

Instruments That Measure Youth Engagement and Facilitator Quality: A Toolkit for Sexual Risk Avoidance Education (SRAE) Grant Recipients

OPRE Report #2023-282
September 2023

Scilla M. Albanese, Lizabeth Malone, Victoria Sims, Rhiannon Jones, and Daisy Gonzalez

This page has been left blank for double-sided copying.

Instruments That Measure Youth Engagement and Facilitator Quality: A Toolkit for Sexual Risk Avoidance Education (SRAE) Grant Recipients

OPRE Report 2023-282
September 2023

Contract Number:
HHSP233201500035I/HHSP23337031T

Scilla M. Albanese, Lizabeth Malone, Victoria Sims, Rhiannon Jones, and Daisy Gonzalez

Project Officers:

Calonie Gray, Office of Planning, Research, and Evaluation
Tia Brown, Office of Planning, Research, and Evaluation
MeGan Hill, Family and Youth Services Bureau
Nakia Martin-Wright, Family and Youth Services Bureau
Administration for Children and Families
U.S. Department of Health and Human Services

Project Director:

Susan Zief
Mathematica

This report is in the public domain. Permission to reproduce is not necessary.

Suggested citation: Albanese, S. M., L. Malone, V. Sims, R. Jones, and D. Gonzalez (2023). Instruments that Measure Youth Engagement and Facilitator Quality: A Toolkit for Sexual Risk Avoidance Education (SRAE) Grant Recipients (OPRE Report No. #2023-282). Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer: The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research, and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This report and other reports sponsored by the Office of Planning, Research, and Evaluation are available at www.acf.hhs.gov/opre.

Connect with OPRE



Acknowledgments. The authors would like to express their appreciation to Calonie Gray and Tia Brown in the Office of Planning, Research, and Evaluation, and MeGan Hill and Nakia Martin-Wright in the Family and Youth Services Bureau for their input as we conceptualized the toolkit and continued guidance throughout the toolkit's development. We thank the Mathematica team, including Susan Zief, Heather Zaveri, Tiffany Waits, Elizabeth Mugo, Effie Metropoulos, Matthew Mariani, and Sheryl Friedlander who contributed to the development of the toolkit. We are also grateful to Sally Atkins-Burnett at Mathematica for her expert advice.

Contents

I.	How to use this toolkit.....	1
A.	What are the definitions of youth engagement and facilitator quality?	1
B.	How were the instruments selected to be in the toolkit?	2
C.	What is included in each instrument profile?.....	3
D.	How do I interpret information in each profile about an instrument’s performance?	4
II.	Getting to know the toolkit instruments.....	7
A.	Introduction to the instruments	7
B.	Questions to consider in choosing an instrument	10
C.	What is the next step if I want to use one of the instruments in the toolkit?	13
III.	Instrument profiles.....	15
	Involvement and Interaction Rating Scale.....	17
	Out-of-School Time (OST) Observation Instrument.....	21
	Participant Satisfaction Survey (also known as the Participant Feedback Survey).....	27
	Program Observation Form for TPP Program Grantees	31
	Program Quality Instrument for Youth (PQI-Youth).....	35
	Tiffany-Eckenrode Program Participation Scale (TEPPS)	41
	Youth Development Assessment Device (YDAD)	45
	Youth Program Quality Assessment (YPQA).....	49
	Appendix A. Defining youth engagement and facilitator quality that promotes youth engagement.....	A.1
	Appendix B. Process for identifying and reviewing instruments	B.1
	Appendix C. Glossary of terms.....	C.1

Exhibits

I.1. Instruments in the toolkit.....	2
I.2. Key dimensions of the instrument profile overview page	5
I.3. Key dimensions of the instrument profile narrative page.....	6
II.1. Overview of the program types of the toolkit instruments.....	7
II.2. Overview of the SRAE toolkit instruments organized by topic and topic elements.....	8
II.3. Overview of the types of instruments found in the toolkit and who completes them	9
II.4. Overview of the performance testing available for the toolkit instruments.....	10
II.5. Steps to take to choose an instrument	12
II.6. Instrument levels of permission	13
A.1. Parameters used in the youth engagement literature search	A.1
A.2. Parameters used in the facilitator quality literature search	A.2
B.1. Key search terms	B.1

I. How to use this toolkit

This toolkit provides information on instruments that Sexual Risk Avoidance Education (SRAE) grant recipients can use to measure youth engagement in SRAE programs and what facilitators do in the classroom that promotes this engagement (which we refer to as facilitator quality that promotes youth engagement, or more simply, facilitator quality). Measuring youth engagement and facilitator quality may inform your program improvement efforts, as an SRAE grant recipient.

- The toolkit focuses on instruments that you can use to measure how youth engage with the program overall, with program facilitators, and their peers—youth engagement.
- It also includes instruments that capture how facilitators work with youth participants and the quality of those interactions—facilitator quality.
- The toolkit does not replace the performance measures you collect through entry and exit surveys.

In this chapter, we describe:

- The definitions of youth engagement and facilitator quality
- The eight instruments in the toolkit and how they were selected
- The contents of each instrument profile
- How to interpret information in each profile about an instrument's performance

It is followed by:

- Chapter II, which summarizes key information across the instruments to help you identify the instrument(s) that fit your needs, and provides you with next steps if you want to use one of the instruments in the toolkit
- Chapter III, which presents profiles of each instrument

A. What are the definitions of youth engagement and facilitator quality?

Youth engagement and facilitator quality are essential components of SRAE programs as they help to ensure youth connection with program content.

Youth engagement and facilitator quality are broad topics comprised of several elements that specify characteristics or actions essential to the topic. For this toolkit, youth engagement and facilitator quality are each defined by three topic elements that are relevant to SRAE programs. Defining youth engagement and facilitator quality in this way will help you identify what you want to measure and see how comprehensive an instrument is in measuring the two topics.

Youth engagement has three elements:

1. **Behavioral.** Youth's attendance at program meetings and sessions, or participating in specific lessons and activities
2. **Emotional or psychological.** Youth's interest in and perceptions of the program, and the relationships they build with the facilitator and peers
3. **Cognitive.** Youth's investment in the program through their understanding and application of the content provided

Facilitator quality that promotes youth engagement has three elements:

1. **Classroom management.** Managing youth behavior and the facilitator's general preparedness for and organization of the class
2. **Supportive climate.** Creating an atmosphere that fosters engagement and promotes positive relationships between facilitators and youth, and among the youth
3. **Content delivery.** Making the class interesting and delivering the program in a clear way that challenges youth

Want to know more? Appendix A details how the definitions were developed.

B. How were the instruments selected to be in the toolkit?

Since youth engagement and facilitator quality span a variety of disciplines, we set clear screening criteria to help identify the eight instruments for the toolkit (Exhibit I.1). We looked for instruments that assess:

- Youth risk prevention programs and youth development programs, which are relevant to SRAE programs that use the positive youth development framework in their programming.¹
- A curriculum or intervention delivered by a facilitator
- Programs taking place in a group setting that includes middle school and high school aged youth
- At least one topic element of youth engagement or facilitator quality

Exhibit I.1. Instruments in the toolkit

Instrument
Involvement and Interaction Rating Scale
Out-of-School Time (OST) Observation Instrument
Participant Satisfaction Survey
Program Observation Form for TPP Program Grantees
Program Quality Instrument for Youth (PQI-Youth)
Tiffany-Eckenrode Program Participation Scale (TEPPS)
Youth Development Assessment Device (YDAD)
Youth Program Quality Assessment (YPQA)

¹ U.S. Department of Health and Human Services, Administration for Children and Families, Family and Youth Services Bureau. "Sexual Risk Avoidance Education Program (General Departmental-Funded)." January 9, 2019. <https://www.acf.hhs.gov/fysb/adolescent-pregnancy-prevention/sexual-risk-avoidance-education>.

Instruments were included if they were:

1. A survey or observation form that met one of the following approaches:
 - Was tested to confirm its performance
 - Used repeatedly within a study, across multiple studies, or as a performance measure for federal grant recipients
2. Developed or used in the past 20 years
3. Developed or used with a U.S. sample
4. Available in a hard copy version with no administration costs
5. Intended to be administered during the program, not after extended follow-up
6. Available with documentation and simple scoring (averages, percentages, etc.)

Want to know more? Appendix B details the instrument search.

C. What is included in each instrument profile?

We use an instrument profile to describe each of the eight instruments in this toolkit. Each profile has two sections: an overview followed by a narrative.

- The overview has basic information about the instrument so you can quickly identify whether it is relevant.
- The narrative includes more details you should know before deciding to use the instrument.

Together, these two sections of the profile cover key dimensions of the instrument, including:

- Its purpose and content, so you can see if it is something you want to use at your program
- How to administer it, so you know how easy it is to use
- Availability, so you know how to access it and what you will need to use it
- Instrument contact, so you can reach out to them with questions
- Information about the instrument's performance in any testing the developers conducted on its reliability and validity
 - **Reliability** refers to how dependable or stable an instrument is over different conditions.
 - **Validity** refers to whether an instrument accurately measures what it was designed to measure.

Ways to test an instrument's performance: Reliability

- **Internal consistency reliability.** How well the items (statements, questions, or tasks) in an instrument measure the same construct (an idea, underlying theme, or topic).
- **Inter-rater reliability.** The extent to which different observers or raters using the same instrument would interpret and record the information in the same way.

Ways to test an instrument's performance: Validity

- **Content validity.** How well items in an instrument measures behaviors important to the construct it is intended to measure. Usually involving review of research-evidence or professional judgement of research and practice experts reviewing the instrument.
- **Construct validity.** How well the instrument assesses the construct it is supposed to measure.
- **Concurrent validity.** How well the instrument results relate to other information collected at the same time.
- **Predictive validity.** How well the instrument results relate to later outcomes.

Exhibits I.2 and I.3 on pages 5 and 6 are examples of what the overview and narrative sections, respectively, of a profile look like and what they include.

D. How do I interpret information in each profile about an instrument's performance?

The reliability and validity sections of the instrument profile summarize the testing the developers conducted on the instrument, including any statistical findings they reported. The following bullets give some general guidance on how to interpret the results.

- Many researchers and instrument developers require that instruments have reliability values of 0.70 or higher. They may use different types of analyses to examine reliability, particularly for inter-rater reliability which may have different acceptable values. The glossary of terms in Appendix C provides additional information on types of analyses noted in the profiles.
- **Correlation** is the degree to which two sets of scores or other data vary together, ranging from -1.0 (a perfect negative relationship) to 1.0 (a perfect positive relationship), with 0 indicating no association. Many researchers and instrument developers consider correlations below 0.40 as small, correlations at 0.40 to below 0.60 medium, and correlations of 0.60 or above as large.²
- An **item** is a statement, question, or task in an instrument.
- Some instrument developers create **subscales**. A subscale is a set of items within a larger instrument that assesses a particular aspect of the trait being measured. Subscales may be specified based on theoretical grounds (grouping items based on their content) or empirical evidence (factor analysis of items in a longer scale may reveal meaningful subscales). The toolkit shares the subscale names and item groupings developed for each instrument and summarizes how they relate to the elements making up youth engagement and facilitator quality.

Want to know more? Appendix C has a full glossary of terms including other terms you might come across in some of the profiles.

² Evans, James D. *Straightforward Statistics for the Behavioral Sciences*. Thomson Brooks/Cole Publishing Company, 1996.

Exhibit I.2. Key dimensions of the instrument profile overview page

Instrument name (Instrument acronym)

Purpose and type of educational program instrument was developed for

Purpose: Options include improvement; monitoring; and/or research

Type of educational program: Options include adolescent pregnancy prevention or positive youth development

SRAE toolkit content found in the instrument

Topics and topic elements: Options include

- Youth engagement (behavioral, emotional/psychological, and/or cognitive); and/or
- Facilitator quality that promotes youth engagement (classroom management, supportive climate, and/or content delivery)

Instrument characteristics

Type of instrument: Options include self-administered survey³ or direct observation

Who completes it: Options include

- For self-administered survey: Person who completes and report type
 - Facilitator (self-report and/or report about youth)
 - Youth (self-report and/or report about facilitator)
- For direct observation: Options include program staff, self-trained; program staff, with publisher training; and/or publisher's data collector

Length/time: Number of individual items (or minutes)

Languages available: Options include English, Spanish, other languages

Instrument performance and the sample it is based on

Reliability: Options include available (meets/some meet the minimum acceptable rating) or not available based on internal consistency for surveys and inter-rater reliability for observations.

Validity: Options include available based on construct/concurrent validity or predictive validity, content validity only, or not available

Study sample: Sample size and demographics, information about setting type, geographical location, and year of development

Availability

Level of permission required: Options include no permission required for use; permission required for use; or permission requirements unknown

Associated costs: List of costs associated with materials or training (as of time of review in 2023)

Instrument website: Full linked URL

Instrument contacts

Name, organization, and email for instrument contact

³ For this toolkit we define a self-administered survey as a survey that a participant completes on their own without help from an interviewer.

Exhibit I.3. Key dimensions of the instrument profile narrative page

Narrative

Description. Summary of what the instrument was designed to assess, for whom, and how that information was collected (type of instrument, type of settings the instrument was developed for and used in, number of items, average administration time, and information about subscales). This section also describes the developer's noted purpose of the instrument, which includes—improvement, monitoring, or research.

- **Improvement.** Used to provide feedback to inform program operations and improvements, or identify professional development needs; usually for self-assessment
- **Monitoring.** Used to monitor program quality, such as with performance measures; usually for external assessment
- **Research.** Used to describe a topic or examine its associations with other topics.

Methods of scoring. Overview of scoring process to include types of responses and how scores are calculated.

Interpretability. Information provided by developers or publishers on how to interpret a score or range of scores.

Training support. Information on training provided by or recommended by developers or publishers to administer the instrument, score it, and interpret results.

Reliability. Information on internal consistency and inter-rater reliability (as applicable, typically for observations).

Validity. Information on indicators of the instrument's validity including content, construct, concurrent, and predictive validity.

Key considerations for using it in SRAE programs. Discussion of how the instrument relates to the toolkit topics and topic elements and aspects of the instrument to consider using in an SRAE education program setting. For an SRAE instrument, notes that the instrument was developed in an SRAE setting. For non-SRAE instrument, notes relevance for SRAE settings with examples and whether the instrument may need minor revisions or adaptations (with developer permission, unless otherwise stated in Overview).

References. Consulted sources for the instrument, manuals, and other information used to complete the profile.

II. Getting to know the toolkit instruments

This chapter previews the toolkit instruments, with tables summarizing key information about them. It also answers some questions you might have about finding the most relevant instruments for your work. It concludes with the next steps to take if you want to use any of the instruments in the toolkit. As noted in Chapter I, each profile gives detailed information about the corresponding toolkit instrument. You will want to read the profiles to really understand an instrument.

A. Introduction to the instruments

Overview of profiled instruments

- The instruments were either created for programs designed to prevent adolescent pregnancy or for programs that promote positive youth development.
- All the instruments measure both toolkit topics—youth engagement and facilitator quality that promotes youth engagement.
- Most are self-administered youth surveys—that is, surveys youth complete on their own without help from an interviewer.
- Most have reliability and validity information available to help you understand their performance.

What types of programs were the instruments designed for?

Exhibit II.1 indicates whether the toolkit instruments were designed for adolescent pregnancy prevention or positive youth development.

Exhibit II.1. Overview of the program types of the toolkit instruments

Instrument	Adolescent pregnancy prevention	Positive youth development
Involvement and Interaction Rating Scale		X
Out-of-School Time (OST) Observation Instrument		X
Participant Satisfaction Survey	X	
Program Observation Form for TPP Program Grantees	X	
Program Quality Instrument for Youth (PQI-Youth)		X
Tiffany-Eckenrode Program Participation Scale (TEPPS)		X
Youth Development Assessment Device (YDAD)		X
Youth Program Quality Assessment (YPQA)		X

Note: An “X” means the program type applies to the instrument.

What content can I find in the SRAE toolkit instruments?

The toolkit is a compilation of instruments whose content covers the topics of youth engagement and facilitator quality that promotes youth engagement. Each of these topics has three distinct elements. Chapter I gives the definitions of the topics and the topic elements. Exhibit II.2 organizes each instrument by its topic and topic elements.

SRAE toolkit topics and topic elements

Youth engagement. Behavioral, emotional/psychological, and cognitive

Facilitator quality. Classroom management, supportive climate, and content delivery

One thing to consider about an instrument's content is the number of items (statements, questions, or tasks) and specific detail on a given topic element vary in an instrument. In other words, if Exhibit II.2 indicates that an instrument has content related to one of the topic elements, that does not imply anything about the amount of content available—it could be one item, or it could be many. The profiles include a description of the content that will provide information on the potential breadth or depth of a topic. The instrument itself should be your final source to determine the level of detail on a given topic element.

Exhibit II.2. Overview of the SRAE toolkit instruments organized by topic and topic elements

Instrument	Youth engagement			Facilitator quality		
	Behavioral	Emotional or psychological	Cognitive	Classroom management	Supportive climate	Content delivery
Involvement and Interaction Rating Scale	X	X	X	X	X	X
Out-of-School Time (OST) Observation Instrument	X	X	X	X	X	X
Participant Satisfaction Survey		X		X	X	X
Program Observation Form for TPP Program Grantees	X		X	X	X	X
Program Quality Instrument for Youth (PQI-Youth)		X	X	X	X	
Tiffany-Eckenrode Program Participation Scale (TEPPS)	X	X	X		X	
Youth Development Assessment Device (YDAD)	X	X	X		X	
Youth Program Quality Assessment (YPQA)		X		X	X	X

Note: An "X" means that the SRAE toolkit content applies to the instrument.

What types of instruments are in the toolkit, and who should complete them?

The toolkit includes self-administered surveys (a survey that a participant completes on their own without help from an interviewer) and direct observations. Exhibit II.3 lists the types of instruments in the toolkit and who completes them.

Exhibit II.3. Overview of the types of instruments found in the toolkit and who completes them

Instrument	Self-administered survey		Direct observation		
	Facilitator	Youth	Program staff, self-trained	Program staff, with publisher training	Publisher's data collector
Involvement and Interaction Rating Scale	X	X	n.a.	n.a.	n.a.
Out-of-School Time (OST) Observation Instrument	n.a.	n.a.	X		
Participant Satisfaction Survey		X	n.a.	n.a.	n.a.
Program Observation Form for TPP Program Grantees	n.a.	n.a.	X		
Program Quality Instrument for Youth (PQI-Youth)^a	X	X	n.a.	n.a.	n.a.
Tiffany-Eckenrode Program Participation Scale (TEPPS)		X	n.a.	n.a.	n.a.
Youth Development Assessment Device (YDAD)		X	n.a.	n.a.	n.a.
Youth Program Quality Assessment (YPQA)^b	n.a.	n.a.	X	X	X

Note: An "X" means the administration characteristic applies to the instrument.

^a A version of the Program Quality Instrument for Youth (PQI-Youth) can be completed by facilitators, but that is not the toolkit's focus because it did not have accompanying information on instrument performance.

^b The publisher recommends that the assessment team leader participate in a publisher-developed online training, after which the assessment team leader can train the other members of the assessment team.

n.a. = The characteristic is not applicable to the specific instrument type.

What do we know about the instruments' performance?

To measure a construct (an idea, underlying theme, or topic) well, you need an instrument with strong measurement properties. An instrument's performance is based on its reliability (how dependable or stable it is over different conditions) and validity (whether an instrument accurately measures what it was designed to measure). We looked at different types of reliability and validity (as defined in Chapter I) for an instrument. For the profiles, we assigned ratings for reliability and validity based on the following:

- Internal consistency information for surveys or inter-rater reliability information for direct observations was available.
- Construct (concurrent or predictive) validity information was available.

Exhibit II.4 indicates whether reliability and validity performance is available for each toolkit instrument. Although the two adolescent pregnancy prevention instruments do not have accompanying performance information, they were chosen for this toolkit because both instruments are used by their respective program grantee recipients as

part of monitoring and assessing their programs, and the available information shows they have repeated use within the field. Their accompanying information gives users the information they need to administer the instruments, score them, and interpret their results.

Exhibit II.4. Overview of the performance testing available for the toolkit instruments

Instrument	Reliability	Validity
Involvement and Interaction Rating Scale	X	Content validity only
Out-of-School Time (OST) Observation Instrument	X [†]	X [†]
Participant Satisfaction Survey	n.a.	—
Program Observation Form for TPP Program Grantees	—	—
Program Quality Instrument for Youth (PQI-Youth)	X	X
Tiffany-Eckenrode Program Participation Scale (TEPPS)	X	X
Youth Development Assessment Device (YDAD)	X	X
Youth Program Quality Assessment (YPQA)	X	X

Note: An “X” means this type of information is available for the instrument. An “—” means this type of information is not available for the instrument. We assigned ratings for reliability based on internal consistency information for surveys and inter-rater reliability information for direct observations. For validity, we assigned ratings based on construct (concurrent or predictive) validity information.

[†] Information available presents items in different subscales than published instrument does.

n.a. = not applicable; survey only presents individual item scores.

B. Questions to consider in choosing an instrument

The following section has answers to some questions you might have that may help you choose an instrument.

Q1. What if I want to use a survey to measure youth engagement?

A. All five of the surveys in the toolkit will allow you to measure youth engagement. The Involvement and Interaction Rating Scale, Tiffany-Eckenrode Program Participation Scale, and Youth Development Assessment Device will allow you to measure all topic elements of youth engagement. The Program Quality Instrument for Youth (PQI-Youth) measures two topic elements. The Participant Satisfaction Survey has content related to one of its topic elements.

Q2. What if I want to use a direct observation to measure youth engagement?

A. All three of the direct observation instruments will allow you to measure youth engagement. The Out-of-School Time Observation Instrument will allow you to measure all topic elements of youth engagement. The Program Observation Form for TPP Program Grantees measures two topic elements. The Youth Program Quality Assessment has content related to one element of the youth engagement topic.

Q3. What if I want to use a survey to measure facilitator quality?

A. Five of the instruments are surveys of facilitator quality based on youth reports. The Involvement and Interaction Rating Scale and Participant Satisfaction Survey have content related to all three elements of facilitator quality, and the other three surveys touch on one or two elements. The Involvement and Interaction Rating Scale measures facilitator quality through facilitator self-reports. The Program Quality Instrument for Youth has another version that can be completed by facilitators.

Q4. What if I want to use a direct observation to measure facilitator quality?

A. You can use any of the direct observation instruments to measure all elements of facilitator quality.

Q5. Which instrument could I use if I want to measure as many of the topic elements as possible?

A. If you would like to use a survey, the Involvement and Interaction Rating Scale measures all youth engagement and facilitator quality topic elements. If you would like to use a direct observation, the Out-of-School Time Observation Instrument measures all elements of youth engagement and facilitator quality, and the Program Observation Form for TPP Program Grantees measures two elements of youth engagement and all elements of facilitator quality.

Q6. Which of the instruments are available in Spanish?

A. The Program Quality Instrument for Youth, a survey, is available in English and Spanish. The Youth Program Quality Assessment, a direct observation, is also available in both languages. If you are interested in using an instrument that is not available in Spanish, contact the developer about the translation.

Q7. Which instruments may be the best fits for SRAE program settings?

A. The two instruments used by adolescent pregnancy prevention programs would not require any changes to use. The Program Observation Form for TPP Program Grantees measures two elements of youth engagement and all elements of facilitator quality. The Participant Satisfaction Survey measures one element of youth engagement and all elements of facilitator quality. The other toolkit instruments may require some changes, such as minor word changes to individual items, or potentially removing items that may not apply to SRAE programs. We have provided guidance on steps to take in the next section and more details about potential changes with examples in the individual instrument profiles.

Q8. How easy are the instruments to administer?

A. All of the surveys are easy to administer. They are available on paper and only require program staff to hand them out to participants. The direct observation instruments would require some level of training to become familiar with the items.

Q9. Which instruments would require the least amount of time to administer?

A. The self-administered youth surveys will take the least amount of time because they each have fewer than 30 individual items to answer.

The direct observation instruments will take longer to administer than the surveys. You will need to familiarize yourself with the instrument before conducting the observation. Aside from the training time, the OST Observation Instrument may take as little as 15 minutes to complete, whereas the Program Observation Form for TPP Program Grantees and Youth Program Quality Assessment would require observing an entire program session.

Q10. Which instruments are the easiest to score and interpret the data on?

A. All of the toolkit instruments have simple scoring techniques. Most instruments involve adding or averaging responses across items to calculate total and subscale scores. Two instruments do not require any calculations—the Participant Satisfaction Survey and the Program Observation Form for TPP Program Grantees—and present individual item scores only.

Q11. Do any instrument developers offer formal trainings that I can take to learn more about administering the instrument?

A. By design, the toolkit only includes instruments that do not have administration costs associated with their use. That said, the developers of the Youth Program Quality Assessment offer trainings and other supports for an added cost.

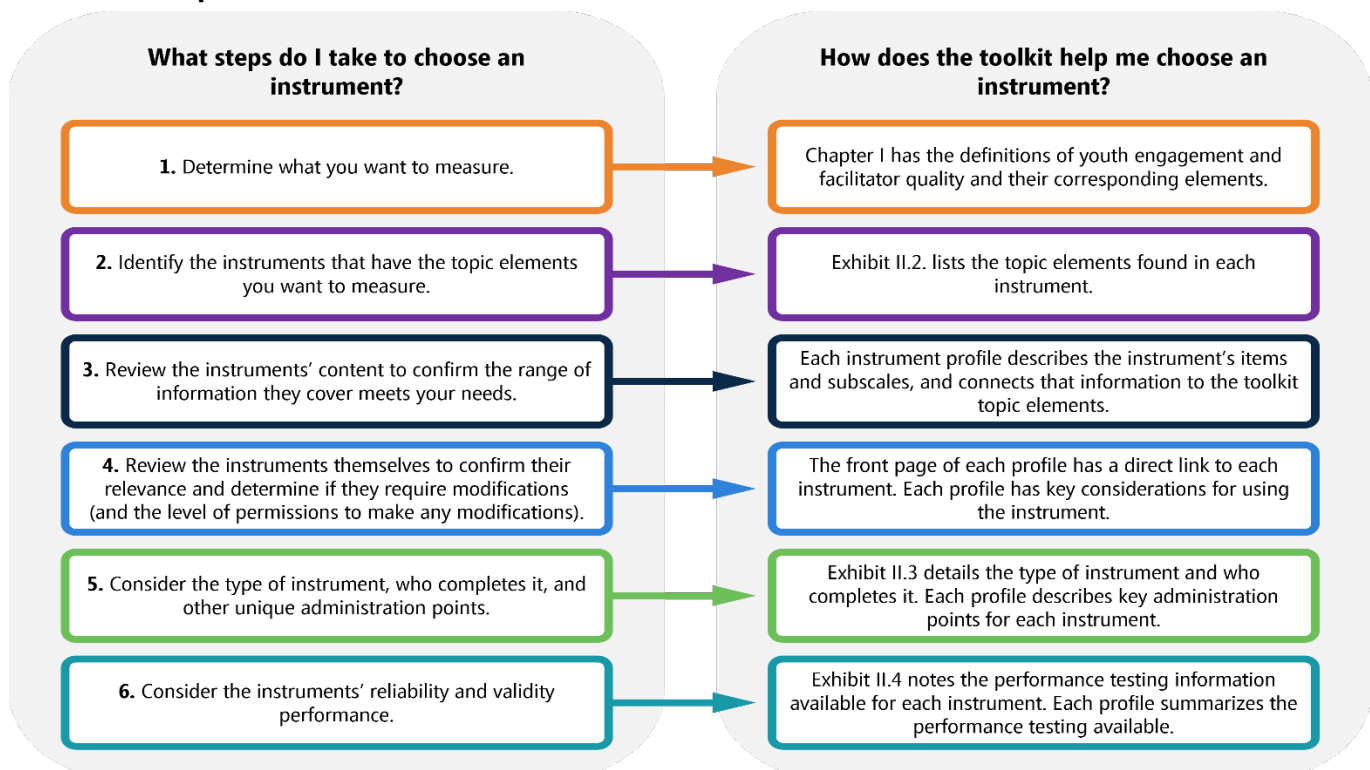
Q12. Should I use an instrument that does not have any information available on its reliability and validity?

A. The two adolescent pregnancy prevention instruments do not have accompanying reliability and validity information. However, both instruments are used by their respective program grant recipients as part of monitoring and assessing their programs. The Program Observation Form for TPP Program Grantees is one of the performance measures that teen pregnancy prevention program grantees report on to the Office of Population Affairs in the U.S. Department of Health and Human Services. The Participant Satisfaction Survey is used by North Carolina Teen Pregnancy Prevention Initiative grantees as part of their required program evaluations. It is this repeated use by state and federal grant recipients that allowed the instruments to meet the inclusion criteria for the toolkit. Additionally, both instruments have accompanying information that gives users the information they need to administer the instruments, score them, and interpret their results. Finally, as SRAE grant recipients, you and your staff have the experience in the field to review these instruments for their relevance to SRAE programs.

Q13. How do I choose an instrument to use?

A. Every SRAE grant recipient has different research and evaluation needs, which will result in SRAE grant recipients choosing different instruments to use. There are a few steps you can take to help you choose an instrument to use. Exhibit II.5 summarizes these steps and demonstrates what parts of the toolkit will help you make decisions along the way.

Exhibit II.5. Steps to take to choose an instrument



C. What is the next step if I want to use one of the instruments in the toolkit?

If you would like to use one of the instruments in the toolkit, your next step is to review the instrument's level of permission. You will find this information in the Availability box in the profile overview, and in the section titled "Considerations for using it in SRAE programs" in the profile narrative. Exhibit II.6 describes the different levels of permission associated with the toolkit instruments and what they mean if you want to use the instrument.

Exhibit II.6. Instrument levels of permission

If the instrument's level of permission is...	Then you...
No permission required for use	May use the instrument as is. Some instruments may have conditions (such as including a copyright statement or sharing findings). The profile includes notes on such conditions.
Permission required for use	Will need to contact the instrument developer or publisher. When available, the profile includes links to request forms.
Permission requirements unknown	Will need to contact the instrument developer or publisher.

For instruments in the toolkit, the following practices apply:

- **All toolkit instruments are for noncommercial use only**, which means educational or nonprofit use by researchers or practitioners.
- **When using the instrument, please credit the original source using the citation we provide in the Instrument resources section of the profile.** Cite the original source on any hard copies of the instrument you print for your data collection and in any reports you prepare that present findings based on the instrument.
- **In reviewing the instrument, you may want to delete or change items. Check with the instrument's contact before you do.** Each instrument is unique—some allow adaptations and some do not; some allow you to adapt freely, and others require permission. Look at the instrument's profile for guidance. The instrument contacts may be able to tell you how these changes might impact subscale or total scores that can be calculated for some instruments.

This page has been left blank for double-sided copying.

III. Instrument profiles

Involvement and Interaction Rating Scale.....17

Out-of-School Time (OST) Observation Instrument.....21

Participant Satisfaction Survey (also known as the Participant Feedback Survey).....27

Program Observation Form for TPP Program Grantees 31

Program Quality Instrument for Youth (PQI-Youth).....35

Tiffany-Eckenrode Program Participation Scale (TEPPS) 41

Youth Development Assessment Device (YDAD) 45

Youth Program Quality Assessment (YPQA).....49

This page has been left blank for double-sided copying.

Involvement and Interaction Rating Scale

Purpose and type of educational program instrument was developed for

Purpose: Improvement

Type of educational program: Positive youth development

SRAE toolkit content found in the instrument

Topic and topic elements: Youth engagement (behavioral, emotional/psychological, cognitive); facilitator quality that promotes youth engagement (classroom management, supportive climate, content delivery)

Instrument characteristics

Type of instrument: Self-administered survey

Who completes it: Facilitator (self-report and report about youth participants); youth (self-report and report about facilitator)

Length/time: 26 individual items

Languages available: English

Instrument performance and the sample it is based on

Reliability: Available (meets the minimum acceptable rating)

Validity: Content validity only

Study sample: 108 participants of 4-H Cooperative Extension Services programs from 10 rural communities in Connecticut, Delaware, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Vermont, and West Virginia, and two urban communities in Pennsylvania

51% of participants were youth ages 13 to 18, and 49% were adults*; 69% were female, and 31% were male; 48% were White, 34% were Black, 6% were Hispanic, 5% were Asian, 1% were Native American, and 6% were another race; 41% lived in rural areas, 44% lived in urban areas, and 16% lived in suburban areas.

** The instrument documentation refers to adult leaders as adult participants; for this profile within the SRAE context, adult participants are referred to as facilitators.*

Availability

Level of permission required: No permission required for use, include copyright statement found on last page of instrument

Associated costs: No known costs

Instrument website: <http://www2.ca.uky.edu/agcomm/pubs/4mo/4mo05oo/4mo05oo.pdf>

Instrument contacts

Kenneth R. Jones, University of Kentucky

kenrjones@uky.edu

Narrative

Description. The Involvement and Interaction Rating Scale is a survey that measures youth-adult relationships in structured positive youth development programs or other youth-adult partnerships, that work with youth ages 13 and older. It can be used for program improvement to assess youth and adult program perceptions and experiences to better understand youth-adult relationships, overall program strengths and areas for improvement.

Content of instrument. The survey evaluates the perceptions and experiences of youth and adult interactions, using 26 individual items in three subscales:

1. Youth Involvement (eight items) measures youth's ability to act as decision makers and play an active role in the programming, including openly sharing their thoughts, helping peers, and their enthusiasm about the program.
2. Adult Involvement (eight items) measures adults' partnership with youth and support of youth involvement and leadership in the program, as well as their attitudes towards and commitment to the project.
3. Youth-Adult Interaction (ten items) measures collaboration, respect, and rapport between youth and adults.

Administration points

- The survey is designed to be administered to both youth and adults to capture the entire group's perceptions.
- It should be administered near the middle of the program. This ensures participants have varied experiences to reflect on when answering the survey. It also ensures there is enough time remaining in the program to make adjustments based on the survey results.
- A person who is not facilitating the program should administer the survey. This could be an individual in a leadership role or someone who manages day-to-day operations for the program, such as a lead educator or program administrator.

Methods of scoring

- Items are scored on a 10-point scale. This scale consists of two opposite statements, with 1 representing a negative perception statement and 10 representing a positive perception statement. For example, the developers include an item about taking initiative in the Youth Involvement subscale. On one end, "little initiative" is associated with a score of 1, while "lots of initiative" is associated with a score of 10.
- Subscale scores for individual participants are calculated by averaging the items.
- Then, program-level mean subscale scores can be created from the participant-level mean subscale scores.

Interpretability

- Low scores are associated with negative perceptions and experiences and high scores are associated with positive perceptions and experiences.
- Mean subscale scores between 1 and 5.4 are classified as "low" and mean subscale scores between 5.5 and 10 are classified as "high."
- Once each subscale mean score has been classified as low or high, these three subscale classifications are used to determine the type of youth-adult relationship.

- The developers provide guidance on all potential combinations of subscale scores, whether on the high or low ends, and describe each associated relationship type (see Table 1, Appendix A; Jones 2006).
- For example, if all three subscales are classified as high, that would indicate a Youth-Adult Partnership relationship, whereas if the Adult Involvement subscale is classified as high and the two other subscales are classified as low, that would indicate an Adult-Centered Leadership relationship.

Training support. Jones 2006 (Appendix A) has instructions on when and how to administer the instrument, specific scoring steps, and guidance on how to interpret scores and use the Continuum of Youth-Adult Relationships.

Reliability

Internal consistency reliability. The developers reported Cronbach's alpha coefficients for a 38-item version of the instrument and each subscale.

- Instrument-level across all items: 0.94
- Subscales: Ranged from 0.83 to 0.87

Inter-rater reliability. Not applicable.

Validity

Content validity. The developers used literature and measurement work on youth leadership, youth development, youth program assessment, mentoring, and youth-adult partnerships to inform the instrument, taking into account the Continuum of Youth-Adult Relationships. An expert panel of faculty and graduate assistants with backgrounds in survey design or youth-adult relationships and a practitioner-based evaluation panel helped refine the instrument.

Construct/concurrent validity

Construct validity. No information available.

Concurrent validity. No information available.

Predictive validity. No information available.

Key considerations for using it in SRAE programs

- SRAE toolkit topic elements found in the instrument:
 - The Youth Involvement subscale contains items related to the behavioral, emotional/psychological, and cognitive elements of youth engagement.
 - The Adult Involvement subscale contains items related to the supportive climate and content delivery elements of facilitator quality.
 - The Youth-Adult Interaction subscale contains items related to the emotional/psychological element of youth engagement and the classroom management, supportive climate, and content delivery elements of facilitator quality.
- The developers initially created the instrument for positive youth development programs that have the same kinds of characteristics as SRAE programs, such as structured group activities with an adult leader.

- Since this instrument was originally designed to operate in the context of a 4-H program in which youth and adults worked together on a community project, adult leaders in the program are viewed as “adult participants.” For application in SRAE programs, SRAE facilitators would be considered serving the same role as “adult participants.” The difference in terminology does not impact the instrument’s relevance to SRAE, as an SRAE program would still be classified as one of the five relationship types on the continuum.
- Although the items in the instrument are not tied to specific program content, it might be necessary to make minor word changes to items in the instrument to use it in the SRAE program setting, such as changing “adults” to “facilitator” or “project” to “program.” Check with the developers to determine if these minor changes can be made.
- The SRAE program context and goals should be considered when interpreting scores. The developers designed the survey to be used alongside the Continuum of Youth-Adult Relationships (Jones and Perkins 2005), the developers’ conceptual framework. This continuum assesses the existing youth-adult relationships within a program and categorizes them into five types.
 - On one end of the continuum is Adult-Centered Leadership, and on the other end is Youth-Centered Leadership.
 - SRAE programs are likely to exist on the Adult-Centered Leadership end of the continuum because of their curricular and structured approach.
- Some items in the three subscales may not be relevant to SRAE programs, such as items in the Youth Involvement and Adult Involvement subscales that ask about youth and adult participants’ concerns with community change. If you want to remove any items, check with the developers to determine if these items may be removed, and if new subscale scores may be calculated and interpreted. However, the information presented here on reliability and validity would not apply if changes are made.
- You do not need to contact the developers to use the instrument, but you must include the copyright statement found on the last page of the instrument if you use it.

References

Instrument resources

Jones, K. R. “Youth-Adult Partnerships: Are You There Yet? How to Evaluate Your 4-H Youth Development Program.” Pub no. 4MO-0500. University of Kentucky, Cooperative Extension Service, 2006.
<http://www2.ca.uky.edu/agcomm/pubs/4mo/4mo0500/4mo0500.pdf>.

Background research

Jones, K. R., and D. F. Perkins. “Determining the Quality of Youth-Adult Relationships Within Community-Based Youth Programs.” *Journal of Extension*, vol. 43, no. 5, October 2005, pp. 1–10.
<https://archives.joe.org/joe/2005october/a5.php>.

Jones, K. R., and D. F. Perkins. “Youth and Adult Perceptions of Their Relationships Within Community-Based Youth Programs.” *Youth & Society*, vol. 38, no. 1, September 2006, pp. 90–109.

Out-of-School Time (OST) Observation Instrument

Purpose and type of educational program instrument was developed for

Purpose: Improvement; research

Type of educational program: Positive youth development

SRAE toolkit content found in the instrument

Topic and topic elements: Youth engagement (behavioral, emotional/psychological, cognitive); facilitator quality that promotes youth engagement (classroom management, supportive climate, content delivery)

Instrument characteristics

Type of instrument: Direct observation

Who completes it: Observer (program staff, self-trained)

Length/time: 28 individual items/ 15 minutes to observe one activity as an observation cycle (with the possibility of conducting more than one observation cycle per observation, see Narrative description for details)

Languages available: English

Instrument performance and the sample it is based on

Reliability: Available (meets the minimum acceptable rating)*

Validity: Available*

Study sample: 35 out-of-school time (OST) program sites in New York and New Jersey

Sample 1 (2005): 31 observation cycles in 10 New York City OST programs serving grades K–8

Sample 2 (2006 and 2008): 40 observation cycles in 15 New York City OST programs serving grades K–12

Sample 3 (2006 and 2007): 30 observation cycles in 10 New Jersey OST programs serving grades K–8

** Information available presents items in different subscales than published instrument does.*

Availability

Level of permission required: No permission required for use

Associated costs: No known costs

Instrument website: <https://www.policystudies.com/out-of-school-time>

Instrument contacts

Christina Russell, Policy Studies Associates, Inc.

crussell@policystudies.com

Narrative

Description. The Out-of-School Time (OST) Observation Instrument is a direct observation instrument that measures program features associated with positive youth development. It can be used in schools and other settings with youth in kindergarten through grade 12. The developers created the OST Observation Instrument for research to understand the quality of OST programs. It can also be used for improvement to provide program administrators objective information about the program, youth participants, and program staff.

Content of instrument. The OST Observation Instrument describes quality of interactions between participating youth and program staff using 28 individual items in five subscales:

1. Youth Relationship-Building (five items) measures how youth show positive affect toward one another and program staff, and how youth treat one another.
2. Youth Participation (five items) measures how youth are involved in the activity and show leadership.
3. Staff Relationship-Building (seven items) measures how program staff are use positive management techniques, encourage participation and positive interactions among youth, and treat youth with respect.
4. Staff Instructional Strategies (seven items) measures how program staff communicate with youth, involve and motivate youth, and use different teaching strategies.
5. Activity Content and Structure (four items) measures how the content is organized, how challenging and analytical it is, and how it involves a progression of skills.

It also contains information on:

- Types of program activities that engage participating youth such as academics (not homework), sports, arts and crafts, and college or career preparation
- Program characteristics that support activities, such as staffing, number of participants, grade level of participants, supervision, and materials

Administration points

- The OST Observation Instrument can be administered by program staff.
- Each observation cycle can be completed in 15 minutes. After 10 minutes of observation, still within the observation cycle, the observer can begin scoring items.
- The number of observation cycles may vary. The developers highlight that program administrators can determine the appropriate number of observation cycles based on how many different types of activities are part of a session. For typical three-hour OST programs, the developers recommend four to six observation cycles of different activity types.

Methods of scoring

- Items are scored on a 7-point scale of 1 (exemplar is not evident), 2 (exemplar may occur momentarily but is not sustained), 3 (exemplar may occur once or twice but is rarely evident throughout the observation), 4 (exemplar occurs more frequently but is not fully evident), 5 (exemplar is evident or implicit), 6 (exemplar is strongly evident or implicit, but could be more consistent), and 7 (exemplar is highly and consistently evident).
- Although the developers do not provide instructions on how to calculate subscale scores, the validation report presents means across items within a subscale.
- The developers recommend completing more than one observation cycle within a session but do not provide details on how to combine the item or subscale scores obtained across the different observation cycles.

Interpretability. The developers note that high-quality OST programs will have items rated between 5 and 7.

Training support

- The developers describe the training that observers completed during their research before conducting the OST Observation Instrument in the field but do not provide information on the type of training recommended for nonresearchers, such as program administrators.
 - The research study training included practice observation sessions with video-recorded after-school activities, discussions to understand the instrument's theoretical framework and terminology, and in-person observation sessions at OST program sites (Pechman et al. 2008).
- Policy Studies Associates 2011 provides more details on how to select activities to observe and the number of observation cycles to complete.

Reliability

Internal consistency reliability. The developers reported Cronbach's alpha coefficients for subscales that differ from the published instrument. Cronbach's alpha coefficients ranged from 0.73 to 0.88 for the subscales in the analysis that most resembled the final instrument (Sample 1, 2005).

Inter-rater reliability. The developers assessed inter-rater reliability at the instrument level using two different techniques. They presented the results for the instrument subscales except for the Activity Content and Structure subscale.

- They reported intraclass correlation coefficients for two of the samples.⁴
 - Instrument-level across all items: 0.73 and 0.76
 - Subscales: Ranged from 0.52 to 0.86
- They calculated Pearson's correlation coefficients between a pair of observers conducting the OST Observation Instrument.
 - Instrument-level across all items: 0.73 to 0.83
 - Subscales: Ranged from 0.63 to 0.91. Staff Instructional Strategies did not meet the minimum acceptable rating for either sample (0.63 and 0.64).

⁴ 2008 Sample 2 and 2007 Sample 3

Validity

Content validity. The developers used literature on positive youth development and research on OST program quality to inform the instrument.

Construct/concurrent validity

Construct validity. No construct validity information is available for the five subscales in the final instrument.⁵

Concurrent validity. The developers examined rank-order correlations of program-level average scores between observation scores and youth survey data on similar constructs to examine “alignment” (Pechman et al. 2008 p. 14). The observation subscales examined (listed below) are slightly different from those in the published instrument. The youth survey data included exposure to new experiences, sense of belonging, interactions with staff, interactions with peers, and academic benefits.⁶

- Activity Content and Structure subscale: Not associated with any of the survey scales.
- Relationships subscale: Ranged from 0.01 (academic benefits) to 0.59 (interactions with peers) across two years, but were not consistent across time. For example, the correlation with academic benefits was within this range for another year (0.52).
- Instructional Strategies subscale: Examined for the first three survey scales and ranged from -0.20 (sense of belonging) to 0.39 (interactions with staff and exposure to new experiences) across the two years, but were not consistent across time. For example, the correlation with sense of belonging was 0.37 in another year.

Predictive validity. No information available.

Key considerations for using it in SRAE programs

- SRAE toolkit topic elements found in the instrument:
 - The Youth Relationship-Building and Youth Participation subscales contain items related to the behavioral, emotional/psychological, and cognitive elements of youth engagement.
 - The Staff Relationship-Building and Staff Instructional Strategies subscales contain items related to the classroom management, supportive climate, and content delivery elements of facilitator quality.
 - The Activity Content and Structure subscale contains items related to the classroom management and content delivery elements of facilitator quality.
- The developers initially created the instrument for OST programs that have the same kinds of characteristics as SRAE programs, such as program staff leading structured activities and delivering content.
- The length and variation of activities in SRAE programs will determine the number of possible observation cycles. For example, if only one activity is conducted in a session, you would only conduct one 15-minute observation cycle.
- The developers do not report consistent subscales across the sample studies. Review the items associated with the subscales in the published instrument to confirm content validity. Considering the various subscale combinations, the subscales focused on staff demonstrate stronger internal consistency.

⁵ They did explore whether the items would fit into a different framework (see Pechman et al. 2008).

⁶ Combined 2006 Samples 2 and 3, and 2007 Sample 3.

- Trainings should focus on how to score items within the subscales, paying close attention to the Staff Relationship-Building and Staff Instructional Strategies subscales that had different scores in the research study depending on the observer. See the Inter-Rater Reliability section for more details.
- Although the items in the instrument are not tied to specific program content, it might be necessary to make minor word changes to items in the instrument to use it in the SRAE program setting, such as changing “adult” or “staff” to “facilitator.” Check with the developers to determine if these minor changes can be made.
- Some items in subscales may not be relevant to SRAE programs, such as an item in the Activity Content and Structure subscale that asks how the content of an activity involves a progression of skills. If you want to remove any items, check with the developers to determine if these items may be removed, and if new subscale scores may be calculated and interpreted. However, the information presented here on reliability and validity would not apply if changes are made.
- You do not need to contact the developers to use the instrument as-is.

References

Instrument resources

Policy Studies Associates. “OST Observation Instrument Observer Procedures.” March 2011.

[OUT+OF+SCHOOL+TIME+\(OST\)+procedures.pdf \(squarespace.com\).](#)

Policy Studies Associates. “Out of School Time (OST) Observation Instrument Cover Sheet.” N.d.

[OUT+OF+SCHOOL+TIME+\(OST\)+cover+sheet.pdf \(squarespace.com\).](#)

Background research

Birmingham, J., E. M. Pechman, C. A. Russell, and M. Mielke. “Shared Features of High Performing After-School Programs: A Follow-Up to the TASC Evaluation.” Policy Studies Associates, 2005.

Durlak, J. A., and R. P. Weissberg. “The Impact of After-School Programs That Promote Personal and Social Skills.” Collaborative for Academic, Social, and Emotional Learning (CASEL); University of Illinois at Chicago, 2007.

Kim, J. H., T. D. Miller, E. R. Reisner, and K. P. Walking Eagle. “Evaluation of New Jersey After 3: First-Year Report on Programs and Participants.” Policy Studies Associates, 2006.

Pechman, E. M., M. B. Mielke, C. A. Russell, R. N. White, and N. Cooc. “Out-of-School Time (OST) Observation instrument: Report of the Validation Study.” Policy Studies Associates, February 2008.

Russell, C. A., M. B. Mielke, and E. R. Reisner. “Evaluation of the New York City Department of Youth and Community Development Out-of-School Time Programs for Youth Initiative: Results of Efforts to Increase Program Quality and Scale in Year 2.” Policy Studies Associates, 2008.

Russell, C. A., E. R. Reisner, L. M. Pearson, K. P. Afolabi, T. D. Miller, and M. B. Mielke. “Evaluation of the Out-of-School Time Initiative: Report on the First Year.” Policy Studies Associates, 2006.

This page has been left blank for double-sided copying.

Participant Satisfaction Survey (also known as the Participant Feedback Survey)

Purpose and type of educational program instrument was developed for

Purpose: Improvement; research

Type of educational program: Adolescent pregnancy prevention

SRAE toolkit content found in the instrument

Topics and topic elements: Youth engagement (emotional/psychological); facilitator quality that promotes youth engagement (classroom management, supportive climate, content delivery)

Instrument characteristics

Type of instrument: Self-administered survey

Who completes it: Youth (self-report and report about facilitator)

Length/time: 15 individual items

Languages available: English

Instrument performance and the sample it is based on

Reliability: Not applicable

Validity: Not available

Study sample: North Carolina organizations administering teen Personal Responsibility Education Program (PREP) projects through federal funding awarded to them by the state Department of Human Services, and North Carolina organizations implementing adolescent pregnancy prevention programs with state funding from the Department of Human Services

Availability

Level of permission required: No permission required for use

Associated costs: No known costs

Instrument website: <https://teenpregnancy.dph.ncdhhs.gov/prepproj.htm>, see instrument titled Participant Satisfaction Survey

Instrument contacts

North Carolina Teen Pregnancy Prevention Initiatives (TPPI)

See contacts for this instrument, such as team lead or evaluator, at this link: <https://teenpregnancy.dph.ncdhhs.gov/contact.htm>

Narrative

Description. The Participant Satisfaction Survey (also known as the Participant Feedback Survey) is a survey for youth in middle and high school used by the North Carolina Teen Pregnancy Prevention Initiatives (NC TPPI). The NC TPPI funds programs across the state that use education and health services to prevent a first teenage pregnancy and prevent additional unintended pregnancies in teen parents (NC TPPI 2019). The survey measures youth satisfaction with the facilitator and their program experience. Grant recipients (organizations that receive funding from NC TPPI) use the survey as part of their required program evaluations, submitting summaries of the results to NC TPPI annually (NC TPPI 2022b). Grant recipients can reflect on how survey results align with their expectations, use the insights they gain to identify opportunities for program improvements and develop plans and accountability measures to make these improvements.

Content of instrument

- The Participant Satisfaction Survey contains 10 individual close-ended items:
 - Eight items asking youth to provide feedback on the facilitator’s communication skills, leadership, enthusiasm, interactions with youth, presentation skills, and ability to answer questions
 - Two items about the program’s environment and whether youth would recommend the program to a friend
- The survey also contains five open-ended items in which youth participants can report on their program experiences. These responses may include feedback on the facilitator or how much the participants engaged with the program.

Administration points. The survey is self-administered and allows youth program participants to report on themselves and their facilitator at the end of a program.

Methods of scoring

- Close-ended Items are scored on a 3-point scale of 0 (no), 1 (somewhat), and 2 (yes).
 - Raw counts and percentages for each response can be calculated for each item across youth.
 - Means for each item can also be calculated across youth.
- NC TPPI recommends programs review the responses to the five open-ended items to identify themes and then count the number of responses that fit into each theme.

Interpretability

- Response percentages for each question can provide a high-level overview of what participants thought of the program. Higher percentages of “Yes” responses indicate youth had a more positive perception of each item.
- Mean scores can be used to interpret participants’ agreement with each statement.
 - A mean score close to 2 indicates strong group agreement with the statement.
 - A mean score close to 1 indicates either a lack of consensus among the group about the statement or that the group moderately agrees with the item statement.
 - A mean score close to 0 indicates a strong group disagreement with the item statement.
- The NC TPPI resource, “Instructions for Completing Participant Satisfaction Survey Responses Summary” contains information to help interpret percentages and mean scores.

Training support. NC TPPI provides:

- A template and instructions on how to score and complete a survey response summary; see “Participant Satisfaction Survey Responses Summary Template” and “Instructions for Completing Participant Satisfaction Survey Responses Summary”
- A spreadsheet to assist with scoring calculations; see “Example Participant Satisfaction Survey Formulas”
- A worksheet providing guidance on how to use the survey results; see “Using Satisfaction Survey Results”

Reliability

Internal consistency reliability. Not applicable, as only individual item scores are reported.

Inter-rater reliability. Not applicable.

Validity

Content validity. No information available.

Construct/concurrent validity

Construct validity. No information available.

Concurrent validity. No information available.

Predictive validity: No information available.

Key considerations for using it in SRAE programs

- SRAE toolkit topic elements found in the instrument:
 - Items related to the emotional/psychological elements of youth engagement
 - Items related to the classroom management, supportive climate, and content delivery elements of facilitator quality
- No changes required for item wording. The instrument is used to measure youth’s satisfaction with and get youth’s feedback on adolescent pregnancy prevention programs, including programs that use SRAE curriculum. However, if you want to make any changes, you must contact NC TPPI.
- There are two versions of the survey—one that uses “instructor” and another version that uses “facilitator.” You can choose the version most appropriate for your program.
- You do not need to contact the developers to use the instrument.

References

Instrument Resources

Version of the instrument that uses “facilitator”

North Carolina Department of Health and Human Services, Division of Public Health, Teen Pregnancy Prevention Initiatives. “Adolescent Pregnancy Prevention Program Participant Feedback Survey.” n.d.(a). <https://teenpregnancy.dph.ncdhhs.gov/docs/apppp/forms/ParticipantFeedbackSurvey.pdf>.

Version of the instrument that uses "instructor"

North Carolina Department of Health and Human Services, Division of Public Health, Teen Pregnancy Prevention Initiatives. "Teen Pregnancy Prevention Initiatives Participant Feedback Survey." n.d.(b). <https://teenpregnancy.dph.ncdhhs.gov/docs/prep/4-ParticipantSatisfactionSurvey.pdf>.

Scoring and Interpretation

North Carolina Department of Health and Human Services, Division of Public Health, Teen Pregnancy Prevention Initiatives. "Example Participant Satisfaction Formulas." N.d.(c). <https://teenpregnancy.dph.ncdhhs.gov/docs/prep/7-ExampleParticipantSatisfactionFormulas.xlsx>.

North Carolina Department of Health and Human Services, Division of Public Health, Teen Pregnancy Prevention Initiatives. "Instructions for Completing Participant Satisfaction Survey Responses Summary." N.d.(d) <https://teenpregnancy.dph.ncdhhs.gov/docs/prep/5-InstructionsforCompletingParticipantSatisfactionSurveyResponsesSummary.pdf>.

North Carolina Department of Health and Human Services, Division of Public Health, Teen Pregnancy Prevention Initiatives. "Participant Satisfaction Survey Responses Summary Template." N.d.(e) <https://teenpregnancy.dph.ncdhhs.gov/docs/prep/6-ParticipantSatisfactionSurveySummaryTemplate.docx>.

Background Resources

North Carolina Department of Health and Human Services, Division of Public Health, Teen Pregnancy Prevention Initiatives. "Adolescent Pregnancy Prevention Program Manual." September 2022a. <https://teenpregnancy.dph.ncdhhs.gov/docs/appp/APPPManual.pdf>.

North Carolina Department of Health and Human Services, Division of Public Health, Teen Pregnancy Prevention Initiatives. "Request for Applications RFA #A404 Adolescent Pregnancy Prevention Program." September 2022b. <https://wicws.dph.ncdhhs.gov/docs/A404/RFAA404-AdolescentPregnanc%20PreventionProgram.pdf>.

North Carolina Department of Health and Human Service, Division of Public Health, Teen Pregnancy Prevention Initiatives. "Teen Pregnancy Prevention Initiatives." September 2019. <https://teenpregnancy.dph.ncdhhs.gov/>.

North Carolina Department of Health and Human Services, Division of Public Health, Teen Pregnancy Prevention Initiatives. "Using Satisfaction Survey Results." N.d.(f). <https://teenpregnancy.dph.ncdhhs.gov/docs/prep/3-UsingSatisfactionSurveysGuidance.pdf>.

Program Observation Form for TPP Program Grantees

Purpose and type of educational program instrument was developed for

Purpose: Improvement; monitoring; research

Type of educational program: Adolescent pregnancy prevention

SRAE toolkit content found in the instrument

Topic and topic elements: Youth engagement (behavioral, cognitive); facilitator quality that promotes youth engagement (classroom management, supportive climate, content delivery)

Instrument characteristics

Type of instrument: Direct observation

Who completes it: Observer (program staff, self-trained)

Length/time: 14 individual items/ entire program session

Languages available: English

Instrument performance and the sample it was based on

Reliability: Not available

Validity: Not available

Study sample: Organizations participating in the teen pregnancy prevention (TPP) grant program funded by the Office of Population Affairs (OPA) within the U.S. Department of Health and Human Services

Availability

Level of permission required: No permission required for use

Associated costs: No known costs

Instrument website: The instrument can be found on the Reproductive Health National Training Center website:
<https://rhntc.org/resources/program-observation-form-tpp-program-grantees>

Instrument contacts

Office of Population Affairs, United States Department of Health and Human Services

opa@hhs.gov

Narrative

Description. The Program Observation Form for Teen Pregnancy Prevention (TPP) Program Grantees is a direct observation instrument that measures a facilitator's program delivery and the overall quality of a program session. It was developed for the federal TPP program, which funds nationwide programs for adolescents that aim to "prevent unintended teen pregnancies" (OPA 2023). The instrument is used for monitoring as one of the performance measures that TPP grant recipients (organizations that receive funding) report on semiannually to the Office of Population Affairs (OPA) in the U.S. Department of Health and Human Services. It is also used for research as indicated on the OPA website list of TPP grant recipients evaluations, some of which have used the observation form to study youth engagement and staff-participant interactions (OPA n.d.). Grant recipients also use it for internal improvement to help assess progress and inform work toward sustainability.

Content of instrument

- The Program Observation Form for Teen Pregnancy Prevention (TPP) Program Grantees contains 11 individual close-ended items:
 - One item each related to clarity of instructions, time management, pace of presentation of materials, participant understanding, and active participation
 - One item each related to facilitator qualities including knowledge, enthusiasm, composure, communication skills, and addressing questions
 - A final item that asks observers to score the session's overall quality
- The form also contains three optional open-ended items to support program improvement:
 - One item about implementation problems
 - Two items about strengths or weaknesses of the facilitator's delivery of the session

Administration points

- The Program Observation Form for TPP Program Grantees should be administered by someone who is familiar with the program but is neither a facilitator nor a participant.
- During the observation session, observers can take notes, but the form should be completed after observing the entire session.

Methods of scoring

- Close-ended items are scored on a 5-point scale ranging from 1 to 5, with three anchors that vary by question. A rating of 1 is negative, such as "poor," and a rating of 5 is positive, such as "excellent." The middle anchor (3) is always "average."
 - Items are reported separately. The OPA website lists several evaluations of TPP grant recipients, some of which have combined subsets of items for summary scores (OPA n.d.).
- The form does not include specific instructions for how to score the open-ended items.

Interpretability. For the item on overall program quality, OPA considers scores of 4 or 5 to demonstrate high quality. The form does not include specific instructions for how to interpret scores on the other items. However, higher scores generally indicate higher quality attributes.

Training support. For 10 of the 11 close-ended items, there are examples of what types of observations would result in a score of 1, 3, or 5. Under the close-ended item on overall program quality, there is a description of what excellent and poor sessions look like.

Reliability

Internal consistency reliability. Not applicable, as only individual item scores are reported.

Inter-rater reliability. No information available.

Validity

Content validity. No information available.

Construct/concurrent validity

Construct validity. No information available.

Concurrent validity. No information available.

Predictive validity. No information available.

Key considerations for using it in SRAE programs

- SRAE toolkit topic elements found in the instrument:
 - Items related to the behavioral and cognitive elements of youth engagement
 - Items related to the classroom management, supportive climate, and content delivery elements of facilitator quality
- The instrument is used to measure program delivery and overall quality of TPP programs that have the same kinds of characteristics as SRAE programs even though the curricula may be different. For example, both are structured to have a facilitator deliver a sexual education curriculum to a group of youth.
- Although the instrument is used in TPP programs, the items in the instrument are not tied to specific program content. Therefore, no changes are required for item wording; the Program Observation Form for TPP Program Grantees can be used as-is for SRAE programs. However, if you want to make any changes, you may do so without needing to contact OPA for permission.
- You do not need to contact the developers to use the instrument.

References

Instrument resources

Office of Population Affairs (OPA). Program Observation Form for TPP Program Grantees. U.S. Department of Health and Human Services, March 2022.

https://rhntc.org/sites/default/files/resources/opa_fidelity_program_observ_form_3-17-2022.pdf.

Background research

Office of Population Affairs (OPA). "FY 2010-2014 TPP Grantee Evaluations." U.S. Department of Health and Human Services, n.d. <https://opa.hhs.gov/research-evaluation/grantee-evaluations-fy2010-2014>.

Office of Populations Affairs (OPA). "Performance Measures Snapshot." U.S. Department of Health and Human Services, February 2023. <https://opa.hhs.gov/research-evaluation/teen-pregnancy-prevention-program-evaluations/tpp-performance-measures>.

This page has been left blank for double-sided copying.

Program Quality Instrument for Youth (PQI-Youth)

Purpose and type of educational program instrument was developed for

Purpose: Improvement; monitoring; research

Type of educational program: Positive youth development

SRAE toolkit content found in the instrument

Topics and topic elements: Youth engagement (emotional/psychological, cognitive); facilitator quality that promotes youth engagement (classroom management, supportive climate)

Instrument characteristics

Type of instrument: Self-administered survey

Who completes it: Youth (self-report and report about facilitator)*

Length/time: 19 individual items

Languages available: English; Spanish

* *There is also a version of the survey that facilitators can complete; it does not have accompanying information about instrument performance.*

Instrument performance and the sample it is based on

Reliability: Available (meets the minimum acceptable rating)

Validity: Available

Study sample:

2021–2022 Sample: 511 youth being served by eight Children, Youth, and Families At-Risk Initiative (CYFAR) programs in nine states; ages 10 to 18; 56% were male, and 41% were female; 65% were White, 17% were Black, 14% were American Indian or Alaska Native; 5% were Asian, 4% were Native Hawaiian or other Pacific Islander; and 9% did not report racial identity; 16% identified as Hispanic or Latino (CYFAR Team 2023).

2011–2016 Sample: 4,259* youth being served by CYFAR grant recipients, grades 3–12; demographics of all participants (including those in grades below 3 and past 12) collected at the beginning of the program showed about two-thirds in middle school or high school; about half were male, about half were White; one-third were Hispanic or Latino (CYFAR Evaluation).

*Sample size reflect the beginning of program year, with no information on remaining sample at the end of program year

2011–2018 Sample: 320 middle school and high school youth participating in five geographically diverse youth programs being delivered by CYFAR grant recipients (Ballard et al. 2021).

Availability

Level of permission required: No permission required for use

Associated costs: No known costs

Instrument website: https://cyfar.org/sites/cyfar.umn.edu/files/2023-08/administration_guide_-_cyfar_program_quality_8-1-23.pdf

Instrument contacts

Lynne M. Borden, University of Minnesota

lborden@umn.edu

Narrative

Description. The Program Quality Instrument for Youth (PQI-Youth) is a survey that measures the quality of youth programs designed for participants ages 10 to 17. It was developed for Children, Youth, and Families At-Risk (CYFAR), an initiative supported by the United States Department of Agriculture that funds programs nationally. The programs provide quality programming for “vulnerable, at-risk, low-income, and low-resource children, youth, and families to promote positive life outcomes” (University of Minnesota 2023a). The survey is used for monitoring as one of the common measures that Sustainable Community Projects (also known as CYFAR grant recipients, or organizations that receive CYFAR funding) are required to report to the CYFAR Professional Development and Technical Assistance (PDTA) Center⁷ at the end of each program to assess their impact. It is also used for research as part of broader CYFAR PDTA Center evaluations conducted to understand program impact and in local program evaluation efforts that CYFAR grant recipients may undertake with the support of the CYFAR PDTA Center. Lastly, it is used for program improvement to increase positive youth outcomes.

Content of instrument. The PQI-Youth captures aspects of program quality using 19 individual items in six subscales:

1. Supportive Relationships (three items) measures how program staff treat youth.
2. Opportunities to Belong (three items) measures youth’s feelings of acceptance and personal interactions with program staff.
3. Skill Building and Efficacy (four items) measures how youth are encouraged to be leaders, and what they learn from the program.
4. Community and Family Integration (four items) measures youth’s opportunity to choose the activities they want to do, and family and community involvement.
5. Positive Norms (two items) measures how youth treat each other.
6. Structure and Safety (three items) measures youth’s general sense of safety at the program and the program’s rule structure

Administration points

- The PQI-Youth is self-administered and allows youth program participants to report on themselves and the facilitator.
- It was developed in English, with a Spanish version available.
- The developers recommend administering the survey at the end of the program.
- The CYFAR common measures include versions of the survey for program staff and other participants (adults ages 18 and older, and children ages 9 and younger). Those versions do not have accompanying information on instrument performance.

⁷ The CYFAR Professional Development and Technical Assistance (PDTA) Center is a collaboration between the University of Minnesota, Pennsylvania State University, and coaches from eight other universities. It provides professional development training and technical assistance to CYFAR grant recipients. (University of Minnesota 2023b)

Methods of scoring. Items are scored on a 5-point scale of 1 (never), 2 (rarely), 3 (sometimes), 4 (a lot) and 5 (always). The administration guide provides scoring information to calculate total and subscale mean (or average) scores, which range from 1 to 5.

Interpretability. The developers do not provide specific instructions. However, higher scores generally indicate higher program quality because all items use the same scale and are positive statements.

Training support. The administration guide, although designed for CYFAR grant recipients, includes instructions on how to prepare for survey administration, scripts to use to discuss the survey with youth, and scoring guidance.

Reliability

Internal consistency reliability. The developers reported Cronbach's alpha coefficients for the total score and each subscale (2021–2022 Sample; CYFAR Team 2023).

- Total score: 0.93
- Subscales: Ranged from 0.72 to 0.84

Inter-rater reliability. Not applicable.

Validity

Content validity. The developers used research identifying eight elements to support youth program quality to inform the instrument. These elements appeared in Eccles and Gootman's (2002) National Research Council book, *Community Programs to Promote Youth Development*.

Construct/concurrent validity

Construct validity. The developers used exploratory factor analysis (EFA) to determine the structure of the 2013 version of the PQI-Youth instrument. The EFA results indicated a six-factor solution that the developers used to establish the current 19-item instrument with six subscales (2021–2022 Sample; CYFAR Team 2023).

Concurrent validity

- In the 2016 CYFAR Evaluation Report (2011–2016 Sample), the developers calculated correlations between the 2013 version of the PQI-Youth (with slightly different subscales from those in the latest instrument) and five study-developed “core competency” scales on caring, decision making, critical thinking, social conscience, and personal values. These core competencies are behaviors or attitudes common to the goals of CYFAR that one may expect positive correlations between program quality and these competencies.
 - Total score and core competency scales: 0.38 (caring) to 0.48 (personal values)
 - PQI-Youth subscales (2013 version of instrument) and core competency scales:
 - Safety: 0.18 (critical thinking) to 0.30 (caring)
 - Relationships: 0.31 (caring) to 0.46 (social conscience)
 - Social Norms: 0.33 (caring) to 0.41 (personal values)
 - Support for Efficacy: 0.29 (caring) to 0.41 (personal values)
 - Skill Building: 0.35 (caring) to 0.48 (personal values)

- In that report, the developers also presented correlations between the 2013 version of the PQI-Youth instrument (again with subscales slightly different from those in the latest instrument) and short-term outcomes of the program. These outcomes were measured by study-developed scales related to nutrition, science, technology, workforce preparation (middle and high school youth only), civic engagement (high school youth only), and parenting (teen parents only).
 - Total score and the short-term outcomes: 0.12 (nutrition, middle school only) to 0.50 (civic engagement)
 - PQI-Youth subscales (2013 version of instrument) and short-term outcomes:
 - Safety: -0.06 to 0.37
 - Relationships: 0.03 to 0.40
 - Social Norms: -0.04 to 0.44
 - Support for Efficacy: 0.05 to 0.51
 - Skill Building: 0.06 to 0.47

Predictive validity. Ballard et al. (2021) focusing on high school participants (2011-2018 Sample), found that the Skill-Building and Social Norms subscales of the 2013 version of the PQI-Youth instrument predicted increases in civic engagement skills.

Key considerations for using it in SRAE programs

- SRAE toolkit topic elements found in the instrument:
 - The Supportive Relationships, Opportunities to Belong, and Positive Norms subscales contain items related to the emotional/psychological element of youth engagement and the supportive climate element of facilitator quality.
 - The Structure and Safety subscale contains items related to the classroom management element of facilitator quality.
 - The Skill Building and Efficacy subscale contains some content related to the cognitive element of youth engagement.
 - The Community and Family Integration subscale is not within the scope of this toolkit.
- The developers initially created the instrument for CYFAR grant recipients that offer programs that have the same kinds of characteristics as SRAE programs, such as curriculum-based instructional components and group-centered activities.
- Although the items in the instrument are not tied to specific program content, it might be necessary to make minor word changes to items in the instrument to use it in the SRAE program setting, such as changing “adults” to “facilitator.” Check with the developers to determine if these minor changes can be made.
- Some items in subscales may not be relevant to SRAE programs, such as an item in the Skill Building and Efficacy subscale that asks whether youth receive encouragement to be leaders. If you want to remove any items, check with the developers to determine if these items may be removed, and if new subscale scores may be calculated and interpreted. However, the information presented here on reliability and validity would not apply if changes are made.
- You do not need to contact the developers to use the instrument.

References

Instrument resources

Children, Youth, and Families At-Risk Professional Development and Technical Assistance Center (CYFAR PDTA Center) Team. "CYFAR Program Quality Instrument, Youth (PQI-Youth) Survey and Administration Guide." University of Minnesota and Pennsylvania State University: CYFAR PDTA Center, 2023.
https://cyfar.org/sites/cyfar.umn.edu/files/2023-08/administration_guide_-_cyfar_program_quality_8-1-23.pdf.

Background research

Ballard J., L. Borden, and D.F. Perkins. "Program Quality Components Related to Youth Civic Engagement." *Children and Youth Services Review*, vol. 126, 106022, 2021.

Ballard, J., M. Otto, S. Grant, S. van den Hoogenhof, and L. M. Borden. "Construct Validity of the Program Quality Instrument for Youth." Poster for National Council on Family Relations. University of Minnesota, 2023.

Children, Youth and Families At-Risk (CYFAR) Team. "Construct Validity of the Program Quality Instrument for Youth." Preliminary technical information provided to Mathematica SRAE toolkit authors, June 14, 2023.

Children, Youth and Families At-Risk Professional Development and Technical Assistance (CYFAR PDTA) Team. "2016 CYFAR Evaluation Report." 2017.

Children, Youth and Families At-Risk Professional Development and Technical Assistance (CYFAR PDTA) Team. "CYFAR Youth Common Measures 2022." January 18, 2023. https://cyfar.org/sites/cyfar.umn.edu/files/2023-08/cyfar_2022.share_accessible.final_.pdf.

University of Minnesota. (2023a). "About CYFAR." June 14, 2023. <https://cyfar.org/about>.

University of Minnesota. (2023b). "About CYFAR Professional Development." June 14, 2023. <https://cyfar.org/about-cyfar-professional-development>.

University of Minnesota. (2023c). "CYFAR Sustainable Community Project Grantees." 2023.
<https://cyfar.org/community-grantee>.

This page has been left blank for double-sided copying.

Tiffany-Eckenrode Program Participation Scale (TEPPS)

Purpose and type of educational program instrument was developed for

Purpose: Improvement; research

Type of educational program: Positive youth development

SRAE toolkit content found in the instrument

Topics and topic elements: Youth engagement (behavioral, emotional/psychological, cognitive); facilitator quality that promotes youth engagement (supportive climate)

Instrument characteristics

Type of instrument: Self-administered survey

Who completes it: Youth (self-report and report about facilitator)

Length/time: 20 individual items/ less than 10 minutes

Languages available: English

Instrument performance and the sample it is based on

Reliability: Available (meets the minimum acceptable rating)

Validity: Available

Study sample: 331 youth ages 12 to 18 attending after-school programs at eight community-based agencies in New York City in 2008; 62% of participants were female, 38% were male, and <1% were transgender male-to-female; 42% were Black non-Hispanic, 12% were Black Hispanic, 3% were other race Hispanic, 7% were other race non-Hispanic, 26% were more than one race Hispanic, 9% were more than one race non-Hispanic, and <1% were unknown race.

Availability

Level of permission required: No permission required for use: adaptations allowed, inform developers of use and adaptations, request sharing findings and feedback with developers

Associated costs: No known costs

Instrument website: <http://www.bctr.cornell.edu/wp-content/uploads/2012/07/TEPPS-pencil-and-paper-version.pdf>

Instrument contacts

Jennifer Sarah Tiffany, Cornell University

jst5@cornell.edu

Narrative

Description. The Tiffany-Eckenrode Program Participation Scale (TEPPS) is a survey that measures youth participation. It was administered to youth ages 12 to 18 in programs that promote positive youth development, including after-school programs. The developers created the TEPPS for research to help researchers and practitioners gain a more comprehensive understanding of the quality of participation in youth programs by focusing on aspects of participation that go beyond intensity, duration, and breadth of youth involvement. It can also be used to understand links between youth participation and positive outcomes in youth development. It was part of a study to evaluate how participation in after-school programs promoting positive youth development could enhance adolescent sexual health and reduce HIV risk. It can also be used for program improvement of staff practices that are linked to participation.

Content of instrument. The TEPPS captures the quality of youth participation using 20 individual items in four subscales:

1. Personal Development⁸ (seven items) measures youth's perspectives on how challenging and helpful the program was and their perspectives on how much support they received from program staff.
2. Voice/Influence (four items) measures youth's ability to make program decisions and involvement in activities.
3. Safety/Support (four items) measures youth's general sense of safety and their relationships with program staff and other youth participants.
4. Community Engagement (five items) measures how the program connects to the youth's school and family, and how youth are involved and treated in their neighborhood and community.

Administration points. The TEPPS is self-administered and allows youth program participants to report on themselves and their facilitator.

Methods of scoring

- Items are scored on a 5-point scale of 1 (not at all true for me), 2 (not really true for me), 3 (I'm neutral about this), 4 (sort of true for me), and 5 (very true for me).
- Tiffany et al. (2012b) show items are summed to get subscale scores. Subscale scores can range from 4 to 20 for the Voice/Influence and Safety/Support subscales, 5 to 25 for the Community Engagement subscale, and 7 to 35 for the Personal Development subscale.
- The total score is also a sum of the items and can range from 20 to 100.

Interpretability. The developers note that higher total or subscale scores indicate higher levels of program participation quality.

Training support. The developers state minimal training is needed.

⁸ Developers also refer to this subscale as "Personal Involvement."

Reliability

Internal consistency reliability. The developers reported Cronbach's alpha coefficients for the total score and each subscale.

- Total score: 0.87
- Subscales: Ranged from 0.66 to 0.82 (with Voice/Influence falling below 0.70)

Inter-rater reliability. Not applicable.

Validity

Content validity. The developers used ecological-developmental literature and existing youth instruments on positive practices in youth development to inform the instrument. They conducted a pilot study and a validation study to refine the instrument.

Construct/concurrent validity

Construct validity

- The developers conducted exploratory and confirmatory factor analysis (EFA and CFA) to confirm the instrument's subscales. The EFA results indicated that a one-factor solution and a four-factor solution could both be plausible. A four-factor CFA model produced acceptable results for all fit statistics.
- Correlations between the subscales ranged from 0.14 to 0.36, where lower correlations indicate the subscales measure distinct constructs.

Concurrent validity. The developers examined correlations between the TEPPS total and subscale scores with study-developed scales of intensity, duration, breadth of program participation, and family connectedness as measures of convergent validity (to see if positively correlated with similar constructs).

- Total score: 0.03 (duration) to 0.27 (family connectedness)
- Subscales:
 - Personal Development: -0.03 (duration) to 0.27 (family connectedness)
 - Voice/Influence: -0.01 (intensity) to 0.31 (breadth)
 - Safety/Support: 0.05 (breadth) to 0.25 (intensity)

Predictive validity. The developers also examined longitudinal associations between variables related to positive youth development and HIV risk-reduction practices. Youth in programs with lower TEPPS scores showed decline in HIV risk-reduction practices, whereas those in programs with higher TEPPS scores maintained their level of practices.

Key considerations for using it in SRAE programs

- SRAE toolkit topic elements found in the instrument:
 - The Personal Development subscale contains items related to the cognitive element of youth engagement, and the supportive climate element of facilitator quality.
 - The Voice/Influence subscale contains items related to the behavioral element of youth engagement.
 - The Safety/Support subscale contains items related to the emotional/psychological element of youth engagement.
 - The Community Engagement subscale is not within the scope of this toolkit.
- The developers initially created the instrument for positive youth development programs that have the same kinds of characteristics as SRAE programs, such as having an instructional component led by adult staff and offering the program to youth in a group setting.
- Although the items in the instrument are not tied to specific program content, it might be necessary to make minor word changes to items in the instrument to use it in the SRAE program setting, such as changing “adult” or “staff” to “facilitator.” The TEPPS developers acknowledge that organizations make minor changes to the instrument when they use it.
- Some items in the subscales may not be relevant to SRAE programs, such as an item in the Voice/Influence subscale that asks participants about their influence in program decisions. If you want to remove any items, check with the developers to determine if these items may be removed, and if new total and subscale scores may be calculated and interpreted. However, the information presented here on reliability and validity would not apply if changes are made.
- You do not need to contact the developers to use the instrument. The developers request that you let them know if you are using the instrument, including if you adapted it, and share any findings and feedback (Tiffany 2023).

References

Instrument resources

Tiffany, J. S., D. Exner-Cortens, and J. Eckenrode. “A New Measure for Assessing Youth Program Participation.” *Journal of Community Psychology*, vol. 40, no. 3, 2012(a), pp. 277–291.

Background research

Tiffany, J. S., email discussion, June 2023.

Tiffany, J. S., D. Exner-Cortens, and J. Eckenrode. “Longitudinal Associations Between HIV Risk Reduction and Out-of-School Time Program Participation.” *Journal of Adolescent Health*, vol. 52, no. 6, 2013, pp. 795–797.

Tiffany, J. S., D. Exner-Cortens, and J. Eckenrode. Tiffany-Eckenrode Program Participation Scale (TEPPS) Supplement to “A New Measure for Assessing Youth Program Participation.” *Journal of Community Psychology*, vol. 40, no. 3, 2012(b), pp. 277–291.

Youth Development Assessment Device (YDAD)

Purpose and type of educational program instrument was developed for

Purpose: Improvement

Type of educational program: Positive youth development

SRAE toolkit content found in the instrument

Topic and topic elements: Youth engagement (behavioral, emotional/psychological, cognitive); facilitator quality that promotes youth engagement (supportive climate)

Instrument characteristics

Type of instrument: Self-administered survey

Who completes it: Youth (self-report and report about facilitator)

Length/time: 21 individual items

Languages available: English

Instrument performance and the sample it is based on

Reliability: Available (meets the minimum acceptable rating)

Validity: Available

Study sample:

2004 Sample: 579 youth from 12 neighborhood youth centers in Connecticut; 67% were male, and 33% were female; 70% were in grades 6–8, and 30% were in grades 9–12; 42% were Black, 42% were Hispanic or Latino, and 4% were White; 73% reported receiving free school lunch (Sabatelli et al. 2009).

2005 Sample: 495 youth from 12 neighborhood youth centers in Connecticut; 60% were male, and 40% were female; 64% were in grades 6–8, and 36% were in grades 9–12; 46% were Black, 38% were Hispanic or Latino, and 3% were White; 76% reported receiving school lunch (Sabatelli et al. 2009).

2008 Sample: 682 youth being served by 19 neighborhood youth centers in Connecticut; 56% were male, and 44% were female; 48% were in grades 5–8; 49% were in grades 9–12; 63% were Black, 19% were Hispanic or Latino, 8% were White, 4% were multiracial, 1% were Asian, 1% were Native American, and 5% were another race; 69% reported receiving free or reduced cost school lunch (CARHD 2009).

2009 Sample: 1,090 youth being served by the same 19 neighborhood youth centers in Connecticut; 56% were male, and 44% were female; 53% were in grades 5–8, 47% were in grades 9–12; 53% were Black, 25% were Hispanic or Latino, 9% were White, 6% were multiracial, 1% were Asian, and 5% were another race; 69% reported receiving free or reduced cost school lunch (CARHD 2009).

Availability

Level of permission required: No permission required for use

Associated costs: No known costs

Instrument website: None

Instrument contacts

Ronald M. Sabatelli, University of Connecticut

Ronald.sabatelli@uconn.edu

Narrative

Description. The Youth Development Assessment Device (YDAD) is a survey that measures program quality through youth experiences in the program. It was administered to youth ages 12 to 18 in programs that promote positive youth development, including neighborhood youth centers. The developers created the YDAD for improvement to inform the planning and delivery of programs.

Content of instrument. The YDAD captures the developmental quality⁹ of positive youth development programs using 21 individual items in three subscales:

1. Emotional Safety and Well-Being (seven items) measures how physically safe and emotionally valued youth feel in the program environment.
2. Supportive Environment (seven items) measures the guidance, emotional support, and practical support that program staff provide to youth program participants.
3. Challenge and Involvement (seven items) measures youth's sense of belonging, collaboration, and the extent to which program activities are constructively challenging.

Administration points. The YDAD is self-administered and allows youth program participants to report on themselves and their facilitator.

Methods of scoring

- Items are scored on a 7-point scale of 1 (strongly disagree), 2 (disagree), 3 (slightly disagree), 4 (neither agree or disagree), 5 (slightly agree), 6 (agree), and 7 (strongly agree).
- Although the developers do not provide scoring instructions, Center for Applied Research in Human Development (2009) provides item-level means to understand how the items performed and show that items can be averaged to obtain subscale scores. Subscale scores can range from 1 to 7.

Interpretability. The developers do not provide specific instructions. However, higher scores generally indicate higher quality because all items use the same scale and are positive statements.

Training support. No information available.

Reliability

Internal consistency reliability. Cronbach's alpha coefficients ranged from 0.83 to 0.95 for the three subscales (2008 and 2009 Samples; CARHD 2009).

Inter-rater reliability. Not applicable.

⁹ "Developmental quality is the extent to which a program provides a set of program components that have been found to facilitate positive youth development and academic achievement" (Sabatelli et al., 2009, p. 362).

Validity

Content validity. The developers used research on the developmental quality of youth programs to inform the instrument. Items were reviewed by graduate students, professors, and program providers.

Construct/concurrent validity

Construct validity

- The developers conducted exploratory and confirmatory factor analysis (EFA and CFA) to determine the instrument's structure using a 63-item survey version with a 4-point scale. The EFA results indicated a three-factor solution among of subset of items administered (2004 Sample; Sabatelli et al. 2009). The CFA established acceptable model fit with a three-factor model with 21 items (2005 Sample; Sabatelli et al. 2009).
- Correlations between subscales on the 4-point scale ranged from 0.67 (between Challenge and Involvement and Emotional Safety and Well-Being) to 0.86 (between Supportive Environment and Challenge and Involvement). Though the subscales are strongly correlated, they provide distinct information. The developers tested competing models, and the three-factor model produced the best model fit results.

Concurrent validity. No information available.

Predictive validity. No information available.

Key considerations for using it in SRAE program

- SRAE toolkit topic elements found in the instrument:
 - The Emotional Safety and Well-Being and Supportive Environment subscales contain items related to the emotional/psychological element of youth engagement and the supportive climate element of facilitator quality.
 - The Challenge and Involvement subscale contains items related to the behavioral, emotional/psychological, and cognitive elements of youth engagement.
- The developers initially created the instrument for positive youth development programs that have the same kinds of characteristics as SRAE programs, such as life skills training and enrichment and leadership development opportunities.
- Although the items in the instrument are not tied to specific program content, it might be necessary to make minor word changes to items in the instrument to use it in the SRAE program setting, such as changing "center" to "classroom" and "staff" to "facilitator." Check with the developers to determine if these minor changes can be made.
- Some items in subscales may not be relevant to SRAE programs, such as an item in the Emotional Safety and Well-Being subscale related to youth's opportunities to design programs offered at a center. If you want to remove any items, check with the developers to determine if these items may be removed, and if new subscale scores may be calculated and interpreted. However, the information presented here on reliability and validity would not apply if changes are made.
- You do not need to contact the developers to use the instrument.

References

Instrument resources

Center for Applied Research in Human Development (CARHD). "2007-2009 Process Evaluation Report: Governor's Urban Youth Violence Prevention Grant." University of Connecticut, 2009. <https://portal.ct.gov/-/media/OPM/CJPPD/CjJyd/JjydPublications/GUYVP20072009FinalReport91909pdf.pdf>.

Background research

Sabatelli, R. M., S. A. Anderson, I. Kosutic, J. Sanderson, and S. Rubinfeld, "A Validation Study of the Youth Development Assessment Device." *Family Relations*, vol. 58, 2009, pp. 361–372.

Youth Program Quality Assessment (YPQA)

Purpose and type of educational program instrument was developed for

Purpose: Improvement; monitoring; research

Type of educational program: Positive youth development

SRAE toolkit content found in the instrument

Topic and topic elements: Youth engagement (emotional/psychological); facilitator quality that promotes youth engagement (classroom management, supportive climate, content delivery)

Instrument characteristics

Type of instrument: Direct observation

Who completes it: Observer (program staff, with publisher training [at least one staff]; additional program staff, self-trained; or publisher's data collector)

Length/time: 63 individual items/ time to administer depends on the number of observers and length of the sessions (each observer should observe an entire program session), another 3 hours for scoring; see description for details.

Languages available: English; Spanish

Instrument performance and the sample it is based on

Reliability: Available (some meet the minimum acceptable rating)

Validity: Available

Study sample: 59 youth-serving organizations in Michigan, primarily Detroit and Grand Rapids. Most were in urban and suburban areas, with a few in rural regions. Majority of programs were after-school programs; some in school, summer, and residential programs were also included. About 40% of the participants were male, and 60% were female. The majority reported they achieved mostly Bs or better in school.

Wave 1 (2003–2005): 335 youth from 22 programs in 13 organizations; participants had a mean age of 14; 40% were Black, 11% were Latino, 33% were White, and 16% were other races.

Wave 2 (2004–2005): 1,300 youth from 116 programs within 46 organizations; participants had a mean age of 13; 38% were Black, 7.5% were Latino, 40% were White, and 15% were other races.

Availability

Level of permission required: Permission required for use, complete the request page with user agreement:

<https://form.jotform.com/203156864824157>

Associated costs: PQA Basics Online Course (Youth) (\$125/person with at least one staff to complete); Youth PQA Handbook (\$12; one per observer); Online Scores Reporter Access (\$75/year); all of these supports are available, see training support section for further details

Instrument website: <https://forumfyi.org/weikartcenter/assessments/>

Instrument contacts

David P. Weikart Center for Youth Program Quality, a unit of the Forum for Youth Investment

info@cypq.org

Narrative

Description. The Youth Program Quality Assessment (YPQA) Form A is a direct observation instrument that measures the quality of programs delivered to youth in grades 4 to 12. The instrument was designed for use in after-school programs, community organizations, and schools. Form A can be used to research program quality and monitor quality over time. It can also be used for program improvement to identify opportunities for staff training and program improvement for promoting youth motivation and engagement.

Content of instrument

- The YPQA Form A captures youth program quality using 63 individual items organized within subscales, which are then grouped into four domains:
 1. Safe Environment (19 items) measures how the program supports youth’s emotional and physical safety, including having clean and secure conditions.
 - Subscales: Emotional Safety, Healthy Environment, Emergency Preparedness, Accommodating Environment, and Nourishing
 2. Supportive Environment (24 items) measures how staff supports youth’s educational and personal development and growth, including facilitating opportunities for active participation and learning, fostering positive relationships, and developing new skills.
 - Subscales: Warm Welcome, Session Flow, Active Engagement, Skill-Building, Encouragement, and Reframing Conflict
 3. Interaction (12 items) measures youth’s experience with their peers, including peer-to-peer support, sense of inclusion and belonging, and engagement and ability to express ideas in group activities. This also includes how staff support and encourage this positive environment and provide opportunities for youth to partner with the staff.
 - Subscales: Belonging, Collaboration, Leadership, and Adult Partners
 4. Engagement (8 items) measures youth’s ability to play an active role in their experience and learning, including being given opportunities to plan, act as decision makers, practice reflection, and identify lessons learned.
 - Subscales: Planning, Choice, and Reflection
- The YPQA also includes a Form B—Organization Practices and Policies. It is an interview-based instrument conducted with program administrators and focused on organizational support quality. This instrument is not within the scope of this toolkit.

Administration points

- The YPQA Form A is used to observe a session of a particular program (which the publisher refers to as a program offering such as a class or club, with potentially multiple offerings at a site or organization).
- The YPQA Form A can be administered by program staff for self-assessment or by publisher-trained observers for external assessments. This profile focuses on using the YPQA Form A for self-assessment. For additional information about how the YPQA Form A is administered by publisher-trained observers, refer to the YPQA handbook.

- The publisher recommends conducting observations as a team with one self-assessment leader (usually the program administrator) and at least two other staff acting as observers. The YPQA handbook includes more information on how to choose observers.
- During the observation session, observers should take notes. The form should be completed after the observation session.
- Each observer should observe a full program session. After all observers complete their observations, the observation team meets for a scoring session that lasts about three hours, where everyone completes the PQA Form A together.

Methods of scoring

- Items are scored on a 5-point scale with three anchors of 1, 3, and 5. The instrument includes descriptors for each point on the scale. The observer can record an "X" instead of a number if no opportunity arises during the observation to score the item. The observer can also record an "NS" (for Not Scored) for an item if before the observation the observation team determines that the item is not relevant to the program. The YPQA handbook provides details on how marking items X or NS impacts subscale and domain scores. Generally, an observer must be able to rate at least half of the items in a subscale and half of the subscales within a domain to create these scores.
- Once each item is scored, subscale scores are calculated by averaging the item scores within the subscale.
- In turn, domain scores are calculated by averaging the subscale scores within the domain.
- Finally, a total score is calculated by averaging the four domain scores.

Interpretability. Ratings of 1 typically indicate the absence of a practice or implementation of a poor practice. Ratings of 3 indicate the informal or inconsistent implementation of a practice. Ratings of 5 indicate high quality practices with "intentional delivery" (David P. Weikart Center for Youth Program Quality 2020). These ratings can be used to track a program's progress over time and promote programs of optimal quality by highlighting areas for improvement, which in turn leads to better outcomes for youth participants.

Training support. The Center for Youth Program Quality (CYPQ) has an online store (<https://cypq.myshopify.com/collections/all>) with several tools available to purchase:

- The publisher recommends that at least the self-assessment leader complete the PQA Basics Online Course (\$125/person) before using the PQA. It is a one-hour virtual training that prepares participants to use the YPQA. A live version of this training is also available periodically for \$200 per person.
- The publisher offers a Youth PQA handbook (\$12 per observer) that provides step-by-step guidance on the self-assessment process. The publisher recommends the observation team consult the handbook while conducting observations and scoring the items (David P. Weikart Center for Youth Program Quality 2020). The handbook also describes how the self-assessment leader can conduct a Take It Back training for the other members of the observation team to familiarize them with the PQA Form A and help them understand how the team-based self-assessment works.
- The publisher offers scoring support through its Online Scores Reporter (\$75 annually/person). The publisher recommends programs use this tool to enter data and calculate scores. It can also produce reports based on observation scores, and record and track program improvement plans and assessments.

Reliability

Note. Changes were made between Wave 1 and Wave 2 to many subscales for all domains except Safe Environment (Smith and Hohmann 2005). Some information is only available for Wave 1; we present information on both Wave 1 and Wave 2 when available for full context.

Internal consistency reliability. Smith and Hohmann (2005) calculated Cronbach's alpha coefficients for the total score and its four domains for Wave 1 and Wave 2:

- Total score: 0.84 and 0.74, respectively
- Domains: Ranged from 0.38 to 0.85 across the two waves. The following domains did not meet the minimum acceptable rating for at least one wave of data collection:
 - Safe Environments (0.38 and 0.43)
 - Interaction (0.72 and 0.64)

Test-retest reliability. Pearson-r correlation coefficients ranged from 0.81 to 0.98 for observer pairs visiting the same program across three time points within three months.

Inter-rater reliability. Smith and Hohmann (2005) assessed inter-rater reliability using two different techniques.

- Intraclass correlation coefficients (ICCs) were calculated based on paired observations conducted at three time points each for 22 programs (Wave 1).
 - Total score: 0.66
 - Domains: Ranged from 0.48 to 0.83
- Percent perfect agreement, based on 44 observer pairs (Wave 1), ranged from 48 to 80 percent across subscales (calculated as the averaged percent item agreement in a subscale).¹⁰

Validity

Content validity

- The YPQA development began in the 1990s and was based on best practices in the field of youth programming.
- During the validation study, expert practitioners, researchers, and trainers/data collectors provided feedback on the instrument.
- The instrument was revised during the validation study (Smith and Hohmann 2005) and in 2012 and 2020 (David P. Weikart Center for Youth Program Quality 2020).

¹⁰ Some subscales have only two or three items. This can lead to low percent agreement when only one item lacks perfect agreement. Three subscales' agreement were not calculated. These subscales allow raters to skip items when certain circumstances are not present (for example, food and drink are not served, or conflict does not occur).

Construct/concurrent validity

Construct validity

- Smith and Hohmann (2005) examined the factor structure of the instrument using subscale scores associated with three of the four domains—Supportive Environment, Interaction, and Engagement. They excluded Safe Environment because it had little or no variance, with most programs scoring at the high end. Subscales generally contributed strongly to the domain as associated with the published instrument. However, some subscales shared greater variance with other domains other than the one in the published instrument. Therefore, the structure of the instrument domains may not be consistent for other researchers.
- The correlations between the three domains ranged from 0.61 to 0.62, indicating they are generally measuring distinct constructs.

Concurrent validity

- The validation study examined Pearson’s correlation coefficients between the YPQA Form A total and domain scores with the Youth Development Strategies, Inc. Youth Survey (YDSI) fielded at the same time for two waves of data collection. The YDSI is a survey that measures youth experiences in a program, and it may be expected to correlate with the YPQA.
 - Total scores of YPQA Form A and YDSI: 0.75 and 0.47 (Wave 1 and Wave 2, respectively)
 - Domain score with its most relevant YDSI subscale across Waves 1 and 2:
 - Safe Environment and YDSI physical safety: 0.38 and 0.42
 - Supportive Environment and YDSI belonging: 0.44 and 0.29
 - Interaction and YDSI peer knowledge: 0.69 and 0.44
 - Engagement and YDSI decision making: 0.48 and 0.32
 - Domain scores also predicted YDSI scores, after accounting for youth background and program attendance, as follows:
 - Interaction predicted YDSI decision making
 - Engagement predicted YDSI youth giving back to their communities and YDSI youth growth through program attendance
- The validation study also compared the YPQA Form A with other indicators of quality or best practices.
 - Total score was correlated with expert ratings as follows:
 - 0.59 with organization youth-centeredness
 - 0.52 with organization availability of resources
 - Leaders trained on a youth programming approach had higher total scores than untrained leaders.

Predictive validity. No information available.

Key considerations for using it in SRAE program

- SRAE toolkit topic elements found in the instrument:
 - The Safe Environment domain contains items related to the supportive climate element of facilitator quality.
 - The Supportive Environment domain contains items related to the classroom management, supportive climate, and content delivery elements of facilitator quality.
 - The Interaction domain contains items related to the emotional/psychological element of youth engagement and the classroom management and supportive climate elements of facilitator quality.
 - The Engagement domain contains items related to the content delivery element of facilitator quality.
- The publisher initially created the instrument for a wide range of youth programs, including school-based extracurricular and after-school programs. These programs have the same kinds of characteristics as SRAE programs, such as offering programming led by adult staff to youth in a group setting, including structured group activities. Some youth programs may also include curriculum-based instruction.
- Although the items in the instrument are not tied to specific program content, it might be necessary to make minor word changes to items in the instrument to use it in the SRAE program setting, such as changing “staff” to “facilitator.” Check with the publisher to determine if these minor changes can be made.
- Some items or subscales may not be relevant to SRAE programs, such as the Nourishment subscale that reports foods and drinks provided during the session. Before conducting observations, you can decide not to score these items or subscales. The YPQA handbook provides details on how not scoring individual items or subscales impacts domain and total scores.
- You must receive permission from the publisher to use the instrument. You can obtain permission by completing the request page with the user agreement found on the publisher’s website.

References

Instrument resources

David P. Weikart Center for Youth Program Quality. “Youth Program Quality Assessment (PQA).” High/Scope Educational Research Foundation, 2020.

Smith, C., T. Akiva, M. Jones, A. Sutter, B. Hillaker, L. Wallace, and G. McGovern. “Program Quality Assessment Handbook: Youth Version.” Forum for Youth Investment, David P. Weikart Center for Youth Program Quality, Revised 2018.

Background research

Smith, C., and C. Hohmann. “High/Scope Youth PQA Technical Report: Full Findings From the Youth PQA Validation Study.” High/Scope Educational Research Foundation, 2005.

Links to additional resources (Accessed May 2023)

For information on purchasing training supports, see <https://cypq.myshopify.com/collections/all>.

For information on training workshops, see <https://forumfyi.org/weikartcenter/workshops/>.

For information on live workshops for program observers, see <https://forumfyi.org/event/pqabasics-042023/>.

For information on the Online Scores Reporter, see <https://forumfyi.org/weikartcenter/srguide/>.

Appendix A. Defining youth engagement and facilitator quality that promotes youth engagement

Our plan for the measures toolkit began with developing definitions of youth engagement and facilitator quality that promotes youth engagement (facilitator quality). Appendix A summarizes the findings from a literature search of key terms and details the approach the team took to develop working definitions of youth engagement and facilitator quality that guided the search for and screening of instruments for the toolkit.

A. Youth engagement

The study began with a literature search focused on the term “youth engagement” in sexual risk avoidance education (SRAE) and positive youth development programs. (Exhibit A.1 gives the parameters used in the literature search.) The search yielded no literature that discussed youth engagement in educational group settings per the defined scope, but one article and two agencies attempted to define the term.

- In a study of teacher narratives in programs designed to prevent drug use, Miller-Day et al. (2015) rated engagement as the degree of student attention to and participation in the lesson.
- The Children, Youth, and Families At-Risk Initiative (CYFAR), a national initiative that funds community projects, developed a youth engagement survey that measures “concentration, interest, and enjoyment in the program” (University of Minnesota 2019).
- [Youth.gov](#), which supports youth development programs, has a tool to assess youth involvement and engagement in the community (Youth.gov n.d.). The section on youth involvement asks program leaders how youth take initiative in project work, participate in meetings and events, share ideas, and help one another, and whether they are excited about their involvement in the program.

Exhibit A.1. Parameters used in the youth engagement literature search

Search parameter	Terms and resources
Keywords	Youth*[ti] OR student*[ti] OR adolescen*[ti] OR 12-17 years[ti] OR high[ti] OR middle[ti] AND engage*[ti] AND classroom*[ti] OR school*[ti] OR educat*[ti] OR develop*[ti] program*[ti] OR (((prevention[ti] OR avoid*[ti] OR educat*[ti] OR sexual risk[ti] OR pregnan*[ti] OR drug[ti] OR alcohol[ti] OR tobacco[ti] OR bully
Disciplines and fields targeted	Youth development programs OR education
Relevant databases	Academic Search Premier, PubMed, ERIC, Education Research Complete, CINAHL with Full Text, EBSCO Database

Note: [ti] = terms in the title.

To gain deeper understanding of youth engagement in SRAE, the literature search was expanded to include educational settings for a related topic of “school engagement.” Five reviewed articles revealed that many researchers define school engagement as made up of two to four elements.

- There was a general agreement within the field that elements of school engagement include behavioral engagement and emotional or psychological engagement. Behavioral engagement considers students’ actions, such as their participation in school activities and their conduct. Emotional or psychological engagement refers

to students' feelings about school, their teachers, and their peers, and considers students' sense of belonging and the relationships they build at school (Li et al. 2011; Lee 2012).

- Fredricks et al. (2004) described a third element of cognitive engagement, which refers to students' investment in school and their willingness to try to understand complex activities and work through difficult tasks.
- Other researchers such as Appleton et al. (2008) and Yonezawa et al. (2009) noted a fourth element of academic engagement that focuses on youth's academic behavior and performance, such as credits earned toward graduation, time invested on academic tasks, and homework completion.

The SRAENE grant recipient instrument toolkit defines youth engagement using the school engagement framework that connects to three elements found in the youth development literature: behavior (participation), emotional or psychological (interest), and cognitive (investment). A fourth element, the academic element, was omitted because it pertains to youth completing activities and not their active involvement with the program, which is the focus of the toolkit. The final definition of youth engagement is in Chapter I.

B. Facilitator quality

Evaluations and implementation studies of youth risk prevention programs or youth development programs do not directly define facilitator quality, but three articles considered several elements in understanding the role facilitators play in administering programs. (Exhibit A.2 gives parameters for the literature search.)

- Elder et al. (1994) rated individual health facilitators' performance based on their organization, classroom management, ability to make the class interesting, communication skills, clarity of instruction, and ability to get along with students.
- Jarrett et al. (2009) considered facilitator characteristics in a school-based smoking cessation program: trustworthiness, care for students, confidentiality, being nonjudgmental, preparedness, and not nagging or preaching.
- A review of youth development literature by Roth and Brooks-Gunn (2016) demonstrated the importance of program experiences a facilitator could directly contribute to and thereby influence youth engagement. A facilitator might affect program experiences such as:
 - Consistent, positive, and supportive facilitator–youth relationships
 - A welcoming and supportive atmosphere

Exhibit A.2. Parameters used in the facilitator quality literature search

Search parameter	Terms and resources
Keywords	facilitat*[ti] OR teach*[ti] AND qual*[ti] AND youth*[ti] OR student*[ti] OR adolescen*[ti] OR 12-17 years[ti] OR high[ti] OR middle[ti] AND classroom*[ti] OR school*[ti] OR educat*[ti] OR develop*[ti] OR program*[ti] OR (((prevention[ti] OR avoid*[ti] OR educat*[ti] OR sexual risk[ti] OR pregnan*[ti] OR drug[ti] OR alcohol[ti] OR tobacco[ti] OR bully
Disciplines and fields targeted	Youth development programs OR education
Relevant databases	Academic Search Premier, PubMed, ERIC, Education Research Complete, CINAHL with Full Text, EBSCO Database

Note: [ti] = terms in the title.

Similar to the exploration of youth engagement, education research was consulted for definitions of teacher quality that could serve as comparable key terms in the search. This search revealed that journal articles tend to focus on teachers' educational background, experiences, and credentials, rather than on how they work with students to promote school engagement. As a result, the search was expanded to include literature related to "instructional quality" to see whether definitions of that topic could fill the gaps. In educational settings, instructional quality encompasses multiple elements, some of which are not within the scope of this task, such as the fidelity of curriculum delivery. However, the search yielded seven articles that described instructional quality elements that can apply to SRAE programs.

- Blömeke et al. (2016) and Nilsen et al. (2016) noted broad agreement that instructional quality includes classroom management, which Kane and Cantrell (2012) and Decristan et al. (2015) described as management of student behavior and classroom procedures.
- Other elements Nilsen et al. (2016) mentioned were cognitive activation (or cognitively challenging learning opportunities [Baumert et al. 2010]), clarity of instruction (teacher actions that support the understanding of information [Ferguson 2010]), and a supportive climate (such as evidence of positive teacher–student relationships [Decristan et al. 2015]).

Based on the youth development and educational research, the SRAENE grant recipient instrument toolkit defines facilitator quality using three elements: classroom management, supportive climate, and content delivery. Cognitive activation was intentionally excluded from the definition, for the same reasons the academic element for youth engagement was excluded. Cognitive activation seems more tied to the content and fidelity of the curricula than to the strategies employed to actively involve youth in the program. The final definition of facilitator quality is in Chapter I.

References

Youth engagement

- Appleton, James, Sandra Christenson, and Michael Furlong. "Student Engagement with School: Critical Conceptual and Methodological Issues of the Construct." *Psychology in the Schools*, vol. 45, no. 5, 2008, pp. 369–386. doi:10.1002/pits.20303.
- Fredricks, J.A., P.C. Blumenfeld, and A.H. Paris. "School Engagement: Potential of the Concept, State of the Evidence." *Review of Educational Research*, vol. 74, no. 1, 2004, pp. 59–109. <https://doi.org/10.3102/00346543074001059>.
- Lee, J. "The Effects of the Teacher-Student Relationship and Academic Press on Student Engagement and Academic Performance." *International Journal of Educational Research*, vol. 53, 2012, pp. 330–340. <https://doi.org/10.1016/j.ijer.2012.04.006>.
- Li, Y., A. Doyle Lynch, C. Kalvin, J. Liu, and R.M. Lerner. "Peer Relationships as a Context for the Development of School Engagement During Early Adolescence." *International Journal of Behavioral Development*, vol. 35, no. 4, 2011, pp. 329–342. <https://doi.org/10.1177/0165025411402578>.
- Miller-Day, M., M.L. Hecht, J.L. Krieger, J. Pettigrew, Y. Shin, and J. Graham. "Teacher Narratives and Student Engagement: Testing Narrative Engagement Theory in Drug Prevention Education." *Journal of Language and Social Psychology*, vol. 34, no. 6, 2015, pp. 604–620. <https://doi.org/10.1177/0261927X15586429>.

University of Minnesota. "Youth Engagement Survey." Children, Youth, and Families At-Risk, 2019.

<https://cyfar.org/node/926>.

Yonezawa, S., M. Jones, and F. Joselowsky. "Youth Engagement in High Schools: Developing a Multidimensional, Critical Approach to Improving Engagement for All Students." *Journal of Educational Change*, vol. 10, 2009, pp. 191–209. <https://doi.org/10.1007/s10833-009-9106-1>.

Youth.gov. "Youth Involvement and Engagement Assessment Tool." n.d. <https://youth.gov/youth-topics/positive-youth-development/how-do-you-assess-youth-involvement-and-engagement>. Accessed December 5, 2022.

Facilitator quality

Baumert, J., M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan, U. Klusmann, et al. "Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress." *American Educational Research Journal*, vol. 47, no. 1, 2010, pp. 133–180. <https://doi.org/10.3102/0002831209345157>.

Blömeke, S., R.V. Olsen, and U. Suhl. "Relation of Student Achievement to the Quality of Their Teachers and Instructional Quality." In *Teacher Quality, Instructional Quality and Student Outcomes*, edited by T. Nilsen and J.E. Gustafsson. Switzerland: Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-41252-8_2.

Decristan, J., E. Klieme, M. Kunter, J. Hochweber, G. Büttner, B. Fauth, A.L. Hondrich, et al. "Embedded Formative Assessment and Classroom Process Quality: How Do They Interact in Promoting Science Understanding?" *American Educational Research Journal*, vol. 52, no. 6, 2015, pp. 1133–1159. <https://doi.org/10.3102/0002831215596412>.

Elder, J.P., S.I. Woodruff, J.F. Sallis, C. de Moor, C. Edwards, and M.B. Wildey. "Effects of Health Facilitator Performance and Attendance at Training Sessions on the Acquisition of Tobacco Refusal Skills Among Multi-Ethnic, High-Risk Adolescents." *Health Education Research*, vol. 9, no. 2, 1994, pp. 225–233. <https://doi.org/10.1093/her/9.2.225>.

Ferguson, R. "Student Perceptions of Teaching Effectiveness." National Center for Teacher Effectiveness and the Achievement Gap Initiative, Harvard University, 2010. <https://www.yumpu.com/en/document/read/22107301/student-perceptions-of-teaching-effectiveness-harvard-graduate->

Jarrett, T., K. Horn, and J. Zhang. "Teen Perceptions of Facilitator Characteristics in a School-Based Smoking Cessation Program." *Journal of School Health*, vol. 79, no. 7, 2009, pp. 297–303. <https://doi.org/10.1111/j.1746-1561.2009.00413.x>.

Kane, T., and S. Cantrell. "Gathering Feedback for Teaching. Combining High-Quality Observations with Student Surveys and Achievement Gains." MET Project Research Paper. Seattle, WA: Bill & Melinda Gates Foundation, 2012. <http://files.eric.ed.gov/fulltext/ED540960.pdf>.

Nilsen, T., J.E. Gustafsson, and S. Blömeke. "Conceptual Framework and Methodology of This Report." In *Teacher Quality, Instructional Quality and Student Outcomes*, edited by T. Nilsen and J.E. Gustafsson. Switzerland: Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-41252-8_1.

Roth, J.L., and J. Brooks-Gunn. "Evaluating Youth Development Programs: Progress and Promise." *Applied Developmental Science*, vol. 20, no. 3, 2016, pp.188–202. <https://doi.org/10.1080/10888691.2015.1113879>.

Appendix B. Process for identifying and reviewing instruments

A. Searching for and screening instruments

The key terms in Exhibit B.1 were used to search for instruments of youth engagement and facilitator quality that promotes youth engagement.

Exhibit B.1. Key search terms

Search category	Terms
Topic	Facilitator quality OR teacher quality or educator quality OR instructor quality OR classroom quality OR program quality OR quality OR delivery OR preparedness OR organization OR management OR support OR relationship OR content delivery OR program delivery OR interesting OR clear OR satisfaction OR Engagement OR involvement OR attendance OR participation OR interest OR perception OR perceive OR understand OR invest OR attitude OR contribute OR relationship OR satisfaction OR experiences
Approach	Study OR evaluation OR measure OR survey OR observation OR assess OR tool OR scale
Youth programming content	Prevention OR sexual risk avoidance OR pregnancy OR drug and alcohol use OR bullying OR tobacco OR reproductive health OR sexual health OR health OR literacy OR positive development
Programming type	Program OR curricula OR initiative OR class
Population	Youth OR middle school OR high school OR adolescents OR 12–17 year-olds OR teen OR adolescent

The screening criteria described in Chapter I were used to identify instruments for the toolkit. Five different searches were conducted and found 16 potential instruments for the toolkit:

1. A library search of key terms (Exhibit B.1) to identify published instruments and studies of youth risk prevention or youth development programs that were published in the last 25 years. This search produced more than 140 journal articles related to youth engagement and 200 journal articles related to facilitator quality. The title and abstract for each journal article were reviewed to determine if it included information about instruments to measure youth engagement or facilitator quality. Studies conducted outside the United States were immediately removed from consideration. After the initial title and abstract review, 39 journal articles that mentioned an instrument and were within the scope of the toolkit were selected for a full-text review. Four instruments were identified for further consideration after this search.
2. A search of key terms within .gov websites to identify studies of youth risk prevention or youth development programs. This search found instruments used by federal grant recipients across different government agencies, including the Administration for Children and Families, and a Youth.gov assessment tool. Four instruments were identified for further consideration from this search.
3. A search of key terms in the Family and Youth Services Bureau's Adolescent Pregnancy Prevention Exchange website. This search did not reveal any instruments that were not already found through the .gov website search.

4. A scan of websites of publishers of adolescent pregnancy prevention curricula to collect information about studies or instruments they may have developed to work alongside their curricula. Over 50 publisher websites were scanned. A majority of them shared research study reports. However, most reports described instruments not within the scope of the toolkit or did not provide enough information about an instrument's content or use to include it. Five instruments were identified for further consideration from this search.
5. A scan of websites of youth-serving organizations to collect information about studies or instruments they may have developed to work alongside their curricula. To identify websites of youth-serving organizations a Google search was conducted on the phrase "youth serving organization." Over 20 websites of youth servicing organizations were scanned. Large youth servicing organizations like YMCA study the impact of their programs, but information about the instruments they use is generally not published on their websites. Three new instruments were identified from this search.

B. Reviewing instruments and making final decisions on what to include

For each of the 16 instruments, a Google search was conducted to collect more information about them and confirm they met the selection criteria.

1. **Instruments needed to stand alone, and not be small sections of a larger instrument.** Instruments that contained few questions related to youth engagement or facilitator quality within a larger instrument were excluded. Four instruments were excluded for this reason.
2. **Instruments needed to contain the information needed to complete the profiles.** Instruments lacking detailed information on their development, use, or interpretation were excluded. Three instruments were excluded for this reason.

Following this approach, 9 of a possible 16 instruments were selected for the toolkit.

Appendix C. Glossary of terms

Construct. An idea, underlying theme, topic, or trait to be assessed (for example, behavioral engagement or building positive relationships). The construct is a concept or characteristic of an individual that an instrument is supposed to measure.

Concurrent validity. A type of validity evidence that is collected to determine if an instrument measures what it is supposed to measure. Concurrent validity involves looking at the association (usually measured as a correlation) between a score on a given instrument and performance on another measure obtained at approximately the same time. The other measure may assess the same construct or a similar one (known as convergent validity). If the instrument has the expected positive association with the other measures, that suggests they are measuring the same thing. Another way to examine concurrent validity is to determine whether the construct has a weak or negative association with a measure that assesses a different construct (known as divergent or discriminant validity). Concurrent validity of an instrument can also include examining the association between the instrument and an outcome or characteristic assessed at approximately the same time to see if the instrument predicts the outcome.

Construct validity. A type of validity that is a determination of how well an instrument measures what it is supposed to measure. There are different ways to see if an instrument has construct validity, to include factor analysis of the items and concurrent validity (both defined in this glossary). We can also assess construct validity by looking for evidence of an association between the instrument and an independent measure of something related to the construct, such as youth report of lower frequency of risk behaviors or facilitators demonstrating high observed quality (known as criterion-related validity).

Content validity. A type of validity used to examine evidence that items in an instrument measure the behaviors needed to adequately measure a construct. No statistics are associated with content validity. Instead, it is based on the professional judgment of research and practice experts who review the items to verify that the instrument adequately represents the behaviors or attitudes that are important to the construct, and that the items provide variety and a range of difficulty. Reviews of research evidence used to develop the items are also cited as evidence of content validity.

Correlation. Degree to which two sets of scores or other data vary together, ranging from -1.0 (a perfect negative relationship) to 1.0 (a perfect positive relationship), with 0 indicating no association. Many researchers and instrument developers consider correlations below 0.4 as small, correlations at 0.4 to below 0.6 as medium, and correlations of 0.6 or above as large (Evans 1996). Different types of correlations include Pearson's correlation coefficients, rank-order correlations, or Spearman's rho correlations.

Cronbach's alpha coefficient. An estimate of internal consistency reliability that assesses how well groups of items in an instrument "hang together" or measure a particular trait or characteristic because of common factors among them. The greater the covariance among items, the higher the reliability is (and thus the higher the value of Cronbach's alpha coefficient). Values of the alpha can range from -1.0 to 1.0, with higher values indicating stronger internal consistency. Negative values suggest that an item is not contributing to reliability or needs to be reverse-coded.

Factor analysis. Statistical analysis that examines the pattern of relationships between items in related groups, done to measure underlying latent constructs (that is, unobservable abstract concepts), using correlations or a covariance matrix. Factor analysis may be exploratory (looking at how items group together in the data) or

confirmatory (examining whether the relationships between items are consistent with a predetermined hypothesized factor structure).

For confirmatory factor analysis, model fit of the factor structure can be assessed with various fit statistics. To summarize model fit in a profile, we focus on the following indicators: (a) Comparative Fit Index (CFI; 0.90 or above, Brown 2015), (b) Tucker Lewis Index (TLI; also referred to as nonnormed fit index or NNFI; 0.90 or above, Brown 2015), (c) Root Mean Square Error of Approximation (RMSEA; 0.05 or above for good fit, MacCallum et al. 1996), and (d) Standardized Root Mean Square Residual (SRMR; less than 0.08, Hu and Bentler 1999).

Internal consistency reliability. A measure of the reliability of a score, derived from the relationship between items of a single measure and the extent to which they measure the same construct. Internal consistency reliability is presented as the correlation between groups of items or between all items. For example, split-half reliability refers to the correlation between the odd- and even-numbered items in an instrument. Another measure of internal consistency reliability is based on the correlations between all individual items, such as Cronbach's alpha.

Inter-rater reliability. A way to assess whether different raters or observers using the same instrument would interpret and record the information in the same way. It can include agreement on scoring of items, administrative procedures, or observation of a given behavior. It is usually reported as either the correlation between the scores or ratings obtained by two observers or the percentage of items on which the two agree. Developers may also examine an intraclass correlation or Kappa. In research, inter-rater reliability is often a certification criterion for observers to meet at the conclusion of training and during in-field data collection.

Intraclass correlation (ICC). An estimate of how much of the variance in a measure can be attributed to different factors. As a measure of inter-rater reliability, it provides an indicator of how well different observers or raters agree on their scores or ratings when observing the same thing. The ICC can be affected by the sample of individuals surveyed. For example, if the sample's scores do not show much variation, the ICC may be lower than in samples whose scores have more variation. The ICC is the ratio of the variance due to or explained by the independent variable (trait) divided by that explained variance plus the residual variance due to rater differences and measurement error. Researchers consider ICCs of below 0.5 as poor inter-rater reliability (Koo and Li 2016). ICCs of at least 0.50 can indicate moderate agreement among the raters.

Item. A statement, question, or task in an instrument.

Pilot study. As part of the instrument development process, a pilot study is early-stage work to help draft the instrument and examine how items are working by administering it to a small sample. Study data collection may include developers asking participants questions to see how they interpreted the items.

Predictive validity. A type of validity that is used to determine if an instrument measures what it is supposed to measure by predicting something in the future or comparing it to a "gold standard." Specifically, it is an indicator of a type of criterion-related validity that demonstrates how accurately scores from an instrument can predict scores on another measure or criteria gathered in the future. If the correlation of an instrument with another measure obtained at a later time is strong enough (defined here in Correlation), evidence of predictive validity is established. If, for example, a measure of youth engagement in the fall is highly correlated with youth attitudes toward reduced risk behaviors in the spring, the instrument could be said to have evidence of predictive validity. In some cases, researchers use other activities or events as the criterion, rather than another measure. For example, researchers might show a positive correlation between a measure of facilitator quality and participant attainment of a program training certificate as evidence of predictive validity.

Reliability. The extent to which scores obtained from a measure or group of measures are consistent over one or more possible sources of error, including time, raters, items, environment, and sample groups of a population. Indicators of reliability assess how dependable an instrument is for the purpose it is being used for. Reliable instruments are stable over time and include items that measure the same thing in different ways. Statistical measures of reliability are typically reported as coefficients, which range from 0 to 1.0, with a greater value reflecting greater reliability. Many researchers and developers require that reliability measures have internal consistency or correlation values of 0.7 or higher with the gold standard measure of the construct. Typical indicators of reliability include internal consistency, inter-rater agreement (with intraclass correlation values acceptable at 0.5), and test-retest. An unreliable instrument cannot be valid.

Sample. A selection of a specified number of individuals from a larger set of people who are called the population.

Self-administered. Describes an instrument a participant completes without the assistance of an interviewer.

Subscale. A set of items within a larger instrument that assesses a particular aspect of the trait being measured. Subscales may be specified based on theoretical grounds (grouping items based on their content) or empirical evidence (factor analysis of items in a longer scale may reveal meaningful subscales). Subscales should be related to the whole instrument as an overall scale.

Test-retest reliability. The stability of an instrument's results over time. Evidence of test-retest reliability involves testing the same group of individuals at least twice with a relatively short interval between administrations—usually no more than a few days or weeks apart. The reliability coefficient is then obtained by correlating both sets of scores. The higher the test-retest reliability, the more stable the instrument is considered to be. Longer periods between administrations of the same measure will reduce the reliability, partly because the individual's situation (for example, skill and experiences) can be expected to change.

Validation study. As part of the instrument development process, that is a later stage of work that takes place once an instrument is developed. A validation study tests the instrument (typically with a larger sample) and examines the evidence of reliability and validity, usually compared with an established measure and/or outcome. Validity evidence should be based on collecting data from samples of the population the instrument will be used with and for the use that the instrument is intended for. Over time, additional validity evidence is collected.

Validity. The degree to which an instrument accurately measures what it is designed to measure. Validity is often assessed by comparing an instrument to other instruments established to measure the same or similar behavior or traits. Types of validity include content, construct, concurrent, and predictive. An instrument cannot be valid if it is not reliable.

Sources

The definitions in this glossary were adapted from Malone et al. (2010; 2021) with the following exceptions: acceptable levels of intraclass correlations for inter-rater reliability and definitions for pilot study, self-administered, and validation study.

Brown, T.A. *Confirmatory Factor Analysis for Applied Research*. Guilford Publications, 2015.

Hu, L., and P.M. Bentler. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling*, vol. 6, 1999, pp. 1–55.

Koo, T.K., and M.Y. Li. "A guideline of selecting and reporting intraclass correlation coefficients for reliability research." *Journal of Chiropractic Medicine*, vol. 15, no. 2, 2016, pp.155–163.

MacCallum, R.C., M.W. Browne, and H.M. Sugawara. "Power Analysis and Determination of Sample Size for Covariance Structure Modeling." *Psychological Methods*, vol. 1, 1996, pp. 130–149.

Malone, L., S. Albanese, C. Jones, Y. Xue, G. Kirby, and A. Douglass. "Compendium of Existing Measures for Understanding Leadership in Early Care and Education." OPRE Report #2021-220. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2021.

Malone, L., C. Cabili, J. Henderson, A. Mraz Esposito, K. Coolahan, J. Henke, S. Asheer, M. O'Toole, S. Atkins-Burnett, and K. Boller. "Compendium of Student, Teacher, and Classroom Measures Used in NCEE Evaluations of Educational Interventions. Volume II: Technical Details, Measure Profiles, and Glossary (Appendices A-G)." NCEE 2010-4013. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, April 2010.

This page has been left blank for double-sided copying.

Mathematica Inc.

Our employee-owners work nationwide and around the world.

Find us at mathematica.org and edi-global.com.

Mathematica, Progress Together, and the "spotlight M" logo are registered trademarks of Mathematica Inc.

