



PROGRAM EVALUATION

Second Step for Middle School

Purpose of This Guide

This guide is written specifically for people who want to evaluate a school's or district's implementation of the Second Step Middle School Program but aren't trained in program evaluation and aren't working with a professional evaluator.

This isn't a general guide to evaluating school-based programs—it's written specifically with Second Step in mind.

Why Evaluate?

People evaluate their use of Second Step for a variety of reasons. In general the goal is to show funders that the resources put into the program are paying off. Another important audience is parents and community members.

Many people choose to evaluate the program to see how it's working. Evaluation evidence can increase staff motivation and commitment to implementing the program fully and well. Evaluation can also help schools see how implementation might be affecting outcomes and how it might be improved to ensure students are benefiting fully. In addition evaluation is useful for tracking progress toward desired program goals and outcomes over time.

Evaluating Implementation

What Am I Evaluating?

One of the keys to successful, effective evaluation is to be sure you know just what you're evaluating. Every school and district purchases the same Second Step Program. However, what students actually receive can vary widely. You can make your Second Step evaluation more powerful and useful by looking at how the program is being implemented in your school or district. Remember, you're evaluating the

intervention your students actually get and, depending on implementation, that might be more or less like the exact program you purchased.

What Information Should I Gather?

What does a school need to know to be able to include implementation in its evaluation? Assessing implementation primarily means gathering information on how the program is being taught in your setting or settings.

- At the school or district level, how many students are receiving Second Step lessons? All? Only certain grades or classrooms?
- At the school level, what else is being done outside formal lessons to reinforce Second Step skills and concepts, both in the classroom and throughout the school?
- At the classroom level, are all the lessons being taught? If not, how many are? Are lessons being taught the way they're written, or are they being changed significantly?

What's Implementation Fidelity?

Surveying staff on how the program is being taught can also go beyond examining how many students are receiving how many lessons. Implementation evaluation can also look at the "fidelity" of implementation. Fidelity basically means the extent to which the program is taught as written.

A full implementation ideally means students are receiving all the lessons in order and all the content in each lesson. For a variety of reasons, staff sometimes only teach parts of lessons and skip others, teach lessons out of order, or change some of the content. These are all examples of low fidelity. Obviously it's possible to change lessons in ways that don't harm or might even improve outcomes, but it's also possible to change lessons in ways that reduce program effectiveness. Committee for Children recommends implementing the program with as much fidelity as possible. It can be useful in an evaluation to know the fidelity with which the program was taught.



Types of Evaluation Design

It might be helpful to think about your Second Step evaluation as falling somewhere along a spectrum of evaluation rigor. The most rigorous approach is an experimental design, in the middle is what's called quasi-experimental design, and the least rigorous approach is a non-experimental design. Each of these designs and their pros and cons are described below.

Experimental Design

One of the main challenges in program evaluation is determining whether any effects you find were in fact caused by the program you're evaluating. In any given school, Second Step is only one of many factors affecting students' attitudes and behaviors. The purpose of an experimental design is to increase your confidence that changes you find in students were caused by their exposure to Second Step.

This is primarily accomplished through random assignment. Random assignment means you determine which students will be involved in the study (your study population), and each of those students has an equal chance of either being taught the program or not. Random assignment is a powerful way to create two groups that are as likely as possible not to be significantly different. This goes a long way toward ruling out differences in outcomes being due to initial differences in the students being studied.

For complicated technical reasons, random assignment for evaluating a program like Second Step requires assigning entire schools to either implement the program or not (the ones that don't implement serve as non-intervention controls). In addition, for statistical reasons, a large number of schools must be involved in the evaluation. Scientifically valid experimental design evaluations of Second Step commonly involve 30 to 60 or more schools. A study this large is typically not feasible for a school district, and since experimental design requires randomizing entire schools, this approach can't be done by an individual school.

Quasi-Experimental Design

Quasi-experimental designs are a way to try to assess program effects when random assignment isn't possible. Rather than a randomly selected control group, a quasi-experimental design includes a comparison group. Comparison groups are made up of students who are not receiving the program. The key to creating a good comparison group is attempting to match the students as closely as possible to those receiving Second Step lessons. The more alike the two groups are, the more useful the comparison group data will be. The most common way to match comparison group students (or classrooms or schools) to those getting Second Step lessons is by using demographics, such as age, race or ethnicity, gender, income, etc.

The drawback to the quasi-experimental approach is you ultimately have less certainty that the students in the two groups you're comparing are alike to begin with than with random assignment, and differences between the two groups that don't have to do with Second Step may be part of the cause of differences you find in outcomes. However, this approach is a reasonable way to increase the strength of an evaluation.

Non-Experimental Design

A non-experimental design means gathering data on children who receive Second Step only, without any control or comparison students involved. This approach is often the most feasible for many schools and districts. Just keep in mind that it can't tell you whether any outcomes you find were actually caused by Second Step. It may be that Second Step is causing the changes you find, or it could be that schools using Second Step are also doing other things that benefit children and cause the changes you're finding. The clear advantage of not including control or comparison groups in your evaluation is that it's simpler and relatively inexpensive.

The primary approach used in non-experimental Second Step evaluation is to collect data before and after the program is implemented. This information is often called



pre- and post-test data. Getting this information typically involves surveying students and/or staff in the fall and again in the spring.

Although it's difficult to know how much of the change (positive or negative) from fall to spring was caused by Second Step, there are ways to make this evaluation approach stronger and more informative. Student behavior typically changes from the beginning to the end of the school year, regardless of what programs you're using. Students often start the school year out on their best behavior, but by the end of the year their behavior can look worse—even if you implement the program and it's working. It may be that students are having more conflicts and problems by the end of the year, but without Second Step lessons, those increases would have been much larger.

Strengthening Your Non-Experimental Evaluation

One way to tease out these types of effects and strengthen a simple pre/post evaluation is to collect data across multiple years. It can be particularly useful, once a fall baseline is established, to collect data each spring. It often takes time for staff to become familiar with the program, so implementation quality can improve over time, yielding better outcomes when the program has been in place longer. More importantly, tracking data across multiple years allows you to see the cumulative effect of students receiving a larger dose of the program. Second Step isn't intended as a one-year intervention. It's carefully designed so each year's lessons build on those that came before. Collecting data on outcomes across multiple years allows you to capture that growth.

A final way to strengthen a non-experimental approach to evaluation is to examine implementation. In some schools and districts, implementation will vary—some students will get more lessons than others, some staff will implement the lessons more fully than others, and some staff will reinforce skills more than others. If you're collecting data from staff on implementation, you may be able to compare outcomes for students who received different amounts, or doses, of the program. If students who received more lessons or more reinforcement show better outcomes, that can help you see how to increase outcomes for more students.

Using School Data Evaluation

Schools collect data as part of their everyday operations, and the most commonly used school data is probably discipline referral data. Many schools look at their disciplinary referrals over time as a way to see whether implementing the Second Step Program has resulted in fewer problem behaviors. One of the advantages of this approach is that schools can often compare the number of referrals for the year before they implemented the program to the number once Second Step has been in place.

It's also possible to track referrals over time to see whether the program results in fewer students having behavioral problems once it's been in place for multiple years. Check the Evaluation Design section for information on how different evaluation designs affect the strength of the connection between Second Step and any outcomes you find.

Although it's possible to look at other types of school data for evaluation purposes, disciplinary referrals are the most common and safest source of information on Second Step outcomes. Things like attendance, grades, and test scores can be affected by the program, but its effect on those outcomes is less direct and can be harder to see.

Using Evaluation Findings

Positive Outcomes

Congratulations! Your evaluation has shown that your Second Step implementation has improved outcomes for students. This is the time to ensure that your school or district continues to teach the program and supports what students are learning in Second Step lessons throughout the school day and school environment. Remember that ongoing support for the program by building leaders has been shown to be the number one factor that drives continued successful implementation over time. Share the good news with school staff, district staff, parents, and the community, so your efforts continue to be sustained and supported.



Poor Outcomes

With No Implementation Evaluation

If your evaluation suggests students aren't benefiting sufficiently from Second Step, a natural place to look for causes is implementation. As discussed in the Evaluating Implementation section, how the program is implemented is very important and has been shown to affect outcomes. If you haven't examined Second Step implementation as part of your evaluation, doing so may provide you with ideas for how to improve the effect on students. See Outcome Measures for tools you can use to examine Second Step implementation in your school or district.

With No Implementation Evaluation

If your evaluation included implementation, then poor outcomes indicate that it's important to look closely at how the program is being implemented to see where there's room for improvement that may increase program effects. Keep in mind that high-quality Second Step implementation goes beyond teaching the lessons. Just like with academics, Second Step skills have to be reinforced and practiced in order to be mastered. Look for ways staff can cue students to use Second Step skills throughout the school day and school environment, and find ways to reinforce students' skill use.

If it appears that Second Step implementation in your setting has been done well, it can be harder to know where to turn if you're not finding sufficiently positive outcomes from your evaluation. Keep in mind that a truly rigorous evaluation requires random assignment of a large number of schools, and that quasi- or non-experimental evaluations can make it hard to separate Second Step effects from other factors. Also recall that positive program outcomes may be lost in a one-year pre/post evaluation, because behaviors typically worsen from fall to spring. A lack of findings may result from changes in student behavior throughout the school year, despite positive program effects.

If your one-year evaluation produces disappointing results, remember that the program is designed to have a cumulative effect across multiple years, and that teaching it, like anything else, takes time to master. A one-year evaluation does not necessarily capture program effects well,

and it may be that data collected across more than one year will tell a different and more positive story.

Outcome Measures

It's important to choose carefully developed and tested tools for your Second Step Program evaluation. The basic approach to looking at data from surveys is to compare averages across surveys administered at different times. The following are outcome measures we recommend.

PANORAMA EDUCATION

The Panorama Education suite of surveys covers a wide range of topics in addition to social-emotional learning topics, including school- and classroom-level student experiences, school climate, and student-teacher relationships. Surveys are available for students in grades 3 through 12 as well as for teachers and parents. The company will compile survey results into a user-friendly data dashboard for a fee. Schools can also integrate their own data, such as on discipline referrals, into the Panorama system. All Panorama Education surveys are available online free to schools.

Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaire is a brief behavioral screening questionnaire for use with 3- to 16-year-olds. It asks about 25 attributes, some positive and some negative, on five different scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior.

The Social-Emotional Assets and Resilience Scales (SEARS)

The SEARS is a strengths-based measure designed to assess students' assets and resilience, taking into account problem-solving skills, interpersonal skills, the ability to make and maintain friendships, the ability to cope with adversity, and the ability to be optimistic when faced with adversity. The SEARS includes surveys for students in grades 3 through 6 and 7 through 12 as well as teacher and parent surveys. The SEARS surveys are available in paper form or online for a fee, and scoring and data reporting services are available for a fee.