general findings can help these leaders recognize the strengths and challenges in their own program implementation, evaluation design, and reporting processes.

This year's review revealed a number of significant strengths, as well as challenges in the selected AERs. Some of the most notable strengths are summarized in Section II below. Areas in which reporting quality more often could have been strengthened are highlighted in Section III.

It should be noted that this summary is not a comprehensive report of this year's AER reviews; it focuses on sections with findings and recommendations we believe offer high leverage for the implementers and consumers of evaluation to reflect on quality and improvement.

1 As noted in the Evaluation Manual.

II. Notable Strengths of the AERs

Across the sample, two report sections stood out as consistently strong and well formulated: Point of Service Observations and Conclusions and Recommendations

A. Point of Service Observations

Across programs, reporting on the first annual program site visit focused on a variety of different topics, ranging from program management and administration, to fidelity of program activities to the original proposal, to assessments of the quality of instruction and level of student engagement. This variability was a direct result of the fact that SED, MI and the local evaluators collectively agreed that the original purpose of the first site visit as stated in the Evaluation Manual – to inform the evaluability process – did not apply in the same way after the first year, but clear guidance on what should take its place was not provided in time for the fall 2018 visits. The different foci of these various reports were undoubtedly useful to individual programs, but since consistency across programs had not yet been established, these site visits were not closely assessed during MI's Year 2 AER reviews.

For the second round of visits (generally spring 2019), however, all programs appropriately focused on conducting point of service observations. The large majority of programs continued to use the Out of School Time (OST) observation instrument or an adaptation thereof, while most of the remainder used some other, locally developed observation instrument.

While the scope of the local evaluations, given budget limitations and time constraints, was not sufficient to obtain observations of a representative sample of activities, the observations that were conducted were mostly of very high quality. Local evaluators did observe a range of different types of activities, to the extent that time and scheduling allowed. Typically, they provided numeric ratings and/or qualitative assessments on a number of site- and activity-specific characteristics (many of which were derived from the OST instrument). Assess characteristics included participation level, scheduling issues, environmental context, youth and staff relationship building, instructional strategies, youth participation and engagement, and activity content and structure, among others. In most cases where numeric ratings were used, they were also accompanied by narrative that provided context, depth and relevance to the numbers. Most observers reported circulating through the space and observing and/or talking with individual students or small groups as they worked.

Frequently, the narrative part of the observations often included several recommendations, such as how a particular activity could be adapted to provide more targeted support to meet needs of student sub-populations. Typically, such suggestions were also linked to broader findings reported in the Conclusions and Recommendations sections of the AERs.

The few shortcomings noted in the Observations section of these reports included some cases where numeric ratings were not accompanied by narrative. Numeric ratings in the absence of rich and meaningful narrative can be harder to interpret and are potentially less useful for the consumers of the report. In other cases it was noted that the formal observation process did not appear to have utilized a structured protocol. The Evaluation Manual and Addendum require that point of service observations must use a structured protocol (preferably the OST), to ensure that all observations within a program are captured and documented consistently and provide results that can be compared and summarized across activities. Ideally, all point of service observations should utilize a structured protocol, and observation report summaries should provide *both* numeric ratings and narrative descriptions.

B. Conclusions & Recommendations

The Conclusions & Recommendations sections in the reviewed reports represent another area of consistently high quality. This section invited evaluators to synthesize the data and summarize program highlights and achievements, reflect on what was learned from the year's evaluation activities, and recommend changes to improve program implementation, program outcomes, and potentially the evaluation itself, as needed. Prevalent themes that emerged from the reviewed recommendations are presented below. This is followed by a discussion of the characteristics of these narratives that distinguished them.

1: **Most Common Themes** among Conclusions & Recommendations:

The Conclusions and Recommendations sections of the Year 2 AER template required a discussion of how recommendations from Year 1 were implemented and what, if any, impact they had on the program. About three-fifths (58%) of reviewed reports included a discussion of the status and impacts of Year 1 recommendations. This section was also used to report on recommendations resulting from Year 2 evaluation activities. All reviewed reports (100%) included recommendations for improvements based on Year 2 findings. The most commonly occurring themes among these recommendations are summarized in Table 1, below.

TABLE 1

***Most Common Themes** Emerging from the Recommendations in the Year 2 AER Sample*

Recommendations*	Percentage of AERs
Increase parent and family involvement in the program	76%
Increase stakeholder participation in the Advisory Board	43%
Improve data systems and recording processes	38%
Improve survey tools, survey administration and/or survey response rates	33%
Increase student input in programming	29%
Increase student recruitment and enrollment	24%
Provide more challenging activities to participants, and include more time for discussions	24%
Increase students' program attendance	14%
Develop framework for sharing best practices, within a program and/or among other 21C programs	14%
Improve staff communication with students and behavior management strategies	10%
Include more college and career focused activities	10%
Improve or increase community partnerships	10%
Improve utilization of the program logic model and/or evaluation plan	10%
Improve program alignment with the school day	10%

** Recommendations were content analyzed and grouped by construct or theme.*

2. Quality of Recommendations

In virtually all cases, recommendations emerged directly from the evaluation findings and identified clear aspects the program could focus on improving next year.

Many reports distilled recommendations into concrete, specific and actionable suggestions – *e.g.*, proposed solutions or change opportunities that program staff could consider implementing. This approach exemplifies NYSED's expressed intention that local evaluation be accessible for program staff (consumers) and utilization-focused to provide actionable, developmental, improvement-targeting feedback.*

For example, in one case in response to low survey participation rates, one report recommended that:

> *...clear expectations with program staff about the survey should be set to ensure that administration across all students is done. Digital surveys should be considered if distributing, collecting, and delivering completed surveys is difficult in paper format.*

In other cases, the reports featured more general recommendations, framed as gaps or need areas identified in the assessment of the data that invite program staff to reflect and explore change possibilities.

For example, in a report for a program that had not administered the required participant surveys at all, the recommendation was stated concisely as:

> *Implement year-end surveys.*

Such contrasts in style might reflect differences in the nature of the evaluation contract, the relationship between the evaluator and program manager, and/or the context of the recommendation (as in the second example, above). Neither style is necessarily right or wrong, although some degree of specificity is generally desirable, where possible and appropriate. However, even if concrete detail is not included in the AER recommendations, evaluators should provide such support to their client through activities such as participation in advisory committee meetings and action planning process, and/or through other modes of (perhaps less formal) communication.

*Sub-grantee RFP #: GC17-001, pp. 14-15; Elizabeth Whipple, NYS Project Manager, Statewide Conference Presentation; Schenectady, NY; November 20, 2019.

III. Notable Opportunities for Improving AERs

This section provides examples of AER content that follow best practice for each of several characteristics; in each case, it also contrasts these best practices with one or more examples that needed improvement to ensure clarity. It should be noted that all quoted text (such as PIs, objectives, etc.) is taken verbatim from one of the reviewed AERs. Where such text is not in quotes, it was paraphrased from one or more examples from the reviewed reports and was written to exemplify a recurring theme.

A. Usefulness of Performance Indicators (PIs)

Goals need to be clearly defined for the process to effectively demonstrate strong implementation/outcomes and support program improvement. PIs should possess all five elements of SMART goals: Specific, Measurable, Achievable, Relevant and Time-bound.

TABLE 2

Examples of the Different Degrees to which PIs reflect SMART Goal Elements

Examples demonstrating best practice	Examples needing improvement
Sample PI #1: **"50% of regular attendees will show a 5-point increase in ELA, math and science grades from first to last marking periods."** This PI is… ✓ **Specific** - indicates specific area of improvement for specific group (ELA, math and science for all regular program participants) ✓ **Measurable** - a 5-point increase is clearly defined, also clear about what percentage will reach this goal ✓ **Achievable** - realistic for a program focused on ELA, math and science to accomplish this ✓ **Relevant** - academic outcomes are relevant to the project ✓ **Time-bound** - specifies gains from fall to spring	*Sample PI:* **"100% of students will have a comprehensive portfolio; 80% will maintain or improve class performance; 80% will improve homework completion & class participation."** This PI is… ✘ **Not Specific** - The terms "comprehensive" and "class performance," are not defined. ✘ **Not Measurable** - Can't measure % with a "comprehensive portfolio" because not defined. There is no indication of how "improved performance" or "participation" will be measured. ✘ **Not achievable** - 100% target is generally unrealistic for most indicators. Compound PIs such as this (one PI includes multiple, separate criteria) are much harder to achieve because they require all criteria to be met.

TABLE 2 continued

Examples demonstrating best practice	Examples needing improvement
Sample PI #2: **"80% of regular attendees who had at least 2 disciplinary referrals during prior academic year will have reduction in # of disciplinary referrals during the current academic year."** This PI is… ✓ **Specific** - indicates specific area of improvement for specific group (reduced disciplinary referrals for all regular program participants who had at least 2 prior referrals) ✓ **Measurable** - Comparison in counts of referrals ✓ **Achievable** - this may be an optimistic goal but not unrealistic for a targeted and high quality social-emotional program to achieve. ✓ **Relevant** - behavioral outcomes are relevant to the project and to the federal and state indicators. ✓ **Time-bound** - specifies gains from Y1 to Y2	ℹ **Relevant** - Indicators appear relevant to 21st CCLC program but cannot determine for certain without information about specific goal being assessed. ✘ **Not Time-bound** - Does not indicate by what date students will have completed their portfolios, nor the time frame during which class performance, homework completion and class participation will improve.

Table Key: | ✓ meets standards/expectations | ✗ does not meet standards/expectations | ℹ unclear/need more information |

B. Appropriateness of Assessment/Measurement Tools and Alignment with Performance Indicators

Just as all performance indicators (PIs) must be defined in ways that are *measurable* (see discussion of SMART goals above), the assessments used to measure them must be appropriately chosen to obtain data that are reliable and aligned with the performance indicator. Goals/indicators need to be meaningfully assessed for the process to effectively demonstrate strong implementation/outcomes and support program improvement.

TABLE 3

Examples of the Different Degrees to which ***Assessments/Measures Align with PIs*** *and are* ***Appropriately Selected***

*Organized Into *Focus* Areas*

Focus: ***The data source must generate data that responds directly to both the content and the metric of the PI***

Sample PI #1: "70% of students in K-5 who attend this activity at least 75% of the time achieve gains in social-emotional development."

Examples demonstrating best practice	Examples needing improvement
✓ The survey used to assess gains includes questions that clearly address social-emotional development. ✓ The survey utilizes a balanced, 5-point scale to assess change. ✓ The survey is administered in both fall and spring to the entire targeted population OR assesses retrospective impression of change. ✓ Only those students who attended the activity at least 75% of the time are included in survey results.	✗ Survey questions were developed separately from the activity, and do not directly address questions about social-emotional development. ✗ The survey is administered in the spring only and assesses current beliefs or skills at that single point in time with no baseline reference. ✗ The survey is administered fall to spring but uses a 3-point scale (e.g. definitely, kind of, not really), so that smaller gains cannot be detected.

TABLE 3, continued

Sample PI #2: **"70% of students who attend this activity at least 75% of the time achieve gains in ELA performance, as measured on standardized test results."**	
✓ Using an assessment that compares pre-post (e.g. spring to spring) gains. ✓ Alternative measures, such as report card grades, surveys, teacher reports, or focus groups for participating students who are not tested both years (e.g. grades K-3).	✗ Using an assessment that only looks at end of year test results (e.g. 75% achieve Performance Level 3 or above on spring English regents). ✗ Students who are not tested (e.g. grades K-3) are not included in reporting of results (no alternate assessments used).
Sample PI #3: **"100% of students will participate in enrichment activities."**	
✓ Measured directly through attendance records from enrichment activities.	✗ Assessed using the OST Observation tool; this tool only addresses descriptions and quality of activities/participation but does not measure attendance.

Focus*:** ***The data source(s) must be appropriate for, and must address, the entire target population.

Sample PI: **"70% of students in K-5 who attend this activity at least 75% of the time achieve gains in social-emotional development."**

Examples demonstrating best practice	Examples needing improvement
✓ The survey is linguistically appropriate for K-5 students. ✓ The survey is only appropriate for upper elementary, but data is gathered for K-2 students through focus groups or observation.	✗ The survey language is not appropriate for K-2 but is given to them anyway. ✗ The survey is not administered to grades K-2, but no alternative measure is used for those grades.

Focus*:** ***The data source needs to generate enough information to be at least reasonably representative of the target population. Low response rates/data availability should be taken into consideration when interpreting results.

Examples demonstrating best practice	Examples needing improvement
✓ Survey responses are received from at least 50% of the target population. For program improvement purposes, interpretation of results takes into consideration the fact that half the population is not represented.	✗ Very low response rate (e.g. less than 20%) makes it impossible to attribute the results to the whole population, but the PI result is stated without caveats.

Table Key: | ✓ meets standards/expectations | ✗ does not meet standards/expectations | ⓘ unclear/need more information |

C. *Alignment Among Performance Indicators, Supporting Activities and Goals/Objectives*

Regardless of how well written the Performance Indicators are and how well designed the assessment measures are, ultimately the PIs must align with and inform the success of program activities and goals/objectives.

TABLE 4

Examples of the Different Degrees to Which There was ***Alignment Between PIs, Objectives, and Activities*** *Organized into "Focus" Areas*

Examples demonstrating best practice	Examples needing improvement
Sample Objective: ***"Regularly participating students will improve their college readiness"***	*Sample Objective:* ***"Run youth development, social-emotional learning and recreation activities"***
PIs for this Objective: • "At least 80% of regularly attending participants will report a higher level or sustained above average level of motivation and expectation to attend college." • "At least 90% of graduating seniors will submit at least 6 college applications." • "At least 80% (who are eligible) will complete FAFSA forms." **Listed activities to support this objective:** ✓ Individual high school choice and college access counseling sessions and group high school choice and college access workshops.	**PIs for this objective were** • "525 students will perform in an exhibition, play or concert." • "260 will participate in specialized programming for ELLs." Listed activities to support this objective: ✘ Physical activity, healthy living instruction, project-based learning clubs, and arts instruction for grades K-12.

TABLE 4, continued

Focus*:** ***Activities need to support the objective			
✓	Cited counseling and workshop sessions seem appropriate to developing college readiness	ⓘ	It is not clear which of these activities would be counted as SEL.
Focus*:** ***PIs need to be aligned with entire objective			
✓	PIs address the whole targeted population (not just seniors).	✗	The first PI, which seems to apply to the whole program population, does not address SEL.
		✗	The second PI, which only addresses ELLs, seems to imply that all listed activities include adaptations or special classes for ELLs but that is not explicitly stated, and it is not explicit which of these might support SEL.
Focus*:** ***PIs need to align with the activities			
✓	Indicated activities can reasonably be expected to result in improvements in these PIs.	ⓘ	It is not clear whether indicated activities related to the participation targets listed in the PIs. (participation in performances, and in activities for ELLs)

Table Key: | ✓ meets standards/expectations | ✗ does not meet standards/expectations | ⓘ unclear/need more information |

D. Reporting the Status of PIs

When reporting the status of the Performance Indicators, results of the PI should be described in detail, making appropriate use of the available terms: Met, Not Met, Partially Met, Data Pending, or Not Measured. (Note that while "Not Measured" should be used where appropriate, the expectation is that all PIs should be measured). The explanation should be specific enough to support the result reported so the program can clearly see where they are excelling and where they need to improve. The results need to specify the number/proportion of participants meeting the Performance Indicator, the assessment used, and the timeframe represented by the results.

Although there is currently no official category for it, it is perfectly appropriate to use the "Explain [PI Status]" column to describe progress the program is making towards the PI even if it has not yet been met.

TABLE 5

*Examples of the Different Degrees to Which There Was **Alignment Between PIs, Objectives, and Activities***

Examples demonstrating best practice	Examples needing improvement
Focus*:** ***Status of data availability must be accurately represented, and results updated when data become available.	
Sample PI specifies gains **"as measured by standardized tests."**	
✓ Results for the current year are reported as "Data Pending" because spring standardized test data are not provided until after AER is due. ✓ Prior year results for the same PI are updated.	✗ Results for the current year are reported as "Not Available" because standardized test data are not provided until after AER is due. It should have been reported as "Data Pending." ✗ No updates are provided for prior year results. The PI should have been updated with prior year results since Year 1 spring test data would have been available by then.

TABLE 5, continued

<table>
<tr><th>Examples demonstrating best practice</th><th>Examples needing improvement</th></tr>
<tr><td colspan="2">*Focus*: The data sources and outcomes that were used to establish that a PI was met must be described.</td></tr>
<tr><td>PI: “60% of enrolled students will earn 3 or more credits towards HS graduation.”

✓ Results reported as “Met”: “96% of all enrolled students earned 3 or more credits.”</td><td>PI: “90% of students will participate in an internship.”

✗ Results reported as “Met”, but no explanation provided. ”Explain” column needs to be used to justify the conclusion of “Met.”</td></tr>
<tr><td colspan="2">*Focus*: The results should use the available terms (Met, Not Met, Partially Met, Data Pending, Not Measured) appropriately, as defined.</td></tr>
<tr><td>PI: “The program will survey students annually about activity preferences.”

✓ Results reported as “Partial”: “Site 1 submitted results from student surveys, but no data were received from Site 2.”</td><td>PI: “At least one parent/adult family member [of each participant] will attend classes held at the program or will participate in class or workshop.”

✗ Results reported as “Partial”: “Parents representing some students participated.” Program has only one site; result should have been “Not Met” since the “Partial” designation can only be used when there are multiple sites and one or more has met the Performance Indicator threshold.</td></tr>
<tr><td>PI: “At least 70% of regular attendees will maintain or improve positive school day behavior.”

✓ Results reported as “Not measured”. A brief teacher survey was sent out in February, with many teachers responding. However, with only one administration of the survey completed, the objective [improvement] cannot be measured.”</td><td>PI: “100% of students will report that the program helped them with decision making, leadership skills, and technology skills.”

✗ Results reported as “Not quite”: PS xxx: 87% said they think about how others are affected by their decisions. PS yyy: 76%.” Should have been “Not Met,” even if it was close. Note also that the explanation of the results does not address all parts of the Performance Indicator (especially technology skills).</td></tr>
</table>

TABLE 5, continued

Examples demonstrating best practice	Examples needing improvement
Focus*:** ***Response rates should be robust enough to determine whether or not a PI was met.	
Sample PI: "50% of regular attendees will demonstrate improved school engagement and self-responsibility."	
✓ Results reported as "Not measured." Only 1.5% of parents completed a survey assessing their children's behavior. The response rate is too low to accurately assess whether or not the PI was met.	✗ Results reported as "Met": 50% of parents indicated regular attendees demonstrated improved responsibility/confidence." This was based on a response rate of 1.5%, much too low to attribute the results to the whole population. Moreover, the PI includes reference to school engagement, which does not seem to be measured. Should have been "Not Measured."

Table Key: | ✓ meets standards/expectations | ✗ does not meet standards/expectations | ⓘ unclear/need more information |

E. Communications Between Evaluators and Stakeholders

Section V of the Year 2 AER Template requires the evaluator to describe the processes and products they used to communicate formative and summative findings to program stakeholders.

Some, but not all, of these communications are explicitly required. The Evaluation Manual states that Annual Evaluation Reports are to be completed in collaboration with the 21st CCLC program director. The Manual also states that evaluators are required to attend at least four advisory meetings annually, and to report on evaluation activities during these meetings. Furthermore, the Round 7 RFP states that "Program and partnering administrative staff, school and partnering agency representatives, students, parents and community members should be represented on the program advisory committee." If the program director has provided for such representation, it would create an opportunity for evaluators to communicate with these stakeholders as well.

1: Reported Communications

A tabulated summary of reported communications activities is presented in Table 6 below. Percentages in this table indicate the proportion of reviewed AERs that reported that the evaluator used a particular communication method with a particular stakeholder group (% Yes), or the proportion for which the information was not provided (% Don't Know).

TABLE 6

Types of Communications Used by Evaluators to Engage With and Inform Stakeholders

LEGEND: Yes · Don't Know

Types of Communication From Evaluators[(a)]	Program director, site director, and/or CS director	Front line staff (service providers)	Principal and/or other school leaders	CBO partners	Parents	Students	Advisory Council
Annual Evaluation Reports	76% / 14%	19% / 29%	29% / 33%	29% / 24%	14% / 38%	24% / 38%	24% / 19%
Interim Evaluation Reports	71% / 14%	19% / 24%	24% / 23%	19% / 24%	10% / 28%	10% / 28%	14% / 14%
Memos/Weekly or Monthly reports	48% / 5%	19% / 10%	19% / 10%	24% / 10%	10% / 5%	10% / 5%	5% / 10%
Advisory Group meetings[(b)]	76% / 19%	52% / 19%	47% / 24%	57% / 24%	42% / 29%	29% / 29%	
Other meetings	81% / 0%	38% / 5%	19% / 5%	48% / 0%	19% / 10%	10% / 5%	5% / 0%
Email/phone	90% / 0%	14% / 0%	14% / 0%	24% / 0%	100%	100%	

[(a)] Additional communications noted in the AERs included Observation/Site Visit Narrative Reports, Annual Kickoff Meetings, and YouthServices.net training. These were most often provided to program directors, sometimes frontline staff, and in one case the Principal, CBO partners, parents, and students.

[(b)] Among NYC sub-grantees that are part of the NYCDOE Office of Community Schools, OCS meetings took the place of 21st CCLC-specific Advisory meetings.

- In the large majority of reviewed programs, evaluators maintained very active communication with program directors using multiple methods (including the AERs).
- While the Advisory Council is required to include Principals, parents, and age-appropriate students, it appears from the reviewed sample that this may not always happen. In fact, building leader, parent, and student participation was reported in fewer than half of the reports reviewed – limiting opportunities for evaluators to communicate with these stakeholders.
- Principals are not required to receive the AER but almost one-third of principals appear to have seen it.
- Frontline staff were often not included in formal, explicit communications about program status. The mostly commonly reported engagement of frontline staff, according to about half of the reports reviewed, was through Advisory Council meetings.
- In up to 2 out of 5 cases, reports did not specify who was included in each type of communication. Without this information, it is impossible to fairly assess the extent to which stakeholders are participating in critical communications about the status of the program.

IV. Summary & Recommendations

The large majority of AER reports which we reviewed for Year 2 provided very high-quality summaries of point of service observations and, perhaps most importantly, provided targeted, useful and often quite specific recommendations to support program improvement. Unfortunately, we found a considerably larger number of problems in the Evaluation Plan and Results section, as well as the Engagement and Communications section. The above findings about these sections lead us to the following recommendations that will help make future Annual Evaluation Reports more useful to stakeholders at all levels of the system.

Observations:

- All point of service **observations are required to utilize a structured observation protocol**. The OST is preferred, but an adaptation thereof, or another instrument of similar quality can be used instead.

- Ideally, **observation summaries should include both numeric ratings and a narrative summary**, as the narrative helps explain and provide depth to the ratings. A well-written narrative can serve the purpose by itself, but generally a numeric rating that is not accompanied by a narrative explanation is of limited usefulness.

Conclusions and Recommendations:

- As appropriate to circumstances, **include sufficient detail in recommendations** that make it clearer to program staff how to implement them. Too much specificity may actually be limiting if recommendations are prescriptive, and impractical if they do not incorporate current realities of a program's resources and capacity to implement change. However, considering the State's explicit intent of using local evaluation findings to improve programming during the life of the grant, attempts should be made to frame recommendations in actionable, practical terms so that busy leaders can use them to gain greater clarity and insight into real-time change opportunities.

- Evaluators need to be sure to **follow up with any recommendations** through active participation in the Advisory Committee and by maintaining active communications with all stakeholders.

Performance Indicators (PIs):

- **Performance Indicators must meet all five SMART goals.** They should be written in such a way that they help inform program improvement, and that anyone not familiar with the program would know exactly how they should be measured. They should define outcomes that would be expected to be impacted by the cited activities. PIs can be re-written through modification requests if any of the originals are inadequate.

Tools for Assessing the Status of PIs:

- Assessment and measurement tools must be **valid and reliable and** must be **aligned with the PIs** as written.

- Assessment and measurement tools must **address the entire target population cited in the PI**. For example, assessments for native English speakers in grades 3-8 may not be available for, or not appropriate for, students in other grades, or students who are not native speakers of English.
- Assessments must **generate high enough response rates** to be at least reasonably representative of the population. Extremely low response rates should not be considered as having measured the PI. Middling response rates (e.g. between 50%-69%) that may be considered a meaningful measure of the PI should nevertheless be interpreted very cautiously with the caveat in mind that they do not reflect as much as half the target population.

Alignment among PIs, Activities, and Objectives:

- **All aspects of a given objective must be assessed by at least one PI.** This is often an issue when the objective includes different sub-populations (e.g. grades, or language groups -- see above). It is also relevant when "Compound PIs" are used that cite more than one target.

- Because it is more difficult to meet all the targets in a compound PI, it is always preferable to **split Compound PIs into multiple, separate PIs.**

- PIs must **define outcomes that would reasonably be expected to be impacted by the cited activities**.

- **Objectives must reasonably be expected to be impacted by the cited activities.**

Accurately Report Status of PIs:

- PIs that require State or District **data that have not been released in time for the AER should be reported as "Data Pending,"** *not* as "Not Measured" (and they should be measured in the following year once data are released).

- **PIs can only be reported as "Partial" if they were *fully* met in at least one site** in a multi-site program.

- **A result that almost meets the target should still be reported as "Not Met".** However, the fact that results were very close is important to interpreting them for program improvement; evaluators are strongly encouraged to explain this situation in the "Explain" column in the Evaluation Plan on the AER template, as well as in the Conclusions section.

- **A result that meets some but not all of the components of a Compound PI must be reported as "Not Met."** Separate it into multiple PIs instead.

- Measures that generate **extremely low response rates are not representative and should be reported as "Not Measured"** (see Tools for Assessing PIs, above).

Engagement & Communications

- **Keep track of all communications, organizing them by stakeholder group and by type of communication.** Maintaining communication among all stakeholders is essential to program improvement and cannot be assessed if it is not documented.

- **Share AERs, or at least summaries, with stakeholders other than the program manager more frequently** -- particularly with principals/school leaders, and front-line staff.

- **Explore strategies for improving advisory group participation.** While the cases represented here likely reflect multiple different situations, fewer than half of the reviewed programs were known to have involved school leaders, parents or students in their advisory groups, and only slightly over half were known to have involved frontline staff or service providers at these meetings.

AER Review Template 2020 (Year 2 AERs)

AER Review template 2020 (Y2 AERs)

GOALS - NEW TAB

I. PROJECT INFO

- Move drop down list of Reviewer names to row 10 instead of 9

II EVAL PLAN & YR 2 RESULTS

- Remove "goals/objectives" from cell A4 and A5
- Add NOTE about assessments in Row 12
- Change response rate question from Y/N to "What was the response rate?" in A 28
- Add question in 32D/E for reviewer to identify whether challenge was directly state or interpreted
- Add Y/N dropdown to BC23
- Add N/A to BC25
- Add NOTE about results in Row 36
- Add "Partial" as an option for goal results in Row 42 and 56
- Add "English and/or Math" goal category in between Math and Other Academic (Column F/G)
- Move "fidelity" from Implementation Quality to Level of activity (H/I to J/K)
- Move "Comments on Assessments" to 33A
- Add "Comments on Results" to 64A
- Add "How many goal ratings changed by reviewer" to Row 48 and 62

III. OBSERVATION RESULTS

- Add fidelity definition as outputs to A5
- Add "Modified Out of School Time (OST) Protocol" to Row 11
- Slight change in wording on A14

IV. LOGIC MODEL

- Add "Do connections between logic model components make sense?" to Row 10

V. ENGAGEMENT & COMMUNICATION

- Add "Advisory Coucil" in column H
- Add NOTE in Row 11

VI. CONCL & RECOMMENDATIONS

- Add "General Comments on AER Overall" to A4

Performance Indicator(s) (PI) of success	**Target Population(s)**	**PI Meets SMART Criteria? (Y/N)**	**Activity(ies) to support this program objective**	**PI Measures** data collection instruments & methods (Indicate title if published)	**Describe the analysis conducted, including specific results that directly address the PI.** Include any longitudinal assessments conducted beyond one program year.	**Response Rate** if applicable):	**Was this PI Met? (Yes, No, Partial, Data Pending, Not Measured)**	**EXPLAIN:** **If Yes, No or Partial:** present results (expressed in the same metric as the PI) **If Partial,** indicate # of sites where PI was fully met. **If data pending,** indicate when data expected. **If not measured,** explain why not.
COPY/PASTE EVAL CHART CONTENTS STARTING HERE								

Is PI fully aligned with goal/objective? (Y/N)	Does PI fully assess all components of goal? (Y/N)	Is the PI SMART? (Y/N)	If not SMART, which aspect not met?	Goal Type/Category	Goal Result (Met, Not Met, etc.)	Goal Result changed by Reviewer?

I. Project Information

Project number	
Program Name	
Lead Agency	
Name of Program Director	
Number of program sites	
Grade Levels Served	
Local Evaluator Name	
Local Evaluator Company	
Reviewer	

II. Evaluation Plan: SMART GOALS

"SMART" = Specific: targets a specific area of improvement; Measurable: has a defined target that can be assessed (can include qualitative assessment); Achievable: realistic given baseline conditions and available resources [note this may be difficult for State Evaluator to assess]; Relevant: aligned to program mission, program activities, school day academics, GPRA indicators, etc.; Time-bound: specifies when the goal will be achieved

	Rating		Comments	Examples
Objective 1: Do the formative performance indicators meet ALL criteria for "SMART" goals?		Indicate which aspect of SMART was most commonly not met and explain why		
Objective 2: Do the summative performance indicators meet ALL criteria for "SMART" goals?		Indicate which aspect of SMART was most commonly not met and explain why		
For goals that do not meet all SMART criteria, were limitations addressed?		If yes, how were they addressed? Discuss for both formative and summative goals.		
Are performance indicators aligned with goals/ objectives?		For performance indicators not fully aligned, indicate common areas of misalignment. Discuss for both formative and summative Pis.		
Do performance indicators fully assess all components of goals?		For goals not fully assessed, indicate common gaps (scheduling, attendance, targeted population, activity structure, activity characteristics, quality, other) Discuss for both formative and summative goals.		
Do activities align with stated goals/ objectives /performance indicators? (i.e. do activities have a reasonable expectation to achieve the stated goals?)		For activities not fully aligned, indicate common areas of misalignment		
Does the local evaluator's SMART goal ratings match with the reviewer's ratings?		If not fully matched, indicate areas of misalignment		

II. Evaluation Plan: ASSESSMENTS

NOTE: Include all assessments used or mentioned anywhere in the AER, even if not part of Evaluation plan; do make a note in the Comments if not part of Evaluation plan

	Rating		Comments	Examples
List all types of assessments used				
surveys				
skills assessments [other than State]				
observations				
interviews				
focus groups				
report cards				
attendance				
behavior/ disciplinary records				
State Assessments				
other				
If standardized/published instrument used, give name		Identify which goals/PIs the instrument was used to measure		
For locally developed assessments, was any validation done?		Identify which goals/PIs the instrument was used to measure		
Is the program measuring all of their goals?		If No or Partial, which goals are not being measured?		
What was the response rate, for any survey(s) used?		List survey(s)		
Any limitations of assessments, as identified by evaluator?		If Yes, for which assessments and for which goals?		
Any limitations of assessments, as identified by reviewer?		If Yes, for which assessments and for which goals?		
For assessments with limitations identified by reviewer or evaluator, were limitations addressed?		If Yes, how were they addressed?		
Were there any types of goals that were more challenging to assess?		If Yes, which type(s)? Was the challenge clearly statde in report or interpreted by reviewer?		
Comments on Assessments				

How many of the related <u>formative</u> goals/PIs were met?

NOTE: Partial rating may only be used by programs with more than one site; school attendance outcomes to be placed under "Behavioral" category; Parent outcomes can be formative or summative, follow where it was placed in AER if reviewer generally agrees

**Official attendees defined as at least 30 hours in programming; Regular attendees defined as at least 90 hours in programming*	Official attendees*		Regular attendees		Population Served		Implentation quality		Level of activity/fidelity		Other formative (write-in)		Other formative (write-in)	
	Target # students	Actual # students	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)
# related goals														
Met														
Asserted														
Partial														
Not Met														
Not measureable/defined														
Not measured														
Data pending														
Results not presented														
How many goal ratings changed by reviewer														

***Assessment types** : surveys, skills assessments, observations, interviews, focus groups, report cards, attendance, behavior/disciplinary records, State assessments, other*

How many of the related <u>summative</u> goals/PIs were met?

	English		Math		English and/or Math		Other academic		Social/emotional		Behavioral		Parent		Community	
	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)	# goals…	Assessment Type(s)
# related goals																
Met																
Asserted																
Partial																
Not Met																
Not measureable/defined																
Not measured																
Data pending																
Results not presented																
How many goal ratings changed by reviewer																

Comments on Results	

IIIA. Fall Observation Visit (fidelity)			
Was each site visited at least once during the Fall? (Y/N)		Comments:	
Fidelity reported? (Y/N)		Comments:	
Is the fidelity reported aligned with relevant PIs? (fidelity defines as outputs/level of activity)		Comments:	
Comments on Fall Observations			
IIIB. Spring Observation Visit (quality)			
Was each site visited at least once during the Spring? (Y/N)		Comments:	
Observation protocol used for point of service observations:			
Out of School Time (OST) Protocol			
Modified Out of School Time (OST) Protocol			
Other observation protocol		If Other, enter Name:	
Was quality reported?		Comments:	
Was observation listed as a measurement tool in the Evaluation plan?		Comments:	
Comments on Spring Observations			

Is a logic model included with the report? (Y/N)			
If NO, what was the reason given (if any)?			
Logic Model components	**Does the logic model show connections among each of the following components?**	**Are the activities aligned with the activities as described in Evaluation Plan?**	**Are the short and long-term outcomes aligned with the goals as described in Evaluation Plan?**
Inputs			
Activities			
Outputs			
Short-term Outcomes			
Long Term Outcomes/ Impacts			
Do connections between logic model components make sense?		**If less than ALL, explain/give an example**	
Comments on Logic Model			

V. Engagement & Communication

	Program director, site director, and/or community school director	Front line staff (service providers)	Principal and/or other school leaders if it is a school-based program	CBO partners (may be the same as front line staff – depends on program setup)	Parents	Students	Advisory Council	How often?
Annual Evaluation Reports								
Interim Evaluation Reports								
Memos/Weekly or Monthly reports								
Advisory Group meetings								
Other meetings								
Email/phone								
Other (specify)								

NOTE: Don't Know option should only be used when a group is named and the members of that group are not clearly identified (for example. Advisory Council); If more than one meeting or report is listed within one category, use the most frequent frequency for "How Often" and include all stakeholders that were mentioned in any of the reports/meetings of that category

VI. Conclusions & Recommendations

Were recommendations for action included in the AER?		
If yes, summarize recommendations and whether/how evaluation findings were used for program improvement. Include discussion of formative vs. summative findings, as appropriate.		
General Comments on AER Overall:		