

Purpose of This Guide

This guide is written specifically for people who want to evaluate an early learning center's implementation of the *Second Step* program, but are not trained in program evaluation and are not working with a professional evaluator.

This isn't a general guide to evaluating early learning programs—it's written specifically with the *Second Step* program in mind.

Why Evaluate?

People evaluate their use of the *Second Step* program for a variety of reasons. In general the goal is to show that the resources put into the program are paying off, so one of the most common audiences for evaluations is funders. 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 early learning centers see how implementation might be affecting outcomes and how it might be improved to ensure students are benefitting fully from the program. Evaluation is also 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 early learning center purchases the same *Second Step* program, but 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 center. Remember, you're evaluating the intervention your students actually get, which, depending on implementation, might be more or less like the exact program you purchased.

What Information Should I Gather?

What would an early learning center need to know to be able to include implementation in its evaluation? Assessing implementation primarily means gathering information on how the *Second Step* program is being taught in your setting or settings. In particular:

- How many children are receiving the *Second Step* program
- How many of the Weekly Theme activities are being taught
- To what extent the program is being taught as written
- What else is being done outside of the Weekly Theme activities to reinforce *Second Step* skills, both in the classroom and throughout the early learning center

How Do I Gather It?

Collecting data on what children are receiving typically involves having staff complete a simple survey that covers the questions listed above. There are two surveys for this purpose available on SecondStep.org and listed below. Surveys should be filled out by the relevant staff.

- The Implementation Survey for the *Second Step* early learning program collects information about program implementation as experienced by those teaching and reinforcing program skills. It should be completed by staff who teach Weekly Theme activities and/or reinforce program content. You can use this survey to assess the implementation process during the year and as a record of implementation experiences at the end of the year.
- The early learning Weekly Theme Completion Checklist is a one-page survey that covers the Weekly Themes in each unit. The survey allows you to gather data on how many of the themes were taught, how much and what kind of changes were made, and how much reinforcement was done for each theme.

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 Weekly Themes. 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 Weekly Themes in order and all the content in each Weekly Theme. For a variety of reasons, staff sometimes only teach parts of Weekly Themes and skip others, teach them out of order, or change some of the content. These are all examples of low fidelity.

Obviously it's possible to change the Weekly Themes in ways that don't harm or might even improve outcomes, but it's also possible to change them in ways that reduce program effectiveness. Committee for Children recommends implementing the program with as much fidelity as possible, and 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 is 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 early learning center, the *Second Step* program 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 the *Second Step* program.

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 the *Second Step* curriculum requires assigning entire centers 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 centers must be involved in the evaluation. Scientifically valid experimental design evaluations of the *Second Step* program commonly involve thirty to sixty or more centers in one study. A study this large is typically not feasible for an early learning center to carry out, and since experimental

design requires randomizing entire centers, this approach can't be done by an individual center.

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* Weekly Themes. 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 centers) to those getting *Second Step* Weekly Themes 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 the *Second Step* program 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 the *Second Step* program only, without any control or comparison children involved. This approach is often the most feasible for many early learning centers. Just keep in mind that it can't tell you whether any outcomes you find were actually caused by the *Second Step* program. This method gives you relational findings that tell you how related your outcomes are to the *Second Step* program, but not what is causing that relationship. For example, it may be that the *Second Step* program is causing the changes you find, or it could be that early learning centers using the *Second Step* program 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 the *Second Step* program, there are ways to make this evaluation approach stronger and more informative. Keep in mind that student behavior typically changes from the beginning to the end of the school year, regardless of what programs you're implementing. The simple pre/post evaluation approach can be tricky, because students often start the school year out on their best behavior, but by the end of the year their behavior can look worse than they did at the beginning—even if you implement the *Second Step* 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* Weekly Themes those increases would have been much larger.

One way to tease out these types of effects that 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.

A final way to strengthen a non-experimental approach to evaluation is to look at implementation. In some early learning centers, implementation will vary—some students will get more Weekly Themes than others, some staff will implement the Weekly Themes 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 Weekly Themes or more reinforcement show better outcomes, that can help you see how to increase outcomes for more students.

Evaluation Surveys

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.

Here are some recommended assessments you could use to examine various outcomes of interest. Information about them can be found on SecondStep.org.

Devereux Early Childhood Assessment (DECA) for Preschoolers, Second Edition

The DECA-P2 is designed to measure three social-emotional skills important to a child's well-being: initiative, self-regulation, and attachment/relationships. Scoring the survey will give you a total protective factors (TPF) score for each child, as well as specific scores for each of the three protective factors. The DECA-P2 also produces a behavioral concerns screener score.

Minnesota Executive Function Scale (MEFS) by Reflection Sciences

The MEFS is designed to assess executive-function skills of children ages 2 and up. It is a tablet-based game (iPad and Android) that takes 5 minutes to complete. The MEFS starts at an age-dependent level and adapts to each child's ability.

Using Data for Evaluation

As part of their normal operations, early learning centers may collect data on children that can contribute to evaluation of the *Second Step* program. Any information gathered and tracked over time could be considered for this purpose. For example, many early learning centers record information based on observations of students that could be used to examine program effects. Data on behavior problems could be tracked over time as a way to see whether implementing the *Second Step* program has resulted in fewer problem behaviors. Just be sure to check the evaluation design section of this paper for information on how different evaluation designs affect the strength of the connection between the *Second Step* program and any outcomes you find.

One of the advantages of this approach is that early learning centers can often compare the number of referrals or other data for the year before they implemented the *Second Step* program to the number once the program has been implemented. In addition, it's possible to track referrals or other data over time to see whether the program results in fewer children having behavioral problems once the program is in place for multiple years.

Using Evaluation Findings

Positive Outcomes

Congratulations! Your evaluation has shown that your implementation of the *Second Step* program has improved outcomes for your children. This is the time to ensure that your early learning center or regional or school district program continues to teach the program to children and supports what children are learning in it throughout the day and the early learning environment. Remember that ongoing support for the program by center directors has been shown to be the number one factor that drives continued successful implementation over time.

Share the good news with your center staff, the regional or school district staff, funders, parents, and the community so your efforts continue to be applauded and supported.

Poor Outcomes

With No Implementation Evaluation

If your evaluation suggests that children are not benefitting from the program, a natural place to look for reasons and positive actions that can be taken is program 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 have not examined program implementation as part of your evaluation, doing so may provide you with ideas for how to strengthen the program and improve its impact on children. The Evaluating Implementation section discusses tools you can use to examine implementation of the program in your early learning center, regional, or school district early learning program.

With Implementation Evaluation

If your evaluation included information on implementation, then poor outcomes indicate the importance of looking closely at how the program is being implemented to see where there is room for improvement that may increase program effects.

Keep in mind that high quality program implementation goes beyond teaching the Weekly Themes. Just like with numeracy and literacy, what children learn in the *Second Step* program has to be reinforced and practiced in order to be mastered. Look for ways staff can cue children to use the skills taught

in the program throughout the day and across the early learning environment, and find ways to reinforce children's skill use.

If it appears that program implementation in your setting has been done well, it can be harder to know where to turn if your evaluation is not finding positive outcomes. Keep in mind that a truly rigorous evaluation requires random assignment of a large number of early learning centers and that quasi- or non-experimental evaluations can make it hard to separate out the effects of the *Second Step* program from other factors in your setting.