Updated Findings from the HHS Teen Pregnancy Prevention Evidence Review—May 2023

September 2024

By Emily Forrester, Connor J. Rooney, Tareena Musaddiq, Jean Knab, Julieta Lugo-Gil, Russell Cole, Lexi Ouellette, Erich Denk, Juliette Henke, Brandon Hollie, Lara Hulsey, and Jamie Manzer

Mathematica

KEY FINDINGS

- This brief presents findings from the updated Teen Pregnancy Prevention Evidence Review (TPPER), which includes research released through May 2023.
- Studies of seventeen programs were reviewed during this round of the review:
  - Ten programs were reviewed for the first time, and five of those programs met the review criteria for evidence of effectiveness (2gether, All4You2!, Choosing the Best JOURNEY, MyPEEPS Mobile, and Your Move).
  - New findings were reviewed for seven previously reviewed programs (some of which previously showed evidence of effectiveness and some of which previously did not). None of these newly available findings met the review criteria for evidence of effectiveness, resulting in no change to the categorizations of these previously reviewed programs.
- Nine programs previously shown to have evidence of effectiveness have been moved to the list of inactive programs because the research evidence is more than 20 years old or the program is no longer distributed (see Table 5).
- Forty-eight active programs now meet the review criteria for evidence of effectiveness: Forty-three programs identified in earlier rounds of the review (that have recent evidence of effectiveness and are still available for implementation) plus five new programs with evidence of effectiveness.
- The TPPER expanded its eligibility criteria to include studies on the impact of well-defined components or combinations of components from programs intended to reduce unintended teen pregnancy, sexually transmitted infections, or associated sexual risk behaviors. Two program components met the review criteria for evidence of effectiveness. In the future, after the TPPER compiles a strong evidence base for components, TPPER users will be able to select the most effective components for a specific program aim or population so that they can tailor their program implementation to the unique needs of their community. Refer to page 10 for information about how to interpret findings from studies of program components.

INTRODUCTION

What is the Teen Pregnancy Prevention Evidence Review? Since 2009, the U.S. Department of Health and Human Services (HHS) has sponsored a systematic review of research on teen pregnancy prevention to identify programs with evidence of effectiveness that favorably impact (1) rates of unintended teen pregnancy and sexually transmitted infections (STIs) and (2) sexual behaviors. The HHS Teen Pregnancy Prevention Evidence Review (TPPER) was created in response to the 2010 Consolidated Appropriations Act, which states that teen pregnancy prevention programs must be "proven effective through rigorous evaluation to reduce teenage pregnancy, behavioral risk factors underlying teenage pregnancy, or other associated risk factors." The TPPER is sponsored by the HHS Office of the Assistant Secretary for Planning and Evaluation, the Office of Population Affairs (OPA) within the HHS Office of the Assistant Secretary for Health, and the Family and Youth Services Bureau (FYSB) in the HHS Administration for Children and Families.

What is considered evidence of effectiveness? The TPPER's review criteria require programs to show evidence of at least one favorable, statistically significant impact on at least one outcome of interest reflecting sexual behavior (for example, whether teens have ever had sex) or the consequences of sexual behavior (for example, STIs, HIV, or pregnancy). In addition to having significant and favorable findings, the supporting research studies must meet established criteria for the quality and execution of their research designs. The review team follows prespecified criteria to assess study design, sample attrition, baseline equivalence, reassignment of sample members, and confounding factors. More details about the TPPER review procedures and criteria can be found in Appendix A and in the Review Protocol 7.0 on the TPPER website.

What's new this round? This round of the TPPER – the seventh review of evidence since 2009 – introduced expanded eligibility criteria. For the first time, studies that focus on the impact of well-defined components or combinations of components for programs intended to reduce unintended teen pregnancy, STIs, or associated sexual risk behaviors were eligible for review. Program components are the elements and activities that make up a program. For TPPER, components are elements common to TPP programs including but not limited to: content (for instance, content about STIs), delivery mechanism (for instance, role play or lecture), format (for instance, computer based), staffing (for instance, single facilitator versus co-facilitation), dosage (for instance, two-week delivery window versus 12-week delivery window), context (for instance, clinic versus at home), and intended population characteristics (for instance, age of youth). By permitting studies of components to be eligible for the TPPER, we can begin compiling information about which individual components or combinations of components are effective on their own or in combination with a specific program. Expanding the eligibility criteria to include tests of program components helps to highlight the opportunity for continued research on program component effectiveness, which evaluators can contribute to by reporting program component contrasts when available. Review findings for studies on components are summarized separately from the studies of programs in this brief, but they present comparable information.

What are evidence-based components and how should they be used? Evidence-based program components are elements or activities of a program that have been tested and shown to have evidence of effectiveness in improving sexual behavior outcomes, with that effectiveness distinct from the effects of any associated program. Evidence-based program components may have been implemented independently, in conjunction with, or integrated into a TPP program. Component studies included in the TPPER provide information about which components of existing programs are important for achieving outcomes, and whether there are components that could potentially be added to a program as a supplement. In the future, after the TPPER compiles a strong evidence base for components, TPPER users will be able to select the most effective components for a specific program aim or population so that they can tailor their program implementation to the unique needs of their community.

What's in this brief? This brief presents an update to the evidence base for teen pregnancy prevention. The first section presents the evidence for **new programs** that were reviewed in 2023. The second section presents the evidence for **new program components** that were reviewed at the same time. The final section presents the programs that are now considered to be inactive because the research is more than 20 years old or because they are no longer distributed. A complete list of programs with evidence of effectiveness, including those reviewed in the prior six rounds of review, is available on the TPPER website. Appendix B provides details about the newly available study-level and program-level findings for previously reviewed programs.

PROGRAMS REVIEWED

The TPPER reviewed studies of 17 programs during this round of the review. Ten of the programs were newly identified, meaning that studies of these programs have not previously been reviewed by the TPPER. These 10 programs are the first 10 programs listed in Table 1. Five of the newly identified programs had studies that met the TPPER criteria for evidence of effectiveness, and five of the newly identified programs did not. The next seven programs listed in Table 1 had evaluation studies that were reviewed under a previous round of the TPPER. Additional publications presenting findings for those programs were identified and reviewed in this round. However, none of the new publications met criteria for evidence of effectiveness. Details about these publications are available in Appendix B.

Table 1. Programs reviewed during this round of the TPPER

Newly identified programs
New programs with evidence of effectiveness
2gether
All4You2![a]
Choosing the Best JOURNEY
MyPEEPS Mobile
Your Move
New programs without evidence of effectiveness
All4You2! + service learning
eHealth Familias Unidas
FatherWorks
Project Prepared
Sexual risk reduction intervention (no formal name)
Previously reviewed programs with newly reviewed findings that do not show evidence of effectiveness
Evidence-based programs: Programs previously determined to be evidence-based (the new findings did not change this assessment)
Familias Unidas
Girl2Girl
It's Your Game … Keep It Real!
Love Notes (for ages 14–24)
Reducing the Risk
Non-evidence-based Programs: Programs not previously determined to be evidence-based (the new findings did not change this assessment)
Be yoU, Talented, Informed, Fearless, Uncompromised, and Loved (BUtiful)
Wise Guys

Note: [a] Previously, All4You2! and All4You! were considered versions of the same program, so the findings from both studies were pooled together when describing the evidence for the program. However, based on an updated assessment, HHS and Mathematica determined that the programs appear to be substantively different from each other and therefore All4You2! is no longer be classified as an adaptation of All4You! It is characterized as a distinct program. Therefore, we have re-reviewed the study of All4You2! under the current version of the TPPER protocol and summarized its evidence separately from the previously reviewed evidence for All4You!

Evidence for new programs

We reviewed evaluation studies of 10 newly identified programs. The programs varied in terms of their focus, intended population, setting, and content. Studies that were federally funded by FYSB or OPA were prioritized for review. Three of the 10 newly identified programs were implemented or tested with grant funding from FYSB or OPA. We present the findings for each of the 10 newly identified programs in Table 2.

The first five new programs listed in Table 2 below had evidence of effectiveness. Each of these five programs had evidence of a favorable, statistically significant impact on at least one outcome of interest that reflects sexual behavior or its consequences[1], and this evidence came from one high quality study for each program.

- 2gether
- All4You2!
- Choosing the Best JOURNEY
- MyPEEPS Mobile
- Your Move

Studies of the next five newly identified programs shown in Table 2 did not meet the review criteria for evidence of effectiveness. Specifically, the study of All4You2! + service learning, which examined the impacts of the All4You2! curriculum in combination with a service learning program component, met the high quality study criteria but had no favorable, statistically significant impact on an outcome of interest. The last four programs in Table 2 had supporting impact studies that received a study quality rating of low. Therefore, the review team did not assess evidence of effectiveness from these studies because the studies do not provide credible estimates of program impacts.

Table 2. Newly identified programs: Evidence by outcome domain and study rating

Program	Manuscript reviewed	Outcome domains					Study quality rating[a]
		Sexual activity	Number of sexual partners	Contra-ceptive use	STIs or HIV	Pregnancy	
Programs with evidence of effectiveness							
2gether	Kottke et al. (2022)				●	○	High
All4You2!	Coyle et al. (2013)	○	○	●			High
Choosing the Best JOURNEY[b]	Floren & Floren (2023)	●					High
MyPEEPS Mobile	Schnall et al. (2022)		○	●	○		High
Your Move[b]	Potter et al. (2023)		○	●			High

[1] Measures meeting this definition fall into the following five domains (or categories): (1) sexual activity; (2) number of sexual partners; (3) contraceptive use; (4) STIs or HIV; and (5) pregnancies. Within each domain, there can be multiple eligible measures. For instance, sexual activity includes measures like sexual initiation and recent sexual activity (e.g., in the past week or month or three months).

Program	Manuscript reviewed	Outcome domains					Study quality rating[a]
		Sexual activity	Number of sexual partners	Contra-ceptive use	STIs or HIV	Pregnancy	
Programs with studies that met review standards, but had no favorable effects							
All4You2! + Service learning	Coyle et al. (2013)	○	○	○			High
Programs with studies that did not meet review standards							
eHealth Familias Unidas	Estrada et al. (2017)	n.a.	n.a.	n.a.	n.a.	n.a.	Low
FatherWorks[b]	Mogro-Wilson et al. (2020)	n.a.	n.a.	n.a.	n.a.	n.a.	Low
Project Prepared	Bauman et al. (2021)	n.a.	n.a.	n.a.	n.a.	n.a.	Low
Sexual risk reduction intervention (no formal name)	Schmiege et al. (2021)	n.a.	n.a.	n.a.	n.a.	n.a.	Low

Note: Empty cells indicate that the study or studies did not examine program impacts on measures in that outcome domain. Programs are listed alphabetically within each evidence category.

n.a. = not applicable; following the procedures specified in the review protocol, when the supporting impact study received a study quality rating of low the review team did not assess evidence of effectiveness from that study.

● Evidence of a favorable finding with no evidence of adverse findings across one study reviewed.

○ No statistically significant findings across one study reviewed.

Refer to the key in Appendix Table A.6 for additional details about the program-level rating symbols in this table.

[a] The review team established ratings for the supporting impact studies by following pre-specified criteria to assess study design, attrition, baseline equivalence, reassignment of sample members, and confounding factors. Appendix Table A.1 has a more detailed description of the study quality ratings.

[b] Programs were implemented or tested with federal grant funding from FYSB or OPA.

Summaries of new programs with evidence of effectiveness

Next, we describe the five new programs that met the review criteria for evidence of effectiveness, along with a summary of the research evidence on these programs.

2gether

Program summary. 2gether is a 12-month, multicomponent, clinic-based intervention for African American females from ages 14 to 19. It is designed to increase the use of dual protection (DP) strategies, which focus on simultaneously preventing STIs and unintended pregnancies. One component of the primary intervention is two individual clinic visits that offer programming on DP strategies and one-on-one health education counseling. The second component is four individual phone counseling sessions delivered by nurse educators to reinforce the correct and consistent use of the DP strategies.

Research summary. The program was implemented in a clinical setting in the southeastern United States and tested with a randomized controlled trial. Participants were 714 sexually active African American female adolescents who were HIV negative, were not pregnant, and hoped to avoid pregnancy for at least the next 12 months. Participants were randomly assigned to either a

treatment condition that received the 2gether intervention or a control condition that received the regular standard of care. After participants were enrolled and participated in the first clinic visit, self-reported behavioral data were collected during a three-month follow-up call, a six-month follow-up visit, a nine-month follow-up call, and a 12-month final follow-up visit. Urine specimens for STI and pregnancy testing were collected during enrollment, at a six-month follow-up, and at a 12-month follow-up. In addition, STI and pregnancy test data for all participants were collected from medical record reviews and, when possible, from outside clinics' records, if participants reported having tested at an outside clinic for pregnancy or STI. The study found that 2gether program participants had a significantly lower prevalence rate of testing positive for the STI trichomonas than the control condition during the three- to 12-month follow-up period (effect size = -0.27). The study showed no statistically significant impacts on any other assessed sexual behavior outcomes. The study met the review criteria for a high quality rating.

All4You2!

Program summary. All4You2! is a school-based, sexual health education curriculum for youth, ages 14 to 18, in alternative high school settings. The program addresses youth attitudes, beliefs, and perceived norms about contraceptive use and risk of STI. The program consists of around 12.5 hours of instruction presented in 15 sessions, which are taught two to three times per week during school hours. All4You2! is delivered by classroom teachers or community-based educators.

Research summary. The All4You2! program was evaluated in a cluster randomized controlled trial involving 24 classrooms (430 youth) drawn from 11 district-run continuation high schools in Northern California. Classrooms were randomized either to a treatment group that received All4You2! or a control condition that received content on nutrition and physical education. The study included two additional treatment groups that were not relevant to the test of effectiveness of the All4You2! program. They are described elsewhere in this brief. Surveys were administered at baseline and at six and 18 months after baseline. The study found that All4You2! program participants had significantly lower frequency of unprotected sexual intercourse (vaginal intercourse without a condom) than participants in the control condition at the six-month follow-up (effect size = -0.33). The study showed no statistically significant impacts on any other assessed sexual behavior outcomes. The study met the review criteria for a high quality rating.

Choosing the Best JOURNEY

Program summary. Choosing the Best JOURNEY is a sexual risk avoidance program for youth in grades 9 and 10. It comprises a classroom-based curriculum that focuses on relationships, goal setting, STIs, teen pregnancy, and overcoming pressure to have sex. The program content is delivered by a trained facilitator in eight weekly 45-minute sessions that include video vignettes related to lesson topics and opportunities for students to practice skills through role-playing activities.

Research summary. The program was evaluated with a cluster randomized controlled trial involving 1,143 ninth-graders in six schools in two Georgia school districts. The schools were randomly assigned to a treatment group that received either Choosing the Best JOURNEY or a control group that received abstinence-based classroom lessons. Surveys were administered immediately before the program (baseline), at the end of 9th grade (nine months after the baseline), and again at the start of 10th grade (14 months after the baseline). The study found that nine months after baseline, youth participating in the program were significantly less likely than youth in the control group to report ever having had sex (effect size = -0.29). The study showed no statistically significant impacts on any other assessed sexual behavior outcomes. The study met the review criteria for a high quality rating.

MyPEEPS Mobile

Program summary. MyPEEPS Mobile is a sexual health education program delivered through a web-based application. It was designed to reduce rates of HIV and increase condom use among racially diverse, same sex–attracted males, ages 13 to 18. There are 21 sequential modules in MyPEEPS Mobile that deliver information about sexual health, relationships, and emotional health. Each module includes a variety of activities, including role-plays, scenarios, or games. It takes between six and nine hours to complete all 21 modules. Participants progress through the modules asynchronously. The full program is intended to be completed within three months.

Research summary. The program was evaluated using a randomized controlled trial involving 763 individuals, ages 13 to 18, who identified as male and were attracted to males. The participants lived in cities in Alabama, Illinois, New York, and Washington. Participants were randomly assigned to a treatment group that received MyPEEPS Mobile or a control group that was offered no intervention and received no contact during the study (however, the control group was offered access to MyPEEPS Mobile after the conclusion of the study). Surveys were administered to both groups immediately at baseline and at three, six, and nine months after baseline. The study found that three months after baseline, young men participating in the intervention reported a statistically significant decrease in the number of condomless anal sex acts with partners compared to the control group (effect size = -0.07). The study showed no statistically significant impacts on any other assessed sexual behavior outcomes. The study met the review criteria for a high quality rating.

Your Move

Program summary. Your Move is a sexual health education program that combines in-person, group-based learning and individual, online learning. The program is designed for females ages 14 to 19 and focuses on sexual decision-making skills. The program is designed to be implemented in schools, clinics, or community settings once per week to groups of four to 16 participants. Your Move is delivered in seven 75-minute sessions. Each session includes 60 minutes of facilitator-led, group-based activities followed by 15 minutes of personal reflection activities.

Research summary. The program was evaluated with a cluster randomized controlled trial. The participants were young women ages 14 to 19 who were recruited from community-based organizations or charter schools in Delaware, Missouri, New Jersey, New York, Ohio, and Pennsylvania. A total of 104 groups, ranging from four to 14 participants (n = 808), were randomly assigned to either a treatment group that received the Your Move program or a control group that received a similarly formatted alternate program, Eat Smart, that was focused on promoting healthy eating. Programming occurred in community-based organization settings. Surveys were administered immediately before the intervention (baseline) and three months after it ended. The study found that three months after the end of the intervention, Your Move participants were significantly less likely to report having sex without using condoms or effective birth control in the past three months compared to those in the control group (effect size = -0.37). The study showed no statistically significant impacts on any other assessed sexual behavior outcomes. The study met the review criteria for a high quality rating.

PROGRAM COMPONENTS REVIEWED

This brief presents findings from the first round of the TPPER that program components or combinations of components were eligible for review. Program components are the elements and activities that make up a program. To be eligible for review under Version 7.0 of the TPPER protocol, a component must be (1) a clearly defined practice, procedure, policy, support, or organizational structure, potentially with documented steps for implementation with fidelity to facilitate replication, and (2) (theoretically) capable of being implemented independently, in conjunction with, or integrated into a TPP intervention. For TPPER, components are elements common to TPP programs including but not limited to: content (for instance, content about STIs), delivery mechanism (for instance, role play or lecture), format (for instance, computer based), staffing (for instance, single facilitator versus co-facilitation), dosage (for instance, two-week delivery window versus 12-week delivery window), context (clinic versus at home), and intended population characteristics (for instance, age of youth). Although it is possible to conceptualize each individual component separately in a granular way (such as content about STIs), in practice, individual components are combined with each other in different ways (such as content about condoms, delivered by health teachers, through a demonstration, for a dosage of 10 minutes).[2]

In effectiveness research, it is possible to isolate the effects of a single component or combination of components. This might be accomplished through tests of such components delivered in conjunction with a TPP program, and compared to the same TPP program without the component(s). However, it is also possible that there are unobservable interaction effects or contextual effects that influence the findings of these tests of components, and thus, the observed impacts from evaluations of components are tied to the evaluation context or setting. Over time, as the body of evidence about components grows, we will be able to better understand whether a component or combination of components are effective on their own or whether they are only effective in specific contexts or in conjunction with specific programs. Because effectiveness research on components is only beginning to emerge in the TPPER, we cannot make general conclusions about whether the components with evidence of effectiveness identified through the TPPER would be effective if implemented in conjunction with different programs than they were evaluated with.

The TPPER reviewed studies of eight program components[3] that varied in their focus, intended population, setting, and content. In each case, the component supplemented teen pregnancy prevention content delivered through a program or was used as a method or mode to deliver that content. The components are organized into categories, such as "booster sexual health

[2] For more information about components of TPP programs generally, see Forrester and Cole (2023), and for information about the individual components present across the TPPER identified evidence-based programs, refer to Forrester et al. (Forthcoming).

[3] Some of the studies currently categorized as tests of program components were reviewed in previous rounds of the review as studies of programs. The current TPPER protocol has more formally defined eligibility criteria for tests of components. The studies that are now presented as testing the effectiveness of individual components were re-reviewed using the current version of the TPPER protocol.

counseling" to attempt to distill the general component being tested from the specific context and setting in the evaluation. Specific component names and descriptions for components with evidence of effectiveness are presented in the component summaries after Table 3. Two of the newly identified program components had studies that met TPPER criteria for evidence of effectiveness, while six of the newly identified components did not. Although the evidence for program components is currently limited, we hope that new studies will provide more evidence of effectiveness for program components.

How to interpret findings from studies of program components

Studies of program components provide information about the unique, added benefit of elements and activities that make up a program. Although the impact evaluations of program components attempt to isolate the effects of each component, it is important to interpret these effects in the context of the programs they are delivered with. We do not have enough evidence to indicate whether the tested components would work the same way if they were implemented with a different program.

Program components with and without evidence

The findings in Table 3 indicate that for components that were found to be effective, such as sexual health risk reduction counseling implemented as part of HORIZONS, the inclusion of the specific sexual health risk reduction counseling component to the HORIZONS program yields additional impacts on sexual behavior outcomes. **The results *do not* mean that adding sexual health risk reduction counseling to any program would be considered evidence based.**

Similarly, the addition of an alcohol content component to a sexual risk reduction intervention (no formal name) did not yield additional impacts on sexual behavior outcomes, and the results *do not* mean that alcohol content components are ineffective, generally. As we review more evidence for this component over time, we'll get a better understanding of whether booster sexual health risk reduction counseling is broadly effective, or whether it was the specific combination of what was offered in the HORIZONS study that was necessary for the effect.

Counterfactual conditions in studies of program components

Unlike studies of whole programs, which typically compare a group of participants who received the program to a group of participants who did not receive the program (often, a business-as-usual comparison group), studies of program components often present the impacts of one program component as it compares to another program component. Therefore, it is more important to understand and describe the nature of the comparison condition in evaluations of components. For instance, Gryczynski and colleagues (2021) examined impacts of the delivery format component of a brief intervention by randomizing participants to a computer format or in-person format study condition. The authors compared the outcomes of youth experiencing the two different format conditions to each other. So, if one of these formats has a favorable effect, by definition, the other will have an unfavorable effect. This type of comparison answers the question, "Is there a difference in effects on outcomes when the same content is delivered by a person rather than a computer?" As shown in Table 3, the computer format component—one of the components the authors examined for effectiveness—had significant unfavorable effects on STIs and HIV. However, because this effect is based on a comparison to another component (the in-person format), the inverse of this unfavorable result for computer format is a favorable result for the in-person format component. In other words, the

results from this study revealed that when the programming was delivered by a person and not by a computer, there was a favorable effect on STI and HIV outcomes. **Importantly, the results *do not* mean that computer format components are ineffective, generally.** Again, as we review more evidence comparing in-person to computer-based delivery, we'll learn whether in-person delivery is generally more effective than computer-based delivery or under what conditions that is the case.

Evidence for new program components

We present the findings for each of the eight newly identified program components in Table 3. The first two components in that table had evidence of effectiveness and are categorized into the general components: (1) booster sexual health counseling and (2) in-person delivery format. Specific program component names are shown in Table 3. Both program components had evidence from one high quality study of a favorable, statistically significant impact on at least one outcome of interest that reflects sexual behavior or its consequences. Studies of the next five newly identified program components listed did not meet the review criteria for evidence of effectiveness. Specifically, the studies of the following five program components met the high study quality criteria but had no favorable, statistically significant impacts on an outcome of interest:

- alcohol program content
- alcohol and cannabis program content
- cannabis program content
- computer-based delivery format
- service learning

The last component that was reviewed, text messaging, had a supporting impact study that received a quality rating of low; therefore, the review team did not assess evidence of effectiveness from this study.

Table 3. Newly identified program components: Findings by outcome domain and study rating

Component type(s) that were tested[a]	General name of component or combination of components tested	Specific name of component or combination of components tested	Counter-factual	Program component studied with[b]	Manuscript reviewed	Outcome domains					Study quality rating[c]
						Sexual activity	Number of sexual partners	Contra-ceptive use	STIs or HIV	Preg-nancy	
Components with evidence of effectiveness											
Delivery mechanism; Content	Booster sexual health counseling	Booster sexual risk reduction counseling implemented as a supplement to HORIZONS	General health promotion counseling	HORIZONS	DiClemente et al. (2014)		●	●			High
Format	In-person format[d]	In-person delivery of a clinic-based brief intervention by nurse practitioners	Computer format	Brief intervention (no formal name)	Gryczynski et al. (2021)	○			●		High
Components with studies that met study quality standards but had no favorable effects											
Content	Alcohol program content	Alcohol program content implemented as a supplement to a sexual risk reduction intervention	No alcohol program content	Sexual risk reduction intervention (no formal name)	Bryan et al. (2018)			○			High
Content	Alcohol and cannabis program content	Alcohol and cannabis program content implemented as a supplement to a sexual risk reduction intervention	No alcohol or cannabis program content	Sexual risk reduction intervention (no formal name)	Bryan et al. (2018)			○			High

Component type(s) that were tested[a]	General name of component or combination of components tested	Specific name of component or combination of components tested	Counter-factual	Program component studied with[b]	Manuscript reviewed	Outcome domains					Study quality rating[c]
						Sexual activity	Number of sexual partners	Contra-ceptive use	STIs or HIV	Preg-nancy	
Content	Cannabis program content	Cannabis program content implemented as a supplement to a sexual risk reduction intervention that also included alcohol program content	No cannabis program content	Sexual risk reduction intervention (no formal name) with alcohol content	Bryan et al. (2018)				○		High
Format	Computer format[d]	Computer delivery of a clinic-based brief intervention	In-person format	Brief intervention (no formal name)	Gryczynski et al. (2021)	○			●		High
Delivery mechanism; Content	Service learning	Service learning preparation, activities, and reflection	No service learning	None	Coyle et al. (2013)	○	○	○			High
Components with studies that did not meet study quality standards											
Format; Delivery mechanism	Text messaging	Booster text messages implemented as a supplement to Teen Outreach Program (TOP)[e]	No text messaging	Teen Outreach Program (TOP)	Bull et al. (2017)	n.a.	n.a.	n.a.	n.a.	n.a.	Low

Note: Empty cells indicate that the study or studies did not examine component impacts on measures within that outcome domain. Components are listed alphabetically within each evidence category.

● Evidence of a favorable finding with no evidence of adverse findings across one study reviewed.

○ No statistically significant findings across one study reviewed.

● Evidence of an unfavorable finding with no evidence of favorable findings across one study reviewed.

Refer to key in Appendix Table A.6 for more information about the program-level rating symbols in this table.

n.a. = not applicable; following the procedures specified in the review protocol, when the supporting impact study received a quality rating of low, the review team did not assess evidence of effectiveness for that study.

[a] A "component type" is the broadest level of categorization of program components (e.g., content, format, delivery mechanism).

[b] Components were tested as an addition to the programs listed in this column. In all but one study (Coyle et al., 2013), the counterfactual condition and the treatment condition received the same base intervention.

[c] The review team established ratings for the supporting impact studies by following prespecified criteria to assess study design, attrition, baseline equivalence, reassignment of sample members, and confounding factors. Appendix Table A.1 has a more detailed description of the review criteria.

[d] This contrast appears in two distinct rows because the study presents findings for each component compared to the other. See the box on page 11 of this brief for more information about the presentation of the counterfactual conditions in the studies of program components.

[e] Indicates components were implemented or tested with federal grant funding from FYSB or OPA.

Summaries of new program components with evidence of effectiveness

As shown in Table 3, studies of two of the eight components met TPPER standards for evidence of effectiveness. Below, we describe the two new program components with evidence of effectiveness, along with a summary of the research evidence on these programs. Refer to the box on page 11 for more information about how to interpret findings from studies of program components.

Booster sexual health counseling

Component summary – sexual health risk reduction counseling implemented as part of HORIZONS. As one component within the evidence-based teen pregnancy prevention program, HORIZONS, booster sexual health risk reduction counseling is a component designed to be used as a prevention maintenance intervention to support behaviors that prevent STIs and HIV and reduce incidence of STIs. Health educators delivered scripted, brief, and tailored counseling during a total of 18 telephone calls, each about 10 minutes long. The calls occurred every eight weeks during a three-year period after the conclusion of the in-person, clinic-based HORIZONS program. To inform the sexual health counseling content delivered during calls, participants conducted a risk appraisal that involved identifying and prioritizing their individual risk factors related to sexual risk behavior. Health educators used participants' prioritized risk factors to tailor their counseling for preventing STIs and HIV. The strategies reinforced concepts covered in the HORIZONS program, with the aim of reducing the risk factors prioritized by the participants.

Research summary. The booster sexual health risk reduction counseling was evaluated using a randomized controlled trial involving 701 young African American women, ages 14 to 20, who were recruited from three health clinics in Atlanta, Georgia. Participants were randomly assigned to receive the booster counseling phone calls or a control condition consisting of an identical dosage of general health promotion counseling phone calls. The general health promotion counseling calls focused on participants' nutrition and physical activity goals. Computer-assisted, self-administered interviews and self-collected vaginal swab specimens were collected at baseline and at six, 12, 18, 24, 30, and 36 months after the primary HORIZONS program ended. The study found that 36 months after the primary HORIZONS program ended (immediately after the phone calls ended), adolescents assigned to receive booster sexual health counseling phone calls reported a significantly higher proportion of condom use in the past 90 days (effect size = 2.2) and in the prior six months (effect size = 2.2) relative to adolescents assigned to receive general health promotion phone calls. Adolescents assigned to booster sexual health counseling also reported significantly fewer vaginal sex partners in the past six months (effect size = -2.2). The study showed no statistically significant impacts on any other assessed sexual behavior outcomes. The study met the review criteria for a high quality rating.

In-person format

Component summary – in-person delivery of a clinic-based brief intervention by nurse practitioners. A brief intervention (BI) designed to address substance use and sexual risk

behaviors among adolescents was delivered via an in-person format. This BI was designed to address substance use and sexual risk behaviors among adolescents; the curriculum included motivational and educational content targeting marijuana, alcohol, and risky sexual behaviors. The in-person format involved licensed nurse practitioners, who were first trained on BI techniques and then received additional booster sessions focused on the educational and interactive components of the curriculum delivery. Nurse practitioners delivered the intervention curriculum to participants in the context of a school-based health center (SBHC). The program was designed to be completed in a single session, but participants were encouraged to continue engaging with the SBHC about health behaviors covered in the BI. On average, the in-person delivery of the intervention took 17 minutes.

Research summary. The delivery format was evaluated using a randomized controlled trial involving 300 adolescents, ages 14 to 18, who were screened as having risky marijuana and/or alcohol use. Participants were recruited from two SBHCs in two urban public high schools and randomized to receive either a computer-delivered or a nurse practitioner–delivered brief, single-session, substance use and sexual risk intervention. Computer-assisted, self-administered interviews were conducted at baseline and at three and six months after the intervention ended. The study found that, for a subgroup of 254 adolescents who reported being sexually active at baseline, participants who received the nurse practitioner–delivered intervention were significantly more likely than those who received the computer-delivered intervention to have been tested for STIs within six months of receiving the intervention (effect size = 0.49). The study showed no statistically significant impacts on any other assessed sexual behavior outcomes. The study met the review criteria for a high quality rating.

INACTIVE PROGRAMS

Nine programs that previously had evidence of effectiveness were moved to the inactive programs list on the TPPER website (Table 4). The evidence for eight of these programs was based on data collected more than 20 years ago. Consequently, these findings are no longer considered eligible for TPPER assessments of program effectiveness. The ninth program was moved to the inactive list because the program is no longer being disseminated, as disclosed by the former distributor during outreach conversations for updates to the implementation information.

Implementation information and reviewed evidence for these programs are still available on the website. Inactive programs may become active in the future if new, eligible evidence of effectiveness is identified and if the program is still being disseminated for implementation.

Table 4. Programs moved to inactive list

Program name	Reason for moving to inactive
Aban Aya Youth Project	Program's only evidence is older than 20 years
Be Proud! Be Responsible!	Program's only evidence is older than 20 years
Be Proud! Be Responsible! Be Protective!	Program's only evidence is older than 20 years
¡Cuídate!	Program is not being disseminated
Draw the Line/Respect the Line	Program's only evidence is older than 20 years
FOCUS	Program's only evidence is older than 20 years
Project TALC	Program's only evidence is older than 20 years
SiHLE	Program's only evidence is older than 20 years
Teen Health Project	Program's only evidence is older than 20 years

REFERENCES

Barbee, A. P., Cunningham, M. R., Antle, B. F., & Langley, C. N. (2022). Impact of a relationship-based intervention, Love Notes, on teen pregnancy prevention. *Family Relations, 72*(5), 2569–2588. Advance online publication. https://doi.org/10.1111/fare.12798

Bauman, L. J., Watnick, D., Silver, E. J., Rivera, A., Sclafane, J. H., Rodgers, C. R. R., & Leu, C.-S. (2021). Reducing HIV/STI risk among adolescents aged 12 to 14 years: A randomized controlled trial of Project Prepared. *Prevention Science, 22*(8), 1023–1035.

Bryan, A. D., Magnan, R. E., Gillman, A. S., Yeater, E. A., Feldstein Ewing, S. W., Kong, A. S., & Schmiege, S. J. (2018). Effect of including alcohol and cannabis content in a sexual risk-reduction intervention on the incidence of sexually transmitted infections in adolescents: A cluster randomized clinical trial. *JAMA Pediatrics, 172*(4), e175621. https://doi.org/10.1001/jamapediatrics.2017.5621

Bull, S., Devine, S., Schmiege, S. J., Hammes, A., Pickard, L., & Shlay, J. C. (2017). Text messaging and teen sexual health behavior: Long-term follow-up of a cluster randomized trial. *Computers, Informatics, Nursing: CIN, 35*(11), 549–553. https://doi.org/10.1097/CIN.0000000000000404

Covington, R. D., Wood, R. G., & Goesling, B. (2019). *Focusing on the boys: The longer-term impacts of Wise Guys in Davenport, Iowa* (OPRE Report # 2019-97). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Coyle, K. K., Glassman, J. R., Franks, H. M., Campe, S. M., Denner, J., & Lepore, G. M. (2013). Interventions to reduce sexual risk behaviors among youth in alternative schools: A randomized controlled trial. *Journal of Adolescent Health, 53*(1), 68–78.

Cunningham, M. R., van Zyl, M. A., & Borders, K. W. (2016). *Evaluation of Love Notes and Reducing the Risk in Louisville, Kentucky* [Final evaluation report].

Estrada, Y., Lee, T. K., Huang, S., Tapia, M. I., Velazquez, M. R., Martinez, M. J., Pantin, H., Ocasio, M. A., Vidot, D. C., Molleda, L., Villamar, J., Stepanenko, B. A., Hendricks Brown, C., & Prado, G. (2017). Parent-centered prevention of risky behaviors among Hispanic youths in Florida. *American Journal of Public Health, 107*(4), e1–e7. https://doi.org/10.2105/AJPH.2017.303653

Estrada, Y., Lee, T. K., Wagstaff, R., Rojas, L. M., Tapia, M. I., Velázquez, M. R., Sardinas, K., Pantin, H., Sutton, M. Y., & Prado, G. (2019). eHealth Familias Unidas: Efficacy trial of an evidence-based intervention adapted for use on the internet with Hispanic families. *Prevention Science, 20*(1), 68–77. https://doi.org/10.1007/s11121-018-0905-6

Forrester, E., D. Aharpour, R. Cole, and J. Knab. "Core Components of Evidence-Based Teen Pregnancy Prevention Programs." Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Forthcoming.

Forrester, E., and R. Cole. "Core Components of Teen Pregnancy Prevention Programs." OPA Brief. Washington, DC: Office of Population Affairs, Office of the Assistant Secretary for Health, U.S. Department of Health and Human Services, 2022.

Floren, M., & Floren, K. (2023). *Impact of the Choosing the Best program on sexual activity among students in communities committed to abstinence education: Additional analysis findings with adjustment for clustering*. Practical Statistics, Choosing the Best Publishing.

Goesling, B. (2015). *Making sense of replication studies: Guidance for the field of teen pregnancy prevention research.* U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

Gryczynski, J., Mitchell, S. G., Schwartz, R. P., Dusek, K., O'Grady, K. E., Cowell, A. J., Barbosa, C., Barnosky, A., & DiClemente, C. C. (2021). Computer- vs. nurse practitioner-delivered brief intervention for adolescent marijuana, alcohol, and sex risk behaviors in school-based health centers. *Drug and Alcohol Dependence, 218,* 108423. https://doi.org/10.1016/j.drugalcdep.2020.108423

Kissinger, P. J., Green, J., Latimer, J., Schmidt, N., Ratnayake, A., Madkour, A. S., Clum, G., Wingood, G. M., DiClemente, R. J., & Johnson, C. (2023). Internet-delivered sexually transmitted infection and teen pregnancy prevention program: A randomized trial. *Sexually Transmitted Diseases,* 50(6), 329–335. https://doi.org/10.1097/olq.0000000000001784

Kottke, M. J., Sales, J. M., Goedken, P., Brown, J. L., Hatfield-Timajchy, K., Koumans, E. H., Hardin, J. W., Kraft, J. M., & Kourtis, A. P. (2023). 2gether: A clinic-based intervention to increase dual protection from sexually transmitted infections and pregnancy in young African American females. *Journal of Women's Health, 32*(1), 29–38.

Mogro-Wilson, C., Drake, A., Coman, E., Sanghavi, T., Martin-Peele, M., & Fifield, J. (2020). Increasing condom usage for African-American and Hispanic young fathers in a community based intervention. *Ethnicity & Health, 25*(3), 408–419.

Peskin, M. F., Coyle, K. K., Anderson, P. M., Laris, B. A., Glassman, J. R., Franks, H. M., Thiel, M. A., Potter, S.C., Unti, T., Edwards, S., Johnson-Baker, K., Cuccaro, P.M., Diamond, P., Markham, C.M., Shegog, R., Baumler, E.R., Gabay, E.K., Emery, S.T.. (2019). Replication of It's Your Game … Keep It Real! in Southeast Texas. *The Journal of Primary Prevention, 40*, 297–323.

Potter, S. C., Coyle, K. K., & Ayyaluru, S. (2023). *Your Move: A cluster randomized controlled trial of a blended learning sexual health program* [Unpublished manuscript].

Schnall, R., Kuhns, L. M., Pearson, C., Batey, D. S., Bruce, J., Hidalgo, M. A., Hirshfield, S., Janulis, P., Jia, H., Radix, A., Belkind, U., Rodriguez, R. G., & Garofalo, R. (2022). Efficacy of MyPEEPS Mobile, an HIV prevention intervention using mobile technology, on reducing sexual risk among same-sex attracted adolescent males: A randomized clinical trial. *JAMA Network Open, 5*(9), e2231853. https://doi.org/10.1001/jamanetworkopen.2022.31853

Schmiege, S. J., Magnan, R. E., Yeater, E. A., Ewing, S. W. F., & Bryan, A. D. (2021). Randomized trial to reduce risky sexual behavior among justice-involved adolescents. *American Journal of Preventive Medicine, 60*(1), 47–56.

Sharma, A., Mitchell, S. G., Nordeck, C. D., Schwartz, R. P., Dusek, K., O'Grady, K. E., & Gryczynski, J. (2022). Sexually transmitted infection testing after brief intervention for risk behaviors in school-based health centers. *Journal of Adolescent Health, 70*(4), 577–583. https://doi.org/10.1016/j.jadohealth.2021.11.013

Ybarra, M., Rosario, M., Saewyc, E., Goodenow, C., & Dunsiger, S. (2023a). One-year follow-up after a pregnancy prevention intervention for LGB+ teens: An RCT. *Pediatrics, 151*(4). https://doi.org/10.1542/peds.2022-059172

Ybarra, M. L., Saewyc, E., Rosario, M., & Dunsiger, S. (2023b). Subgroup analyses of Girl2Girl, a text messaging–based teen pregnancy prevention program for sexual minority girls: Results from a national RCT. *Prevention Science, 24,* 292–299. https://doi.org/10.1007/s11121-023-01493-6

APPENDIX A

Review procedures

For this update to the review's findings, the review team followed procedures similar to those used for earlier rounds. In May 2023, the team released a public call for studies that requested new research for consideration. The team also identified studies through a comprehensive literature search that entailed keyword searches of electronic databases, manual searches of relevant academic journals and professional conference proceedings, and citation reviews of recently published literature reviews and meta-analyses. The team screened the resulting studies against prespecified eligibility criteria. Resource constraints limited the number of studies reviewed. We used a prioritization process (described in the review protocol) to determine which studies to include in this round of the review.

There was one change to the eligibility criterion for this round: In addition to focusing on programs, the review focused on the impact of well-defined components or combinations of components for programs intended to reduce rates of unintended teen pregnancy. To be eligible for review under Version 7.0 of the TPPER protocol, a component must be (1) a clearly defined practice, procedure, policy, support, or organizational structure, potentially with documented steps for implementation with fidelity to facilitate replication, and (2) capable of being implemented independently, in conjunction with, or integrated into a teen pregnancy prevention intervention.

For studies that met the eligibility criteria, the team first assessed the quality and execution of each study's research design (Table 7). The reviewers assigned each study a quality rating of high, moderate, or low according to the risk of bias in the study's impact findings. The higher the risk of bias, the lower the confidence in the credibility of the findings. A detailed description of these ratings is in the Version 7.0 review protocol, available online at the TPPER website.

For the subset of studies that achieved a moderate or high quality rating, the review team extracted information on the features of the program or component being tested, evaluation setting, study sample, and research design as well as detailed information on the program impact estimates. The review team then identified programs and components with evidence of effectiveness, which was defined as having a statistically significant favorable impact (and no adverse effects) on at least one priority outcome measured for either the full analytic sample or a subgroup defined by (1) gender or (2) sexual experience at baseline. The priority outcomes included sexual activity, number of sexual partners, contraceptive use, STIs or HIV (including getting tested), and pregnancy.

Before the start of the current round, we updated the study quality review criteria. The updates are described in this summary of changes document. We then applied the updated criteria when we reviewed the studies of the 17 programs described in this brief. The only change to the study quality review criteria was increased transparency in TPPER review approaches. The current Version 7.0 protocol documents (1) what the TPPER will look for when a study presents findings

from a longitudinal data analysis and/or a statistical technique used to equate treatment and comparison groups at baseline and (2) the circumstances in which the TPPER considers it acceptable for an analysis to use a difference-in-differences approach to statistically adjust for baseline differences. More details on the review process and criteria are available on the TPPER website.

Table A.1. Summary of study quality ratings

Criteria category	Features of studies with the high study rating	Features of studies with the moderate study rating
Study design	Random or functionally random assignment	Random assignment design with high attrition or reassignment; quasi-experimental design with a comparison group
Attrition	Random assignment studies that do not exceed What Works Clearinghouse standards for overall and differential attrition (cautious assumption)	Random assignment studies that exceed What Works Clearinghouse attrition standards; attrition is not assessed in quasi-experimental designs
Baseline equivalence	Not assessed; samples assumed to be equivalent by virtue of random assignment and low levels of sample attrition	The equivalence of the research groups is demonstrated at baseline, and systematically adjusted for in impact analyses
Reassignment	Analysis based on original assignment to research groups	They are not assessed, given the baseline equivalence requirement in the box just above
Confounding factors	At least two subjects or groups in each research group, and no systematic differences in data collection methods	There were at least two subjects or groups in each research group, and no systematic differences in data collection methods

Note: Studies that do not achieve the high or moderate rating are given a rating of low. Refer to the Review Protocol 7.0 for more details on applying the study quality ratings.

APPENDIX B

PREVIOUSLY REVIEWED PROGRAMS WITH NEWLY REVIEWED FINDINGS

In addition to identifying and reviewing studies of new programs and program components, we sought to identify and assess any new findings for programs examined in earlier rounds of the review. As part of this update, we identified seven new publications (manuscripts) on seven programs examined in previous rounds of the review. Three publications were new studies and four presented data (for instance, long-term follow-up data, an analysis of subgroups, or a published version of a previously unpublished report that had different methodological decisions) for the same sample as another publication we previously reviewed. For each of the seven programs, the newly reviewed findings did not show evidence of effectiveness. Six of the seven new publications were reports or journal articles that presented findings from studies funded through FYSB or OPA.

Evidence for previously reviewed programs

Five of the previously reviewed programs are active evidence-based programs, and this assessment did not change that status. In other words, during a previous round of the TPPER, they met the review criteria for evidence of effectiveness—that is, having a rating of high or moderate and showing evidence of a favorable, statistically significant impact on at least one outcome of interest that reflects sexual behavior or consequences of sexual behavior. The new publications for these five programs did not provide evidence of effectiveness. The other two previously reviewed programs were not previously considered evidence-based programs, and the new publications for these two programs did not provide evidence of effectiveness to change that assessment.

Evidence for programs previously determined to be evidence-based

- Four programs—(1) Familias Unidas, (2) Girl2Girl, (3) Love Notes (for ages 14 to 24), and (4) Reducing the Risk—have existing evidence of effectiveness. The findings presented in the new publications did not meet study quality standards, so the assessments of these programs did not change; they remain evidence-based.[4]
- One program, It's Your Game … Keep It Real!, also has existing evidence of effectiveness. The newly reviewed publication met study quality standards but presented null findings, meaning none of the findings were statistically significant; therefore, this additional study did not change the conclusions of the previous reviews. The program remains evidence-based.

[4] Because the TPPER uses the highest rating across manuscripts associated with a single study to describe the evidence for that study, Table A.2 provides the rating of the findings reported in the manuscript we first reviewed (which met evidence standards) and the rating for the findings reported in the manuscripts reviewed during this round.

Evidence for programs not previously determined to be evidence-based

- The other two programs—(1) Be yoU, Talented, Informed, Fearless, Uncompromised, and Loved (BUtiful) and (2) Wise Guys—were not previously determined to have evidence of effectiveness. The new publication for BUtiful had a low study quality rating; therefore, the review team did not assess evidence of effectiveness from that study. The new publication for Wise Guys presented null findings and thus did not meet review criteria for evidence of effectiveness. Therefore, these programs are still considered non-evidence-based.

In the next two sections, we provide the study-level and program-level evidence for the previously reviewed programs with newly reviewed findings.

Study-level findings for previously reviewed programs

In Table A.2, and on the TPPER website, we summarize the findings from newly identified publications that had moderate or high quality evidence. The strength of the body of evidence from each study is represented by both the color and quantity of the icons (refer to the key in Table A.3 for icon descriptions). As shown in Table A.2, the two studies that met TPPER evidence standards each had null findings in the sexual activity outcome domain and did not present impacts on outcomes in the other domains. Therefore, this newly reviewed evidence does not materially affect the existing rating for these programs. The other five studies did not meet the study quality criteria and, therefore, their evidence was not assessed. In Table A.4, the findings, the evidence rating, and the reason for the rating (if the rating is low) are described for each of the seven previously reviewed programs.

Table A.2. Individual study ratings by outcome domain for newly reviewed publications of previously reviewed programs

Program and citation	Outcome domains					
	Sexual activity	Number of sexual partners	Contra-ceptive use	STIs or HIV	Pregnancy	Study quality rating[a]
It's Your Game … Keep It Real!						
Peskin et al. (2019)[b]	○					Moderate
Wise Guys						
Covington et al. (2019)[b]	○					High

Note: Refer to the key in Table A.3 for icon descriptions. Empty cells indicate that the study did not examine program impacts on measures within that outcome domain. Programs are listed alphabetically within each evidence category.

[a] The review team established quality ratings for the publications (manuscripts) listed in this table by following prespecified criteria to assess study design, attrition, baseline equivalence, reassignment of sample members, and confounding factors. Table A.1 has a description of these criteria.

[b] Indicates that the publication presents findings from studies of programs that were implemented or tested with grant funding from FYSB or OPA.

Table A.3. Key 1: Strength of the findings from individual studies of programs

Symbol	Description
	Favorable findings: Two or more favorable impacts and no unfavorable impacts, regardless of null findings
	Potentially favorable findings: At least one favorable impact and no unfavorable impacts, regardless of null findings
	Indeterminate findings: Uniformly null findings
	Conflicting findings: At least one favorable and at least one unfavorable impact, regardless of null findings
	Potentially unfavorable findings: At least one unfavorable impact and no favorable impacts, regardless of null findings
	Unfavorable findings: Two or more unfavorable impacts and no favorable impacts, regardless of null findings

Table A.4. Descriptions of reviewed publications of studies that did not change assessments of previously reviewed programs

Program/publication	Quality rating for the findings from this publication[a]	Description
Evidence-based programs: Programs with prior evidence of effectiveness; new publications did not change that assessment		
Familias Unidas		
Estrada et al. (2017)	Low	Presents findings from a new individual-level randomized controlled trial assessing the impact of Familias Unidas, a parent-centered, family-based intervention focused on reducing alcohol use, drug use, and unsafe sexual behavior among Hispanic adolescents. The findings received a low quality rating because the publication only presents findings for endogenous subgroups—adolescents who reported being sexually active at any of the study's follow-up time points.
Girl2Girl		
Ybarra et al. (2023a)	Low	This publication presents 12-month follow-up findings from an individual-level randomized controlled trial for which the immediate post-intervention findings were already reviewed (Ybarra et al. [2021]. Girl2Girl is a teen pregnancy prevention, text-messaging-based program designed for cisgender female youth ages 14–18. The sample in this paper is limited to youth sexually active at baseline. The findings from this publication received a low quality rating because sexual experience was included as a time-varying covariate in each model, and sexual experience post-baseline is endogenous. Analyses that adjust for endogenous covariates do not meet TPPER evidence standards.
Ybarra et al. (2023b)	Low	This publication presents 12-month follow-up findings from an individual-level randomized controlled trial for which the immediate post-intervention findings for the full sample were already reviewed (Ybarra et al. [2021]). This publication assesses the impact of Girl2Girl on a subgroup sample of LGB+ girls based on their sexual experience at baseline. The findings from this publication received a low quality rating because sexual experience was included as a time-varying covariate. Such analyses that adjust for endogenous covariates do not meet TPPER evidence standards.
It's Your Game … Keep It Real!		
Peskin et al. (2019)	Moderate	This publication presents findings from a new cluster-level randomized controlled trial assessing the impact of It's Your Game … Keep It Real!—an HIV, STI, and teen pregnancy prevention middle school program. The study found no evidence of favorable program effects on participants' initiation of vaginal or oral sex at 24 months after the program.

Program/publication	Quality rating for the findings from this publication[a]	Description
Love Notes (for ages 14-24)		
Barbee et al. (2022)	Low	This publication presents 12-month follow-up findings from a three-arm cluster-level randomized controlled trial for which the 3- and 6-month follow-up data were already reviewed (Cunningham et al. 2016). This publication assesses the impact of Love Notes (for ages 14-24), a trauma-informed curriculum that aims to educate youth about healthy relationships and reduce incidence of teen dating violence and unprotected sex, one-year after the end of the intervention. The follow-up analysis presented in this publication received a low study quality rating because the analysis did not account for the differences in probabilities of study condition assignment, and the analytic sample was not equivalent on the race variable.
Reducing the Risk		
Barbee et al. (2022)	Low	This publication presents findings from a 12-month follow-up on a three-arm cluster-level randomized controlled trial for which the 3- and 6-month follow-up data were already reviewed (Cunningham et al. [2016]). This publication assesses the impact of Reducing the Risk, an intervention covering risk behaviors, abstinence, HIV and other STIs, and skills development, one-year after the end of the intervention. The follow-up analysis presented in this publication received a low study quality rating because analysis did not account for the differences in probabilities of study condition assignment, and the analytic sample was not equivalent on the race variable.
Non-evidence-based program: Programs with no prior evidence of effectiveness; new publications did not change that assessment		
Be yoU, Talented, Informed, Fearless, Uncompromised, and Loved (BUtiful)		
Kissinger et al. (2023)	Low	This publication presents 12-month follow-up findings from an individual-level randomized controlled trial for which the 6-month follow-up data were already reviewed (Kissinger et al. [2015]) assessing the impact of BUtiful, a multicomponent, asynchronous online sexual health education intervention for women. The findings for the follow-up analysis presented in this publication received a low study quality rating because there was high attrition at this follow-up and the analysis did not demonstrate baseline equivalence.
Wise Guys		
Covington et al. (2019)	High	This publication presents findings from a new individual-level randomized controlled trial assessing the impact of Wise Guys, a sexual health and relationship curriculum designed for adolescent males. The findings reported in this publication found no evidence of favorable program effects on participants' initiation of sexual intercourse at 24 months after the program.

Program/publication	Quality rating for the findings from this publication[a]	Description
Evidence-based programs: Programs with prior evidence of effectiveness; new publications did not change that assessment		
Familias Unidas		
Estrada et al. (2017)	Low	Presents findings from a new individual-level randomized controlled trial assessing the impact of Familias Unidas, a parent-centered, family-based intervention focused on reducing alcohol use, drug use, and unsafe sexual behavior among Hispanic adolescents. The findings received a low quality rating because the publication only presents findings for endogenous subgroups—adolescents who reported being sexually active at any of the study's follow-up time points.
Girl2Girl		
Ybarra et al. (2023a)	Low	This publication presents 12-month follow-up findings from an individual-level randomized controlled trial for which the immediate post-intervention findings were already reviewed (Ybarra et al. [2021]. Girl2Girl is a teen pregnancy prevention, text-messaging-based program designed for cisgender female youth ages 14–18. The sample in this paper is limited to youth sexually active at baseline. The findings from this publication received a low quality rating because sexual experience was included as a time-varying covariate in each model, and sexual experience post-baseline is endogenous. Analyses that adjust for endogenous covariates do not meet TPPER evidence standards.
Ybarra et al. (2023b)	Low	This publication presents 12-month follow-up findings from an individual-level randomized controlled trial for which the immediate post-intervention findings for the full sample were already reviewed (Ybarra et al. [2021]). This publication assesses the impact of Girl2Girl on a subgroup sample of LGB+ girls based on their sexual experience at baseline. The findings from this publication received a low quality rating because sexual experience was included as a time-varying covariate. Such analyses that adjust for endogenous covariates do not meet TPPER evidence standards.
It's Your Game … Keep It Real!		
Peskin et al. (2019)	Moderate	This publication presents findings from a new cluster-level randomized controlled trial assessing the impact of It's Your Game … Keep It Real!—an HIV, STI, and teen pregnancy prevention middle school program. The study found no evidence of favorable program effects on participants' initiation of vaginal or oral sex at 24 months after the program.

Program/publication	Quality rating for the findings from this publication[a]	Description
Love Notes (for ages 14-24)		
Barbee et al. (2022)	Low	This publication presents 12-month follow-up findings from a three-arm cluster-level randomized controlled trial for which the 3- and 6-month follow-up data were already reviewed (Cunningham et al. 2016). This publication assesses the impact of Love Notes (for ages 14-24), a trauma-informed curriculum that aims to educate youth about healthy relationships and reduce incidence of teen dating violence and unprotected sex, one-year after the end of the intervention. The follow-up analysis presented in this publication received a low study quality rating because the analysis did not account for the differences in probabilities of study condition assignment, and the analytic sample was not equivalent on the race variable.
Reducing the Risk		
Barbee et al. (2022)	Low	This publication presents findings from a 12-month follow-up on a three-arm cluster-level randomized controlled trial for which the 3- and 6-month follow-up data were already reviewed (Cunningham et al. [2016]). This publication assesses the impact of Reducing the Risk, an intervention covering risk behaviors, abstinence, HIV and other STIs, and skills development, one-year after the end of the intervention. The follow-up analysis presented in this publication received a low study quality rating because analysis did not account for the differences in probabilities of study condition assignment, and the analytic sample was not equivalent on the race variable.
Non-evidence-based program: Programs with no prior evidence of effectiveness; new publications did not change that assessment		
Be yoU, Talented, Informed, Fearless, Uncompromised, and Loved (BUtiful)		
Kissinger et al. (2023)	Low	This publication presents 12-month follow-up findings from an individual-level randomized controlled trial for which the 6-month follow-up data were already reviewed (Kissinger et al. [2015]) assessing the impact of BUtiful, a multicomponent, asynchronous online sexual health education intervention for women. The findings for the follow-up analysis presented in this publication received a low study quality rating because there was high attrition at this follow-up and the analysis did not demonstrate baseline equivalence.
Wise Guys		
Covington et al. (2019)	High	This publication presents findings from a new individual-level randomized controlled trial assessing the impact of Wise Guys, a sexual health and relationship curriculum designed for adolescent males. The findings reported in this publication found no evidence of favorable program effects on participants' initiation of sexual intercourse at 24 months after the program.

[a] The review team established quality ratings for the publications listed in this table by following prespecified criteria to assess study design, attrition, baseline equivalence, reassignment of sample members, and confounding factors (See Table A.1).

Findings by program: program-level findings for previously reviewed programs with new publications reviewed in this round

Each of the seven previously reviewed programs have more than one supporting impact study contributing to the summary of their evidence. Therefore, we present findings at the program level in Table A.5 below, which pools the findings across all studies (including eligible studies reviewed this round and in previous rounds) for a given program. Program-level ratings (described in the key in Table A.6) reflect the strength of findings from one or more studies that present findings on the same or different eligible outcomes. The program-level ratings are presented separately for each of the outcome domains on which a program has been evaluated.

In Table A.5, and on the TPPER website, we summarize the available program-level evidence (favorable, potentially favorable, indeterminate, conflicting, potentially unfavorable, and unfavorable) across studies on a particular program in each of five outcome domains. The strength of the body of evidence behind all studies that meet evidence standards for a given program is represented by the color, quantity, and size of the icons (Table A.6). Visitors to the review website may use this information to compare programs that match their outcome domain of interest, the desired program type, population, age range, or implementation setting. Table A.5 only includes program-level findings for the programs assessed in both the current and previous rounds of the review, while the website includes this information for all evidence-based programs.

As shown in Table A.5, although we reviewed new research on five active evidence-based TPP programs, the new research did not provide additional evidence of effectiveness for those programs, either because the study was rated low quality or the study had no statistically significant findings. Specifically, the evidence-based programs—(1) Familias Unidas, (2) Girl2Girl, (3) It's Your Game … Keep It Real!, (4) Love Notes (for ages 14 to 24), and (5) Reducing the Risk—maintained their previous program-level ratings across each outcome domain because the new publications (described in the section on study-level findings above) did not offer new evidence of effectiveness. The previously reviewed programs that were not evidence-based—(1) BUtiful and (2) Wise Guys—maintained their previous program-level ratings because these programs have not yet met TPPER standards for evidence of effectiveness.

Table A.5. Evidence by outcome domain for programs with studies assessed in the current and previous rounds of the review; the newly reviewed findings did not change the overall ratings for any program in this table

Program	Outcome domains				
	Sexual activity	Number of sexual partners	Contraceptive use	STIs or HIV	Pregnancy
Programs with evidence of effectiveness					
Familias Unidas					
Girl2Girl					
It's Your Game ... Keep It Real!					
Love Notes (for ages 14-24)					
Reducing the Risk					
Programs with no evidence of effectiveness					
Be yoU, Talented, Informed, Fearless, Uncompromised, and Loved (BUtiful)					
Wise Guys					

Note: Refer to the key in Table A.6 for icon descriptions. An empty cell means the study did not examine program impacts on measures within that outcome domain. Programs are listed alphabetically within each evidence category.

Table A.6. Key 2: Strength of the body of evidence behind a given program or component

Symbol	Description	Criteria
	Favorable evidence: Strong evidence of favorable findings with no overriding contrary evidence	**1.** Two or more studies show favorable findings and **2.** No studies have inconsistent (both favorable and unfavorable) findings, potentially unfavorable findings, or unfavorable findings.
	Potentially favorable evidence: Evidence of a favorable finding with no overriding contrary evidence	**1.** At least one study shows favorable findings or potentially favorable findings and **2.** No studies have inconsistent findings, potentially unfavorable findings, or unfavorable findings
	Indeterminate evidence: No affirmative evidence of findings; uniformly null findings	**1.** All of the studies show indeterminate evidence.
	Conflicting evidence: Evidence of both favorable and unfavorable findings	**1.** At least one study shows inconsistent findings, or **2.** At least one study shows favorable findings or potentially favorable findings, and at least one other study shows unfavorable findings or potentially unfavorable findings.
	Potentially unfavorable evidence: Evidence of unfavorable findings with no overriding contrary evidence	**1.** At least one study shows unfavorable findings or potentially unfavorable findings and **2.** No studies have inconsistent findings, potentially favorable findings, or favorable findings
	Unfavorable evidence: Strong evidence of unfavorable findings with no overriding contrary evidence	**1.** Two or more studies show unfavorable findings and **2.** No studies have inconsistent findings, potentially favorable findings, or favorable findings

Symbol	Size	Indication
Applies to all program ratings		1 study with a moderate or high study rating that examines outcomes in the domain
Applies to all program ratings		2–4 studies with moderate or high study ratings that examine outcomes in the domain
Applies to all program ratings		5 or more studies with moderate or high study ratings that examine outcomes in the domain