OPA BRIEF

Understanding How Components Of An Intervention Can Influence Outcomes

Teen pregnancy prevention (TPP) interventions often have many program components, the individual parts or distinct features of the broader intervention. Program components can be the parts of an intervention's structure (for example, classroom lessons, weekly text messaging, and/or service learning), its content (for example, content that focuses on positive youth development, sexual risk avoidance, or condom/contraceptive use), or a combination of the two (for example, some classroom lessons could be categorized as "sexual risk avoidance lessons," others as "positive youth development lessons"). Program components are the individual ingredients that make up the whole intervention being implemented.

These individual program components can play a role in influencing the outcomes of participants, such as how often adolescents engage in risky sexual behavior, or less risky behavior, such as how correctly and consistently they use condoms. Because of this connection, several audiences might want to understand which program components matter most in influencing participant outcomes—that is, the core components of a program. Core components are the essential elements or intervention activities that are necessary to produce desired outcomes for participants (Blase and Fixen 2013). Practitioners might want to distinguish key program components so they can adapt the intervention and enhance its critical components and minimize, or even eliminate, less important components. Researchers might want to identify critical program components so they can design studies that rigorously test the effects of individual components on participant outcomes. Policymakers might choose to fund interventions based on whether they contain promising components.

Do program components matter only in an impact evaluation? Are analyses of program components feasible in studies without a control/comparison group?

Understanding the role that individual program components play in a multicomponent intervention is often a secondary goal of an impact evaluation focused on an intervention's effectiveness. The primary goal is usually to test the effects of the intervention as a whole by comparing youth who were offered the intervention against a sample of youth who were not offered it (sometimes by conducting a randomized experiment). Such studies attempt to answer a confirmatory research question such as "What is the impact of [TPP intervention] on sexual behavior outcomes?"

In this brief , we focus on understanding the role of individual program components, not on the effect of the intervention as a whole. The goal is to answer exploratory questions such as "Which program components matter the most in influencing participant outcomes?"

The approaches and methods described in this brief focus on the experiences of the youth exposed to the intervention (their experiences with program components) and program outcomes. As a result, this brief is applicable to studies that do not have a control/comparison group, and such studies might be particularly well suited to the exploratory approaches described below.

This is the second of two related briefs. The first brief discussed how to define and operationalize the program components of the intervention through careful data collection (Cole and Murphy 2016). This brief expands on the first by presenting research questions that practitioners, policymakers, and researchers might ask about how program components affect outcomes. It also briefly outlines the analytic approaches that can be used to answer these questions. For example, this brief highlights the potential of research questions such as the following:

- What is the effect of receiving a strong enough dose of a particular program component?

- In a multicomponent intervention, which program component plays the biggest role in influencing participant outcomes?

This brief is intended for readers who are interested in systematically examining the effects of different aspects of their program, so the information presented here on methods and approaches is more illustrative than technical. References and footnotes are included for readers who require additional details on implementing any of the described analytic approaches.

The following sections of this brief cover (1) the foundational elements of the program components analysis, which were covered in greater detail in Cole and Murphy (2016); (2) a description of potential types of research questions about how program components influence outcomes; and (3) a short description of analytic approaches for answering those research questions. The brief concludes with a section on reporting and interpretation.
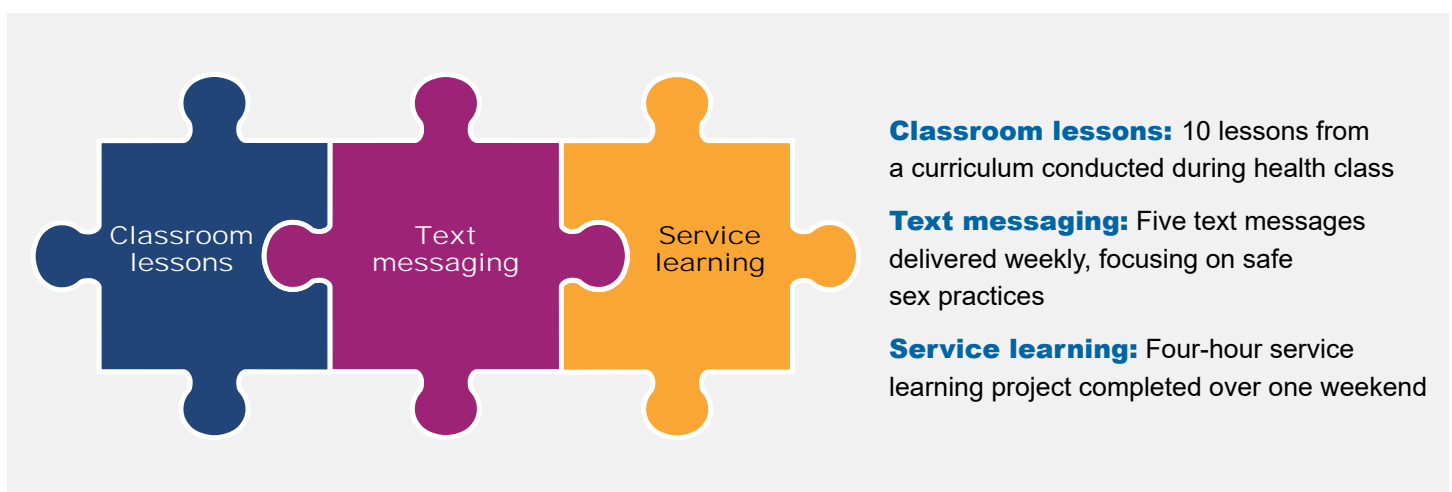
A. The foundation: defining program components, developing a pathway diagram, and collecting data

Three foundational elements are needed to be able to link program components to participant outcomes: (1) a disaggregation and enumeration of an intervention into its program components; (2) a conceptualization of how program components affect participant outcomes (for example, a pathway diagram); and (3) data on the implementation of program components and on participant outcomes. Cole and Murphy (2016) provide details on these three elements. Below, we briefly discuss these concepts to develop a working example to guide readers through the following sections.

1. Disaggregate and enumerate program components

To develop a working example, we focus on structural elements as the program components of a hypothetical intervention. This hypothetical intervention contains three components, presented as an interconnected jigsaw puzzle in Figure 1, where each program component is a puzzle piece: (1) classroom lessons of a sexual health program, (2) informational text messaging, and (3) a service learning project.

Figure 1. Example program components

Classroom lessons: 10 lessons from a curriculum conducted during health class

Text messaging: Five text messages delivered weekly, focusing on safe sex practices

**Service learning:** Four-hour service learning project completed over one weekend

We chose a broad-stroke example of program components, although a more fine-grained approach might be more appropriate in certain contexts. For example, if the intervention were five positive youth development classroom lessons and five sexual risk avoidance education classroom lessons, our "classroom lessons" could be broken into two separate program components. A combination of theory, practical implementation questions, and the type of available data should inform decisions about how nuanced to make the presentation of program components.

2. Link program components to outcomes (pathway diagram)

A pathway diagram shows how each program component of the intervention is expected to influence participant outcomes, including proximal (short-term) and distal (long-term) outcomes. We are interested in the visual representation of the connections between these key elements.

Figure 2 depicts our three hypothetical program components of the intervention as blue rectangles. The proximal outcomes of the intervention are presented as red ovals. In this example, they include (1) participants' attitudes about choosing not to have sex and (2) knowledge about sexually transmitted infections (STIs). Finally, a purple oval shows the distal outcome of the intervention: the frequency of risky sexual behavior (for example, having recent sex without a condom).

Figure 2. Example pathway diagram
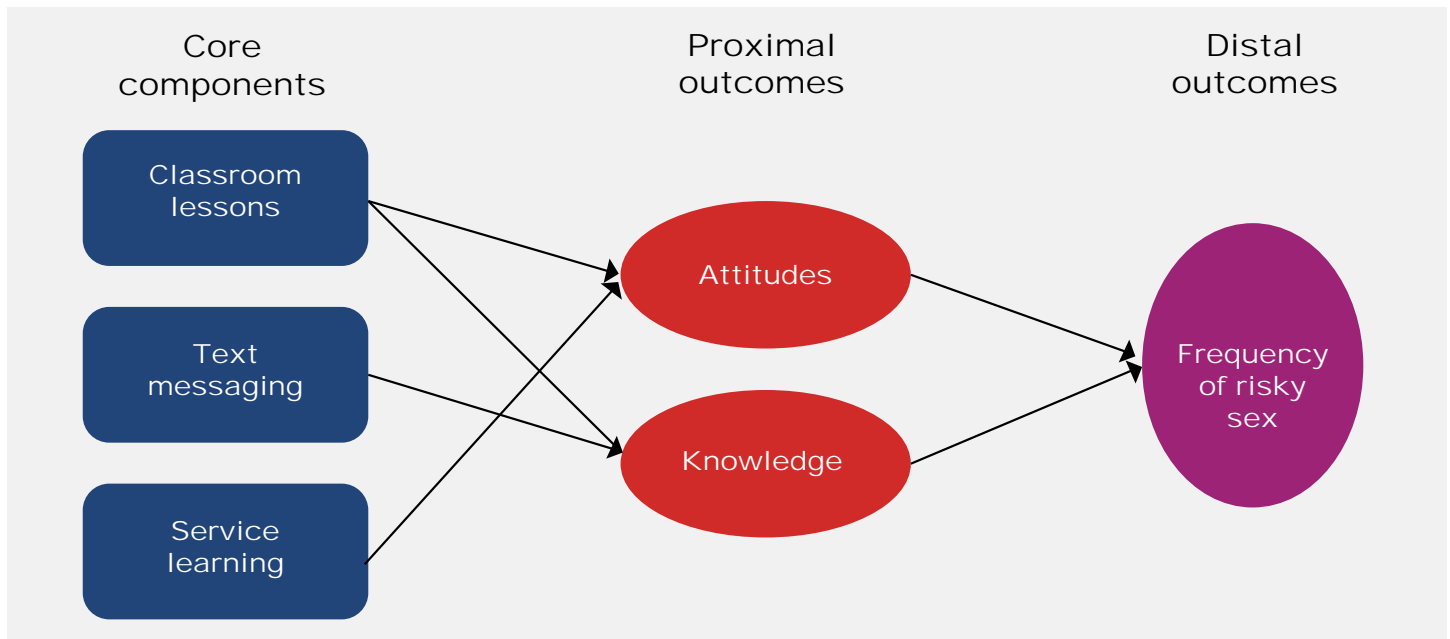
The power of a pathway diagram is in demonstrating how program components influence proximal and distal outcomes. For example, the program component of classroom lessons is hypothesized to influence youth's attitudes about choosing not to have sex and knowledge about STIs (the proximal outcomes), but the text messaging component is only expected to influence knowledge. In this model, the distal outcome, frequency of risky sex, is directly influenced by the two proximal outcomes. The program components indirectly affect this distal outcome, through changes in participant attitudes about choosing not to have sex and knowledge of STIs.

3. Collect data on implementation of program components and proximal and distal outcomes

Data on implementation of program components. Data on different implementation features of program components support richer analyses of program components. As described in Cole and Murphy (2016), this would potentially include the following for each of the program components of interest: information on dosage received, engagement of participants with the program component, quality of implementation, and adherence to the delivery of the intended model (if applicable). It might be infeasible to measure and collect all these potential aspects of program component implementation. In such a situation, only a partial description of the actual experiences that participants have of the program components of an intervention can be incorporated   into the analysis. As a result, the analysis linking program components to outcomes will be somewhat limited, given that potentially important features of implementation will not be measured.

Outcome data. Participant outcomes are likely obtained through surveys administered at program exit or at a longer-term follow-up. Assuming these outcomes are valid and reliable, it is likely they will be sufficient on their own without additional measurements.

Linking implementation and outcome data. The implementation and outcome data must be linked at an individual level to permit exploration of how variation in people's experiences of program components influences their outcomes. It will be important to measure and document how each person exposed to the intervention experiences features of implementation of program components and to then be able to link those implementation data with outcome data to answer the research questions listed in the next section.

B. Illustrative research questions

Articulating a research question (or a series of research questions) about the role of program components can guide a researcher to the most appropriate analytic approach to take. Consider the following broad categories of research questions about how program components influence outcomes:

Estimating the effect of receiving an appropriate dose of the program. One broad category focuses on whether participants received a sufficiently high quality "dose" of a program component that is hypothesized to be necessary to achieve participant outcomes. For example, a program developer or stakeholder might believe the most critical feature of the intervention is the number of classroom lessons a student attends. A *quasi-experimental approach* would provide an opportunity to understand the effect of the classroom lesson program component.[1] To provide an

estimate of the effect of the classroom lessons, the researcher could compare outcomes of youth who regularly attend classroom lessons to outcomes of youth who attend more sporadically.

Identifying the relative importance of each program component. A second broad category pits the individual program components in a statistical "horse race" to understand which ones play the largest role in influencing one particular outcome of interest. In this scenario, a program developer or other stakeholder might wonder which program components contribute the most to participant outcomes. A researcher could take a *correlational approach* to understand which program components matter most. The researcher would use implementation data on each program component to predict an outcome of interest and compare the relative predictive strength of each component.

Understanding the full pathway diagram. A final category estimates all relationships in the full pathway diagram (described above), allowing a better understanding of how program components jointly influence *all* the proximal and distal outcomes. A program developer or stakeholder might want to understand the complex relationship among all the program components and outcomes in the entire pathway diagram, instead of simply looking at a single outcome at a time. This approach allows for an examination of the extent of sequencing among outcomes in the proposed pathway diagram—whether program components influence proximal outcomes that subsequently influence distal outcomes. A *structural equation modeling approach* might be advisable, because all relationships among program components and all proximal and distal outcomes can be estimated in a single model.

Figure 3. Suggested analytic approaches for example research questions

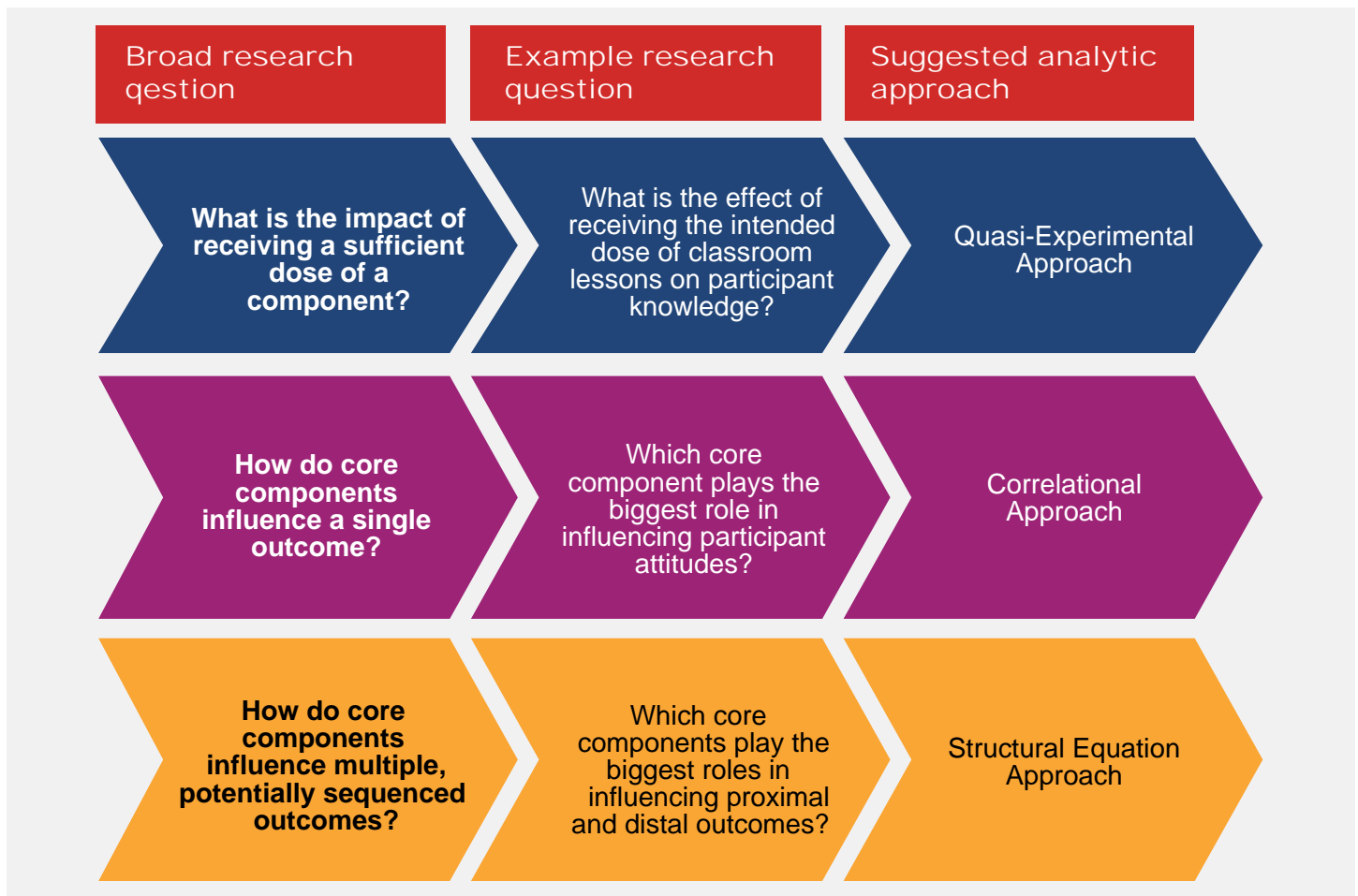| Broad research qestion	Example research question	Suggested analytic approach
What is the impact of receiving a sufficient dose of a component?	What is the effect of receiving the intended dose of classroom lessons on participant knowledge?	Quasi-Experimental Approach
How do core components influence a single outcome?	Which core component plays the biggest role in influencing participant attitudes?	Correlational Approach
How do core components influence multiple, potentially sequenced outcomes?	Which core components play the biggest roles in influencing proximal and distal outcomes?	Structural Equation Approach

Figure 3 shows these three broad categories of research questions, an illustrative research question for that category, and an analytic approach that would be appropriate for answering each type of question.

C. Analytic approaches

In this section, we outline the data preparation steps and methods used to answer research questions about program components.

1. Exploring dosage research questions with a quasi-experimental approach

A quasi-experimental approach allows researchers to estimate the difference in an outcome between youth who received a component of the intervention as intended and those who did not. That is, the approach compares two groups that differ in their experiences of a TPP program. In this illustration, consider assessing the effect of receiving a sufficient dose of classroom lessons on participant knowledge about STIs.

a. Prepare data

The first step in this process is to divide the participants who were offered the intervention into two (or more) groups. The grouping is based on the observed implementation data and on a threshold that defines whether a program component has been implemented in the manner intended. For example, the program developer might assert that attending 75 percent of classroom lessons is the minimum required dosage.[2] As a result, the participants who attended at least 75 percent of the lessons could be considered the treatment group in this quasi-experiment, and their outcomes would be compared to those of participants who attended fewer than 75 percent of the lessons (that is, the comparison group).

b. Understand threats to the validity of testing the effect of the program component

To produce a credible estimate of the effect of receiving a program component as intended, it will first be necessary to convince the reader that the only difference between the two groups is their experience of the program component. Therefore, it is necessary to first demonstrate that the two groups being compared are similar to each other on measurable characteristics that could be related to the outcome of interest. To make this case, at a minimum, we recommend assessing whether the groups are similar at baseline on their demographic characteristics and on a measure of the outcome of interest. These characteristics were previously the baseline equivalence criteria required for studies to be eligible to meet U.S. Department of Health and Human Services evidence standards (U.S. Department of Health and Human Services 2016). Showing the comparability of the two groups on a proxy for motivation would enhance this demonstration of equivalence; presumably, differences in the participant motivation variable will cause differences in attendance/dosage or other measures of implementation.

If this assessment reveals that the groups are not sufficiently equivalent at baseline, it might be necessary to statistically control for the baseline differences to credibly estimate the effect of the program component. Alternatively, if the sample size is large enough, it is possible to trim the analytic sample to only include people who match one another on key variables of interest (see Cole and Agodini 2015 for guidance on this in TPP programs).

c. Compare groups on the outcome of interest

Comparing the two groups on the outcome of interest will produce the potential effect of the program component when received as intended. This analysis can be conducted after adjusting for differences in baseline variables identified above and can produce a *p*-value to help interpret the finding. For example, if the sign and magnitude of the difference in outcomes suggests a large, favorable effect, and if it has a relatively small *p*-value, this is evidence that participants receiving a higher dose of the component had better outcomes.

2. Examining the relative importance of program components using a correlational approach

Researchers can use a correlational approach to estimate the relationship between all measured program components and a single outcome of interest. This approach can identify the program components (or predictor variables) that are most (or least) influential in changing participant outcomes (or dependent variables). For this approach, we build on the pathway diagram developed earlier in Figure 2 and focus on a single outcome of interest: participant attitudes about

choosing to not have sex. In this approach, the predictor variables are the measures of implementation of classroom lessons and the service learning component.

a. Preparing the data

The quasi-experimental approach focused on whether the specified dosage of a program component was received and compared different groups of people based on dosage. For this analytic approach, however, more nuanced information on program implementation might be necessary or valuable. The data for this type of analysis is a single continuous measure of implementation for each component for each person. This might be based on a single measure of implementation (for example, the quality of implementation of a given component), or it might be an aggregation of two or more sources of data (for example, a scale that combines dosage and quality).

b. Conducting the analysis

The correlational approach is commonly conducted through a standard regression analysis, where the outcome of interest is regressed on the implementation data for each program component. If the implementation data and the outcome data vary at the individual level, it might be feasible to use ordinary least squares to conduct this analysis. On the other hand, if certain data vary only at the group level (for example, the measures of implementation of a program component are constant for everyone in a given cluster, like a classroom observation score), then to ensure credible inferences, the analytic approach will need to adjust for this clustering (see Raudenbush and Bryk 2002 for an example of one possible multilevel analytic approach).

When conducting this type of analysis, there are two key statistics to output from the statistical software package and focus on:

i. Partial r^2 for each component. The partial r^2 statistic provides insight into whether changes in the implementation of a program component would be responsible for changes in the outcome of interest. More specifically, it represents the percentage of the variation in the outcome attributable to the program component, after controlling for all other components in the model. This statistic can reveal which of the program components (the predictor variables) play the greatest roles in influencing the outcome (the dependent variable). The program component with the largest partial r^2 statistic is the most influential of the components considered. If the service learning program component had the largest partial r^2 value compared to the partial r^2 values for classroom lessons, this result indicates that service learning is the more influential program component with respect to the outcome.

ii. The standardized regression coefficients (or betas, β). The standardized beta (β) statistic indicates the strength of the relationship between each program component and the outcome of interest after all predictors have been placed on a common scale. The magnitude of each standardized coefficient represents how a one-standard–deviation increase in implementation of each program component influences the outcome of interest. If β = .25, then a one-standard deviation increase in classroom attendance is associated with a 25 percent improvement in favorable attitudes about choosing to not have sex (assuming these attitudes have been measured on or converted to a [0,1] continuum). The largest standardized regression coefficient is the one where a one-standard deviation change has the largest relative effect on the outcome of interest, given a change in the implementation of the given program component.

When reporting these findings, it is important to include both the partial r^2 and standardized regression coefficient results for all program components, along with associated *p*-values. This will give readers the information they need to understand the findings. In addition, we recommend including the overall model R^2 to better illustrate the predictive power of the collection of program components for a given outcome of interest, and conducting several sensitivity analyses to address potential threats to the validity of the finding.

c. Conducting sensitivity analyses

Sensitivity analyses can help convince skeptical readers that the findings from the correlational approach are credible under alternate, but justifiable analytic approaches. For example, if dosage, quality, and engagement of a program component were combined with equal weight in the data preparation stage, an alternative weighting approach might be to use principal components to optimally combine the variables. In addition, to address potential concerns of omitted variable bias, we recommend extending the regression model to include other variables that might influence the outcome of interest, beyond the implementation of the program components. For example, demographics, baseline measures of the outcome of interest, and/or a measure of motivation could be included in the regression model as covariates as a second model specification, above and beyond the inclusion of the program components as focal predictors of interest. By including these additional covariates in the second specification of the analysis, and if the results are substantively similar for the focal predictors, it is more reasonable to conclude that the program components are playing a role in the outcome, given the additional protection against omitted variable bias.

3. Estimating structural equation models

A structural equation model (SEM) provides a way to estimate the program components that play the biggest role in influencing multiple outcomes, including sequential outcomes (for example, distal outcomes that depend on changes in the earlier proximal outcomes). Because this approach allows us to explore more than one outcome of interest at once, it can be regarded as an extension of the correlation approach, which allowed exploration of only a single outcome of interest at a time. Here, we explore a simple application of SEM using observed data to understand how each program component influences multiple outcomes using the pathway diagram described. This is a brief illustration of how SEM can be used to answer questions of how program components influence outcomes.

Figure 4. SEM path diagram depicting the hypothesized theoretical model
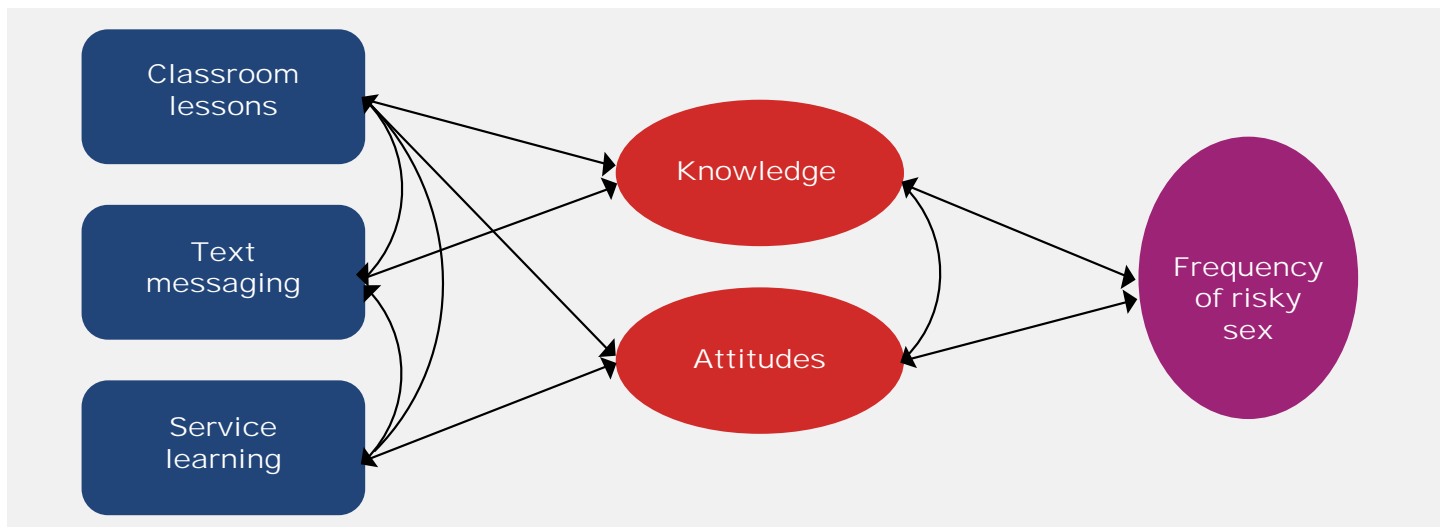
Note: In SEM diagrams, variables on the left side typically predict the variables on the right side. Boxes are used to represent observed or measured variables. Single-headed arrows (paths or regression coefficients) define hypothesized causal relationships in the model, with the variable at the tail of the arrow exerting an effect at the variable at the point. Double-headed arrows indicate covariances or correlations between predictors, without a causal interpretation.

As Figure 4 shows, we hypothesize that the program components of the intervention (lessons, texting, service learning) mutually influence one another, as the bidirectional arrows between them indicate. This is the same (unstated) assumption present in the correlational approach. The program components themselves exert a direct effect on the proximal outcomes (knowledge about STIs and attitudes about choosing to not have sex), and we again assume that these proximal outcomes are correlated. Finally, we assume that the proximal outcomes directly affect the distal outcome (frequency of risky sex). In this figure, the program components only affect the distal outcome indirectly, through the proximal outcomes. In other words, attitudes and knowledge mediate the effect of the program components—classroom lessons, text messaging, and service learning—on the frequency of risky sex.

a. Data preparation

Preparing data for SEM analysis is comparable to preparing data for the correlational approach described above. A single measure of implementation for each program component must be obtained for each participant, along with a single measure of each outcome of interest. As a result, to use the example above, data will be required on each individual participant for six variables of interest: the three program components, two proximal outcomes, and one distal outcome.

b. Method

The typical sequence of steps for an SEM analysis is (1) estimating the model, (2) evaluating and refining the model, and (3) interpreting the model. These steps often require several rounds of refinement before a model is finalized. Below, we briefly outline these steps:

1. **Estimating the model.** The first step in SEM is to use the statistical package of choice to perform the analyses according to a hypothesized model.[3] As with the correlational approach described earlier, if either predictor or outcome data are clustered, a multilevel approach might be required to obtain the correct standard errors and *p*-values from the analysis.

2. **Evaluating and refining the model.** After the model has been estimated, the next step is to examine the fit of the hypothesized model to the data. The ability to test complex theoretical hypotheses about the relationship between program components and outcomes is a key strength of SEM, and the fit of the model to the data helps provide confidence in the findings about program components described later.

   To determine how well the hypothesized model fits the data, it is common to examine several indices commonly produced by statistical software packages (West et al. 2012).[4] If the model fit statistics suggest good fit, it is feasible to move to Step 3, interpreting the model. If one or more fit indices reveal that the SEM is not consistent with the data, the SEM will need to be refined and re-estimated. Often, refining the SEM means adding new relationships among program components and outcomes not previously expected to exist or removing theoretical paths between components of the pathway diagram.[5] Occasionally, refining an SEM to improve model fit can lead to substantive findings (for example, showing that a hypothesized relationship between a program component and an outcome does not exist in the data, or that a program component influences an unexpected outcome).

3. **Interpreting the model.** The statistics of interest from an SEM analysis of program components will be the observed path coefficients (effectively, regression coefficients and standard errors) that link the program components to outcomes. (See Kline 2015 for more information.)

Figure 5 presents hypothetical SEM results for the working example. To simplify presentation for this illustration, only a subset of path estimates is shown. For program components, the hypothesized path model indicates that classroom lessons have a positive effect on knowledge about STIs: a one-standard-deviation increase in the measure of classroom lesson implementation is associated with a .38-unit increase in knowledge of STIs. As the figure shows, this analysis produces *p*-values that can be used to indicate whether the relationship is statistically significant.[6]

Figure 5. Example of an SEM path diagram, with results
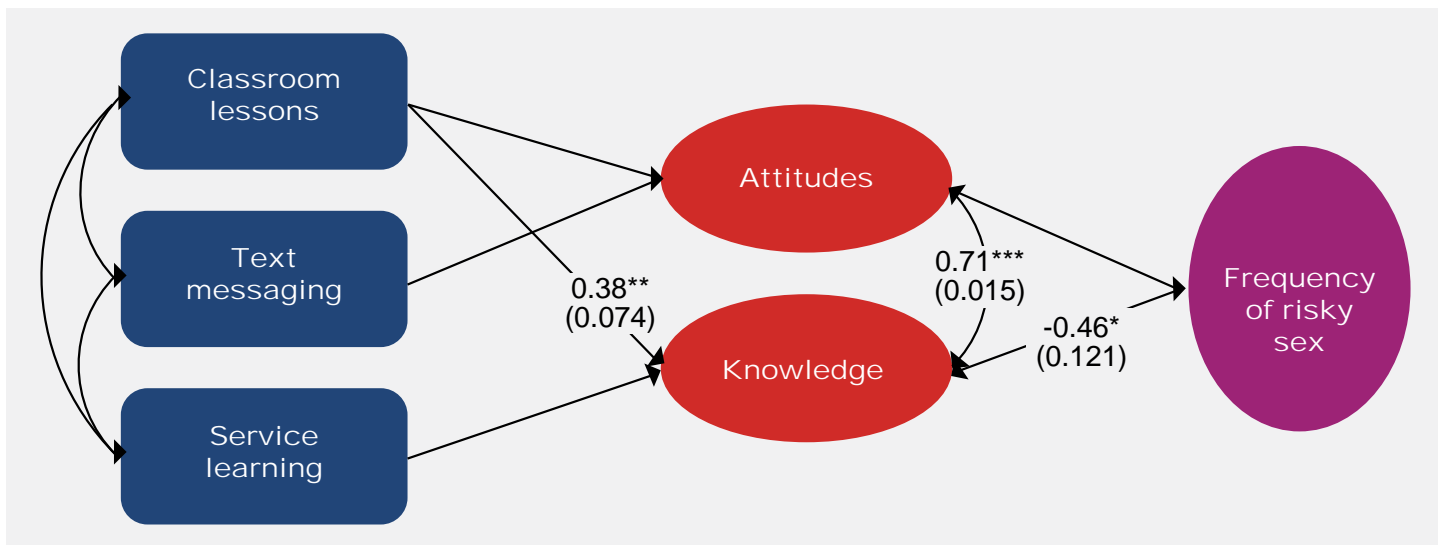
Note: Hypothesized standardized path effects and standard errors (in parentheses) are shown on the diagram.
* = p < .05, ** = p < .01, *** = p < .001

In addition to reporting and interpreting the direct effects of program components on proximal outcomes, the SEM results can provide more nuanced information about how outcomes are related to each other. One notable feature is estimating the extent to which proximal outcomes mediate distal outcomes (the indirect effects of the program components).[7] In Figure 5, we see that attitudes about choosing to not have sex and knowledge about STIs are outcomes that are strongly correlated (r = .71). Furthermore, we see suggestive evidence that one of the proximal outcomes of the intervention, knowledge about STIs, has a strong negative impact on subsequent risky sex behavior. This finding would corroborate our assumptions that increased knowledge about STIs decreases the likelihood that youth would engage in risky sexual behavior (that might put them at risk for STIs). When reporting SEM findings, it is important to include the hypothesized path diagram, the fit indices, the path diagram with results, and any model re-specifications made when finalizing the model.

D. Including program component analyses in a report

In a report evaluating the effects of an intervention, findings on the relationship between program components and the outcome of interest are typically presented as supplemental to the impact findings. The approaches presented in this brief are designed to answer exploratory research questions, not the typical confirmatory impact evaluation research questions; as a result, they should be framed differently. Importantly, these types of exploratory analyses are not producing evidence of the effect of the overall intervention.

In addition to framing these findings as supplemental or separate from impact findings, it is always important to state the limitations of the analyses around these exploratory research questions. Three key limitations are common to nearly all analyses of program components described above:

- The analyses are only as good as the measures of implementation of the program components and outcomes of interest. Without high quality data on implementation or outcomes, it is not possible to credibly link the implementation of program components to the outcomes of interest. As noted earlier, it might not be possible to collect rich data on all aspects of implementation (for example, dosage, engagement, quality, and adherence). As a result, the available data on implementation might not be sufficient to completely describe implementation of each program component of interest.

- Because all three of the analytic options discussed in this brief include only the data from the treatment group, the sample sizes are typically about half as large as those in the impact study. As a result, regardless of the analytic approach, these exploratory approaches offer reduced statistical power compared to the overall impact evaluation.

- Omitted variable bias, or the possibility that a relevant variable was not included in the analyses, will always be a concern in these exploratory approaches. A critical reader might worry that the observed relationships between program components and outcome are actually due to a variable that should have been accounted for in the analytic approach but that was erroneously left out. There are ways to quantify the degree to which omitted variable bias threatens an observed finding (see, for example, Frank 2000), and ways to attempt to mitigate it through sensitivity analyses to show the robustness of the findings to different analytic approaches. At a minimum, it is important to acknowledge this threat to inference, because it is a key limitation for exploratory analyses such as these.

The approaches presented in this brief allow researchers to explore how naturally occurring variation in implementation of program components is associated with variation in program outcomes. The information learned from such analyses might inform future, more rigorous tests of components, where the above limitations might be mitigated. For example, if exploratory analyses suggest that a particular component is particularly important, a stand-alone experiment where youth are randomized to receive the component or not might be a contribution to the field. On the other hand, if exploratory analyses suggest that a particular component is unimportant, a stand-alone experiment where that component is eliminated from the intervention might offer evidence for how to more efficiently deliver an intervention. Finally, a body of research is emerging that focuses on rigorously testing combinations of components as a way to produce an optimal version of an intervention (for example, the multiphase optimization strategy [Collins et al. 2005] or the sequential multiple assignment randomized trial [Murphy 2005]). Information learned from the exploratory analyses described above could allow developers and researchers to create an optimal collection of program components for TPP programs to meet the needs of youth.

References

Baron, R.M., and D.A. Kenny. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology,* vol. 51, no. 6, 1986, pp.1173–1182. Available at http://www.ncbi.nlm.nih.gov/pubmed/3806354. Accessed March 19, 2018.

Betensky, R.A. "The *p*-Value Requires Context, Not a Threshold." *American Statistician*, vol. 73, suppl. 1, 2019, pp.115–117. doi:https://doi.org/10.1080/00031305.2018.1529624.

Blase, K., and D. Fixen. "Core Intervention Components: Identifying and Operationalizing What Makes Programs Work." ASPE Research Brief. Washington, DC: U.S. Department of Health and Human Services, 2013.

Byrne, Barbara M. *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming (Multivariate Applications Series).* Abingdon, UK: Routledge, 2011.

Chou, C.P., and J. Huh. "Model Modification in Structural Equation Modeling." In *Handbook of Structural Equation Modeling,* edited by R.H. Hoyle. New York: Guilford Press, 2012.

Cole, R., and R. Agodini. "Baseline Inequivalence and Matching." Evaluation Technical Assistance Brief no. 4. Submitted to the Office of Adolescent Health and the Administration on Children, Youth and Families Teenage Pregnancy Prevention Grantees. Princeton, NJ: Mathematica Policy Research, November 2015.

Cole, R., and L. Murphy. "Structural Elements of an Intervention." Evaluation Technical Assistance Brief no. 12. Submitted to the Office of Adolescent Health and the Administration on Children, Youth and Families Teenage Pregnancy Prevention Grantees. Princeton, NJ: Mathematica Policy Research, October 2016.

Collins, L.M., S.A. Murphy, V.N. Nair, and V. Strecher. "A Strategy for Optimizing and Evaluating Behavioral Interventions." *Annals of Behavioral Medicine,* vol. 30, 2005, pp. 65–73.

Frank, K.A. "Impact of a Confounding Variable on a Regression Coefficient." *Sociological Methods and Research,* vol. 29, no. 2, 2000, pp. 147–194.

Hoyle, R.H. (ed.). *Handbook of Structural Equation Modeling.* New York: Guilford Press, 2012.

Kautz, Tim, and Russell Cole. "Selecting Benchmark and Sensitivity Analyses." Evaluation Technical Assistance Brief. Submitted to the Office of Adolescent Health. Princeton, NJ: Mathematica Policy Research, September 2017.

Kline, Rex B. *Principles and Practice of Structural Equation Modeling, Fourth Edition.* New York: Guilford Press, 2015.

Murphy, S.A. "An Experimental Design for the Development of Adaptive Treatment Strategies." *Statistics in Medicine,* vol. 24, 2005, pp. 1455–1481.

Raudenbush, S.W., and A.S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition.* Thousand Oaks, CA: Sage Publications, 2002.

Raykov, T., and G.A. Marcoulides. *A First Course in Structural Equation Modeling, Second Edition.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2006.

Schumacker, R.E., and R.G. Lomax. *A Beginner's Guide to Structural Equation Modeling: Third Edition.* Abingdon, UK: Routledge, 2015.

U.S. Department of Health and Human Services. "Identifying Programs That Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors." Review Protocol, Version 5. Washington, DC: DHHS, 2016. Available at https://tppevidencereview.aspe.hhs.gov/pdfs/TPPER_Review%20Protocol_v5.pdf. Accessed March 1, 2018.

Wang, J., and X. Wang. *Structural Equation Modeling: Applications Using MPlus.* West Sussex, UK: John Wiley & Sons Ltd., 2012.

Wasserstein, R.L., and N.A. Lazar. "The ASA's Statement on *p*-Values: Context, Process, and Purpose." *American Statistician*, vol. 70, no. 2, 2016, pp. 129–133. doi:10.1080/00031305.2016.1154108.

Wasserstein, R.L., A.L. Schirm, and N.A. Lazar. "Moving to a World Beyond 'p < 0.05'." *American Statistician*, vol. 73, suppl. 1, 2019, pp. 1–19. doi:10.1080/00031305.2019.1583913.

West, S.G., A.B. Taylor, and W. Wu. "Model Fit and Model Selection in Structural Equation Modeling." In *Handbook of Structural Equation Modeling*, edited by R.H. Hoyle. New York: Guilford Press, 2012.

Endnotes

[1] A quasi-experimental approach is a research design that includes at least two groups (for example, a treatment and a comparison group); however, it does not employ random assignment to allocate participants to groups.

[2] In this example, we assume that dosage alone is sufficient for representing implementation as intended, so we do not include the other features of implementation (such as quality, adherence, or engagement). Alternatively, a researcher could choose any of the features of implementation (for example, quality), or a combination of them (for example, a certain number of high quality sessions attended), to represent whether or not a participant received the intended implementation of the core component.

[3] SEMs can be estimated using either specialized software (Mplus, LISREL, EQS) or general statistical packages such as Stata (via the SEM procedure or GLLAMM package), SAS (via CALIS procedure), SPSS (via AMOS add-on module), and R (via SEM or lavaan packages). For details on how to specify such a model and conduct the analyses, see Byrne 2011; Hoyle 2012; Kline 2015; and Raykov and Marcoulides 2006.

[4] The fit indices give researchers information on whether the hypothesized model is appropriate, or whether one or more revisions to the model are needed before it can be deemed acceptable and appropriate for interpretation. Typical indices to examine and report include the root mean square error of approximation, standardized root mean square residual, comparative fit index, and Tucker-Lewis index.

[5] See Chou and Hug (2012) for details on how to refine the model.

[6] Although this figure shows commonly used thresholds for defining statistical significance, relying solely on statistical significance to determine whether to report a finding is an inappropriate decision rule (Betensky 2019; Wasserstein et al. 2016, 2019). As a result, it is important to focus on additional results from this analysis beyond the *p*-value of a given path coefficient.

[7] See Baron and Kenny (1986) for details about the interpretation of the direct and indirect effects in the SEMs, and Schumacker and Lomax (2015), Byrne (2011), and Wang and Wang (2012) for more guidance on interpreting key features of an SEM.