A Guidebook for Religious Literacy Evaluation:

Resources for Planning and Design

Kate Soules, Ph.D.

Table of Contents

Introduction

Evaluation is a professional field that "involves assessing the strengths and weaknesses of programs, policies, personnel, products, and organizations to improve their effectiveness" (AEA Board, n.d.) and is employed across disciplines and sectors. The following guidebook offers a brief introduction to some of the components of program evaluation with a focus on the evaluation of religious literacy educational initiatives. As an emerging field, there is a need for greater attention to evaluation and assessment of religious literacy education programs. At present, many scholars and programs make large claims about the impacts of religious literacy education, but do not have empirical evidence or clear models to demonstrate those impacts to funders, school administrators, educators, students, or other practitioners in this field. The companion report, The Imperative for Religious Literacy Evaluation: Context, Key Insights, and Recommendations, provides more detail and background about the need for evaluation in religious literacy education and a review of current practices and literature.

This guidebook is intended for anyone working on religious literacy education and interested in assessing their work. It is not a complete primer on evaluation and does not dictate a particular methodology or approach to evaluation. Rather, it provides an introductory evaluation framework with guiding questions at each step to help educators and researchers engaged in religious literacy programs begin to develop and implement an evaluation plan. This guide had been designed to be used during the planning stages of a program to think through the design of the program and develop appropriate evaluation tools to assess impacts. The worksheets found in Appendix A can be used in this process. Appendix B contains sample evaluation designs for several types of programs. While this guide does not go into detail about the processes of data collection and analysis, there are several additional resources and guidelines in Appendix C to support those steps.

Evaluation is about gathering information to develop a better understanding of a program or initiative. Whether you are looking to implement a comprehensive evaluation of your program or just starting to think about how you can evaluate one aspect of the program, this guidebook is intended to support your thinking and planning.

An Evaluation Framework

At its core, evaluation is about answering the questions "Did it work?" and "Are we making a difference?" To answer those questions, it is essential to know what "it" is and what "difference" can reasonably be expected from a given input or activity. Both the design of the program and its evaluation should define the "it" and the "difference." This guidebook is intended to help you, the educator, program director, evaluator, or researcher, answer these questions and design an integrated program and evaluation process. The guidebook's framework is informed by evaluation frameworks from the Center for Disease Control and Prevention (CDC, 1999) and Better Evaluation (2014). For more details about these frameworks and additional guidance, see Appendix C to this Guidebook.

Figure 1. Evaluation Framework

Steps 1 through 3 of this framework should be completed during the planning stages of the program and are the main focus of this guide. When evaluation is added to a program at the last minute or after the program is complete, it is much harder to align the questions and methods with the goals of the program. Integrating these steps into the program planning process also helps to refine the program design. While the steps are presented sequentially, you will likely want to return to each multiple times as you gain clarity about each aspect of the evaluation. Steps 4 through 6 take place during and after the implementation of the program. While this guide includes initial questions to consider for these aspects of the evaluation, these steps are not discussed in detail in this guide.

Step 1: Establish a Baseline and Engage Stakeholders

Evaluation design begins simultaneously with program design. The first steps are to define the context of the evaluation and all people and resources that will be involved. Including stakeholders from the beginning of the process can support longer-term engagement and buy-in to the evaluation process.

> 1. Establish a Baseline & Engage Stakeholders

1A. Identify Available and Needed Resources

The available resources—human, financial, and technical—will have a substantial impact on the types of evaluation that you will be able to do. Map out all of the resources involved in the program as well as the different stakeholders involved in the program and the evaluation.

You will want to revisit this step multiple times as you go through the rest of the evaluation framework. Once you have a clear evaluation design, identify resources needed for data collection, such as an online survey platform, audio or video recording, individuals to conduct interviews or observations. Be sure to account for elements such as data analysis, data management and organization, and dissemination of findings and make sure necessary resources are in place so that elements of the evaluation do not get abandoned due to lack of resources.

1B. Define evaluation priorities and standards

Decide how you will ensure that you are conducting a high-quality evaluation and have taken potential sources of bias into account. Consider who will be conducting the evaluation, how they will conduct their evaluation, and any biases they may bring to the process. Engage multiple stakeholder groups in conversation about the priorities and process of the evaluation. Review any benchmarks, outcomes, or priorities identified by funders or partnering organizations and discuss how those will be addressed in the evaluation.

Possible Stakeholder Groups

These groups are not mutually exclusive

People involved in the program operations	People served or impacted by the program	People who will use the evaluation
• Staff • Collaborators or community partners • Administrators • Funders and sponsors • Invited guest speakers, scholars, religious and community leaders	• Educators • Students • Families • Organizations or corporations • Community Groups	• Program staff • Funders • Staff and administrators at partnering organizations

1C. Review previous evaluations

Locate and review any previous evaluations of this program (or similar programs within the organization). Determine how to include those findings in the program design process. Decide which feature of past evaluations to repeat, which to change, and which to eliminate.

<div style="border: 2px solid purple; padding: 20px;">

Key Questions for Step 1

- What resources are needed for the evaluation?
- Who will be involved in the evaluation?
- Who will be responsible for the evaluation?
- Who will conduct the evaluation?
- Who will be responsible for data analysis and reporting?
- Who will make decisions?
- What community partners should be included in the evaluation process?
- How will the findings of the evaluation be used and shared?

</div>

Step 2: Describe the Program and Define the Evaluation

To design a strong evaluation, you need a clear picture of the components of the program and the intended impacts of the program. It is important to be detailed and concrete at this stage in the process. Outcomes such as decreased bigotry or peaceful co-existence cannot be easily measured and tied to a single program. While these may be long-term desired outcomes, you will need to identify more short-term impacts to measure. The process of outlining the program activities, outputs, short-term impacts, and long-term impacts can clarify what types of questions an evaluation can answer.
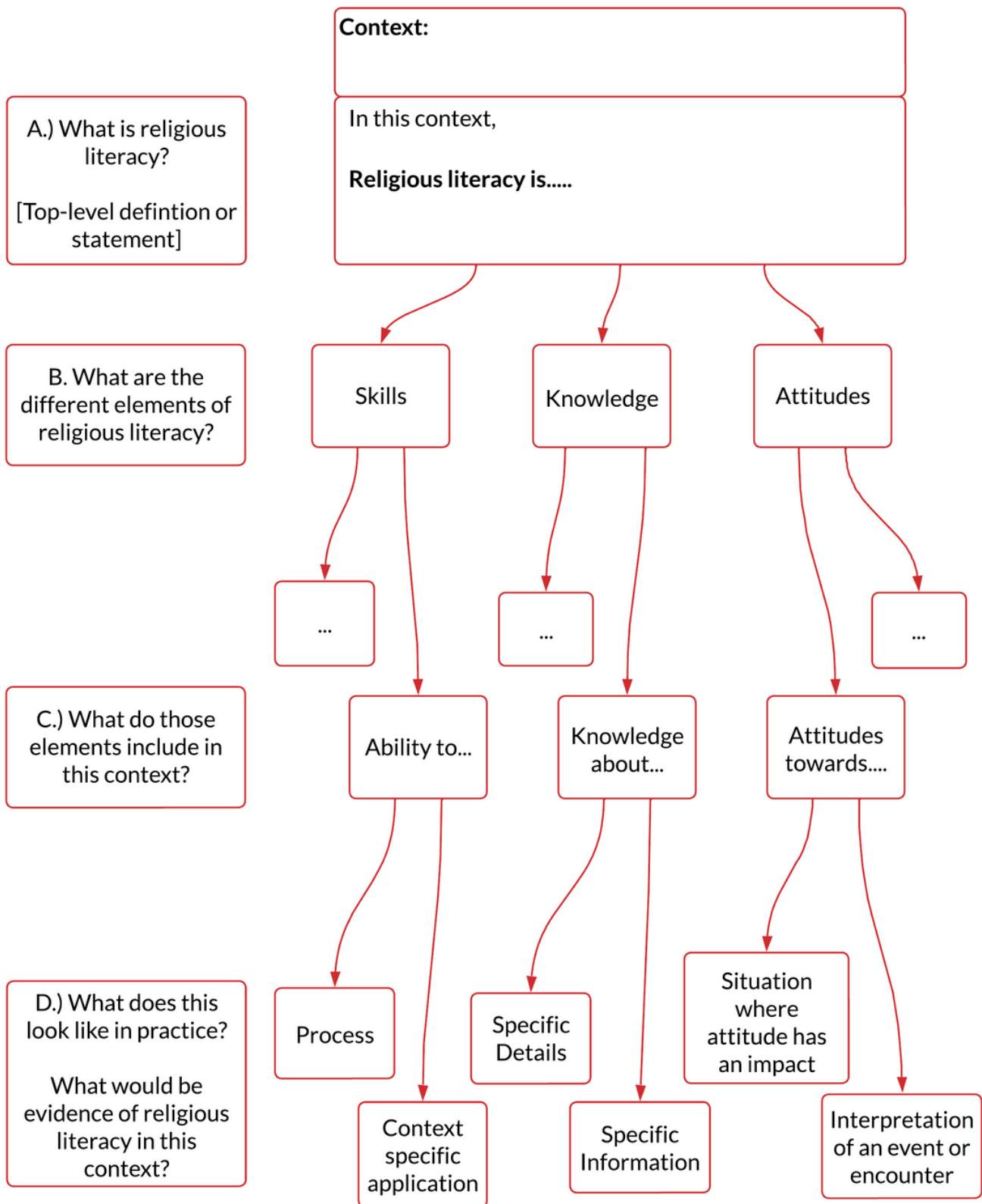
2. Describe the Program & Define the Evaluation

2A. Define Religious Literacy in Your Context

Begin with a working definition of religious literacy that can guide the design of your program activities and your evaluation. This definition does not need to be perfect or comprehensive, but it should be specific to the context in which the program is situated. The goal is to describe what religious literacy will "look like" when applied in your context, not to develop a completely novel definition. You may be best served by prioritizing traditions or issues that are most salient to the particular community in which your program is based (Nord, 2010) or to use frameworks, approaches, and terminology specific to that sector or profession (Soules & Jafralie, 2021; Walker et al., 2021). You can start with an existing definition of religious literacy and then use the steps shown in Figure 2 to work towards more specific details for your context. Clearly articulating your religious literacy definition can also be helpful when sharing your findings and contributing to the development of the wider field.

Figure 2: Defining Religious Literacy

Context:

In this context,

Religious literacy is.....

A.) What is religious literacy?

[Top-level defintion or statement]

B. What are the different elements of religious literacy?

Skills

Knowledge

Attitudes

...

...

...

C.) What do those elements include in this context?

Ability to...

Knowledge about...

Attitudes towards....

D.) What does this look like in practice?

What would be evidence of religious literacy in this context?

Process

Context specific application

Specific Details

Specific Information

Situation where attitude has an impact

Interpretation of an event or encounter

Additional Considerations:
Bias and Internal vs. External Evaluation or Research

Bias is an important factor to be considered in all forms of research and evaluation. The purposes of evaluation can make it more susceptible to bias because the findings can have significant consequences on the future of a program or its funding. Nobody wants to conduct an evaluation to find out that their program has no meaningful impact. It can be easy (intentionally or unintentionally) to design an evaluation and interpret the data in ways that will produce a positive image of the program. Internal evaluations are most susceptible to bias, particularly where the person conducting the evaluation is also responsible for the other aspects of the program.

Working with an external evaluator can reduce bias because they can serve as a more neutral observer. An external evaluator can be involved in some parts of the process, such as providing feedback on your design and conducting the data analysis, or could lead the whole evaluation process from start to finish.

Turning towards a research approach rather than an evaluation approach is another option, though this is usually more logistically complicated and often requires additional external resources (both human and financial) to conduct. With a research approach, the researcher's planning and execution of the research is independent from the program organization, with their own goals and questions. Because the researcher does not have a stake in the success of the program and is pursuing their own research agenda, the findings may be more critical and will not necessarily answer the questions that an evaluation would have asked.

Use the process shown in Figure 2 to start from a broad definition and then break it down into component parts. Clearly define your context and select an overarching definition (Part A). Then, break that definition into its main components (Part B), such as knowledge, skills, and attitudes. From these components, get more specific for the stakeholders and context of your program (Part C). For example, if your program works with public school administrators, the details under Knowledge might include knowledge of the district's policies about religious accommodations, knowledge about which religious traditions are represented in the district, and skills that might include explaining religious accommodation policies to teachers and parents.

Identifying even more specific content can support program design (Part D). Building on the previous example, these details could include specific district policies, demographic data about religion in the school district, examples of issues that have been handled poorly in the past, and student survey data about how they would like administrators to address topics of diversity, equity, and inclusion. From the details in Parts C and D, identify what elements are being addressed in the particular program or initiative that is being evaluated. Remember that a single program may not address every element of religious literacy in your definition. This exercise can help you to prioritize which content to include in both your program and evaluation.

2B. Identify and Align Program Components

There is rarely a direct line between program activities and the types of outcomes often cited in arguments for religious literacy. However, programs are often described as if there were such a direct line (Figure 3).
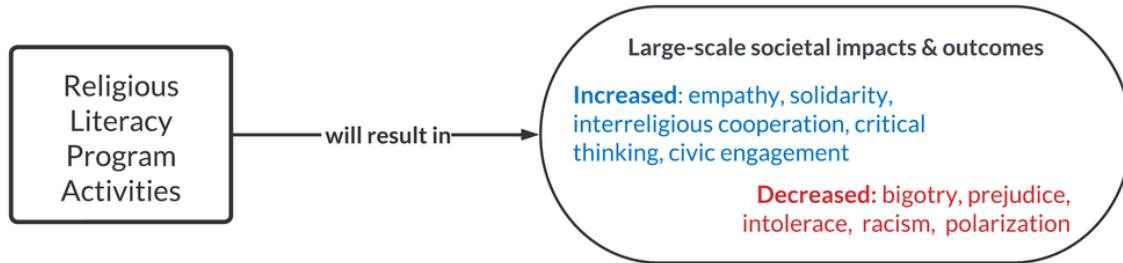

Figure 3: How Religious Literacy Programs are often described

In reality, there is a much more complex relationship between the specific activities of any program and those intended outcomes. Even when trying to map out short-term, intermediate, and long-term outcomes, there are many additional variables to consider. The further away you get from the specific activities of a program the causal relationship to the outcomes becomes much less direct. Figure 4 shows some of this detail, but it is still a simplified version.
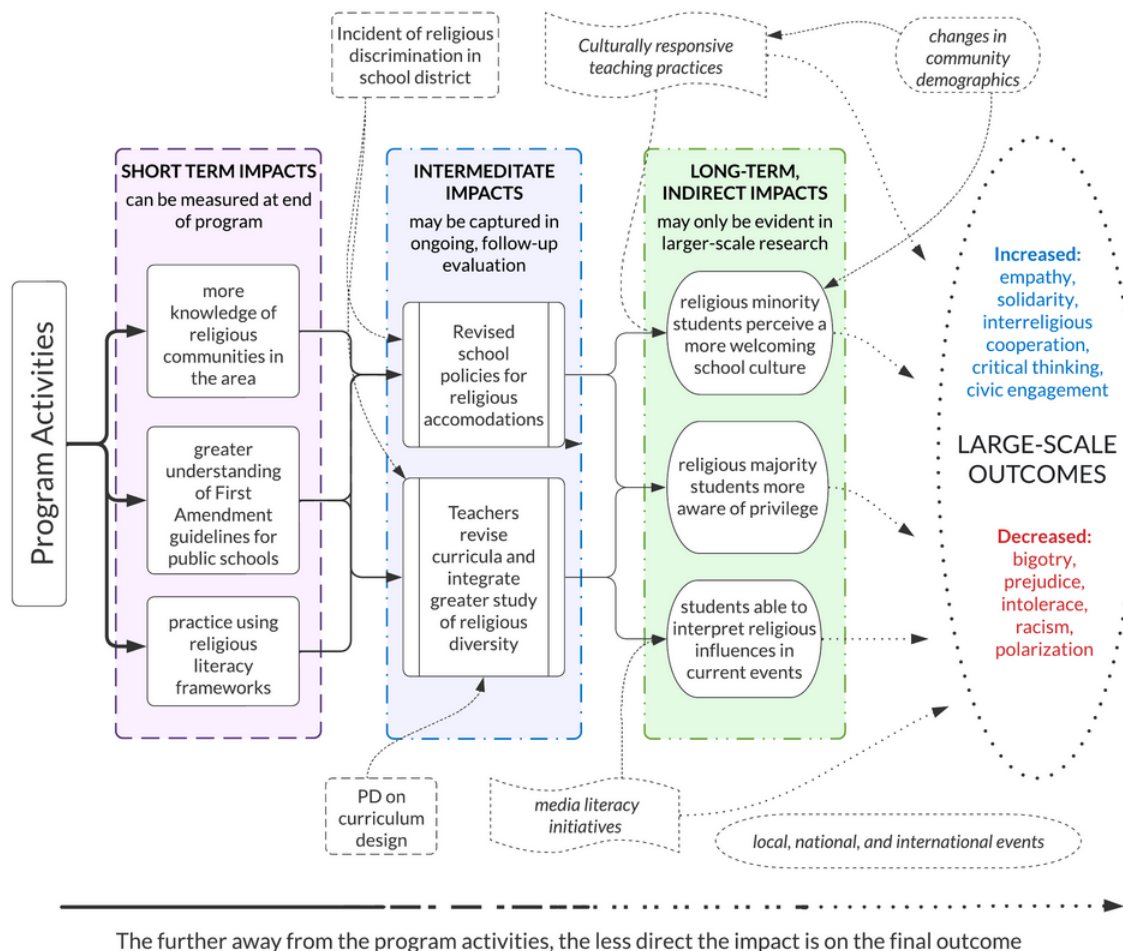

Figure 4. A more realistic model of religious literacy programs

An Evaluation Framework

To align the program components with the outcomes, identify the program inputs, activities, outputs, and impacts at different time frames. A detailed logic model (Figure 5) or theory of change identifies these components of a program and maps how those components lead to specific outputs and short-term impacts. Short-term impacts should then contribute clearly to intermediate and long-term outcomes. You can also use Worksheet 1 in Appendix A of this Guidebook as another way to map out the different components of your program and its outcomes. Worksheet 2 is a blank logic model template similar to the example below.

To identify short-term outcomes, consider the question, "What should participants be able to know or do at the end of this program?" Check for alignment with your program activities to ensure that those activities will enable participants to achieve those outcomes. You may have to work both backward (from your large-scale desired impacts) and forward (from your program inputs and activities) to ensure that you have a logical progression across the program. See Appendix B for examples of aligned and mis-aligned models and a discussion of how to address misalignments. If your program activities and outcomes are not well-aligned, your evaluation will not be able to provide accurate insights into the impacts.
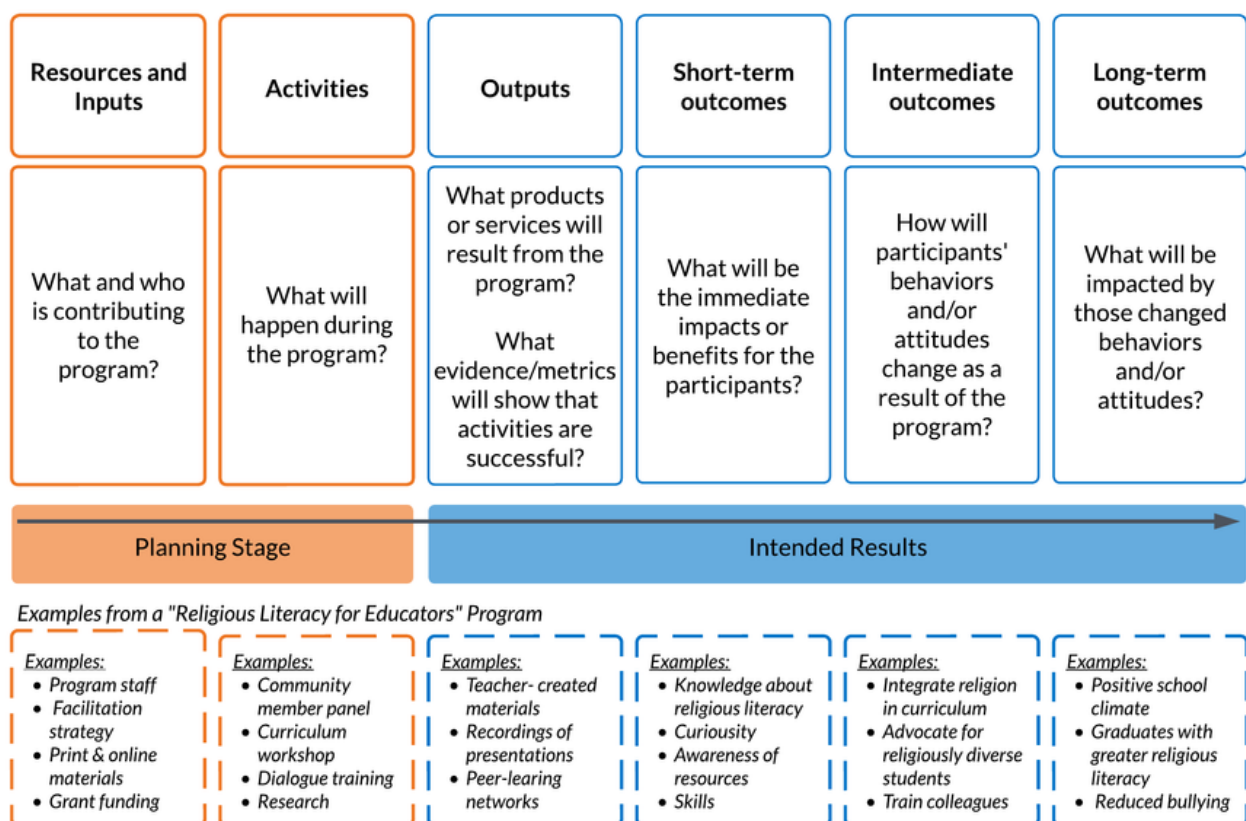

Figure 5. Logic Model template with examples

Individual program evaluations cannot capture all possible outcomes, impacts, or additional variables and contexts that might ultimately lead to large-scale, societal level changes. However, evaluation plans can map out potential pathways and identify which immediate and intermediate impacts could be measured as a part of a program evaluation. Pragmatically, these "pathway to impact" maps can also be useful in funding applications. These diagrams can become complex; having specific evaluation questions can allow you to focus on specific outcomes.

2C. Identify Evaluation Questions

Most evaluation questions can be classified as process questions or impact questions. Process questions ask about the implementation of the program—such as how many people participated, how well the plan was executed, and what worked or did not work in the program. These questions usually relate to the activities and outputs sections of a logic model. Impact questions (sometimes called outcomes questions) examine the effects of the program—such as did the program produce the intended learning, what behavioral changes can be attributed to program activities, and which participants were able to implement what they learned?

Both types of questions are important, but serve different purposes. Many evaluations include a mix of these two types of questions.

Process Evaluation Questions	Impact Evaluation Questions
• How well was the program implemented? • Were the program activities completed successfully? • Who were the participants? Was the targeted audience reached? • What problems or challenges were encountered in the process of implementing the program? • Were there adequate resources available for the program? • How well prepared were the staff to deliver the content of the program? • Were the participants satisfied with the experience?	• Did the program increase understanding about religious literacy? • Were the participants able to implement what they learned during the program? • Did the program have a greater impact with some groups than others? • What behaviors or attitudes changed as a result of the program? • What were the unintended or unanticipated impacts of the program?

Process questions can often be answered with concrete and tangible data, such as the number of participants or the amount of time an activity took to complete. These questions can help to make improvements to the delivery of the program, such as revisions to a presentation the participants found confusing or changing to a day that works better for more people. However, these questions do not generally provide insight into whether or not the program was effective. Outcome questions get to the primary issue discussed in the accompanying report: being able provide evidence to support the claims made about religious literacy. These questions can be more difficult to answer because there are often many more variables to consider. However, these are often the more interesting questions to explore.

Determine which type of questions you want to ask about the program. Return to the priorities identified in Step 1 as needed when developing the evaluation questions. A combination of process and outcome questions can be helpful. Prioritize questions based on:

- what you need to know;
- which questions can be answered give the available evidence;
- your program design;
- the resources you have for the evaluation.

As you determine the design of the evaluation in the next step, these questions may be revised.

Another Type of Evaluation: Needs Evaluation

This guide is mostly focused on process and impact evaluations, which typically happen during and after the implementation of a program. A needs evaluation usually takes place before the design of the program to better understand the current landscape and needs of different stakeholder groups. Some of the same questions and processes can apply in a needs evaluation, but the goals will differ from a process or impact evaluation. The findings from your needs evaluation inform the design of your program and can set up ongoing evaluation.

Key Questions for Step 2:

- What does religious literacy mean, in this context?
- What would religious literacy look like if this were achieved?
- What does success look like in this context?
- At the end of the program, 6 months out, 5 years out?
- What is your theory of change? How will the program activities lead to your desired outcomes?
- What do you need to know?
- What do you want to know?
- What would be nice to know?
- Do you need to report on specific outcomes? (esp. outcomes identified in a grant proposal)
- Which outcomes would be helpful to have data about to secure future funding, attract participants, or encourage other stakeholders to support the program?

Step 3: Design the Evaluation

Designing your evaluation requires selecting determining what data you will collect, by what method(s), and when. Several sample designs are included in Appendix B. There are numerous methods for gathering the evaluation data. Some considerations include time and resources available for data collection, time, skills, and resources available for analysis and reporting, and tools available for data collection and analysis. For example, surveys take far less time than interviews for both program staff and participants.

3. Design the Evaluation

In deciding what types of data to collect, go back to the outcomes defined in Step 2, and ask "How will I know that participants have achieved this outcome?" Start with short-term outcomes and continue to intermediate and long-term outcomes if you plan to include follow-up data collection beyond the term of the program. All methods of data collection have both benefits and drawbacks. Select your data collection methods based on your evaluation questions, the available resources, and balance and benefits and drawbacks of the method.

Pre-Post Evaluation Designs

While it is not always possible to implement a pre-post design (Figure 6), this is a valuable model that enables the collection of both baseline and impact evaluation data. Baseline evaluations are pre-program assessments of the skills, knowledge or attitudes of participants. Post-program evaluations should measure change in those same characteristics. Ideally, this pairing results in a feedback loop. Post-program outcome evaluations alone are somewhat limited in value.
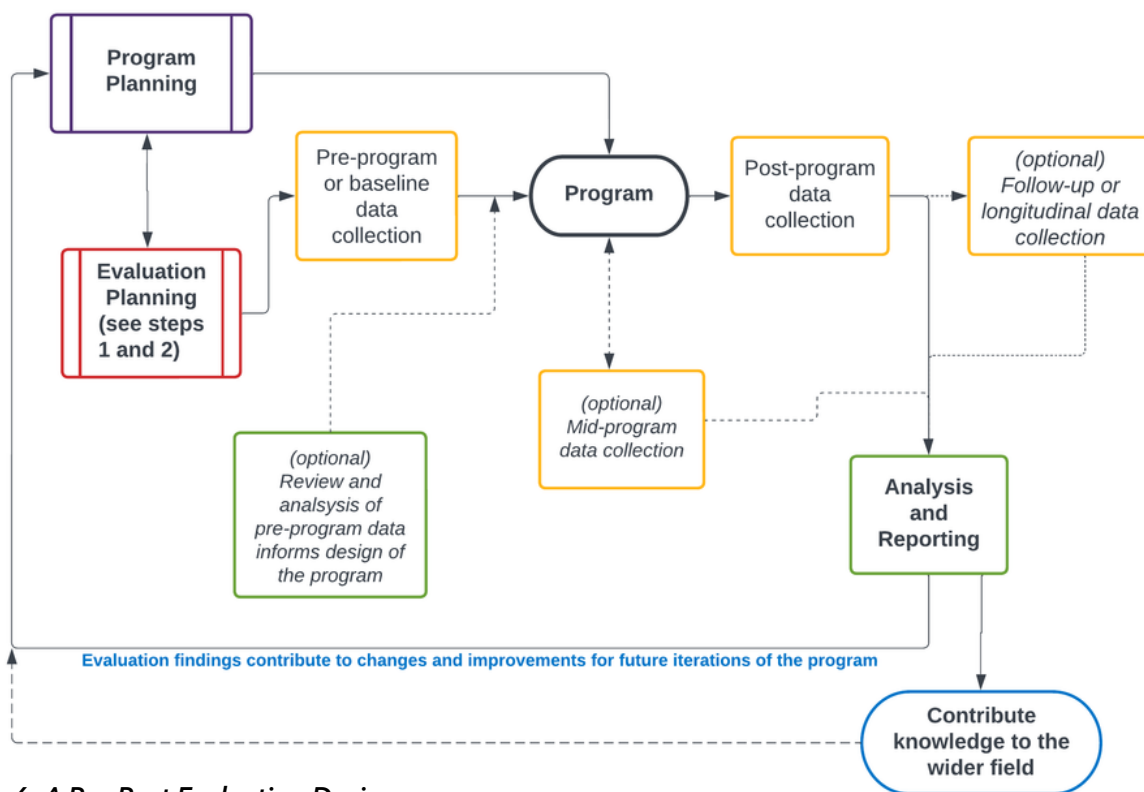

Figure 6. A Pre-Post Evaluation Design

Common Methods of Data Collection		
Method	**Benefits**	**Drawbacks**
Survey	• Can be simple to administer, especially to large numbers • Can provide quantitative data • Anonymous responses • Repeatable across multiple sessions of a program • Collects multiple types of information (opinion, demographics, behaviors) • Easy to summarize data	• Only get answers to the questions you ask • Scope of data can be limited, • Easy to misinterpret results (especially with small samples) • Requires appropriate administration and analysis tools • Cannot ask for clarification on answers • Survey fatigue • Low response-rates on follow-up surveys
Interviews	• Only get answers to the questions you ask • Scope of data can be limited, • Easy to misinterpret results (especially with small samples) • Requires appropriate administration and analysis tools • Cannot ask for clarification on answers • Survey fatigue • Low response-rates on follow-up surveys	• Very time consuming • Hard to analyze • Lack of confidentiality • Can be influenced by interviewer • Hard to quantify results • Cannot reach as many people
Focus Groups	• Efficient for reaching multiple people • Group members generate ideas together • Can lead to unexpected topics • Can ask for clarification • Can observe tone and non-verbal cues	• Can be time consuming • Harder to analyze • Some participants may be influenced by others in group • Lack of confidentiality • Hard to quantify results
Participant Observation	• Can account for the impacts of an unexpected elements of a program (i.e. a disruption during an activity) • Does not require any additional input from participants	• Participants can be influenced if they know they are being observed • Not as effective for large groups or settings • Cannot capture attitudes or opinions of participants
Analysis of Artifacts or Documents	• Does not require additional time from participants • Is not time-dependent (i.e. the document will be the same next week) • Easy to track changes over time • Can be anonymous	• Time consuming • Not all documents are publicly available • Cannot ask for clarification • Can be incomplete or lack important context

Survey Design Tips

- Keep it short: people get asked to take a lot of surveys. Be strategic in your data collection so that you only need to ask the most important questions on the survey.
  - Gather demographic or background information as part of registration or on a pre-survey.
  - Make sure every question aligns to one of the evaluation questions you want to answer
  - Order questions so that the most important items are at the beginning of the survey and less important questions are at the end, when people might be moving faster or skipping more questions.

- Build time into your program to complete the survey: don't count on people completing a survey that gets sent out via an email later.

- Consider accessibility: when selecting a survey platform or format, consider if it will be easy to access for all participants, particularly for online surveys.
  - Does it work well on mobile devices?
  - Will there be adequate internet access at the site?
  - Do you have an alternative option if someone does not have a device or there is a problem with internet access?
  - How will participants access the link to the survey? QR code? Shortened link?

- Design your survey with the end in mind: your survey is only useful if you are able to use the data collected. Think about how you are asking questions that will result in easy-to-use data.
  - Don't use an open-ended field for questions that could be closed-ended, such as number of years teaching, location, or grades taught. Reduce the number of steps you will need to take to be able to analyze the data.

- Just because a survey platform/software automatically generates reports or charts/graphs doesn't mean that those will show you what you need. Will the raw data be useful?

- Be consistent with your questions across surveys and events: developing a set of questions you can use across different surveys and programs creates opportunities for comparison and to be able to aggregate your data for broader insights.

- Test, Test, Test: Test out your survey in multiple ways before using it.
  - Be sure that the questions are clearly worded.
  - Double check that the format and display is clear on multiple devices (phones, tablets, PCs, Macs, etc.).
  - Enter several sample responses and try exporting the data to ensure the responses are being collected appropriately. Test to see if it will be in a format that you'll be able to work with to analyze.

Key Questions for Step 3:

- What types of evidence will answer your evaluation questions?
- What is your capacity?
- What resources can you commit to the evaluation?
- How much time do you have? How much time can participants commit to interviews, surveys, or focus groups?
- Is your evaluation design realistic?
- Are you asking questions that can be answered based on the evidence you will be able to collect?
- How does your evaluation design fit into the program design?
- Have the people who will be delivering the program been involved in the evaluation design?
- How long can you keep participants engaged for long-term follow-up?
- Will participants be willing to be contacted for future evaluations?

Step 4: Gather and Analyze Data

The specifics of collecting and analyzing your data will vary considerably by program and evaluation design. Refer to the additional resources in Appendix C of this Guidebook or consult with a researcher or evaluator to determine the best approach for a given context. Regardless of your program or evaluation design, you will want to consider the following as part of your data collection and analysis:

> 4. Gather and Analzye Data

- Think about timing for data collection in your planning for program activities. If you want participants to complete a survey or interview before the program begins, you will need to make that request with plenty of warning.

- Be aware of how much data you are collecting. You'll want to find a balance between having enough data to answer your questions, but not so much data that it will be burdensome on the program staff or participants to collect or overwhelming to analyze.

- Try to collect data anonymously when possible. Participants are likely to be more open and honest when they are assured their responses will be anonymous. When there is not a way to collect the data anonymously (e.g., interviews), names can sometimes be removed for analysis or analysis can be conducted by a different person.

- Make a plan to keep your data organized and secure. If you collect pre- and post-program data, be sure you will be able to differentiate responses on repeated questions. Back-up your data. Use clear, descriptive file names.

- If you are conducting any type of follow-up data collection after the program, inform the participants about the next stages and ensure you have accurate contact information.

- Make a plan for data analysis before the data is collected. Be sure that you will have the appropriate time, skills, and tools necessary to analyze the data. Your evaluation will not be useful if the raw data sits in a cabinet or a file without anyone able to analyze it. If necessary, consider engaging outside assistance with the analysis.

Key Questions for Step 4

- Have you built time into the program activities to collect data?
- What are you going to do with your data once it is collected?
- Surveys: does data need to be downloaded? What data cleaning will need to happen?
- Interviews and focus groups: do you need to create transcripts from recordings?
- How will you keep data organized and secure? Who will be able to access the data?
- What questions must be answered immediately?
- How quickly do you need to deliver your analysis?

Step 5: Develop Conclusions Based on Analysis of Data

Your evaluation questions, program design, and evaluation design will all impact the types of conclusions you can make from your data. One-off or short-term programs are less likely to show substantial changes or impacts on participants' knowledge or behaviors. A long-term program with follow-up evaluations will be able to provide much more in-depth and nuanced insights into participants' experiences and the impacts of the program. As you develop conclusions based on your data, you will want to be aware of the following points:

5. Develop Conclusions Based on Analysis of Data

- Remember that many outside factors can contribute to a person's experience of a program. Negative comments or a bad experience may not reflect the quality of the program, but could be because that individual was dealing with something else that impacted how they experienced the program.

- Do not read too much into results from small numbers of people, especially for quantitative data. With small samples, a single outlier or difference in a participant's experience can dramatically skew the data.

- Avoid over-stating your findings. Interpret your findings in the context of your program. If the program was a two-hour training, your conclusions will be different than if similar content was discussed over a semester.

- Be aware and transparent about the limitations of the evaluation. If there was a snag in collecting data on time or a change of agenda in the program, consider how these may have impacted your findings.

Key Questions for Step 5

- What does this evaluation say about religious literacy in this context?
- Were there any surprises in what you found?
- Were any of your assumptions confirmed?
- What is this evaluation unable to say?
- What were the limitations of the evaluation?
- How well were you able to answer your evaluation questions?
- What new questions have come up during the analysis?

Step 6: Share and Apply Lessons Learned

The final step in the evaluation process is to share your findings, internally and externally, so that you can put what you have learned into practice. Internally, the time and resources devoted to evaluation are wasted if they are not used to improve future programming.

It is important to share your findings externally as well. Your funders or other people who control the flow of resources may need this impact evidence to justify continued spending on your program. Further afield, the ability to demonstrate the impacts of religious literacy programs may be the best way to grow this field by helping other organizations design better programming, or by adding evidence to the scant research and evaluation data currently available.

Key Questions for Step 6

- What are the best methods for sharing your findings?
- Will you need to create separate internal and external reports?
- Who would most benefit from learning about your findings?
- How will you respond to negative findings?
- How will you modify future iterations of the program based on your findings?
- How will you modify the evaluation for the future?
- How can your findings justify future or ongoing support for this and other religious literacy programs?
- What findings are limited to the context of this program?
- What findings provide broader insights?

Appendix A:
Evaluation Planning Worksheets

Use these worksheets to guide your evaluation design process. Figure 2 (page 5) may also be helpful for developing a definition of religious literacy to use with these processes.

Worksheet 1: Mapping Inputs, Outputs, and Impacts
Use this worksheet to move from large-scale impacts to specific program activities and identify the evidence that can be used to demonstrate success on specific desired outcomes. Alternatively, the program activities are already in place, this worksheet can be used to identify related outcomes and impact.

Worksheet 2: Logic Model
Use the Logic Model worksheet to map the program inputs and activities and draw connections to the desired outcomes and impacts.

Worksheet 3: Selecting Evidence
Use this worksheet to select methods of data collection in alignment with the evaluation questions.

Worksheet 1: Map Impacts, Outcomes, and Inputs

Start with a long-term change or impact based on your definition of religious literacy:
Determine intermediate steps or changes that would lead to that impact. Then, determine what short-term outcomes could lead to those changes. Finally, identify program activities that will lead to the immediate outcomes.
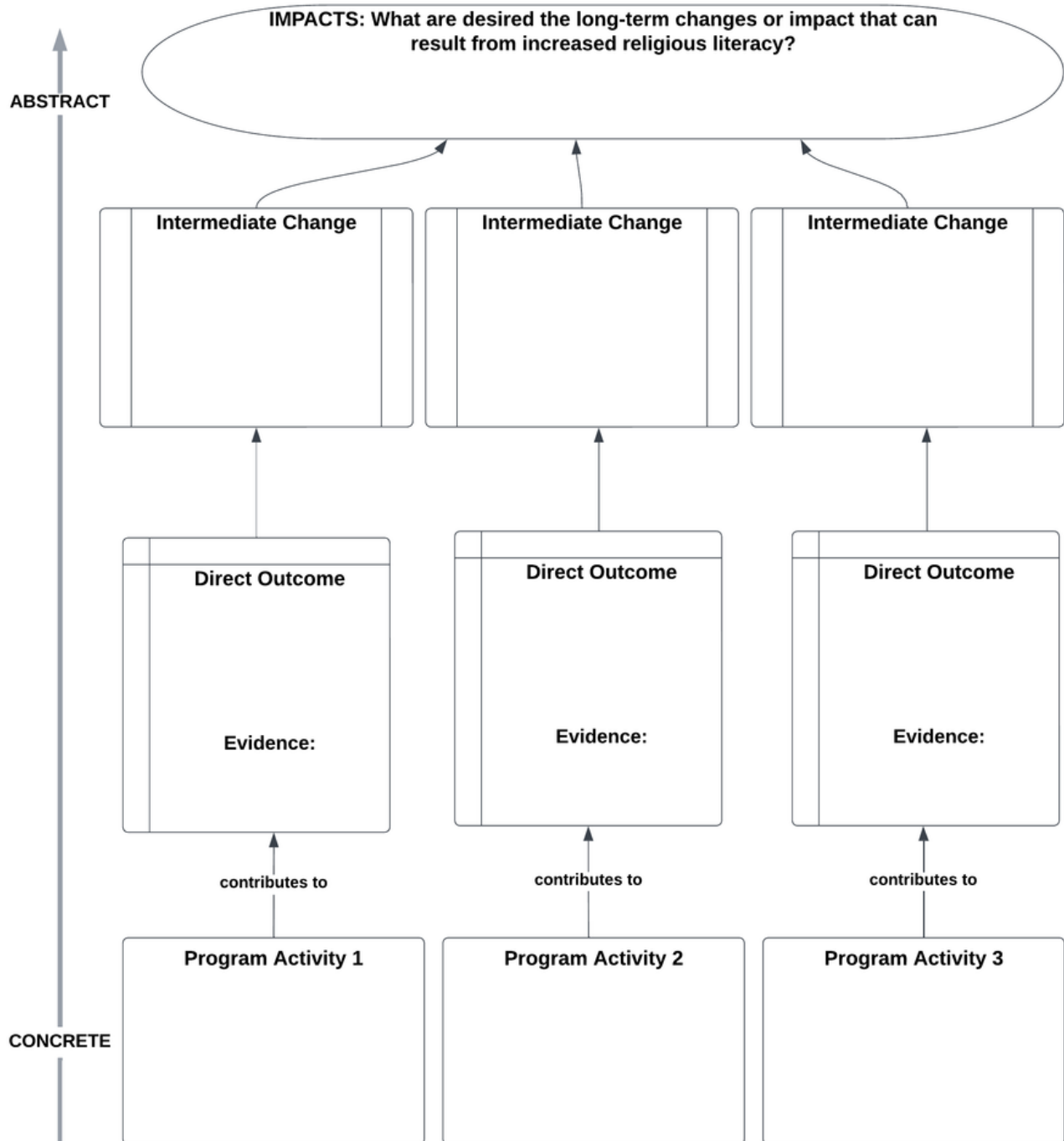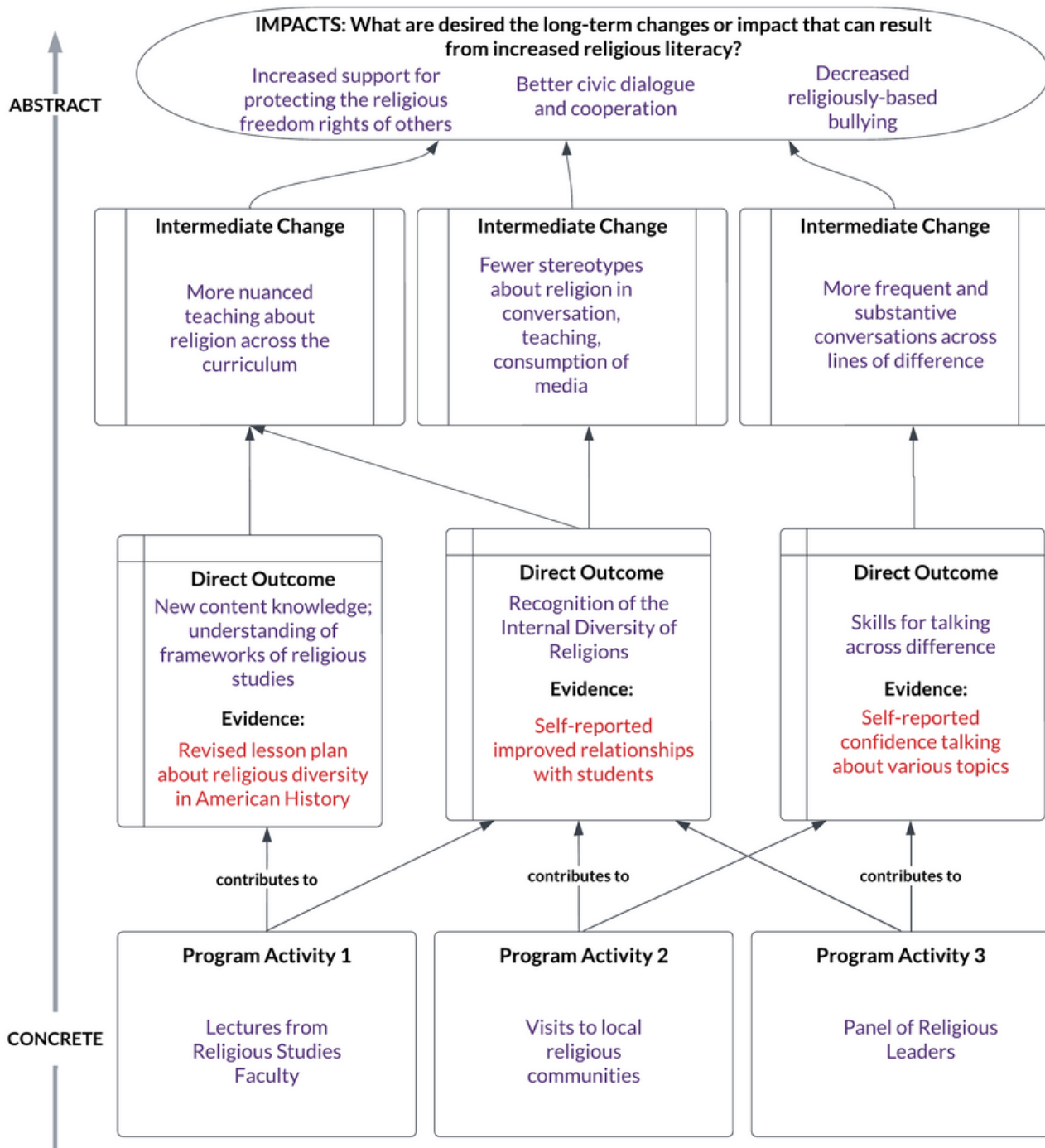

Start from the Program Activities:
Use this worksheet to ensure that all program activities have a direct connection to the desired outcomes (direct and intermediate) and impacts. Identify the types of evidence that could show that you have achieved your direct outcomes.

Worksheet 1: Mapping Inputs, Outputs, and Impacts

Sample Completed Worksheet 1

This planning map shows some sample information for a hypothetical professional development program for high school humanities teachers. For clarity, only some program activities, evidence types, outcomes, and their relationship are shown. You may have additional outcomes and connections between activities, evidence, and intermediate changes and impacts.

Worksheet 2: Logic Model

Planning Stage		Intended Results			
Resources or Inputs	**Activities**	**Outputs**	**Short-term outcomes**	**Intermediate outcomes**	**Long-term outcomes**
What and who is contributing to the program?	What will happen during the program?	What products or services will result from the program? What evidence will show that activities are successful?	What will be the immediate impacts or benefits for the participants?	How will participants' behaviors and/or attitudes change as a result of the program?	What will be impacted by those changed behaviors and/or attitude?

Assumptions: *What beliefs do you have about the program, the participants, and/or how the program will work?*

Evaluation Activities: *What will you do to assess the implementation and outcomes of the program?*

External Factors: *What factors or influences beyond your control might impact this program?*

Worksheet 3: Aligning Evidence and Evaluation Questions

If your evaluation is already designed, start on the left, list all of the evidence that you will collect, and determine which questions that evidence will answer. If the evidence doesn't answer any of your evaluation questions, consider if it is something you need to collect.

If you are still designing your evaluation, start with your questions. For each desired outcome, ask "How will I know that this objective has been met?" and "What evidence will I need?" You can then decide how and when to collect that evidence.

Evidence	How & when will it be collected?	What questions does it answer?	How does it answer that question?

Sample Completed Worksheet 3

In this example, the evaluation is focused on two questions, one process question and one impact question. Not all evidence has been included in this sample.

Evidence	How & when will it be collected?	What questions does it answer?	How does it answer that question?
Survey questions about subjects taught by participants	Pre-survey sent before program	What audience is interested in this training? (Process question)	The most represented subjects may indicate which teachers find the training relevant
Survey question about confidence dealing with a conflict related to religion in the classroom	Pre- and post-surveys	How does the program impact teachers' abilities to navigate religiously diverse classrooms? (Impact question)	Scores on Likert-scale, if scores increase, can infer increase in confidence
Examples of incidents related to conflict in the classroom and teachers' responses	Written responses: Asked on pre-survey Ask in three-month follow-up interview	How does the program impact teachers' abilities to navigate religiously diverse classrooms? (Impact question)	Answers will show how teachers deal with conflict, how it has changed, and if they are more confident after the training.
Teacher responses to students' questions in classroom discussions	Teachers' self-reports in follow-up interview	How does the program impact teachers' abilities to navigate religiously diverse classrooms? (Impact question)	Answers will show if teachers are more confident and capable responding to students' questions. May reveal areas where teachers need additional support in future training.

Appendix B: Sample Program and Evaluation Designs

The following diagrams show some different program designs and hypothetical evaluations that could be implemented with these designs. For clarity, only some of the possible evaluation questions or data collected are included in the diagrams.

Short, One-Time Workshop

Because this program involves a short, one-time workshop, the evaluation questions are simple. These questions are not expected to answer extensive impacts from the single workshop. They instead gather information on the goal of providing information on a particular topic and if the content of the workshop is appropriate for the audience. Asking about the characteristics of the attendee can help assess the effectiveness of the marketing or recruitment strategies. It can also help to align the program content with the types of participants who express interest.
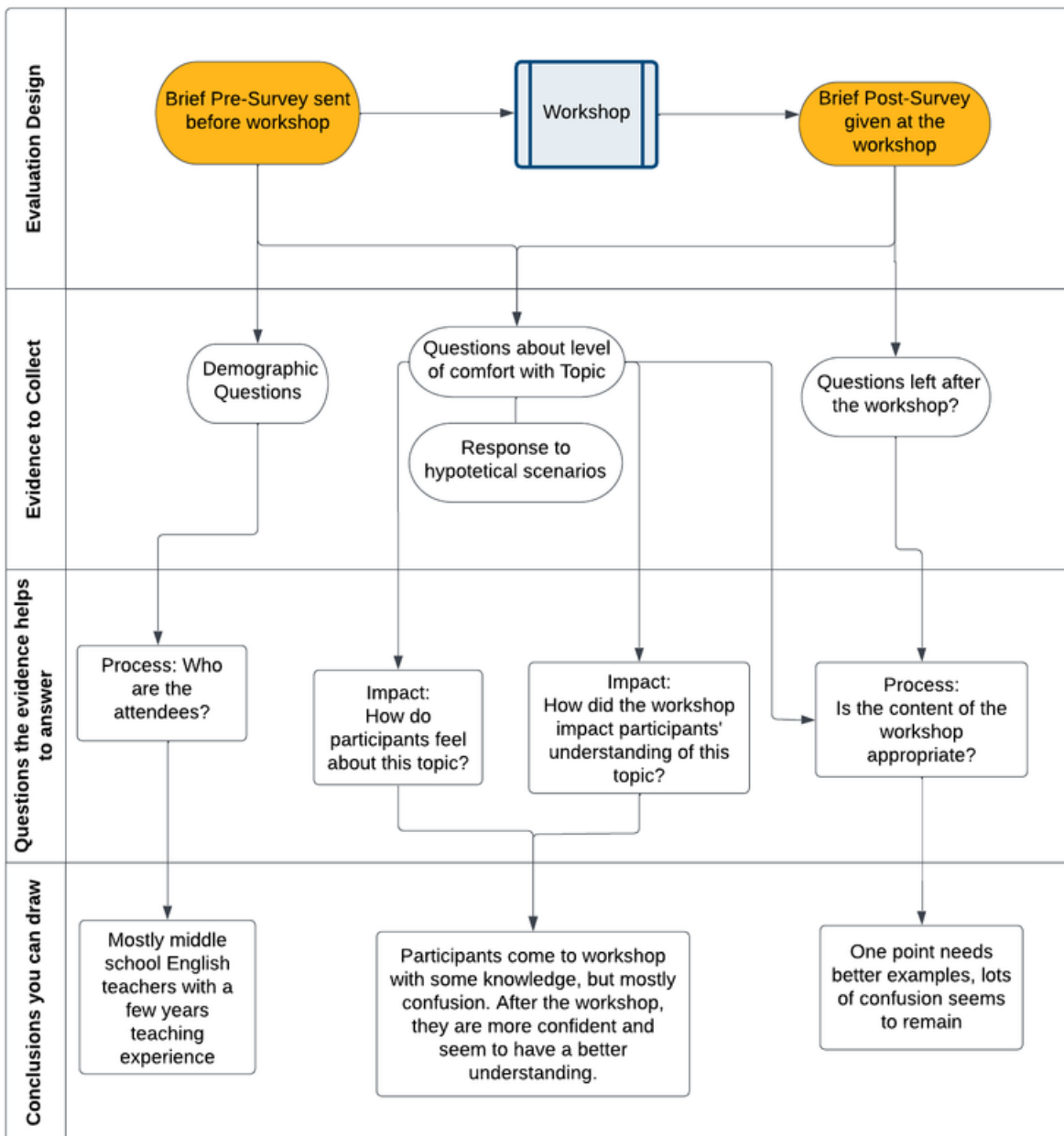
Three-Session Workshop

This program aimed to be responsive to participants' needs and questions and used repeated surveys to modify the content of each session. The program also hoped to offer a second session for the participants based on what they needed following the first session, and therefore followed up after they had time to implement their learning.
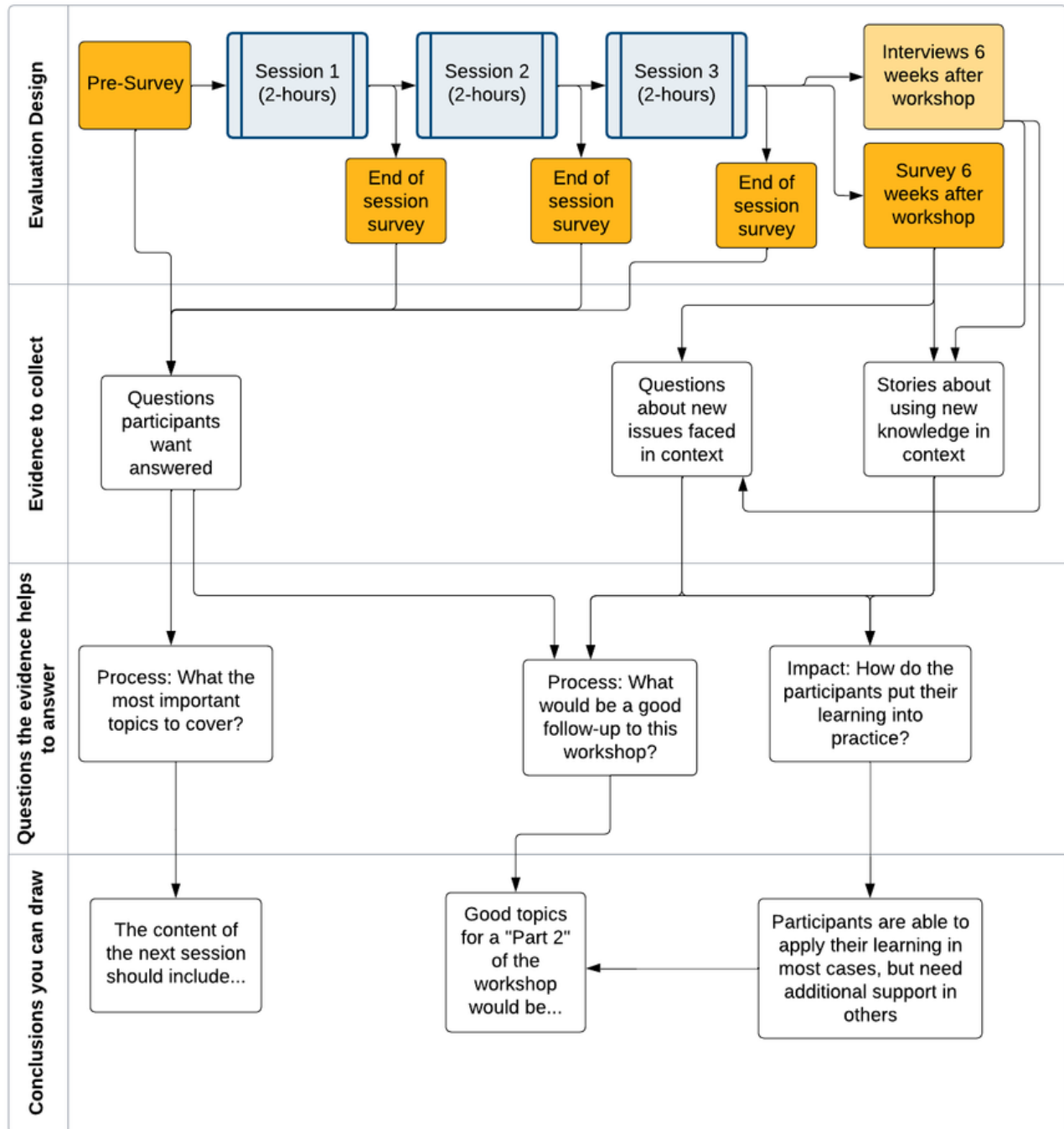
*The evidence and conclusions rows are left off the remaining diagrams for clarity.*

Week-long program

This week-long intensive program focuses on changing participants' attitudes towards a topic and building their skills to be able to teach about that topic. The pre- and post-surveys were designed to measure changes in attitudes. Collecting data by participant observation helped to gather evidence about what was happening during the workshop itself, such as interactions between participants, non-verbal cues, or the types of questions asked. These would be hard to capture in a survey or interview. By analyzing the lessons that the participants created over the course of the workshop, the evaluation can see how they put the content of the program into practice. If this evaluation were extended, it could include follow-up interviews or observations of how the participants implemented their lessons in their contexts after the workshop.
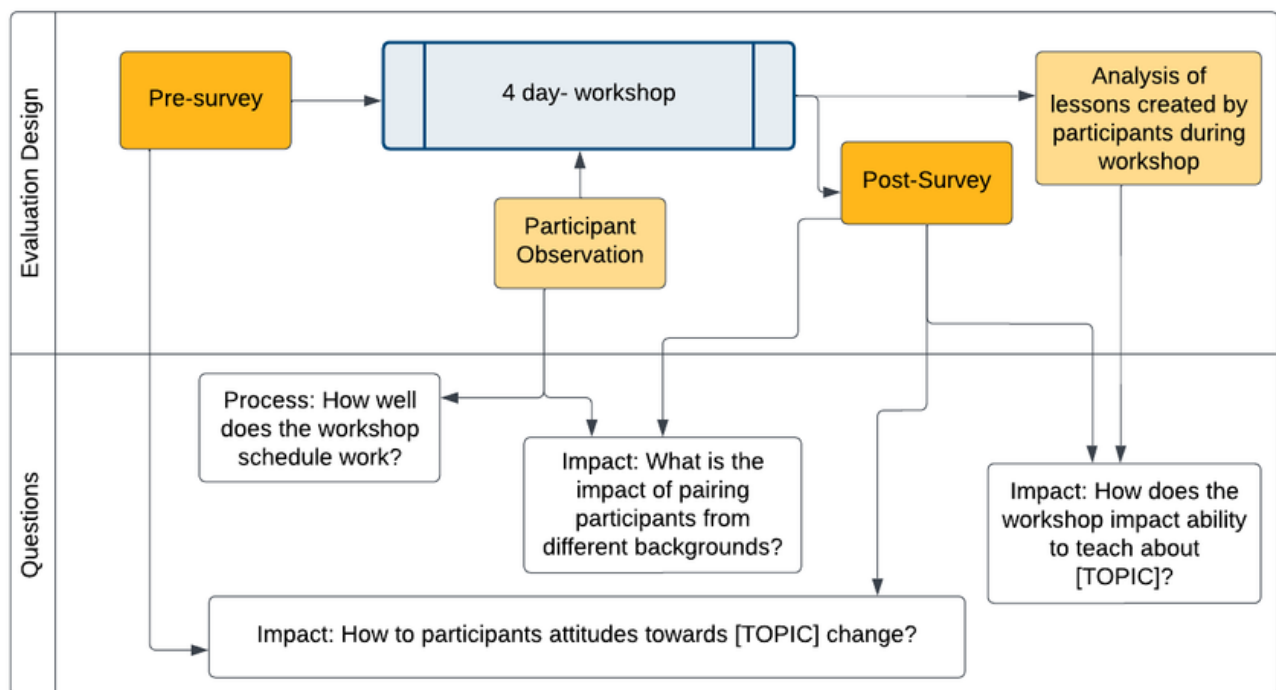
Year-Long Cohort Program

This year-long program was focused on building relationships between the participants and helping them develop skills that they could put into practice. Similar to the previous example, data gathered via participant observation captured a wide range of insights, such as non-verbal cues, interactions, and informal conversations between formal presentations and during meals. The interviews and focus groups enabled the evaluation to develop a deeper understanding of each participant, their contexts, and how they changed over the year. By including three follow-up focus groups at 3, 6, and 12 month intervals after the program, the evaluation was able to look at the longer-term impacts of the learning, as well as determine what additional support could be provided for participants during and after the program.
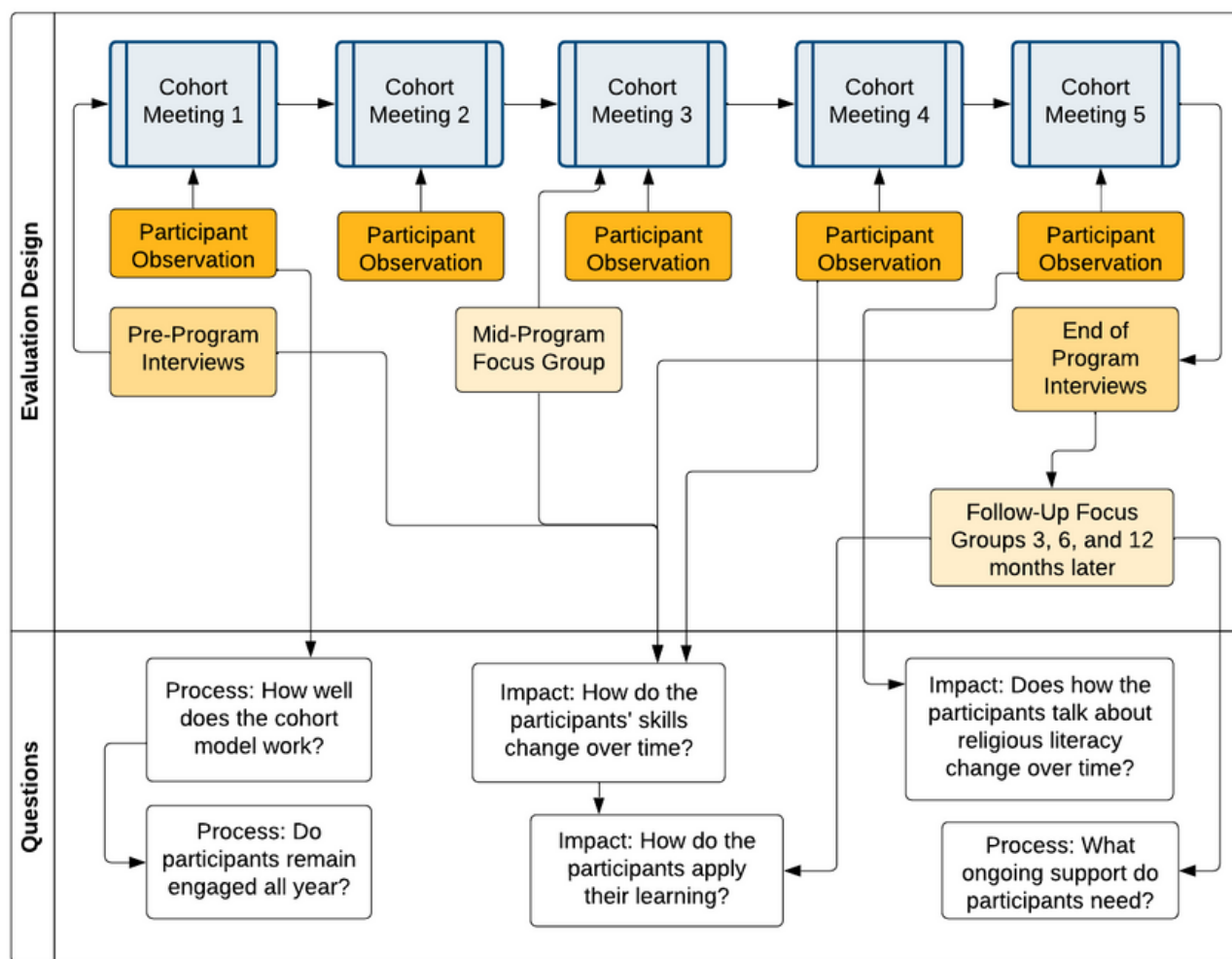
For clarity, lines are only drawn from one instance of participant observation.

Alignment Between Program Activities and Outcomes

This simplified logic model shows mismatched program activities and desired outcomes. The four activities on the left are unlike to produce the outcomes shown on the right. If this program were to be evaluated based on these desired outcomes, it would likely show that the program failed to meet its objectives.
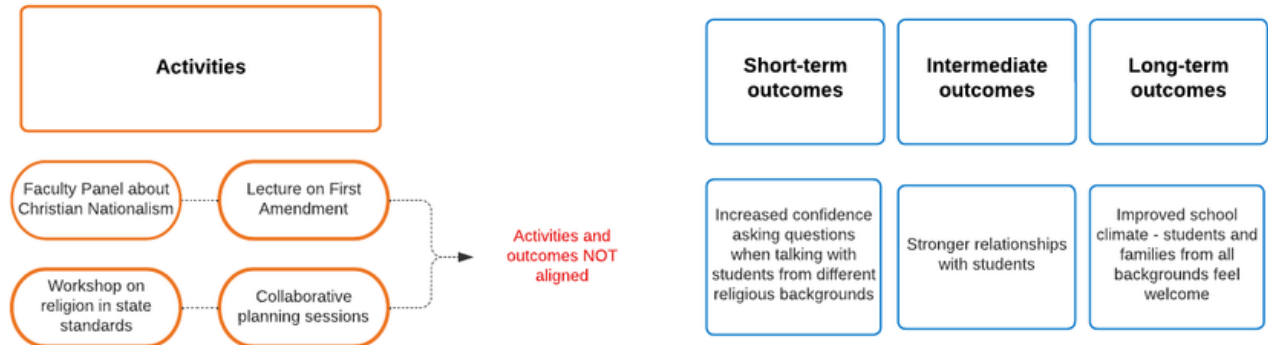

There are two main ways to fix this alignment issue. The first is to adjust the desired outcomes at the short-term and intermediate time frames. This is helpful if you are working with a set of resources or experts that would be difficult to change or replace. Examine your activities and determine what outcomes can be reasonably expected from those activities and how those outcomes can be measured.
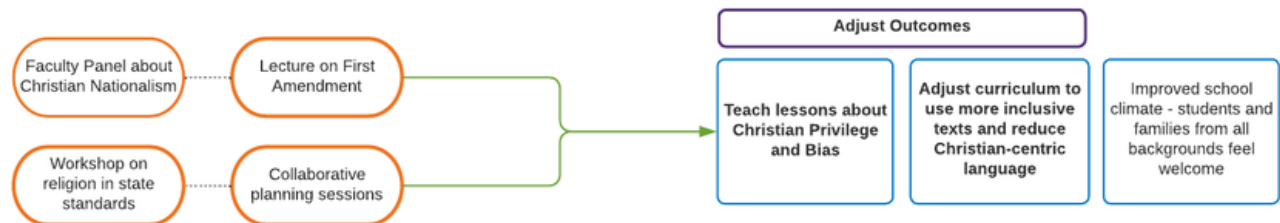
The second option is to adjust the program activities. This may be more possible if you are still in the planning stages, or it may be necessary if your outcomes are less flexible (e.g. you may have stated specific outcomes in a workshop proposal or a grant). You may not have to change all of your activities to improve your alignment to the outcomes.
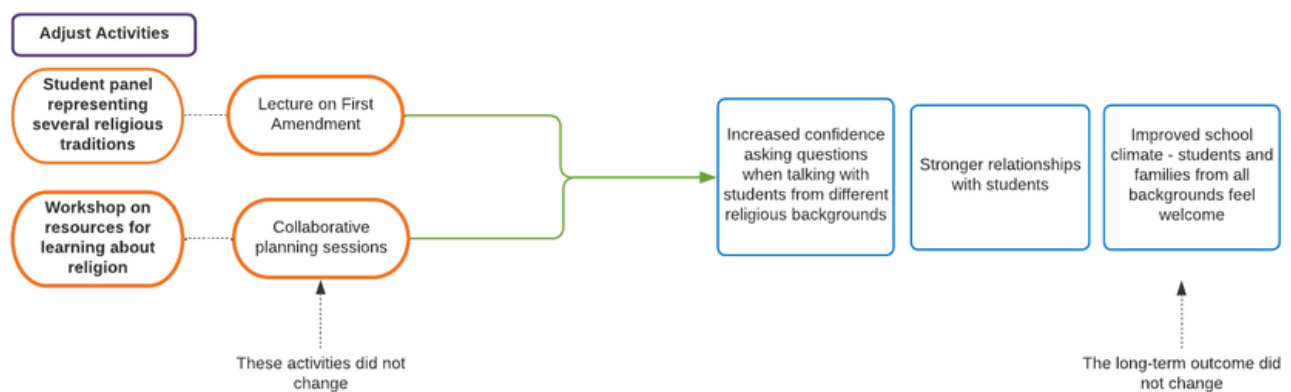
You will notice in both cases the long-term desired outcome, improved school climate, does not change, just the mechanisms for getting to the outcome.

Appendix C: Additional Evaluation Guides and Resources

This is a small section of additional resources for learning about and planning evaluations and assessments. There are many other evaluation guides and references that you may find helpful.

Centers for Disease Control and Prevention. (n.d.). CDC Evaluation Documents, Workbooks and Tools. https://www.cdc.gov/evaluation/tools/index.htm

Ennis, A. (2017). Teaching religious literacy: A guide to religious and spiritual diversity in higher education. Routledge. https://www.taylorfrancis.com/books/9781315206356

How to measure success: A toolkit for the evaluation of interfaith engagement. (2021). Woolf Institute. https://www.woolf.cam.ac.uk/research/projects/toolkit

Institute of Education Sciences Regional Education Laboratory Program (2021) Program evaluation toolkit: A module based toolkit for professional development and program evaluation. https://ies.ed.gov/ncee/rel/Products/Resource/100644/1

Rainbow Framework. (2014). Better Evaluation. https://www.betterevaluation.org/rainbow_framework/downloads

Wiggins, G., & McTighe, J. (2005). Understanding by design (2nd ed.). Pearson.

Woodrow, P., Oatley, N., & Garred, M. (2017). Faith matters: A guide for the design, donitoring & evaluation of inter-religious peacebuilding. CDA Collaborative Learning Projects and Alliance for Peacebuilding. https://www.cdacollaborative.org/publication/faith-matters-guide-design-monitoring-evaluation-inter-religious-action-peacebuilding/

Yarbrough, D. B., Shula, L. M., Hopson, R. K., & Caruthers, F. A. (2010). The program evaluation standards: A guide for evaluators and evaluation users (3rd ed). Corwin Press.