



COMMUNITY OF PRACTICE : ADDRESSING YOUTH DATING VIOLENCE

Research Design

Prepared by Dr. Laura Lambe, PREVNet Research Associate,
and Dr. Wendy Craig
June 2021

Table of Contents

Types of Program Evaluation.....	3
Types of Evaluation Designs.....	4
Non-experimental designs.....	4
Experimental designs.....	5
Quasi-experimental designs.....	6
Randomized-control trial (RTC).....	7
Threats to Evidence and Improving Your Research Design.....	8
Comparing the Strength of Evidence Between Designs.....	9
Program Moderators: Understanding Factors that Alter Program Effectiveness.....	11
Summary and Quick Tips.....	11
References.....	12

Research and evaluation are vital components of program¹ development. Choosing an appropriate research design helps you to gather the evidence necessary to:

- Ensure your program is having the intended effect and meeting objectives
- Ensure your program is not causing unintended harm(s)
- Understand the mechanisms by which a program may be working
- Learn more about how to improve programming
- Include youth voice in your program development and adaptations
- Secure grants and program funding

This PREVNet resource is intended to provide a basic overview of the quantitative research designs that may be used for program evaluation, as well as to outline the pros and cons of each approach. Although there are qualitative and other non-Western approaches to program evaluation, this resource focuses on quantitative program evaluation.

Types of Program Evaluation:

There are many different types of program evaluations that each have different purposes. The most common types of program evaluation are:

- **Formative Evaluation:** typically used when a new program is being designed. It is used to assess whether the program is acceptable to stakeholders and assess feasibility.
- **Process Evaluation:** focuses on whether a program was implemented as intended.
- **Outcome Evaluation:** used to document and evaluate the changes that occurred because of a program in a target population. Outcome evaluation is typically used to determine whether a program met its stated objectives.
- **Impact Evaluation:** used to determine whether a program is successful in achieving its broader goals (e.g., reducing rates of teen dating violence). Impact evaluation goes beyond stated program objectives – this type of evaluation considers broader changes as well as unintended impacts.

The current PREVNet tip sheet is generally focused on outcome evaluation. That said, best practices in program evaluation recommend including a process evaluation whenever an outcome evaluation is being conducted – this way, if you find your program is not meeting stated objectives, process evaluation may help you to answer why this is the case (stay tuned for our PREVNet tip sheet focused on process evaluation!).



¹A program is any structured intervention that is designed to be implemented in a way that is replicable. For example, *The Fourth R* is a youth dating violence prevention program.

Types of Evaluation Designs:

The image below provides a “roadmap” of the types of evaluation designs we will cover briefly in this tip sheet. Each design – non-experimental, quasi-experimental, and randomized-controlled trials – are described in detail below.

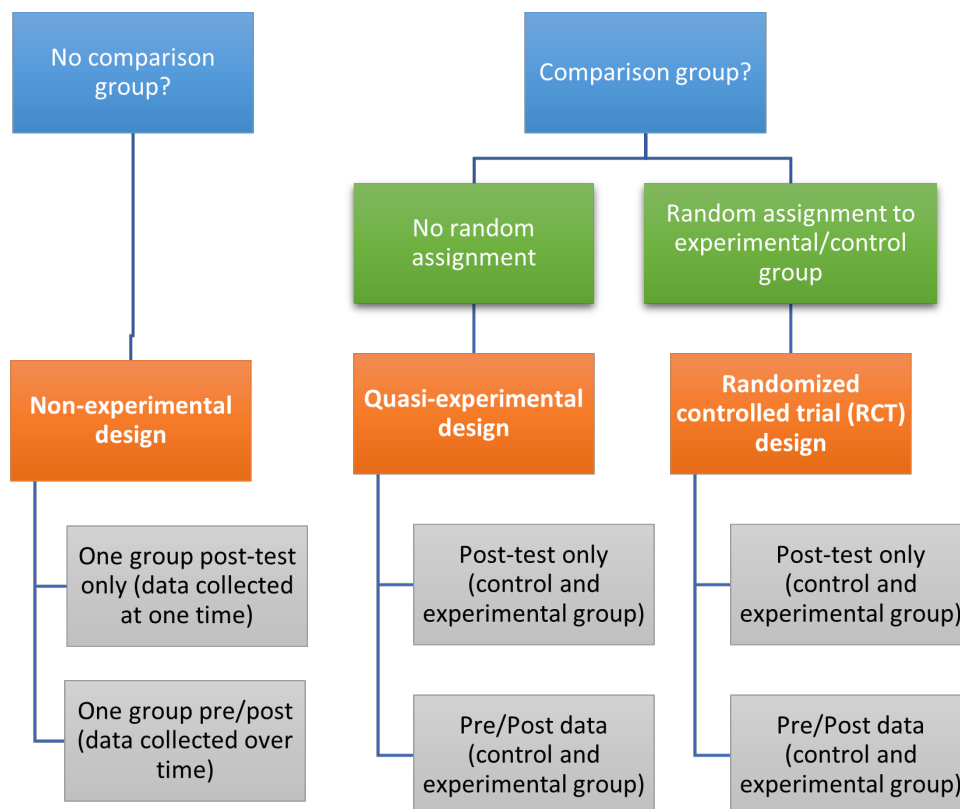


Image by Dr. Deinera Exner-Cortens.

Non-experimental design: These types of research designs are sometimes called pre-experimental designs or descriptive studies. Non-experimental studies can be conducted with post-test data only or with a combination of pre- and post-test data. Some questions that could be answered with a non-experimental study include:

- Who is participating in our program?
- How do youth experience our program?
- Is our program operating as intended?
- Is our program *associated with* a desired outcome?
- Are there factors that alter youths' experience of our program or its intended outcomes? (e.g., their gender, victimization history, etc.)

Note the emphasis on the word “associated with” above. **Non-experimental studies do not have the quality of evidence necessary to make claims about causality.** That is, non-experimental studies CANNOT provide evidence that a program is effective. More on this below.

A special case of research design (which can be used in both experimental and non-experimental studies) is the **retrospective pre-test post-test (RPP) design**. This type of research design collects data about knowledge/attitudes at the *end of the program* by asking participants to answer the same questions considering two timeframes: *before the program*, and *after the program*. Example RPP questions might look something like this:

	Knowledge BEFORE the program (1-4 scale)	Knowledge AFTER the program (1-4 scale)
1. Signs of unhealthy dating relationships		
2. How to get help if I am worried about my dating relationship		

Emerging research suggests that RPP designs can be *more accurate* in measuring change than traditional pre-test post-test designs, at least when it comes to outcomes like attitudes, beliefs, skills, and values. This is because RPP designs overcome the *response shift* bias, in which the comparison standard changes between measurement intervals – for example, participants often over-estimate their knowledge/skills at pre-test, simply because they “don’t know what they don’t know”. In turn, this bias minimizes true program effects. Finally, RPP designs are also efficient and economical because you only need to collect data after the program ends. These types of questions are a great way to improve the quality of program evaluation data, especially in non-experimental designs.

That said, retrospective reports like those used in RPP designs are limited by participants’ memory – human memory is imperfect, which may also inadvertently introduce error into the data (this is known as *recall bias*). In order to minimize the effects of both the response shift bias and recall bias, some evaluators collect both true pre-test data and retrospective pre-test data.

Experimental design: There are two types of experimental designs – **quasi-experimental** and **randomized-controlled trials (RCT)**. Experimental designs differ from non-experimental designs because they include a control group (who does not receive the program) and an experimental group (who does receive the program). There are many different types of control groups, including:

- **Wait-list control group:** receives the program after the evaluation is complete; wait-list group may or may not receive attention/placebo program during the evaluation
- **Attention control group:** the control group does an activity that is not similar to the program in terms of length, social contact, etc.
- **Placebo control group:** the control group receives an activity that mimics the program in terms of length, social contact, etc. but does not receive program content
- **Comparative effectiveness:** the control group receives an existing intervention; useful for comparing the effectiveness of programs

Some questions that can be answered with an experimental study include:

- Does my program *contribute to/cause* a desired outcome? (e.g., a reduction in dating violence)
- What is the strength of the effect of my program? (e.g., does my program reduce dating violence to a greater degree than existing social-emotional learning programs)

The key difference between quasi-experimental and RCT designs, however, is the use of *randomization* – participants are randomly assigned to the experimental or control groups. This randomization helps to ensure that the program effect is truly due to the program, and not the result of some unknown selection bias (i.e., bias that is introduced during participant selection). In other words, the purpose of randomization is to help make sure the experimental group and control groups are approximately equal on all variables except for your program – this way we can have more confidence that any observed changes are the result of the program.

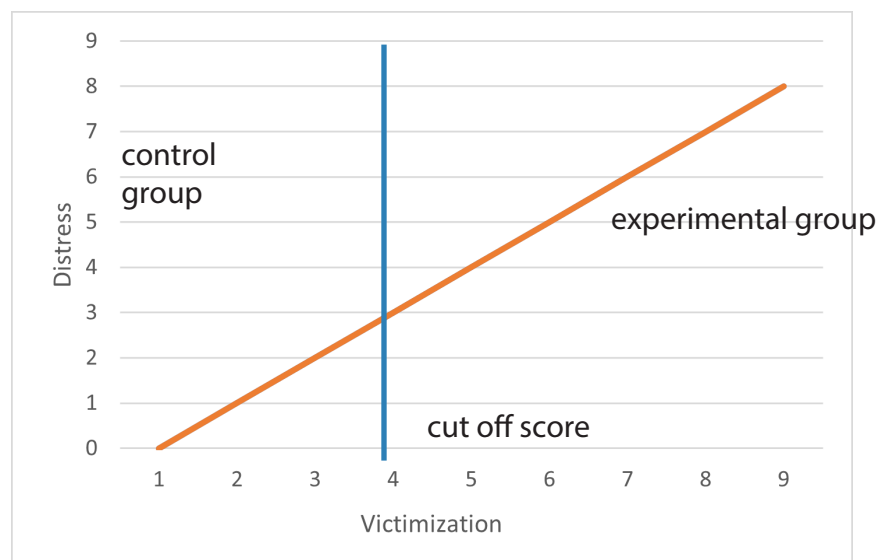
Quasi-experimental designs are used when one aspect of an experimental design is not possible (typically when randomization is not possible/feasible). As you might expect, it is not always ethical or feasible to randomly assign whether or not a group receives a program. For example, you deliver your program to a classroom of teens and compare outcomes to another classroom of teens who did not receive your program. Because these youth were not RANDOMLY assigned for the experimental group or control, this design would be quasi-experimental. Many research designs in the social sciences are quasi-experimental.

Quasi-experimental studies can be conducted with:

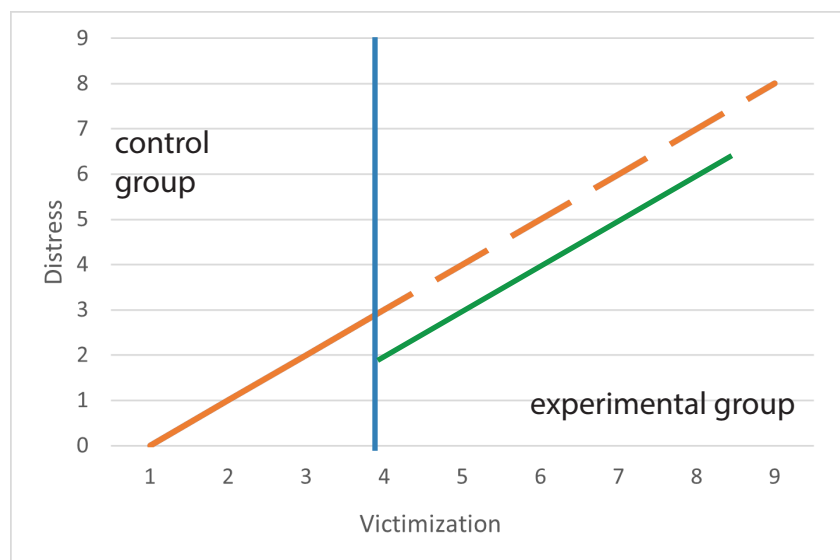
- Post-test data only, with a control group
- Pre- and post-test data, with a control group

A very strong quasi-experimental design is a **regression discontinuity design**. In this design, participants are assigned to either the experimental or control group based on a cut-off score on a pre-program variable. Such designs are especially useful when programs are targeted, and it is ethical to deliver the intervention to those who would benefit the most from it. In this design, evidence of treatment effects are seen when there is a *discontinuity* in the regression line for our variable of interest.

For example, say there is a positive correlation between victimization and distress, and your program is designed to enhance coping strategies and reduce distress (see figure below). Participants with a high level of victimization (scores 4+) are most likely to benefit from your program, so they are assigned to the experimental condition. Those with low victimization scores are assigned to the control group.



After your program is delivered, we would expect that the association between victimization and distress would change for the experimental group. In other words, there would be a discontinuity in the regression line at the cut-point, as shown below:



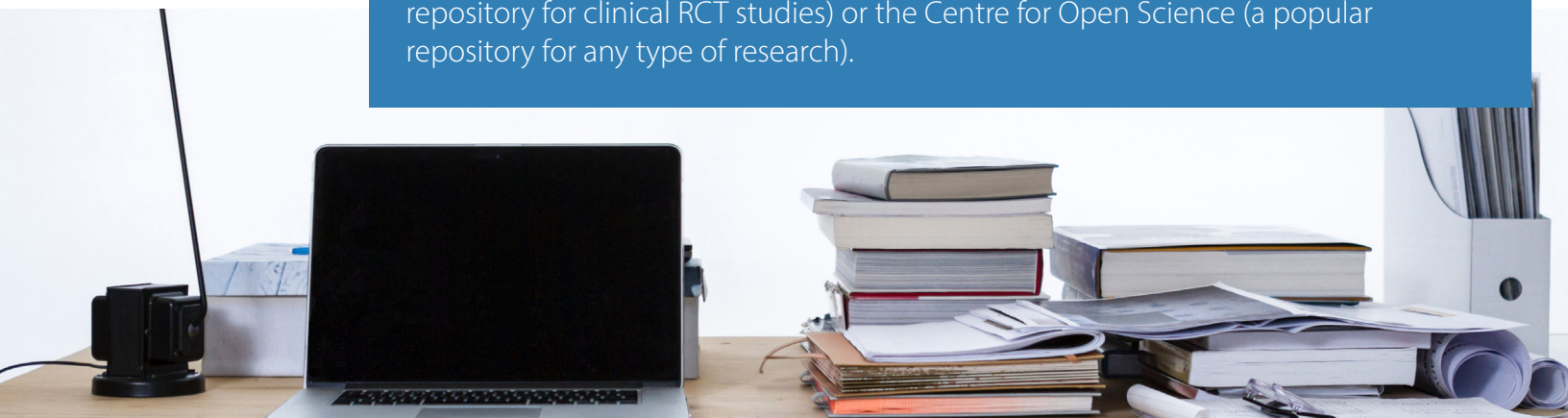
Randomized-controlled trial (RCT) design is the gold standard research design when it comes to assessing causality – that is, that the change in the dependent variable (e.g., teen dating violence) depends on the independent variable (e.g., receiving healthy relationships programming).

True experimental (RCT) designs require:

- Both pre- and post-test data in an experimental and a control group
- Participants who are randomly assigned to either the experimental or control group

Note: RCT designs and strong quasi-experimental designs are often pre-registered before they begin. This means that the study design and hypotheses are published before data are generated. Pre-registration enhances the quality, transparency, and replicability of research evidence because it ensures that data are being used to test hypotheses (and not that data are being used to generate hypotheses). Many top research journals will only publish RCTs if they have been pre-registered.

For more information on pre-registration, check out [ClinicalTrials.gov](https://clinicaltrials.gov) (the largest repository for clinical RCT studies) or the Centre for Open Science (a popular repository for any type of research).



Threats to Evidence and Improving Your Research Design:

Improving the quality of data being collected during program evaluation helps improve the quality of evidence we have to make claims about our programs. More stringent research designs typically have stronger internal validity – the extent to which piece of evidence supports cause-and-effect. The table below outlines some common threats to internal validity, and how you can address them:

Threats to Internal Validity	What to Consider:	How you Could Address This:
History	Did something else happen (besides your program) that could account for change that you are interested in?	<ul style="list-style-type: none"> Collect pre- and post-program data Measure the variables that could also be impacting your outcome Involve a control group
Maturation	Would participants naturally improve on the outcome over time, even without our program?	<ul style="list-style-type: none"> Involve a control group
Subject Bias	Did the way in which you selected your subjects influence the outcome?	<ul style="list-style-type: none"> Random selection of participants to control and experimental group, ensuring they are similar (e.g., same age) Consider any baseline differences in your experimental and control groups
Instrumentation	Did you change the way you measured your variables of interest?	<ul style="list-style-type: none"> Use the same measures to collect pre- and post- data, as well as in the control and experimental groups
Response Bias	Does your questionnaire or research design influence participants to respond in certain ways?	<ul style="list-style-type: none"> Consider whether pre-post data or retrospective pre-post data is most appropriate for your evaluation. Very strong designs collect both!
Program Fidelity / Integrity	Is your program being implemented as intended across groups?	<ul style="list-style-type: none"> Use a program manual and have specific training for program implementers Measure how the program is being delivered (e.g., session checklists)

Comparing the Strength of Evidence Between Designs:

The table below compares various research designs and the quality of evidence that can be obtained from each design. Research designs that produce the highest quality evidence are listed first, in descending order of quality.

Research Design	Pros	Cons
<p>Pre- and post-program data, with randomization and a control group (RCT)</p> <p><i>"Compared to schools in the wait-list control, schools who received our healthy relationships intervention had a 21% reduction in peer victimization over time"</i></p>	<ul style="list-style-type: none"> This is the gold standard approach for gathering evidence in Western research, as we are controlling many threats to internal validity. 	<ul style="list-style-type: none"> RCTs are expensive, time-consuming, and require considerable research skill.
<p>Pre- and post-test data, with a control group but no randomization (Quasi-experimental)</p> <p><i>"Girls with a history of victimization were more likely to report increases in self-esteem after participating in our program, compared to controls"</i></p>	<ul style="list-style-type: none"> This is a very strong research design, as we can look at change over time related to the program while also ruling out that these effects are due to the passing of time alone. 	<ul style="list-style-type: none"> Quasi-experimental designs can also be expensive, time consuming, and required considerable research skill.
<p>Post-data only, with a control group (Quasi-experimental)</p> <p><i>"Compared to youth who did not participate in this program, youth in our program were able to list significantly more ways they can intervene against peer victimization"</i></p>	<ul style="list-style-type: none"> Can still use a control group to gather evidence by comparing the two groups on variables of interest. Can use RPP questions to assess changes in attitudes, skills, beliefs, values. 	<ul style="list-style-type: none"> If we only have post-test data, we cannot make statements about change over time (e.g., change may simply be due to the passage of time). RPP questions may be limited by recall bias.
<p>Post-program data only, with estimates of pre-program data (Retrospective pre-post design, non-experimental)</p> <p><i>"Participants estimate that prior to this program, they now witness peer victimization about 25% less frequently in their schools"</i></p>	<ul style="list-style-type: none"> Retrospective pre-post designs can be very accurate and efficient ways to measure change in attitudes, skills, beliefs, and values. 	<ul style="list-style-type: none"> Non-experimental designs do not have a control group and cannot make claims about causality.

Research Design	Pros	Cons
Pre- and post-program data (Non-experimental) <i>"Participating in our program was associated with a 5-point increase in healthy relationships knowledge on average"</i>	<ul style="list-style-type: none"> With both pre-test and post-test data, we can also gather preliminary evidence about change to explore further using experimental design. 	<ul style="list-style-type: none"> There are many threats to internal validity that impact the quality of this type of evidence. For example, without a control group, it is possible that change on your variables of interest is simply due to the passing of time and not due to your program.
Post-program data only (non-experimental) <i>"10 out of 12 participants in the group reported that they 'strongly agree' that they learned new relationship skills in the program"</i>	<ul style="list-style-type: none"> We can gather useful information about how our program is operating. Can use RPP questions to assess changes in attitudes, skills, beliefs, values. 	<ul style="list-style-type: none"> May not be able to evaluate change if there is no comparison group (e.g., pre-test data or a control group). RPP questions may be limited by recall bias
None; anecdotal evidence only. <i>"I know our program works because I see first-hand how much the participants enjoy it"</i>	<ul style="list-style-type: none"> May be a useful starting point to think about what types of evaluation questions you would like to ask 	<ul style="list-style-type: none"> Data gathered from one person, in a non-systematic way, may be biased (e.g., only remembering the positives) and likely does not represent the program overall



Program Moderators: Understanding Factors that Alter Program Effectiveness

Another important avenue in program evaluation research focuses on understanding the person-in-context: what type of program works for whom, and under what conditions? Current best practices in program evaluation have moved beyond simply understanding whether or not a program works – it is also important to understand the conditions that modify program effectiveness.

A *moderator* is any variable that alters the link between a predictor (e.g., receiving a dating violence prevention program) and an outcome (e.g., attitudes about gender-based violence). For example, participants' gender may moderate the relationship described above, such that girls reported greater change in attitudes after the program than did boys. Moderators can be individual (e.g., gender, age, victimization history) or contextual (e.g., school belonging, school victimization rate).

It is important to gather both quantitative and qualitative data about program moderators. This means that ideally, you will want to collect quantitative data about factors that might impact the desired outcome of your program. Additionally, you will also want to collect qualitative data to understand how your program is experienced by diverse individuals in diverse contexts.

Summary and Quick Tips

Overall, how you design your program evaluation has important implications on the quality of data you are able to collect, which in turn influences the claims you can make about the impacts of your program. Over time, programs that are able to demonstrate converging evidence (e.g., multiple randomized control trials) across various contexts and implementers have the strongest evidence of effectiveness. Each evaluation is an opportunity to contribute to the overall literature base documenting the effectiveness of your program.

The strongest program evaluation designs include both quantitative and qualitative data (this is known as a mixed-methods design) in order to provide both numerical and lived experience support for program evaluation. So be sure to check out the PREVNet tipsheet on qualitative program evaluation!

Before starting your program, here are some simple questions to ask yourself to improve the quality of the evidence you are gathering:

- Can I include pre-program measures?
 - Ideally you want these to be the same as the post-program measures, so you can evaluate change over time (e.g., knowledge, attitudes, behaviours relevant to your program's intended outcome) and limit instrumentation effects.
 - Consider including measures of variables that might also impact the desired outcome of your program (e.g., to determine if the efficacy of your program varies by moderators like age, gender, rate of victimization, etc.).
- Can I plan to include a control group (or wait-list control) group in my design?
 - These participants get the same intervention as the experimental group, just not the program. A wait-list control group would receive the program AFTER both groups have completed post-test (and any follow-up) measures.
 - The control group should be as similar as possible to the experimental group (e.g., same age, gender composition) – this is why randomization is key!

- Is randomization possible with my program?
 - If possible, participants should be randomly assigned to either the experimental or control conditions.
- What do stakeholders want from my program?
 - Consider gathering qualitative data (e.g., a focus group) to determine whether your program is meeting the needs of users and other stakeholders.

If your **program has already started**, you can still evaluate. You can:

- Gather post-test data only:
 - This can still be very useful evaluating how participants experience your program (e.g., quality assurance data).
 - Use a retrospective pre-post design to gather data about changes in attitudes, skills, beliefs and values. Ideally, participants will rate the same question twice using the timeframes before the program and after the program.
 - You can also use question stems like “compared to before this program...” or “before participating in x...” in which participants would answer the question only once. However, the data that is generated from this type of question is NOT as strong as a true RPP design (rated twice) but still may be useful for your program.
 - Consider gathering open-ended, qualitative data from participants about your program. This information can be especially useful for generating new ideas and for providing a lived experience perspective. You can do this by conducting focus groups or interviews or have open-ended questions in your survey. One caution is that participants may simply tell you what they liked about the program and not discuss potential harms.
- Gather post-test data from program and a control group:
 - You can still make some statements about the outcomes of your program if you have an appropriate comparison group. Without pre-test data, however, it is unknown whether the two groups differed before the program.



References

Geldhof, G. J., Warner, D. A., Finders, J. K., Thogmartin, A. A., Clark, A., & Longway, K. A. (2018). Revisiting the utility of retrospective pre-post designs: the need for mixed-method pilot data. *Evaluation and program planning*, 70, 83-89.

Little, T. D., Chang, R., Gorrall, B. K., Waggenspack, L., Fukuda, E., Allen, P. J., & Noam, G. G. (2020). The retrospective pretest–posttest design redux: On its validity as an alternative to traditional pretest–posttest measurement. *International Journal of Behavioral Development*, 44(2), 175-183.

University of Calgary Program Evaluation Toolkit. Available online at: <https://www.ucalgary.ca/mentalhealth/education/program-evaluation-toolkit>

CDC Understanding evidence. Available online at: https://www.cdc.gov/violenceprevention/pdf/understanding_evidence-a.pdf

Jackson, S., Fazal, N., & Giesbrecht, N. A hierarchy of evidence: Which intervention has the strongest evidence of effectiveness? https://www.researchgate.net/profile/Suzanne-Jackson-2/publication/242760598_A_Hierarchy_of_Evidence_Which_Intervention_Has_the_Strongest_Evidence_of_Effectiveness/links/5707da8b08ae2eb9421bdb15/A-Hierarchy-of-Evidence-Which-Intervention-Has-the-Strongest-Evidence-of-Effectiveness.pdf



Public Health
Agency of Canada

Agence de la santé
publique du Canada

The views expressed herein do not necessarily reflect the views of the Public Health Agency of Canada