



ELSEVIER

 JOURNAL OF
 ADOLESCENT
 HEALTH

www.jahonline.org

Commentary

From Mission to Measures: Performance Measure Development for a Teen Pregnancy Prevention Program

 Amy Feldman Farb, Ph.D.^{a,*}, Barri Burrus, Ph.D.^b, Ina F. Wallace, Ph.D.^b, Ellen K. Wilson, Ph.D.^b, and John E. Peele^b
^aOffice of Adolescent Health, Office of the Assistant Secretary for Health, Department of Health and Human Services, Rockville, Maryland

^bRTI International, Research Triangle Park, North Carolina

Keywords: Performance measures; Teen pregnancy prevention; Evidence-based programs; Federal grants

 A B S T R A C T

The Office of Adolescent Health (OAH) sought to create a comprehensive set of performance measures to capture the performance of the Teen Pregnancy Prevention (TPP) program. This performance measurement system needed to provide measures that could be used internally (by both OAH and the TPP grantees) for management and program improvement as well as externally to communicate the program's progress to other interested stakeholders and Congress. This article describes the selected measures and outlines the considerations behind the TPP measurement development process. Issues faced, challenges encountered, and lessons learned have broad applicability for other federal agencies and, specifically, for TPP programs interested in assessing their own performance and progress.

Published by Elsevier Inc. on behalf of Society for Adolescent Health and Medicine. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

The Office of Adolescent Health (OAH) sought to create comprehensive performance measures to appropriately capture the performance of the Teen Pregnancy Prevention (TPP) program, measured through key outcomes, as one of the foundational cornerstones for measuring program performance over time. OAH quickly had to develop a well-thought-out, data-driven performance measurement system based on established priorities that would track progress in implementing this new program. This performance measurement system needed to provide measures that could be used internally (by both OAH and the TPP grantees) for management and program improvement as well as externally to convey program progress to other interested stakeholders and Congress. OAH, with assistance from its

contractor, RTI International, developed performance measures for the TPP program.

The purpose of this article is to share the considerations behind the TPP measurement development process so that it may benefit others who wish to develop a comprehensive system for measuring the performance of a large grant program that addresses adolescent health. Issues faced, challenges encountered, and lessons learned have broad applicability for other federal agencies and, specifically, for TPP programs interested in assessing their own performance and progress.

Developing the TPP program performance measures was not a simple task due to the nature and complexities of the program. The measures needed to be both reliable and valid across 94 grantees and approximately 40 program models. The measures also needed to be valued by experts in the field, and the requirements for collecting the measures needed to be reasonable with regards to time and expense.

Given the newness of OAH and the TPP program, it was important that the performance measures be valuable both for establishing a solid track record of program performance and for identifying areas for program improvement [1]. More broadly, OAH sought a performance measurement system that would facilitate communication, motivation, and resource allocation

Conflicts of Interest: The authors declare no conflicts of interest.

Disclaimer: Publication of this article was supported by the Office of Adolescent Health, U.S. Department of Health and Human Services. The opinions or views expressed in this paper are those of the authors and do not necessarily represent the official position of the Office of Adolescent Health, U.S. Department of Health and Human Services.

* Address correspondence to: Amy Feldman Farb, Ph.D., HHS-Office of Adolescent Health, 1101 Wootton Parkway, Suite 700, Rockville, MD 20852.

E-mail address: Amy.Farb@hhs.gov (A.F. Farb).

[1–3]. As a communication tool, the performance measures needed to reflect indicators of success and be timely enough to make course corrections. As a motivational tool, the performance measures needed to identify and set achievable and realistic trajectories that could provide a sense of accomplishment to OAH and its grantees. In addition, these measures needed to be useful for grantees in describing the implementation of their programs. The focus on trajectories rather than targets acknowledges the importance of incremental advancement toward goals and emphasizes that performance is a process of continual improvement. As a resource allocation tool, the performance measures needed to give managers objective data to identify effective programs and to make informed decisions regarding provision of additional support. The performance measures also needed to provide agency managers with data to demonstrate program value and to secure funding for continued success.

OAH sought to develop a comprehensive performance measurement system that could demonstrate performance within these criteria and inform OAH's work for years to come. A process that was deliberate and well-thought-out, yet fast-paced, was used for the OAH TPP program performance measure development. We conducted a literature review to identify considerations for developing a performance measurement system and to explore existing tested measures. We held meetings with a panel of experts in the TPP and performance measurement fields in which the information obtained through this literature review was used to help identify and operationalize measures, determine appropriate timing of data collection, and brainstorm presentation of measurement data. Finally, we solicited input from TPP grantees and federal program staff who tested the measures and the reporting system, and provided feedback.

Key Challenges in Developing Performance Measures

We faced numerous challenges in developing a comprehensive system for measuring the performance of a complex national TPP program. Arguably the most important challenges were determining appropriate outcomes, accommodating the many and varied TPP program models, and addressing the multiple purposes of the TPP program. It is important to keep in mind that the TPP program is only one factor in a vast array of other social and economic variables that influence adolescents. The grantees' TPP programs are performing against a backdrop in which the adolescent participants are growing older over the course of the program and are, therefore, statistically more likely to engage in sexual behaviors [4]. Secular trends may drive rates for sexual behaviors up or down, independent of any program activities. With this level of "noise" in the program environment, measuring certain outcomes without appropriate comparison groups could result in misleading information and inappropriate conclusions on overall program performance.

Yet another challenge stems from the overall TPP program purpose and focus, meaning that priority measures differ depending on grant type. The TPP program was designed to fund the replication of evidence-based programs for the majority of grantees and the implementation and testing of new and innovative interventions for another group of grantees. Fidelity to the program model was a key consideration for replication grantees, while documentation of process was key for the innovative grantees.

The type, size, and duration of the programs being implemented by the 94 funded organizations (75 organizations

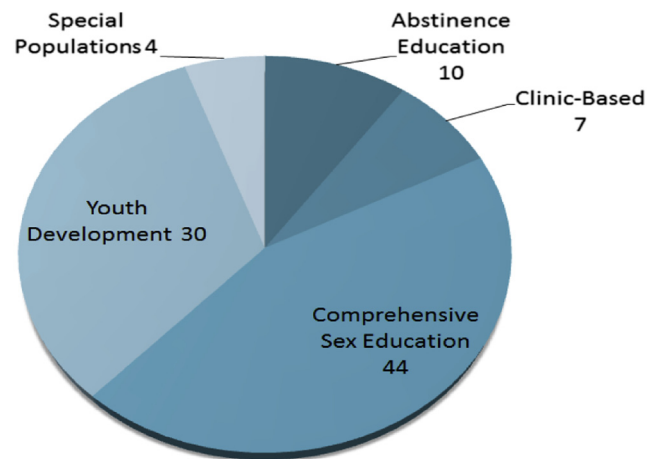


Figure 1. Types of replication programs.

conducting evidence-based replications, referred to as Tier 1, and 19 organizations conducting demonstrations of new and untested programs, called Tier 2) vary markedly at the grantee level. The replication grant program funds many different models, including comprehensive sex education, positive youth development, and abstinence programs. Figure 1 depicts the 95 evidence-based programs being replicated by 75 grantee organizations. One of the funded programs (Sexual Health and Adolescent Risk Prevention [SHARP], formerly known as HIV Risk Reduction among Detained Adolescents) consists of 8 hours of curriculum instruction whereas another funded program (Carrera Adolescent Pregnancy Prevention Program) is delivered every day after school for 4 years, resulting in very different numbers of lessons and activities. Some of the programs take place in schools, serving several classrooms at a time, whereas other programs are provided in clinics in a one-on-one format.

Given these and other variations, development of performance measures at the grantee-level could result in many different types of performance measures. To comprehensively describe the TPP program, it was necessary to settle on common performance measures and to aggregate across grantees. Although not every measure turned out to be compatible across every grantee and program model, the selected measures are meaningful to, and can be aggregated across, the great majority of them.

Members of the expert panel worked to take each of these and other factors into consideration during the measurement development process. In addition, they offered a list of considerations for meaningful selection of performance measures (Table 1). These factors may also be pertinent guidelines to consider for performance measure development in other adolescent health approaches.

The Current TPP Performance Measures

OAH's final set of performance measures for the TPP program include a mix of both grantee-level measures and participant-level measures reflecting the priorities for the replication grants (fidelity) and innovation grants (documentation) as well as the TPP program as a whole (dosage, dissemination). Grantees report on these measures every 6 months. The reporting schedule was designed to parallel the submission of semiannual

Table 1
Factors considered for inclusion as a performance measure

Factor for consideration	Question addressed
Burden	What level of effort would be required for grantees and/or participants to collect the data needed for the measure?
Link to program mission	Is the measure aligned with the program mission? Are the measures being gathered important drivers that help to ensure the program is performing well and ultimately successful in achieving its goals?
Timeliness	Is it reasonable to expect the outcomes to be achieved at or before the point at which the data for the measures are being gathered?
Vulnerability to threats for interpretation	To what extent are the measures likely to be affected by factors other than the intervention activities? For example, what influence would other factors (e.g., social, economic, other programs) have on the measures?
Interpretability	Are the measures clear and understandable to stakeholders, (e.g., members of Congress, program implementers, community members)?
Comparability	Are measures comparable across grantees? If not, can variables that make implementation more difficult for some programs (e.g., implemented with vulnerable populations, sparsely populated areas) also be captured so that, if needed, these factors can be “weighted” in some way as influencing performance?
Usefulness	Are the measures providing useful information for the different audiences? How will different audiences use the data that are collected? see “Metzenbaum, 2009” [6]

progress reports. The following provides more information on these two areas of measurement.

Grantee-level Measures

The performance measures chosen for TPP program implementation were designed to tap a broad range of constructs divided into those concerning program structure and program delivery. The measures are illustrated in Table 2.

Table 2
Grantee-level performance measures

Program structure	Program delivery
<ul style="list-style-type: none"> • Partners: number involved and retained • Training: number of new facilitators trained and number receiving follow-up training • Dissemination <ul style="list-style-type: none"> – Published and submitted manuscripts and presentations – Completed development of pieces of program necessary to package it for replication (Tier2) 	<ul style="list-style-type: none"> • Reach: number served by demographic characteristics • Dosage <ul style="list-style-type: none"> – Median and mean percentage of total intended program services received – Percentage of participants receiving $\geq 75\%$ of program services • Fidelity <ul style="list-style-type: none"> – Adherence to program-specific activities, based on facilitator self-assessment and observation – Adherence to program-specific number of sessions – Quality of implementation, based on observation – System in place to ensure fidelity

Program structure

Program structure measures are designed to provide information about how the program operates to meet its goals. There are three constructs, each with multiple measures: partners, training, and dissemination.

Partners

Partners are entities (public, nonprofit, private business, etc.) that (1) participate in planning for the TPP program; (2) participate in implementing activities related to the TPP program; or (3) help to cosponsor events related to the TPP program. Partners may contribute money, staff time, physical space, or other in-kind donations to the operation of the TPP program. Data on both formal and informal partners are collected. Formal partners are organizations (e.g., schools) with whom the grantee has a memorandum of understanding, contract, or other formal written agreement in place to provide services or other contributions relevant to the TPP program. Informal partners are organizations with whom the grantee does not have a formal written agreement in place. Grantees report the number of formal and informal partners that they are currently working with, the number who are new in any reporting period, and the number lost during the reporting period.

Training

Training measures both the number of facilitators who are newly trained on the intervention and the number who receive follow-up or supplemental instruction, including trainings that are additional to the formal intervention training that will improve the facilitators' delivery of the program, such as a class focused on sexually transmitted infections or anatomy.

Dissemination

Dissemination measures the number of submitted and published manuscripts and the number of presentations by national, regional, and state location. Innovation grantees also report the development of pieces of the program necessary for replication such as the logic model, core components, fidelity monitoring tool, and curriculum manual.

Program delivery

Program delivery measures are designed to assess the extent to which the program was implemented as designed. The key aspects of program delivery are reach, dosage, and fidelity.

Reach

Reach is defined as the number of participants (both youth and others) who are enrolled in the program and receive at least one session or component of the program. The number includes participants who receive the program even if they are not enrolled in the evaluation (i.e., because they do not have consent or because a selection procedure was used for the evaluation). The number served may be different than the number with behavioral data (see Participant Measures section). In the TPP program, reach is tabulated by demographic characteristics,

including sex, age, race/ethnicity, grade, and language spoken at home.

Dosage

Dosage refers to the extent to which participants received the intended program services and is measured by attendance. Grantees report each participant's attendance for each core component. Several performance measures are derived: the median and mean percentage of total intended program services received and the percentage of participants receiving at least 75% of intended program services.

Fidelity

Fidelity addresses how well the implementation adhered to the program's model. Grantees report several measures of fidelity. Because the programmatic core components vary widely across program models, grantees only report on the adherence to the curriculum, an indicator that is common to nearly all program models. Specifically, grantees report on the extent to which they delivered the specific activities that should occur during each session of the program, measured both by facilitator self-report and by observation. OAH requires that 10% of all sessions be observed to verify fidelity. The quality of the implementation of the program's curriculum is also measured by observation. Finally, grantees report on the system that is in place to ensure fidelity with a fidelity process report form, collecting information on orientation and training for facilitators and observers and ongoing fidelity management.

Participant Measures

Participant-level measures are measures of the extent to which the program has had the desired impact on program participants (i.e., a reduction in the risk of teen pregnancy) (see Table 3). They are collected from all grantees who are conducting a rigorous evaluation.

Two key issues are related to the selection and measurement of the participant-level measures: (1) the time frame in which changes could be expected and (2) the need for a control group. Because reducing teen pregnancy is a long-term outcome and may not be measurable within the time frame of the grant

Table 3
Participant-level measures

Behaviors	Intentions
– Any sex: The percentage of grantees whose intervention group reports less sexual activity than the comparison group	– Intentions to have sex: The percentage of grantees whose intervention group reports a lower intention to have sex than the comparison group
– Condom use: The percentage of grantees whose sexually active intervention group reports more condom use than the comparison group	– Intentions to use condoms: The percentage of grantees whose intervention group reports a higher intention to use condoms than the comparison group
– Contraceptive use: The percentage of grantees whose sexually active intervention group reports more contraceptive use than the comparison group	– Intentions to use contraception: The percentage of grantees whose intervention group reports a higher intention to use a contraceptive method than the comparison group

program, the primary selected measures of program impact are shorter-term outcomes that could be expected to lead to reduced teen pregnancy: reduced sexual activity and increased condom and contraceptive use. Because only a small number of pre-teens and early teens are sexually active, for the youngest program participants, any impacts on these shorter-term, behavioral measures are unlikely to be detectable. Therefore, measures of participants' intentions were included as measures that could be assessed in this population (OAH collects data only from participants in the seventh grade or higher).

The primary challenge to using these measures as performance measures is that rates of sexual activity and pregnancy naturally increase over time as youth age. Therefore, reporting program participants' changes over time for these outcomes would suggest that the program is having the opposite effect of its intention. To use these measures, a comparison group is necessary to be able to assess whether the rates of increase are lower for program participants than they would have been without the program. Fortunately, because a number of grantees were conducting rigorous evaluations, intervention and control group data could be collected for a substantial part of the program.

However, many grantees were not conducting rigorous evaluations, and so were not able to report control group data. OAH wanted to be able to report some participant-level data for all of the grantees, so a set of "perceived impact" measures were piloted. Program participants were surveyed at the end of the program and asked whether they thought the program had affected the likelihood that they would abstain from sex in the next year, the likelihood that they would have sex, and, if they were to have sex, the likelihood that they would use condoms or other contraceptives. The pilot revealed that these measures were problematic. The questions were so hypothetical that they were very difficult for some students to understand—especially younger students and those for whom sexual activity felt like a very distant future. There was also an indication that students were either not understanding the questions or not making an effort to respond carefully: nearly one sixth of respondents gave conflicting responses for the impact of the program on the likelihood that they would have sex and the likelihood that they would abstain from sex. Because of these problems, the perceived impact measures were discontinued after a short pilot, and the program now relies only on behavioral outcome data from the rigorous evaluations.

Lessons Learned

This section summarizes lessons learned that may be helpful for others interested in creating comprehensive measures to track the performance of a variety of programs.

Engage stakeholders during development

Engage multiple types of stakeholders in the measurement development process. Stakeholders include performance measurement experts, experts in the field of TPP, funder staff, grantee staff, and the individuals who will ultimately collect and report the data. Experts in the field can help frame issues and identify relevant questions for the field. However, these experts will need to be reminded that performance measures are intended to monitor and improve programs and gauge progress toward goals. It is easy to confuse performance measures with research

outcomes, but these serve different purposes. Including experts in performance measurement helps to expand the thinking beyond the typical “teen pregnancy prevention” measurement strategies that are focused on behavioral outcomes that may be outside the control of the program and not meaningful without the use of a comparison group.

Involving program staff who will actually be using the measures and the reporting system will ensure that the measures are meaningful and feasible. To this end, it is necessary to determine whether there is applicability of the measures across programs. It is also important to ensure that there are no barriers to reporting measures (e.g., measures of sexual behavior) and, if there are, to develop protocols to cover these contingencies.

Including staff from the funding office in initial planning will ensure that the performance measure questions are worded appropriately for the statements the office wants to make about their program. For example, do funders want to know, “How many youth were served each year?” Or is it more important to ask, “How many youth were reached over the life of the program?” The first allows for collection of reach data without concerns of possible duplication of participants from year to year; the second requires collection of information to calculate an unduplicated total count for reach, which is particularly an issue for multiyear programs. After the first round of performance measure data collection, having funder staff review the data reports will provide an opportunity to determine whether the measures, and how they are presented, meet the funder’s needs.

Highlight the value of the performance measure data to those reporting

Individual grantees can also benefit from the information gathered for their own program monitoring and research purposes, but they may need to be shown how the measures can be useful to them. Doing so can potentially deflect some of their resistance to collecting and reporting measures. In the TPP program, performance measures included fidelity and dosage, measures that can be used to both check on their program’s progress and interpret program effects. Having facilitators provide dosage and fidelity on a regular basis to their supervisors allows the supervisory staff the opportunity to discover any implementation issues in a timely fashion. Attendance data can also be used to describe dose-response effects, and fidelity can address whether greater attention to adherence or quality of implementation are associated with better outcomes. Using dosage and fidelity for research purposes could be particularly valuable for grantees who are not conducting a rigorous evaluation.

Make performance measures easy to collect and report

Project directors who manage large grants are often very busy, and meeting the requirement of a performance-measure collection and reporting system can be burdensome. Thus, streamlining the collection and reporting of performance measures to the extent possible will help ease this burden. Providing training ahead of time is one way to proactively address difficulties. The system developers should offer detailed training on an ongoing basis. Although online instructions and user guides are helpful, system users greatly benefit from direct training. Following basic principles of adult learning [5], these trainings need to include opportunities to practice the new skills at or shortly after the training sessions.

Training and individualized technical assistance (TA) with the reporting and data entry should be offered. Grantees often found the reporting requirement frustrating. Having someone readily available to provide tailored answers for specific questions helped to mitigate the frustrations grantees encountered. Different modes of communication (e.g., phone, e-mail, and opportunities for in-person contact) were also helpful. In addition, having a “help desk” where TA requests can be logged and filed allows staff to be sure that requests are handled in a timely fashion; it was important to grantees for their TA requests to be responded to immediately, even if additional time was needed for resolution. Training and TA will need to be provided on an ongoing basis because new staff may not always receive training from their predecessors.

Make reporting as flexible as possible

Grantees may be motivated to provide the data but may have difficulty getting it into a single consistent machine-readable format for data transfer. Developing multiple methods for reporting the data available is one way to address this. In the TPP program, grantees were allowed to enter their data manually within the system and/or upload into Excel spreadsheets that were customized for the required data. Manual data entry worked well for programs serving fewer participants and for those who did not otherwise have a system to track their own data. The Excel spreadsheet option worked better for programs serving large numbers of participants and for those with existing data systems. The latter group exported the data for OAH into an Excel spreadsheet and uploaded it into the reporting system. Attempts to collect aggregate-level data proved unfruitful because the system was designed to link much of the data to unidentified individuals. For many grantees, it would likely be ideal to have a system that can accommodate aggregated participant data, which would permit the planned analyses to be performed.

Ensure high quality data

Mechanisms to check for data accuracy and completeness are essential with such a large number of grantees providing data in the system. Fail-safe mechanisms and consistency checks are needed at the point of system entry to mitigate errors. In addition, checks for completeness need to be run on a grantee-by-grantee basis. Using automated systems that can provide real-time feedback to staff who are entering data can both save time and provide experiential learning.

The TPP Performance Measures development process demonstrates that it is possible to build a meaningful performance measurement system for a complex, multifaceted program composed of grantees with different programmatic goals and objectives. However, care must be taken to ensure that the measures fall within the program’s span of control and represent program performance rather than reflect other “noise” in the implementation environment. In addition to providing accountability at the federal program level, well-thought-out performance measures can provide local program managers with timely guidance on key factors related to program implementation and performance and allow for timely corrective action to be taken as needed. Measuring long-term outcomes related to adolescent sexual behaviors may not be practical, largely due to cost issues. However, proper assessment of key performance indicators, such as those related to reach, fidelity,

and dosage, allow one to expect anticipated outcomes if the theory of change is valid.

Funding Sources

Partial support for this article was provided by the U.S. Department of Health and Human Services, Office of Adolescent Health (contract HHSP23320095651WC).

References

- [1] Behn R. Why measure performance? Different purposes require different measures. *Public Administration Review* 2003;63:586–606.
- [2] Kates J, Marconi K, Mannle TE. Developing a performance management system for a federal public health program: The Ryan White CARE ACT Titles I and II. *Evaluation and Program Planning* 2001;24:145–55.
- [3] U.S. General Accounting Office. Evaluations help measure or explain performance; 2000:1–31.
- [4] Centers for Disease Control and Prevention. Youth Risk Behavior Surveillance—United States, 2011. *Morbidity and Mortality Weekly Report* 2012;61:1–268.
- [5] Queensland Occupational Therapy Fieldwork Collaborative (2005). Adult learning theory and principles. Available at: <http://www.qotfc.edu.au/resource/?page=65375>.
- [6] Metzenbaum SE. Performance Management Recommendations for the New Administration. January 2009. Available at: http://scholarworks.umb.edu/cgi/viewcontent.cgi?article=1010&context=cpm_pubs&sei-redir=1&referer=http%3A%2F%2Fscholar.google.com%2Fscholar_url%3Fhl%3Den%26q%3Dhttp%3A%2F%2Fscholarworks.umb.edu%2Fcgi%2Fviewcontent.cgi%253Farticle%253D1010%2526context%253Dcpm_pubs%26sa%3DX%26scsig%3DAAGBfm0mZveekzxXBivvouFfoF30MUqnsW%26oi%3Dscholar#search=%22http%3A%2F%2Fscholarworks.umb.edu%2Fcgi%2Fviewcontent.cgi%3Farticle%3D1010%26context%3Dcpm_pubs%22. Accessed December 27, 2013.