
Guidance for Designing and Evaluating Behavior Change Programs: Final Report

January 13, 2023

Prepared by:

Joseph Cook, PhD
Washington State University, Pullman, WA

Anand Jayakaran, PhD, PE
Washington Stormwater Center, Puyallup, WA

Aimee Navickis-Brasch, PhD, PE
Evergreen StormH2O

Francesca White, PE
Evergreen StormH2O

Laurie Larson-Pugh
Washington Stormwater Center, Puyallup, WA



Acknowledgements

A Technical Advisory Committee (TAC) played an integral role in this project. The TAC met as a group five times throughout the project. We thank the members for their input and advice on the work products developed for the project and hope these will be useful to their jurisdictions and others. Specific staff who developed and/or supported the development of the manual are as follows.

Aislin Gallagher, Kitsap County

Peggy Campbell, Snohomish County

Andrea Jedel, Ecology

Ann Bustamante, City of Sumner

Cammy Mills, Kitsap County

Laura Haren, City of Kent

Susan McCleary, City of Olympia

Jessica Shaw, City of Wenatchee

Brian Morgenroth, City of Walla Walla

Eric Lambert, Clark County

Trey George, City of Spokane

TABLE OF CONTENTS

1.0	INTRODUCTION	1
1.1	PROJECT OBJECTIVES	1
1.2	REPORT ORGANIZATION	2
2.0	MATCH STORMWATER PROBLEMS TO EFFECTIVE BEHAVIOR CHANGE TOOLS.....	3
2.1	OVERVIEW OF WORK COMPLETE	3
2.2	DELIVERABLES (D) (SEE APPENDIX)	3
3.0	ASSESS THE EFFECTIVENESS OF EXISTING E&O TOOLS NATIONWIDE.....	4
3.1	OVERVIEW OF WORK COMPLETE	4
3.2	DELIVERABLES	5
4.0	WEBSITE DECISION-SUPPORT TOOL	6
4.1	OVERVIEW OF WORK COMPLETE	6
4.2	DELIVERABLES DEVELOPED	6
4.3	GUIDANCE FOR USING WEBSITE	6
4.4	SEARCHABLE DATABASE.....	6
4.5	TRANSITION PLAN FOR MAKING WEBSITE SUSTAINABLE	14
5.0	REPORTING BEHAVIOR CHANGE EVALUATIONS	16
5.1	OVERVIEW OF WORK COMPLETE	16
5.2	DELIVERABLES DEVELOPED	16
6.0	EVALUATION GUIDANCE MANUAL	18
6.1	OVERVIEW OF WORK COMPLETE	18
6.2	DELIVERABLES	19
7.0	APPENDIX: KEY DELIVERABLES	Appendix pg.
7.1	SUMMARY OF KEY STAKEHOLDER INTERVIEWS (D2.1 & D2.2).....	A1
7.2	FINAL WEB SURVEY REPORT OF BEHAVIOR CHANGE PROFESSIONALS (D2.4).....	A6
7.3	ANNOTATED BIBLIOGRAPHY (D2.5).....	A17
7.4	SYNTHESIS AND BIBLIOGRAPHY OF BEHAVIOR CHANGE STUDIES (D3.2).....	A36
7.5	WEBSITE DECISION-SUPPORT TOOL - SURVEY RESULTS (D4.4).....	A49

7.6 SUSTAINING THE WEBSITE DECISION-SUPPORT TOOL (D4.5)..... A52

7.7 BEHAVIOR CHANGE EVALUATION DRAFT REPORT TEMPLATE – PILOT TESTING RESULTS (D5.5)..... A54

7.8 BEHAVIOR CHANGE EVALUATION REPORT CHECKLIST & TEMPLATE (D5.6 & 5.7)... A57

7.9 EVALUATION GUIDANCE MANUAL (D6.6)..... A87

1.0 INTRODUCTION

Public Education and Outreach (E&O) programs strive to increase awareness of stormwater impacts and encourage behavior changes contributing to water quality and quantity problems within a watershed. Washington State Department of Ecology (Ecology) recognizes the potential benefits of these programs on our state’s water bodies and includes E&O as a required component of Stormwater Management Programs (SWMP). More recent Western Washington municipal stormwater permits require permittees to develop targeted programs using Social Marketing methods, evaluate the effectiveness of these programs, and apply results to improve E&O programs.

The overall goal of this study was to provide jurisdictions with tools to identify and select which stormwater problems and behaviors to focus on and guidance on how to conduct and report effectiveness evaluations. These evaluations can then inform and improve future E&O efforts in a positive feedback loop of doing and learning. The study was made possible by the [Stormwater Action Monitoring](#) (SAM) program, funded by permittees and administered by the Ecology.

As the project commenced, the project team and its Technical Advisory Committee (TAC) agreed to narrow the focus solely to behavior change campaigns¹. By “behavior change”, we mean campaigns that seek to get people to change a concrete behavior, such as picking up pet waste or reducing fertilizer use. Many “education and outreach” programs seek to communicate information or change attitudes or programs that provide stewardship and volunteer opportunities without clear quantitative measures.

1.1 Project Objectives

- Determine what types of stormwater problems are amenable to, and best addressed, by behavior change efforts (Chapter 2).
- Identify effective behavior changes tools by conducting a broad literature review of behavior change campaigns, organized by stormwater issue (Chapter 3).
- Build an online tool to guide jurisdictions in prioritizing behavior change efforts and provide guidance for utilizing the tool (Chapter 4).
- Develop a template for reporting results from behavior change evaluations (Chapter 5).
- Develop guidance for evaluating the effectiveness of behavior change campaigns (Chapter 6).
- Assemble and manage the TAC throughout the project.

¹ The original title of the funded project was “Stormwater Education and Outreach: Addressing Challenges Through Behavior Change & Incentives and Guidance for Evaluating the Effectiveness”

1.2 Report Organization

Each chapter is organized by the project objective (above) and uses a similar format. The chapter provides a brief overview of the work completed and deliverables developed as part of this project. All project deliverables are included in the [appendix](#).

2.0 MATCH STORMWATER PROBLEMS TO EFFECTIVE BEHAVIOR CHANGE TOOLS

2.1 Overview of Work Complete

The objective of this task was to “determine what types of stormwater problems are amenable to, and best addressed by, behavior change efforts.” Our approach was both qualitative and quantitative. We began by interviewing 11 behavior change professionals in Washington State, several of whom are considered regional leaders in behavior change campaigns and are trained and enthusiastic about the use of “community-based social marketing” approaches in behavior change campaigns. The purpose of these semi-structured interviews (see Deliverable 2.1 for questions) was to learn more about how behavior change campaigns are chosen, administered, and evaluated. Our report on the conversations (Deliverable 2.2) highlighted six key lessons from the interviews, many of which informed the next stage of work, a nationwide survey of behavior change professionals.

We surveyed staff at public agencies (cities, counties, conservation districts) nationwide who conduct behavior change campaigns in stormwater or water quality to quantify and broaden our understanding of how staff choose, administer, and evaluate their programs outside Washington State. We collected 163 complete responses from staff in 19 states. Deliverable 2.4 provided eight key findings from the survey, including a) the predominance of pet waste and yard care campaigns; b) the desire for additional training on social marketing and program evaluation; c) staff feelings that campaigns are often under-resourced; and d) a need for additional high-quality evaluations of behavior change campaigns.

The last component of this objective was to review the scientific literature that links specific behaviors to water quality, distinct from the review of the social science literature (in Task 3) assessing whether behavior change campaigns are effective in changing behavior in a lasting way. Deliverable 2.5 is an annotated bibliography that provides an overview or road map of stormwater pollutant prioritization by summarizing several critical, peer-reviewed studies from the last 15 years. It also briefly reviews studies that outline the connection between environmental justice and pollutant prioritization, illustrating the need to ensure that the under-resourced communities that have long borne the brunt of stormwater pollution must factor into any pollutant mitigation plan.

2.2 Deliverables (D) (see Appendix)

- D2.1 and D2.2: Key informant interview questions and contact list, and summary of interviews.
- D2.3 (draft) and D2.4 (final, incorporating TAC feedback): Practices and challenges in stormwater behavior change programs: A survey of U.S. professionals.
- D2.5: Prioritizing pollutants: an annotated bibliography.

3.0 ASSESS THE EFFECTIVENESS OF EXISTING E&O TOOLS NATIONWIDE

3.1 Overview of Work Complete

The objective of this task was to synthesize the literature on how effective existing campaigns are in durably changing behaviors related to stormwater or water quality. We conducted a systematic review of both peer-reviewed evaluations of behavior change campaigns globally, as well as those written by consultants, local governments, or NGOs that have not undergone academic peer review (the “gray” literature). The principles used in constructing the criteria for identifying studies were relevance, breadth, and replicability, which we applied to a search of Google Scholar, GreenFile (EBSCO), Academic Search Complete, and Science.gov. Two members of the study team independently coded each identified study, ultimately identifying 56 studies for further review. For each study, we recorded the study year, location, targeted behavior, and target audience. We recorded whether the program specifically targeted under-represented groups, such as non-native English speakers or racial minority groups, and abstracted information on the implementation of the program. We evaluated the research quality of each study by scoring on whether the campaign had a) a well-defined priority audience (target audience), and b) a well-defined target behavior; and the evaluation included c) validated or pretested data collection instruments; d) a comparison group; e) pre-intervention data; f) observational data (rather than self-reported data); g) long-term measurement of change (≥ 1 year); h) a discussion of possible selection bias in uptake among target audience; i) and water quality measurements. We categorized studies scoring below 4 as “*fair*”, studies with scores between 4 and 6 as “*good*”, and scores of 7-9 as “*exemplary*”.

Many evaluations came from organizations in areas near estuaries or lakes, such as municipalities in western Washington, California, Minnesota, and states surrounding the Chesapeake Bay, such as Maryland and Delaware. We found that the evaluations were most commonly focused on lawn and yard care programs, echoing the results from the nationwide survey of behavior change professionals in Task 2. We classified more than half of the collected behavior change evaluations as “good”, scoring between 4 and 6. We ranked one quarter as “fair”, scoring less than 4, and 9% (5 of 56) of studies as “exemplary”. We found that all or most studies identified well-targeted audiences and behaviors, and the majority collected pre-intervention data. However, three-quarters relied on self-reported data, and only 13% measured behavior in a comparison group. Evaluators measured behaviors at time periods longer than one year in 36% of studies.

We conclude the report by recommending that evaluations invest the staff resources in tracking behaviors in comparison groups to improve the evidence base for behavior change campaigns. We also recommend that studies pay more attention to the possibility of selection bias – that people more amenable to behavior change are more likely to participate in campaigns – since this has important implications for whether campaigns can “scale”

effectively with additional resources. The deliverable includes a tabular summary of all collected articles as an appendix. This information also formed the backbone for the website discussed for Task 4, discussed next. The studies are searchable and downloadable on <http://www.waterbehaviorchange.org>.

3.2 Deliverables

- D3.1 (draft) and 3.2 (final, incorporating TAC feedback): A Synthesis and Annotated Bibliography on Stormwater Behavior Change Campaigns.

4.0 WEBSITE DECISION-SUPPORT TOOL

4.1 Overview of Work Complete

The goal of this task was to build a website to help behavior change professionals. It has two main purposes. First, what works? To help jurisdictions assess the effectiveness of existing campaigns around the country, it provides a searchable, downloadable compendium of every known evaluation of a behavior change campaign in stormwater or water quality (from Task 3). Second, it provides jurisdictions with guidance on choosing and evaluating their own campaigns. The site provides a downloadable decision-aide (in spreadsheet form) that jurisdictions can use to decide among several options for new campaigns, discussed more below. An earlier version of the site featured an interactive approach for walking users through a series of questions in choosing a campaign, but user feedback suggested that a downloadable spreadsheet would be simpler.

We began by designing a simple “mockup” of the website and gathering feedback from the TAC on the basic functionalities (D4.1 and 4.2) before proceeding to build and populate the website (D4.3). We then gathered feedback in an online survey of both TAC members and permittees in Washington to inform further changes to the site (D4.4).

4.2 Deliverables Developed

- D4.1 (draft) and D4.2 (final, incorporating TAC feedback): Mock-up of effectiveness archive and decision-support tool.
- D4.3: Website operational; <http://www.waterbehaviorchange.org>.
- D4.4: Report on survey jurisdictions testing website.
- D4.5: Sustaining the behavior change website.

4.3 Guidance for Using Website

The [website](#) has two main functions: a) a searchable database of existing behavior change campaigns (from Task 3) and b) a downloadable decision support tool.

4.4 Searchable Database

The home page of the site displays a map showing evaluations by location. As shown in Figure 1 below, one can click on a “pin” to load a page with the details for a particular study (Figure 2). The page also shows some summary statistics on the universe of campaign evaluations, such as the targeted pollutants, target audience, and behavior change (the user should hover their mouse over pie slices to see labels).

Figure 1. Splash page

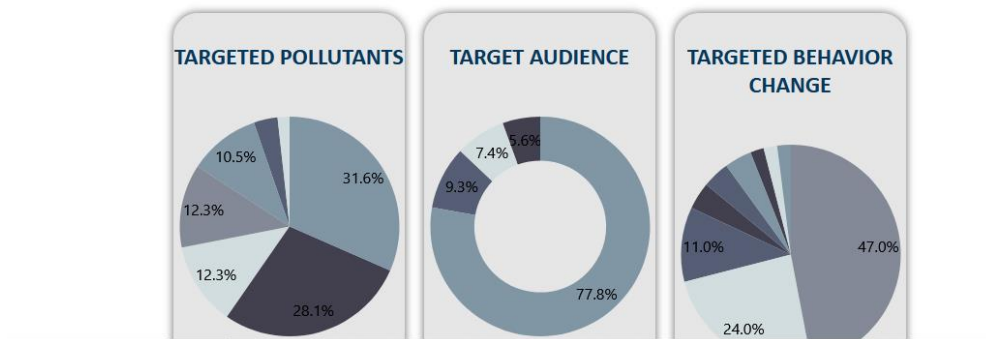
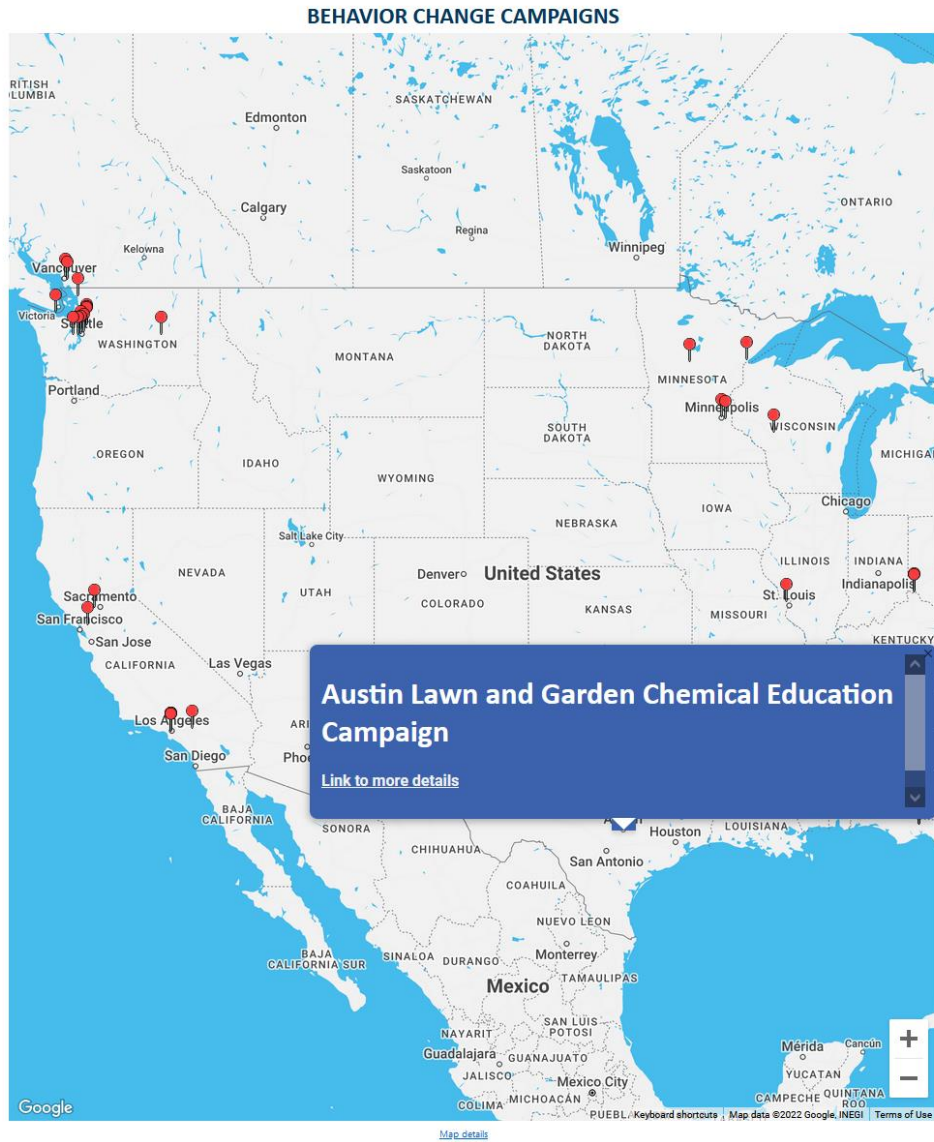



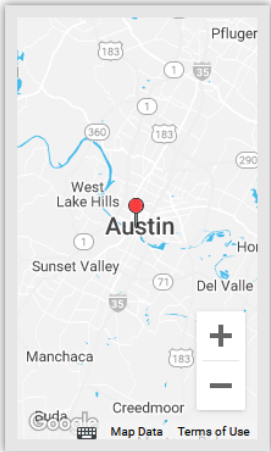
Figure 2. Study details page



Tools and Resources for Effective Behavior Change Campaigns

IN STORMWATER & WATER QUALITY

Home
Add Campaign
Search Campaigns
Campaign Selection
Other Resources
About



Austin Lawn and Garden Chemical Education Campaign

Authors: Shay, Kathy	Published: 2011	Location: Austin, Texas
-----------------------------	------------------------	--------------------------------

Targeted Audience

Homeowners in the middle to high socio-economic group because they are most likely to have funds to purchase landscape chemicals. In addition, the research team also included information from the fertilizer company Scotts/Miracle Gro. As a result, the target audience also included men of 55 years old and older who value green and well-manicured lawn, homeowners who were in their 30s searching for first-time device on landscaping techniques and had not yet developed less earth-wide habits, and women who were purchasing or influencing more than 50% of fertilizer purchases.

Targeted Behavior

Reducing the use of lawn and garden chemicals

Targeted Pollutant

Nutrients, Toxic Chemicals (Pesticide, Household Cleaner, etc.)

- ABSTRACT/ISSUE
- TARGETED AUDIENCE, BEHAVIOR, AND POLLUTANT**
- RESEARCH SCORE
- CAMPAIGN IMPLEMENTATION
- EVALUATION METHODS
- EVALUATION RESULTS
- INCENTIVES
- SOCIAL/RACIAL JUSTICE CONSIDERATIONS
- PDF LINK

The user can access information on the various components of the study by clicking on the components in the blue bar on the left. The final option on the blue bar is to download a pdf of the original study or report.

Under the “Search Campaigns” tab, users can also search by keyword or filter studies by targeted pollutant, audience, research ranking, state, and whether the study addressed environmental justice concerns. When the page first opens, all studies are automatically loaded into the page (Figure 3); the user can click on any of the blue boxes for individual studies to pull up the study’s details page (Figure 2). Figure 4 shows how to filter the studies that target toxic chemicals.

Figure 3: Full set of studies in “Search Campaigns”



Tools and Resources for Effective Behavior Change Campaigns

IN STORMWATER & WATER QUALITY

Home
Add Campaign
Search Campaigns
Campaign Selection
Other Resources
About

Keyword Search

TARGET POLLUTANT:

LID/Infiltration

Metals

Nutrients

Oils

Pathogens (Fecal Coliforms, Bacteria, E. Coli)

Sediment

Toxic Chemicals (Pesticide, Household Cleaner, etc.)

Trash

TARGET AUDIENCE:

Businesses

Contractors

Developers

Engineers

General Public

Land Use Planners

Landscapers

Mobile Businesses

Property Managers

Residents

School-Age Children

RESEARCH QUALITY RANKING:

Exemplary

Good

Fair

CAMPAIGN STATE:

Addresses Environmental Justice Concerns

Addresses Concerns

Campaigns

<p style="font-size: small;">Optimising the development and use of persuasive communication to...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Restoring the Redwood River</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Using a reverse auction to promote household level stormwater control</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Vermont Rethink Runoff Survey 2007</p> <p style="font-size: x-small;">Click for more details</p>
<p style="font-size: small;">Pooches for the Planet</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Education and Changes in Residential Nonpoint Source Pollution</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">2008 Pet-Waste Study</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Austin Lawn and Garden Chemical Education Campaign</p> <p style="font-size: x-small;">Click for more details</p>
<p style="font-size: small;">Blue Water Baltimore Rainwater Harvesting Behavior Change Implementation</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Final Report: Snohomish County Septic System Program</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Clean Lawn Care for Clean Water: A Collaborative Approach</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Behaviour change: Trialling a novel approach to reduce industrial stormwat...</p> <p style="font-size: x-small;">Click for more details</p>
<p style="font-size: small;">Alachua County Scoop the Poop Campaign</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">New York City's Wait... Pilot Program: An Integrated Approach to Water Quality...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">The Neighborhood Water Stewardship Program: An Innovative Approac...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Mobile Contractor Illicit Discharge Education & Outreach Effectiveness Study</p> <p style="font-size: x-small;">Click for more details</p>
<p style="font-size: small;">Education Campaigns to Reduce Stormwater Pollution In Commerical Areas: ...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Implementing a Neighborhood Scale Stormwater Retrofit: Effect of Self-Drainin...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">San Bernardino County Stormwater Public Education Program</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">State of California Integrated Waste Management Board. Grasscycling Follow-...</p> <p style="font-size: x-small;">Click for more details</p>
<p style="font-size: small;">Overcoming barriers to community participation in a catchment-scale...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">City of Bellingham Natural Yard Care Training Post-Program Participant Survey</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Marketing for Behavior Change and Nutrient Reduction</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Residential Stormwater Survey Public Attitudes, Awareness, and...</p> <p style="font-size: x-small;">Click for more details</p>
<p style="font-size: small;">Chesapeake Bay Social Marketing Initiative 2004-2005 Final Report</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Community Science: Process, Procedure and Analysis of a Community...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">NCCDE Great Schools, Clean Streams 2021 Final Report</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">School Education and Model Assessment - Final Report</p> <p style="font-size: x-small;">Click for more details</p>
<p style="font-size: small;">Maintaining the Health of the Nippersink Creek Watershed</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Ontario Healthy Lawns 2004 Homeowner Survey Report to the Healt...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Inspiring Resident Engagement: Identifying Street Tree Stewardship...</p> <p style="font-size: x-small;">Click for more details</p>	<p style="font-size: small;">Identification and Induction of Human, Social, and Cultural Capitals through a...</p> <p style="font-size: x-small;">Click for more details</p>
<p style="font-size: small;">Think About Personal Pollution (TAPP)</p>	<p style="font-size: small;">Storm Water Pilot Test Evaluation</p>	<p style="font-size: small;">Soak It Up Rain Garden and Native</p>	<p style="font-size: small;">Summary Report MS4 Statewide Awareness</p>

Figure 4: Filtered studies

The screenshot displays a web interface for 'Tools and Resources for Effective Behavior Change Campaigns' in the context of 'STORMWATER & WATER QUALITY'. The page features a navigation bar with links for Home, Add Campaign, Search Campaigns, Campaign Selection, Other Resources, and About. A left sidebar contains a 'Keyword Search' field and two filter sections: 'TARGET POLLUTANT' and 'TARGET AUDIENCE'. The 'TARGET POLLUTANT' section has radio buttons for LID/Infiltration, Metals, Nutrients, Oils, Pathogens (Fecal Coliforms, Bacteria, E. Coli), Sediment, Toxic Chemicals (Pesticide, Household Cleaner, etc.), and Trash. The 'TARGET AUDIENCE' section has radio buttons for Businesses, Contractors, Developers, Engineers, General Public, Land Use Planners, Landscapers, Mobile Businesses, Property Managers, Residents, and School-Age Children. The 'Toxic Chemicals' option is selected. The main content area, titled 'Campaigns', shows a grid of 16 campaign cards, each with a title, a brief description, and a 'Click for more details' link. The campaigns include reports and evaluations from various locations such as Bellingham, Chesapeake Bay, Ontario, and Seattle.

Tools and Resources for Effective Behavior Change Campaigns
IN STORMWATER & WATER QUALITY

Home Add Campaign Search Campaigns Campaign Selection Other Resources About

Keyword Search
 Search... 🔍

TARGET POLLUTANT:

- LID/Infiltration
- Metals
- Nutrients
- Oils
- Pathogens (Fecal Coliforms, Bacteria, E. Coli)
- Sediment
- Toxic Chemicals (Pesticide, Household Cleaner, etc.)
- Trash

TARGET AUDIENCE:

- Businesses
- Contractors
- Developers
- Engineers
- General Public
- Land Use Planners
- Landscapers
- Mobile Businesses
- Property Managers
- Residents
- School-Age Children

Campaigns 🗒️

City of Bellingham Natural Yard Care Training Post-Program Participant Survey Click for more details	Clean Lawn Care for Clean Water: A Collaborative Approach Click for more details	Will Soil Testing and Fertilizer Recommendations Reduce Fertilizer Use? Click for more details	Thurston County Natural Lawn Care Education Evaluation Report Click for more details
Chesapeake Bay Social Marketing Initiative 2004-2005 Final Report Click for more details	School Education and Model Assessment - Final Report Click for more details	Austin Lawn and Garden Chemical Education Campaign Click for more details	Clark County's NPDES Phase I Stormwater Management Effectiveness... Click for more details
Ontario Healthy Lawns 2004 Homeowner Survey Report to the Healthy Lawns... Click for more details	North and South Sound Natural Yard Care Education Evaluation Report Click for more details	Snohomish County Natural Yard Care Education Evaluation Report Click for more details	Community Program #1: Hudson / St. Lazare / Notre Dame, Quebec Click for more details
Community Program #2: Halifax Regional Municipality, Nova Scotia Click for more details	Community Program #3: Hamilton, Ontario Click for more details	Community Program #4: North Shore Recycling Program Click for more details	Community Program #5: Chesapeake Bay, Pennsylvania (Harrisburg) Click for more details
Community Program #6: Seattle and King County, Washington Click for more details	Community Program #8: Frejlev Aalborg, Denmark Click for more details		

The site is intended to be a “living” resource for evaluations of behavior change campaigns in water quality and includes a page for authors of a new study to submit it for inclusion into the database (Figure 5). Submitted information includes contact information, study information (target pollutant, target behavior change), data collection procedures, and campaign materials. The submission will not automatically be added to the database. Instead, it is forwarded to an email address monitored by the website administrator (currently Prof. Cook), who vets the study before including it in the database posted on the website.

Figure 5: Add campaign page

WASHINGTON STATE UNIVERSITY

Tools and Resources for Effective Behavior Change Campaigns

IN STORMWATER & WATER QUALITY

[Home](#) [Add Campaign](#) [Search Campaigns](#) [Campaign Selection](#) [Other Resources](#) [About](#)

SUBMIT DATA

If you know of stormwater and/or water quality behavior change campaigns that we do not have listed, please help us by inputting the details through this form. We will follow up to verify information before publicly posting. If you have any questions or would like to suggest corrections to existing entries in the data please contact us at Joe.Cook@WSU.edu.

To begin, enter info in the tabs below.

DATA USAGE

The data collected here is reviewed by our team and added to the working database.

SECURITY

All data is held within a secured database. Contact information is not shared publicly in any capacity.

GENERAL BACKGROUND **STUDY INFORMATION** **DATA COLLECTION** **DATA VERIFICATION** **CAMPAIGN MATERIALS & INSTRUMENTS**

Jurisdiction Name & Department

Jurisdiction Contact

Other Agencies Involved

Other Agency Contact(s)

Study Name

Report Author(s)

Year Study Started
mm / dd / yyyy

Year Study Completed
mm / dd / yyyy

Upload Report

NEXT →

The second main functionality is a decision support tool. The tool – a downloadable Excel file (Figure 6)- helps the user choose a behavior change campaign based on the jurisdiction’s needs. This might be when the jurisdiction is considering adding a new campaign or moving away from a current campaign, perhaps as required by permit. For example, suppose the user

was interested in starting a new campaign on either pet waste OR reducing illicit discharges, or perhaps two different types of campaigns that both targeted pet waste. The spreadsheet steps through the various components of a behavior change campaign to help the user assess which campaign might be the best choice, including questions around target behaviors, target audience, water quality impacts, situational analysis, equity, community impact and the ability to leverage other resources or partnerships. The tool allows the user to compare multiple types of campaigns (the default is two, but one can always add columns). Different jurisdictions will inevitably weigh these factors differently, so the intention is **not** to have the tool automatically produce an answer by selecting the “best” campaign. Instead, it helps the user make sure they are asking the right questions.

Figure 6. Screenshot of decision support tool (Excel)

	A	B	C
1		Campaign Option One	Campaign Option Two
2	Program Purpose and Problem		
3	1. Program Title: List the title of the behavior change program.		
4	2. Problem Addressed: List the problem the desired behavior change addresses, such as water quality issues due to pet waste.		
5	Target Behavior		
6	1. Target Behavior: Briefly describe the targeted behavior that is to be measured as part of the behavior change program.		
7	2. Is the behavior an “end-state” behavior? An end-state behavior is one that cannot be further subdivided. For example, the behavior “practicing proper dumpster management” can be further divided into other behaviors such as picking up garbage around the dumpster or closing the dumpster lid; therefore, “practicing proper dumpster management” is not an end-state behavior. “Closing the dumpster lid” is an example of an end-state behavior. Select yes or no.		
8	3. Can the behavior be measured? Is there a way to measure the behavior so the effectiveness of the campaign can be assessed? Is it feasible to have a third-party evaluator observe the behavior?		
9	4. What gap or need for action will this behavior change fill? List the gap or need.		
10	Situational Analysis		
11	1. Is data available to show percent of people <u>not</u> engaging in the targeted behavior? Select yes or no.		
12	2. Generally, are most people able to enact the behavior? Select yes, maybe, or no.		
13	3. What are the barriers to enacting the behavior? List the barriers that may prevent someone from enacting the behavior.		
14	4. What is the likelihood that barriers could be addressed as part of the program? Very likely, somewhat likely, not likely.		
15	5. If people were aware of the benefits of doing so, to what degree would people be willing to enact the behavior? Select Very willing, somewhat willing, not willing/resistant.		
16	6. Have prior and similar efforts been undertaken in your community? Select yes or		
17	7. What political issues, if any, are surrounding this work? List any political issues surrounding the potential campaign.		

4.5 Transition Plan for Making Website Sustainable

Finally, we estimated the annual workload and cost for maintaining the site (D4.5) to ensure that the site would be maintained and continue to serve as a “living” repository of behavior change campaign evaluations. This information can be used in future grant applications for sustaining the site, though the report notes that in the short run the project Principal Investigator (Cook) has a 37% Extension appointment which can support some of the

website maintenance. The website is housed in an Amazon Web Services domain controlled by Washington State University's Information Technology unit.

5.0 REPORTING BEHAVIOR CHANGE EVALUATIONS

5.1 Overview of Work Complete

The objective of this task was to develop a template for reporting results for a behavior change evaluation. Our approach to developing the template started with creating interview questions as well as a concept report template and evaluation guidance. The concepts and interview questions were discussed with the TAC as a group during TAC meetings to better understand their needs for behavior change reporting and evaluation guidance. The concepts and interview questions were also sent to Ecology to understand their vision for the basic information they would like included in a report to support compliance with the Municipal Separate Stormwater System (MS4) Permit (MS4 Permit, also commonly referred to as the NPDES Permit) sections Western Washington (WWA) Phase I S5.C.11.a.vi-vii and WWA Phase II S5.C.2.a.ii.(e)-(f). Responses from the TAC and Ecology were used to develop a draft template which went through two rounds of reviews and updates to address and incorporate TAC comments.

Once the draft report template was complete, the TAC was asked to pilot test the document. Specifically, the TAC was asked to compare the report template to a previously completed evaluation to determine if information was missing from the template or if information needed to be added or revised to make the template more useful. We developed a rubric of questions about the report template and asked the TAC to fill out the rubric during their pilot testing. Two TAC members returned a completed rubric and Ecology staff provided feedback on the draft report template. Comments received from the TAC and Ecology were incorporated into the final report template.

The use of this template is not required to meet MS4 Permit requirements. The intent of the template is to streamline report writing by identifying what information is required by the MS4 Permit, providing suggestions for content, and highlighting basic information Ecology informally recommended for the final report. The report template suggested content exceeds the Permit reporting requirements and the additional content was included because they are common evaluation steps that can help Permittees meet their MS4 Permit requirements and [tell a more complete story of their evaluation process](#). Instructions for using the report template are included at the beginning of the document and throughout the report template. Permit requirements are denoted in relevant sections. We also developed a report checklist that outlined the contents of the report to provide template users with an overview of potential content for the report.

5.2 Deliverables Developed

- D5.1 Draft Interview Questions
- D5.2 Summary of Vision for Guidance Documents addressing TAC
- Comments

- D5.3 Summary of Instruments/Data/Analysis Methods & Interview responses
- D5.4 Draft Checklist & Report Template; rubric for pilot testing
- D5.5 Summary of the Results from the Pilot Testing
- D5.6 Final Checklist and Report Template with TAC comments addressed

6.0 EVALUATION GUIDANCE MANUAL

6.1 Overview of Work Complete

The objective of this task was to develop an evaluation guidance manual. The intent of the manual is to provide guidance for evaluating changes in understanding and adoption of targeted behaviors as a result of implementing a behavior change campaign. The manual content was identified from the TAC responses to interview questions and the concept evaluation guidance discussed during a TAC meeting as described in Chapter 5. We then developed a vision for the guidance document which was submitted to the TAC for review and comment. The vision included items such as the manual format, types of possible instruments, data, and analysis methods that the TAC might use to conduct an evaluation. The draft vision was developed through a literature search and review of completed evaluations. Responses from the TAC on the document and during TAC meetings were used to guide the development of the manual content. The draft evaluation manual and draft whitepaper were submitted to TAC for review and the final evaluation guidance manual was updated to address TAC comments. The Evaluation Guidance Manual includes the following chapters.

- **Chapter 1 Introduction**
The purpose of this chapter is to provide an overview of the manual purpose and content, identify relevant permit requirements, provide guidance for how to use the manual, provide an overview and resources for community based social marketing and social marketing resources, and provide an overview of evaluation design.
- **Chapter 2 Sample Size Selection**
The purpose of this chapter is to provide an overview of common methods for selecting sample size for an evaluation.
- **Chapter 3 Instruments**
The purpose of this chapter is to provide an overview of common types of instruments and considerations for selecting appropriate evaluation instruments. In the context of behavior change evaluations, an instrument (e.g., surveys, observational data checklist, etc.) is used to measure the target audience's understanding and adoption of targeted behaviors.
- **Chapter 4 Data Types**
The purpose of this chapter is to provide an overview of common types of qualitative and quantitative data that may be collected as part of an evaluation. The chapter also includes suggestions for organizing data in preparation for data analysis.
- **Chapter 5 Analysis Methods**

The purpose of this chapter is to provide an overview of common method used to analyze both qualitative and quantitative data that can be used to evaluate whether there are changes in the understanding and adoption of targeted behaviors as a result of implementing a behavior change campaign.

6.2 Deliverables

- D6.1 (draft) and D6.2 (final, incorporating TAC feedback) Whitepaper (this document)
- D6.3 (draft) and D6.4 (final, incorporating TAC feedback) Fact Sheet
- D6.5 (draft) and D6.6 (final, incorporating TAC feedback) Evaluation Guidance Manual (located in Section 7.8)

APPENDIX – KEY DELIVERABLES

7.1 SUMMARY OF KEY STAKEHOLDER INTERVIEWS

SAM E&O Behavior Change Project, IAA C2100054

WSU-OCI (Joe Cook PI)

October 7, 2021

Deliverable 2.1: Key informant interview questions and contact list

In each of the key informant interviews, we used the following questions as our base for discussion. Not all questions were covered in all interviews because of time constraints.

- What stormwater behavior change campaigns do you currently run or are in planning?
- How did you decide on which campaign to adopt? How much was driven by a particular pollutant or by prior campaigns?
- How did you design the campaign? How were materials devised? Was information from other jurisdictions or campaigns used?
- Do you evaluate the campaign? If so, how? What metrics do you use to measure the campaign? Is the evaluation used in subsequent decisions? If so, how?
- In your view, what are the key constraints now for E&O campaigns?
- In your view, what are the key opportunities for E&O campaigns?
- What would you most want to learn from or about other jurisdictions' E&O campaigns?
- What would you like to see come of this SAM behavior change project?

We contacted the following key staff to request an interview.

Table 1. Staff contacted and interviewed

Name	Jurisdiction	Phase	Interviewed?
Mary Rabourn	King County	Ph. I	Yes
Andrea Logue	Clark County	Ph. I	Yes
Susan McCleary	Olympia	Ph. II	Yes
Emily Hegarty	Bellingham	Ph. II	Yes
Laura Haren	Kent	Ph. II	Yes
Jason Quigley	Skagit County	Ph. II	No
Cammy Mills	Kitsap County	Ph. II	Yes
Christy Lovelace	Shoreline	Ph. II	Yes
Melanie May	Auburn	Ph. II	No
Jessica Shaw	Wenatchee	Ph. II	Yes
Ann Marie Pearce	Thurston County	Ph. I	Yes
Sarah Norberg	Tacoma	Ph. I	Yes
Peggy Campbell	Snohomish County	Ph. I	Yes
Julie Colehour	C&C	n/a	No
Heidi Keller	Heidi Keller Consulting	n/a	No
Nancy Hardwick	Hardwick Consulting	n/a	No

Deliverable D2.2 Summary of interviews

Why: The purpose of the interviews was to learn more about how behavior change campaigns, required under Phase I and Phase II MSR4 permits, are run by cities and counties in Washington State (particularly the Puget Sound). In particular, we asked about how campaigns are chosen, and how they are administered. See “interview questions” on previous page. The responses in these interviews also helped us to develop and refine questions for a nationwide survey of behavior change staff.

Who: Prof. Joe Cook (WSU School of Economics) and Wisnu Sugiarto (PhD student, WSU School of Economics) conducted eleven interviews in June-August 2021. Each interview took approximately one hour. Many interviewees are considered regional leaders in education and outreach programs and are trained and enthusiastic about the use of “community-based social marketing” approaches in behavior change campaigns.

Six key lessons from interviews

- 1. This SAM project should focus squarely on behavior change (BC) campaigns.** The permits require that jurisdictions create “education and outreach” (E&O) programs to a) build general awareness, b) effect behavior change and c) create stewardship opportunities (Phase I permit section S5.C.11; Phase II Western Washington S5.C.2). In most conversations, we discussed jurisdictions’ awareness programs (e.g. storm drain stenciling, K-12 curriculum) and stewardship programs (e.g. volunteer water quality monitoring). Although it is relatively easy to monitor outputs such as storm drains stenciled or volunteer hours logged, their impact on behavior and water quality is harder to quantify. It is likely that behavior change campaigns benefit from higher levels of overall awareness and pro-environmental attitudes: for example, a pet waste campaign is more likely to induce dog owners to scoop waste when they already understand that water quality is an important concern. Indeed, some interviewees felt that the three components (awareness, stewardship, and behavior change) were intertwined. We hope to investigate whether evidence supports this in our review of the literature, but we plan to carefully distinguish behavior change campaigns from awareness and stewardship programs in our survey work. This is also consistent with our scope of work.
- 2. The specific behavior change campaign chosen by a jurisdiction is often selected due to staffing, budget or history considerations -- rather than the result of a pollutant-drive approach.** Some interviewees mentioned that when their permit cycle required a decision on a behavior change (BC) campaign such as expanding an existing campaign or starting a new one, they (or their predecessor) opted to continue with an existing campaign because it was a “known”. This seemed more likely in jurisdictions with fewer FTE devoted to E&O campaigns. Other interviewees mentioned choosing a commercial dumpster lid-closing campaign because of the regional [Dumpster Outreach Group](#) initiative led by Bellevue. As discussed below, some participants believe regional coordination could make campaigns more effective; others participated because it allowed a small staff to “piggyback” on the efforts of the larger group, including the hiring of a social marketing consultant.
- 3. Interviewees viewed the need for evaluation studies differently.** Some expressed the sentiment that “we know it works”. They welcomed further evaluations to help demonstrate campaign effectiveness and build more political and funding support for behavior change campaigns. Others

felt that behavior change campaigns should be more data-driven and based in quantitative social science (discussed more below) with measurable outcomes and well-designed studies. Some mentioned the importance of complementary regulatory enforcement action as the “stick” to behavior change campaigns’ “carrot”. Phase I and Phase II (Western WA) jurisdictions are now required to “follow social marketing practices and methods, similar to CBSM, and develop a campaign that is tailored to the community, including the development of a program evaluation plan.”. Several interviewees saw this requirement as an important opportunity for the field to conduct more high-quality evaluations. Some interviewees mentioned the need for financial and political support to do more long-term follow-up studies to examine whether campaigns caused durable behavior change. Such studies are rare.

4. **Regional collaboration on BC campaigns is both valuable and valued.** The EPA’s Phase II [factsheet](#) encourages this regional collaboration, and it is specifically mentioned as a compliance option in the permits. As already noted, some jurisdictions felt they could benefit from economies of scale in implementing the regional dumpster campaign. As was done in the regional “Don’t Drip and Drive” or dumpster lid campaigns, materials could be developed as a group and the costs could be shared. Another interviewee pointed out that since many residents move between Puget Sound jurisdictions for home, work and recreation, regional collaboration can enhance the effectiveness of messaging and campaigns. This may also be true for the dumpster-lid campaign since businesses may have locations in several jurisdictions.

Regional collaboration also raises the possibility for creative evaluation strategies, as could be used in the dumpster-lid campaign. For example, if all implementing jurisdictions included comparison groups (discussed more below), the regional team could explore different campaign configurations or messaging alternatives in different jurisdictions and assess which had the largest impact. Alternatively, if allowed by permit, campaign implementation could be staggered over time, with monitoring happening both in the jurisdictions where the campaign is underway and those where it is planned in the future. The later-adopting jurisdictions could then serve as a control group used for comparison. The demise of the [Sound Behavior Index](#), implemented by Western Washington University and Puget Sound Partnership, was mentioned by one interviewee as an important missed opportunity for tracking behaviors across the region over time. One interviewee saw a potential downside in regional collaboration if the decision-making process for which campaign to select is not transparent and inclusive.

Finally, many interviewees mentioned that Stormwater Outreach for Municipalities (STORM) is a valuable venue for sharing expertise, including the group’s library.

5. **The quality of existing evaluations could be improved, but this will likely require additional financial and staff resources.** This is a tentative conclusion based solely on conversations and may evolve as our team evaluates public-available evaluations from WA and nationwide. There are three core concerns.

The first is **credible outcome measures**. This was brought up most frequently by interviewees as a challenge. Many studies rely on self-reported behavior, and this is indeed the only path possible for

many types of BC campaigns where it is infeasible to directly observe behavior, like pet owners picking up poop in private yards or contractors dumping carpet cleaning chemicals down the toilet of a residential home. But other studies show that observation is possible but can be labor-intensive. The pet waste campaigns in both Kirkland and Clark County employed direct observation (counts) of poop to examine whether campaigns to provide messaging, poop bags, etc. were effective and for how long. Some jurisdictions in the regional dumpster initiative have staff driving by businesses that have received the BC messaging to directly observe whether dumpster lids are closed. According to social marketing principles, outcome measures should also be non-divisible and “end state” (engaging in the behavior produces the desired outcome).

The second area is tracking **counterfactuals or “control” groups**. Control groups came up during discussions of program evaluation in roughly half of the interviews, particularly among those who had more expertise in CBSM or training in natural sciences. Most evaluations collect baseline data on participants to measure how much the campaign changes behavior. But collecting before- and after- data for comparison groups who did not participate in the campaign helps rule out that other factors that happened concurrently with the BC campaign are what led to behavior change. For example, a natural yard care campaign might ask residents to reduce fertilizer use because of phosphorus runoff contributing to harmful algal blooms in a nearby lake. It might do this through yard care workshops or incentive campaigns with the target group and not with the control group. Suppose, by chance, an algal bloom happened at the same time as these BC activities. Even non-participants (the control group) might reduce fertilizer use because the issue of lake pollution was suddenly much more salient, particularly if the bloom received widespread media attention. Evaluators of the BC campaign might over-estimate the impact of the BC workshops and incentive campaigns.

The third area is that **campaign participants may not be a representative sample of the overall population**. From an evaluation perspective, this is important because it is likely that people who participate in workshops, for example, are those who are already amenable to the change in behavior, perhaps because of pro-environmental attitudes. The risk is in assuming that a campaign that successfully changed their behavior is scalable to a larger section of the population. For example, some interviewees mentioned that some natural yard care projects tend to attract primarily older and relatively affluent homeowners. One interviewee mentioned another reason it is important: the intersection with equity concerns. This interviewee mentioned the need for BC campaigns to include more audiences of color, those with lower-incomes or those who don’t speak English as a first language (which are both mentioned in the NPDES permit). These groups, and local organizations representing them, could also be more involved in choosing which types of behavior change campaigns to implement.

The **benefits of improved evaluation** are clear. One interviewee mentioned that the phrase “measurable impacts” is very “sellable” to elected officials and managers. Another described a virtuous circle where a manager was willing to risk “failing” by pairing a high-quality evaluation with a novel CBSM approach. When the evaluation showed success, it was easier to make the case for hiring more staff with CBSM experience.

But the **costs of evaluation** are also clear. Nearly all interviewees mentioned the staff time necessary to do evaluations, and in some cases the lack of training in techniques. Improving evaluations requires additional resources, which can be difficult to advocate for. One interviewee mentioned that receiving a grant that specifically required more rigorous CBSM methods “tied my hands” to enable a higher-quality evaluation that the interviewee’s boss didn’t support. Another interviewee mentioned an inherent challenge in explaining CBSM concept to upper-level managers and elected officials. CBSM works by focusing on a very specific behavior in a very specific audience by addressing specific barriers to change, but elected officials naturally want campaigns that benefit as many of their constituents as possible. Returning to an earlier point, there is also the possibility that evaluations will show that some BC campaigns are not working or are not very cost-effective ways of improving water quality.

6. Finally, **BC campaigns seem to heavily rely on consulting firms for advice.** Many interviewees mentioned using the services of a consulting firm, particularly to help with campaign evaluation and surveys. Given the shortcoming of existing evaluations described above, it might seem that these firms could be providing more sound advice. We also heard, however, that consultants often encourage jurisdictions to do more rigorous evaluations, but these are ultimately ruled out because of staff time requirements. Improving access to online tools and repositories like [STORM's](#) or [EPA's](#) may reduce reliance on using consultants to design campaigns, and we hope the guidance provided as part of this SAM project will help jurisdictions conduct high-quality evaluations on their own or determine when a consultant and higher-level evaluation is needed. Program evaluation is also now a core skill taught in Masters of Public Administration and Masters of Public Policy programs. Collaboration with other city or county staff with training (but with no knowledge of social marketing, stormwater or water quality) may also enable better program evaluation without a substantial increase in staff time.



WASHINGTON STATE UNIVERSITY
EXTENSION

Practices and challenges in stormwater behavior change programs:

A survey of U.S. professionals

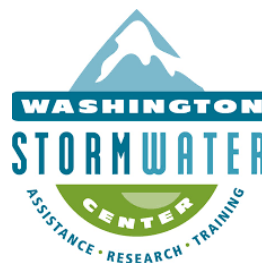
Wisnu Sugiarto and Joseph Cook

School of Economic Sciences

Washington State University

FINAL REPORT

February 27, 2022



Please cite as “Sugiarto, W. and J. Cook. “Practices and challenges in stormwater behavior change programs: a survey of US professionals”. Washington State University Stormwater Center. December 2021.

Background and purpose of this report

We conducted a nationwide survey of staff at public agencies (cities, counties, conservation districts) who conduct behavior change campaigns in stormwater or water quality. This survey was part of a larger project on these campaigns funded by the [Stormwater Action Monitoring](#) (SAM) program (funded by permittees and administered by the WA Department of Ecology). By “behavior change”, we mean programs that seek to get people to change a concrete behavior, such as picking up pet waste or reducing fertilizer use, rather than other “education and outreach” programs that seek to communicate information or change attitudes or programs that provide stewardship and volunteer opportunities. The objective of the survey was to understand how staff chose behavior change programs, how they evaluated them, and the key constraints and opportunities in the increased use of these tools. It was also intended for staff who oversaw consultants who designed, implemented or evaluated behavior change programs.

This report details the procedures we used to collect survey responses and highlights key results. Additional information is provided in Appendix C to this report, which shows the exact questions asked and provides a comprehensive set of response statistics.

Survey elicitation

We know of no possible sample frame (i.e. a list of all behavior-change personnel nationwide) and it was out of our scope of work to construct one. Rather than randomly select participants from such a frame, we sought to solicit as many participants as possible using an open Qualtrics web survey. We disseminated the survey link through Ecology’s listserv, national stormwater associations, and key personnel. We also encouraged word-of-mouth spread of the survey link. We solicited contacts from the Washington State Municipal Stormwater Conference (MuniCon). It is likely that our survey responses suffer from a selection bias and may not be representative nationwide. We would note, however, that we received responses from a majority of regulated entities in western Washington, so our results are somewhat more likely to be representative of that population. The survey opened on October 20, 2021 and ended on November 20, 2021.

Responses and demographics

We collected 224 total responses; 31 of them were discarded because the respondents were not staff members who worked at a city, county, watershed district or conservation district on behavior change campaigns in stormwater runoff reduction or water quality improvement. Among those who reported being qualified to take the survey, 8 respondents did not grant research consent, ending the survey for them. Additionally, we had one test-run after the survey was publicly launched and the test-run response was also excluded from our analysis. Finally, we had 21 responses that reported being qualified and consented to participate in the survey, but no other questions were answered. We excluded the 21 responses from our analysis since they did not yield any information for the study. As a result, we were left with 163 responses that met qualification, consented, and provided information. Survey participants were able to skip questions at any time or submit an incomplete survey, which means the number of

responses varies by question. Our survey respondents came from 19 states. The U.S. map below portrays the number of survey respondents by state with the darker shades representing a greater number of participants. Almost half (78 of 160) were from Washington, with 48 districts or local government agencies represented (refer to Appendix B for the complete list).

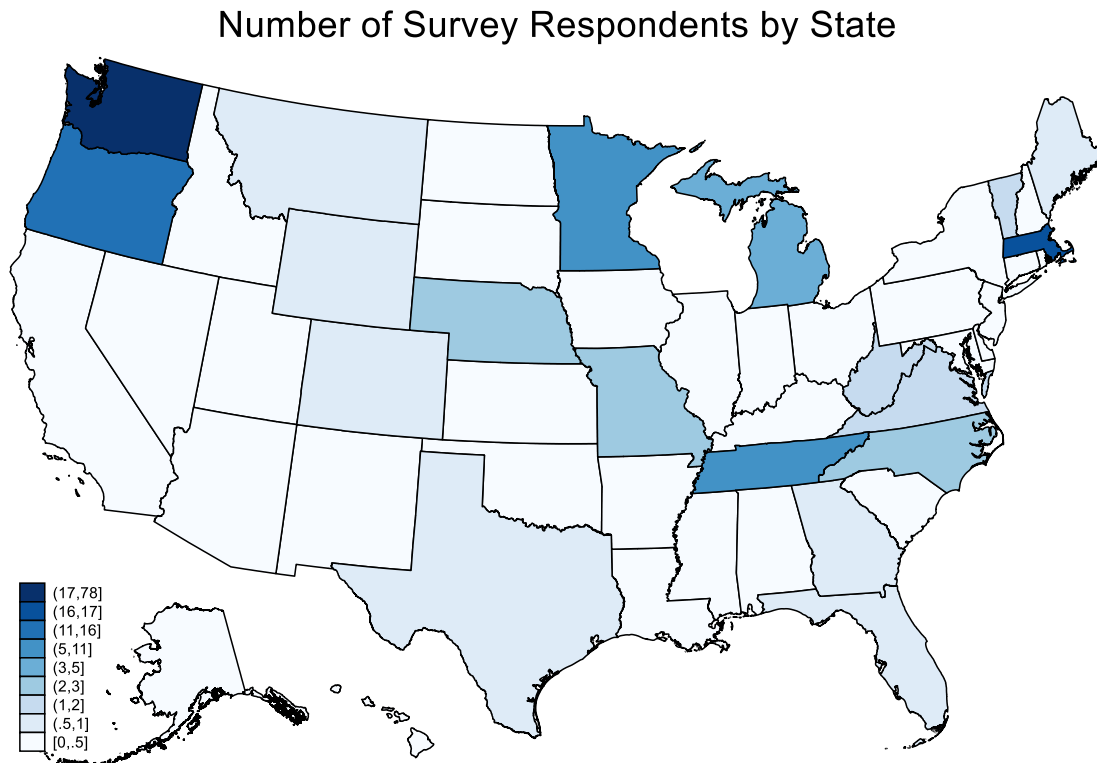


Figure 1: National Distribution of Survey Respondents by State

Half of respondents (56 of 111) had been working on behavior change campaigns in stormwater runoff reduction, water quality or any topic at their current employers for at least five years; 25% (28 of 111) had been working for at least ten years at their current employers. Forty-six percent (49 of 106) of respondents had been working on behavior change campaigns in stormwater runoff reduction, water quality or any topic over the course of their careers for at least ten years. Fourteen percent (15 of 106) had at least 20 years of experience. Over half of respondents (59 of 115) had a bachelor's degree, and 39% (45 of 115) had a master's degree. Twenty-four percent (27 of 112) studied *environmental science, ecology, biology*; 22% (25 of 112) had an *engineering* background; 21% (23 of 112) were in the *field of environmental studies, natural resources*.

Eight key findings

#1 Behavior change campaigns on stormwater reduction or water quality improvement involve multiple views on defining what a behavior change campaign is.

Our definition of behavior change (see above) was clearly stated in the survey preamble and had been reviewed by our expert stormwater colleagues on the Technical Advisory Committee. Nevertheless, we found that some respondents had a broader definition of what a behavior change campaign is than what we defined in the survey preamble. As one individual stated, “In our programming, we do not make a clear-cut distinction between awareness-raising and behavior change campaigns...”. Seven respondents also specifically mentioned *Adopt a Drain* programs, which we would consider a stewardship program that provides volunteer opportunities. A responder specifically wrote “I realize that you may not consider stormdrain adoption as a behavior change program since it does have an element of volunteerism in it; however, the goal of the program is really to get people to participate in the specific behavior of cleaning their storm drains regularly to prevent pollution and flooding...” It is also important to note, however, that we do not separate responses from people who may have defined a behavior change program more broadly.

#2 Combatting pollutants from pet waste and yard care behavior appear to be the priorities in recent behavior change campaigns.

We asked survey respondents to focus on the program that they evaluated most recently if they had evaluated more than one program in the past 5 years. Twenty-four percent (22 of 93) selected *pet waste management and disposal* program; 20% (19 of 93) chose *yard care techniques protective of water quality*. The third most popular evaluation was on *dumpster and trash compactor maintenance* with 8% (7 of 93). We asked stormwater managers which factors were most important in choosing what campaign to implement (Figure 2). Among 102 respondents who described the importance of targeting a specific pollutant or contaminant of concern, 62% (63 of 102) selected *very important* and 27% said *somewhat important*. The second most (57 of 104) *very important* consideration in choosing what campaign to implement was the targeted behavior was something that the responders and their colleagues noticed and believed to be a problem worth addressing. On the other hand, respondents were least likely to say that continuing an existing campaign was an important consideration.

Additionally, we asked a question on which pollutant or contaminant respondents focused on, and 135 participants responded. Thirty-two percent (43 of 135) chose *bacteria (e.g. coliforms)* and 29% (39 of 135) selected *nutrients (e.g. phosphorus)*. We found that fewer respondents, 14% (19 of 135), focused on *toxic chemical (pesticide, household cleaner, etc.)*, 9% (12 of 135) focused on *automotive-related pollutant (tire, oil leaks, cleaning products, etc.)*, and a few people (7 of 135) focused on *heavy metal*. Figure 2 shows survey responses on the most important reason for a pollutant or contaminant to be selected. The most important reason was because it was *listed in the TMDL of a receiving waterbody in their jurisdiction* - 35% (30 of 86). The second most important reason was based on *local water quality data indicating the pollutant was a concern* with 28% (24 of 86) respondents.

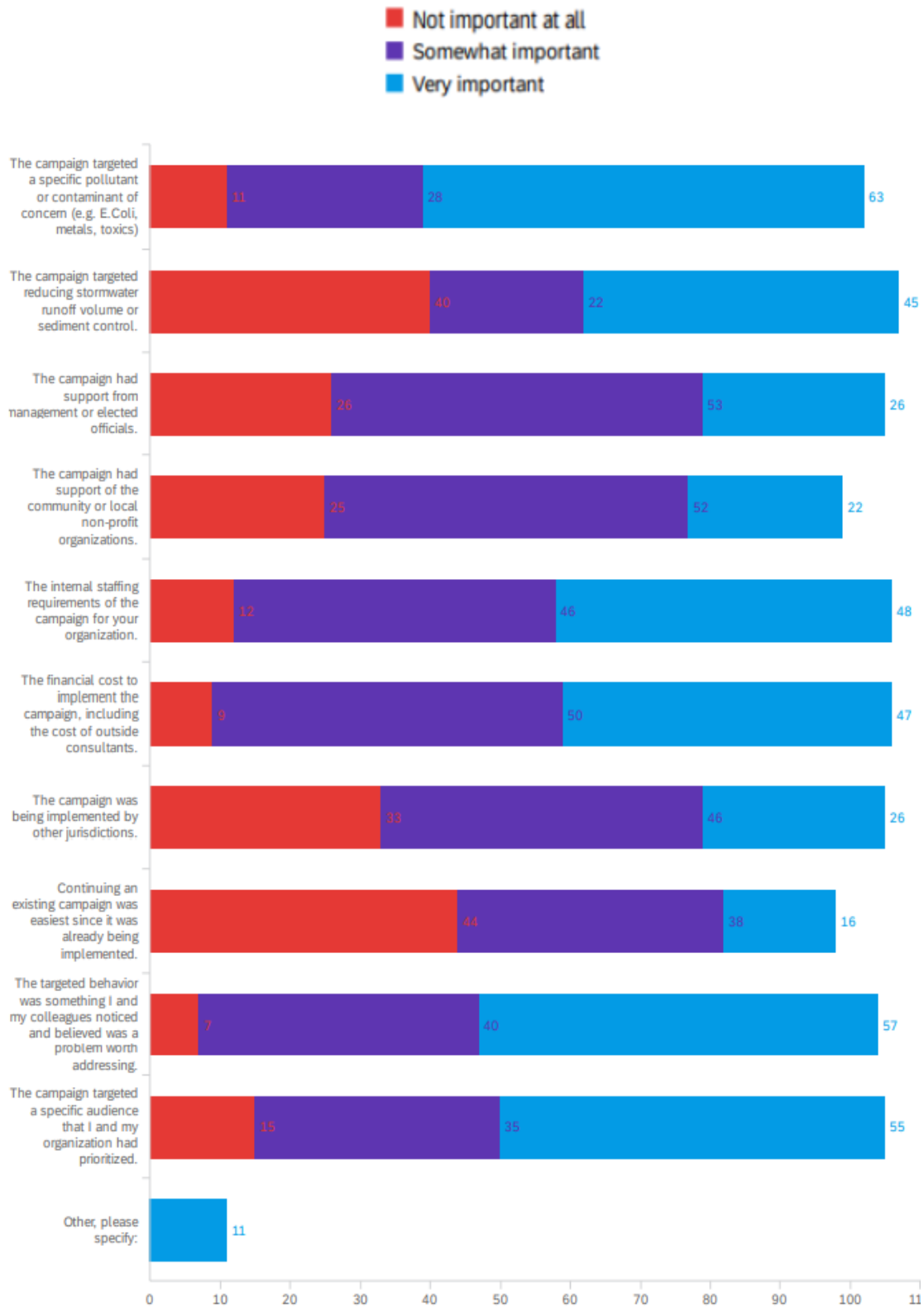


Figure 2: Level of importance of factors that respondents and their colleagues might have considered in choosing a campaign to implement.

Looking at Washington state respondents only, 41 people answered the focused pollutant or contaminant question. The most important reason was also because of TMDL listing which was selected by 24% (10 of 41) of respondents, followed by local water quality data which accounted for 22% (9 of 41) of respondents.

#3 Additional training on social marketing and program evaluation will be helpful in areas where NPDES permits require social marketing.

The majority of respondents, 75% (95 of 126), are required by their NPDES permit to conduct an evaluation of at least one of their behavior change campaigns. However, 38% (44 of 116) had not been trained in “social marketing or community-based social marketing (CBSM)”, 66% (77 of 116) had not had training in program evaluation, and 35% (41 of 116) had not been trained in either. We wanted to gauge respondents’ experience in program evaluation because stormwater professionals who might not have expertise in social marketing could be familiar with methods necessary to help conduct a high-quality impact evaluation, such as using counterfactuals, creating proper survey design, and conducting robust statistical analysis. Only 4% (3 of 69) and 3% (1 of 36) of respondents reported having *expert* skills in social marketing or CBSM and program evaluation, respectively. Twenty percent (14 of 69) and 25% (9 of 36) were *advanced* in social marketing or CBSM and program evaluation, respectively.

We also asked respondents what programs they would find most helpful if training programs were to be offered. Figure 3 reports distribution of respondents’ rankings of training programs. Training programs in *evaluation strategies and metrics* were thought to be the most helpful, as they were the most commonly ranked in the top three. The second most helpful training would be training in *communication strategies*, as one respondent stated “it is very hard to find current behavior change campaign trainings (i.e. have been updated to include the ever changing social media tools, online ads, and other new tools). The regulators are also often not current new methods, processes, and procedures which causes ineffective and/or costly permit requirements that hinder novel approaches.”

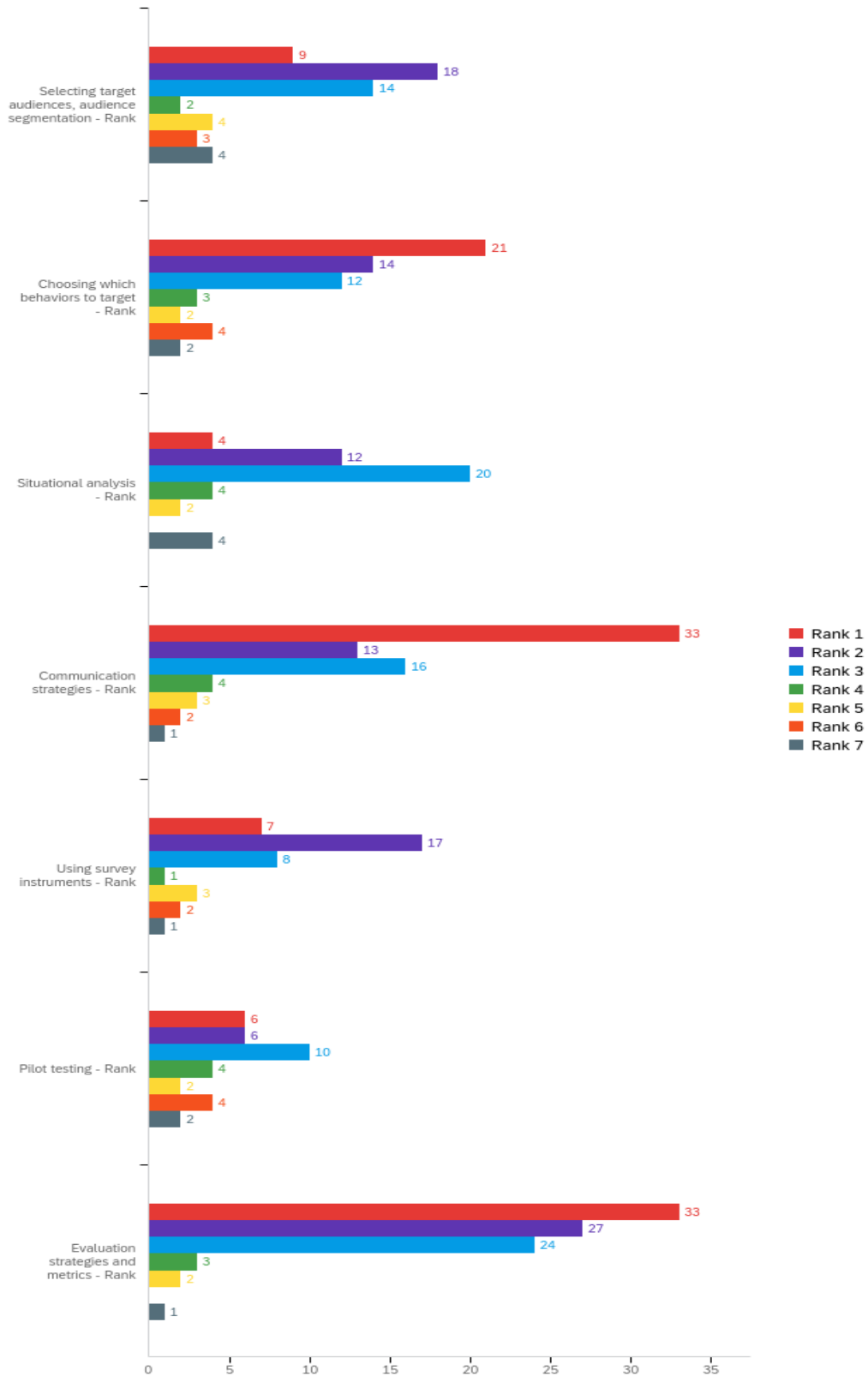


Figure 3: Ranking of preferred training programs

#4 Stormwater professionals feel supported by their organizations, but appear to be short-handed, have competing responsibilities, and need more resources.

Respondents generally felt supported by management in their organizations with 35% (40 of 114) reporting managers were *very supportive* and 33% (38 of 114) reporting *somewhat supportive*. Twenty-one percent (23 of 112) felt local officials were *very supportive* and 24% (27 of 112) felt they were *somewhat supportive*. Twenty-eight percent (31 of 112) felt regulators were *very supportive* and 31% (35 of 112) felt they were *somewhat supportive*. Overall, respondents felt that management in their organizations (35%) were relatively the most *very supportive*, followed by regulators at second (28%) and local officials at last (21%). The following Figure 4 reports the breakdowns in more details. We also asked respondents about obtaining additional funding for a current campaign or a new one, 66% (74 of 113) felt that it was difficult (18% *very difficult* and 48% *somewhat difficult*). One respondent noted, “Our storm water department is severely underfunded, and it is no secret that if the county mayor had a choice the department would not exist. The mindset needs to change beginning at the top in order for the program to receive proper funding and support to do what is needed to really make an impact in our community.”

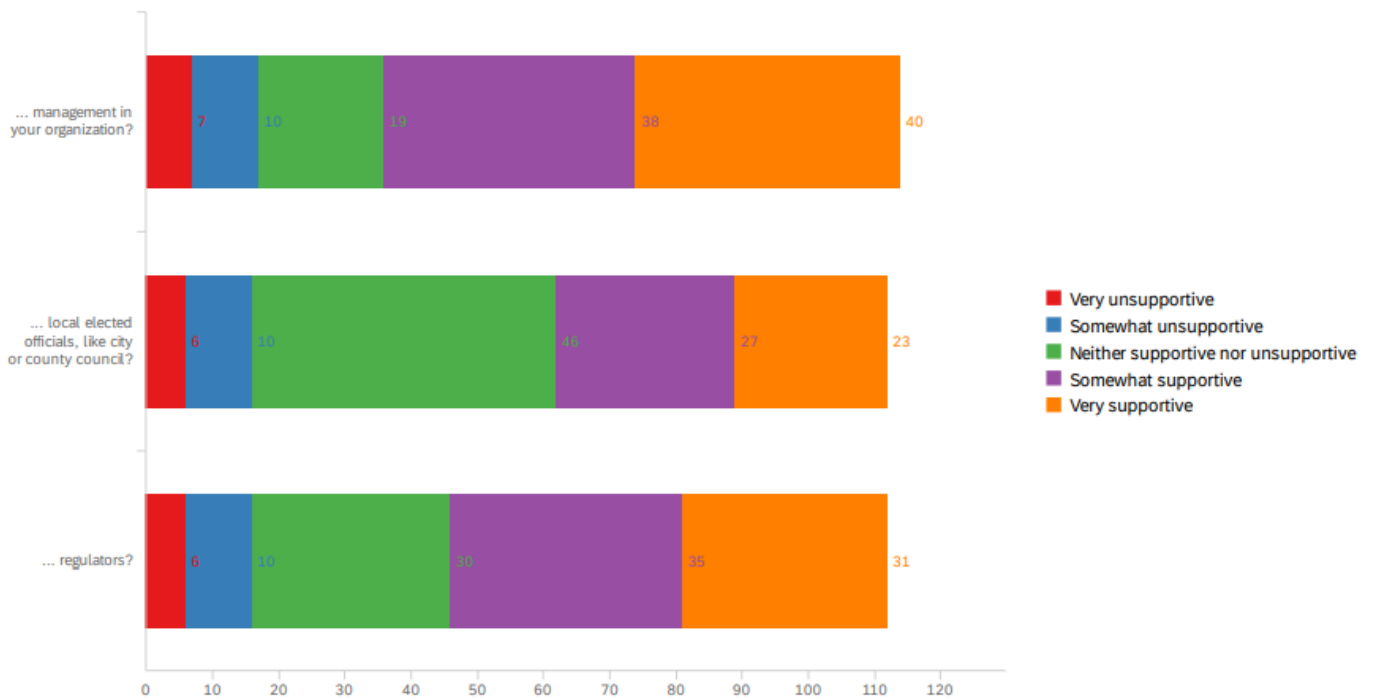


Figure 4: Level of support perceived by respondents

We also asked respondents how many full-time staff (expressed as “full-time equivalents” or FTEs) worked in their organizations on behavior change campaigns on stormwater runoff reduction or water quality improvement. Thirty-eight percent (61 of 160) reported working in organizations with less than one FTE, 33% (53 of 160) reported 1 FTE, and 13% (20 of 160) reported 2 FTE. Forty-five percent (52 of 115) spent less than 10% of their time on behavior change campaigns around stormwater runoff reduction or water quality improvement, and 30% (34 of 115) allocated 10-25%. Meanwhile, most respondents, 92% (105 of 114), also worked on building awareness, fostering stewardship, or both programs. Thirty-seven percent

(39 of 105) spent less than 10% of their time on awareness and stewardship programs, and 35% (37 of 105) used 10-25% of their time for awareness and stewardship programs. Additionally, there were multiple respondents that commented on the difficulties with having insufficient financial resources and personnel. One respondent stated “staffing and funding are our biggest challenge. We have lots of great ideas, but little money or staff to make it happen.” In addition, another respondent wrote “...We tend to do the bare minimum required to meet our MS4 permit requirements. Funding & staff capacity are our biggest restraints.”

#5 Program evaluations on behavior change campaigns could be improved.

We asked whether respondents collected baseline data and data on a comparison group as evaluation measures. Fifty-one percent (49 of 96) did not collect baseline data, and 80% (77 of 96) did not collect data on a comparison group that was not exposed to the campaign materials. A respondent stated “...One of the troubles we ran into in my opinion is a lack of baseline data, a lack of confidence in the with the team that there was value in the campaign and confidence that the behavior change mattered in the big picture...” Another respondent mentioned “my experience working with other professionals is that few understand social science statistics and how to interpret them or appropriate survey design. My other experience is that use of controls is very difficult and expensive. Also that stormwater “impact” in terms of pollution is impossible to measure...” These findings from the survey are consistent with the information that we collected from a review of the literature, where we found that 38% of 47 studies identified did not collect any baseline data, and 89% did not use a comparison group. While the use of baseline data was more common than control groups, both evaluation components could be applied more frequently to make behavior change campaign evaluations more robust.

#6 Staff face challenges with accessing external resources in behavior change campaign implementation

We were interested in learning whether and how stormwater behavior change professionals used external resources. Considering many staff members were untrained in social marketing or CBSM, program evaluation, or both (finding #3), one alternative is to hire external consultants. Fifty-six percent (61 of 110) of respondents had a list of qualified consultants that they could reach out to for help regarding campaign implementation or evaluation. However, jurisdictions might not have sufficient funding for hiring consultants. Most respondents, 79% (103 of 131), did not hire external consultants to help them choose which behavior change campaign to implement and 58% (57 of 98) did not hire any consultants to help with evaluation. While over half of respondents reported having a list of qualified consultants that they could reach out to for help, many respondents did not, with 44% (49 of 110) saying *no*. In addition to external consultants, online materials from the EPA’s Non-Point Source Toolbox are also available to help stormwater professionals in the development, implementation or evaluation of behavior change campaigns. However, 56% (62 of 111) had never used any materials from the online resource. One reason is because staff members might not be aware of the online materials, as one respondent stated “...I wasn’t aware of the EPA Outreach Toolbox...” Another possible explanation is that staff members did not have enough time to explore external resources. They were short-handed and had competing responsibilities (finding #4).

#7 Collaborations with other jurisdictions are common in stormwater runoff reduction or water quality improvement behavior change campaigns

Most organizations have collaborated with other jurisdictions in designing or implementing behavior change campaigns: 79% (88 of 112) of respondents reported coordinating with other jurisdictions in the past five years. As we can see on Figure 5, the most common reason for collaborations was to create consistent messaging across jurisdictions and increase the campaign’s impact by triggering regionally normative behavior (43% of respondents). The second most important reason was to share financial costs such as hiring a common consultant, 25% (22 of 87) of respondents, followed by 16% (14 of 87) said to share expertise. One respondent wrote “All of our campaigns have been through partnership. Small jurisdictions do not have the capacity to implement these campaigns alone. They are very time intensive and typically comprise only a small fraction of the staff’s workload.”

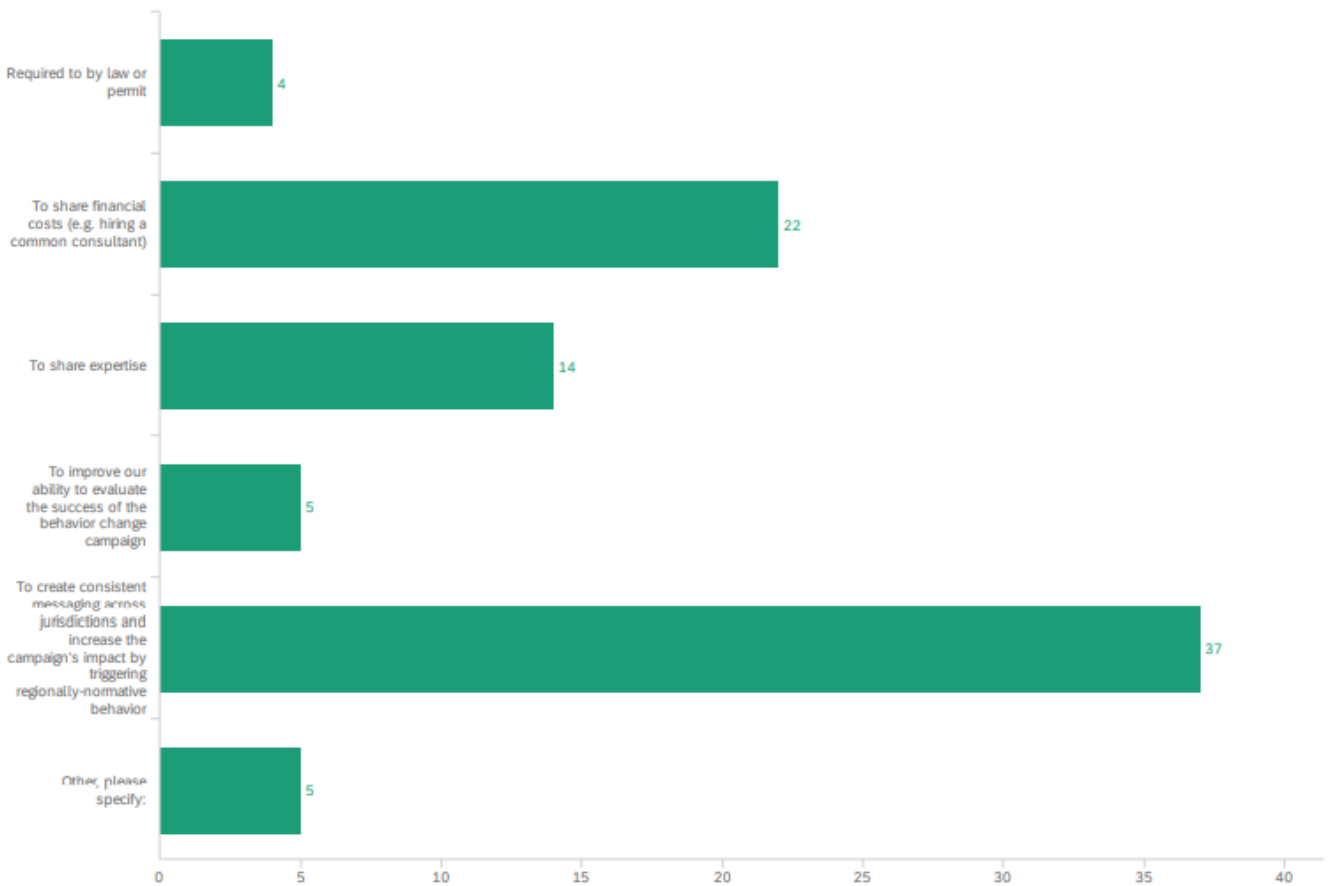


Figure 5: Reason for jurisdiction collaborations

#8 Permit requirements may need adjustments to account for different jurisdictions' sizes

We noticed some of the write-in comments were related to the need for differentiating the permit requirement between small- and large-size jurisdictions. For example, one respondent stated “Behavioral change campaigns are a challenge for small jurisdictions. There is limited staff and financial resources..., these campaigns should be handled at the regional level.” Another respondent wrote “This permit requirement is much better suited to large jurisdictions.” The comments prompted us to conduct some comparisons between large-size jurisdictions and the overall findings. While there were no specific survey questions on determining each jurisdiction’s size, we used the number of full-time equivalents (FTE) employed as a proxy for an organization’s size. We assumed that the respondents who reported working in organizations with 3 or more FTE were in large-size jurisdictions.

Focusing on the presumably large-size jurisdictions, we found 60% (9 of 15) felt that it was difficult (7% *very difficult* and 53% *somewhat difficult*) to obtain additional funding for an existing or a new campaign. The percentages are relatively low when compared to the overall findings, particularly the *very difficult* responses. Additionally, 24% (4 of 17) of respondents spent less than 10% of their time on behavior change campaigns around stormwater runoff reduction or water quality improvement. No respondent reported spending less than 10% of their time on awareness and stewardship programs. The preceding information suggest that organizations with 3 or more FTE also appear to have more resources than the overall findings. However, we needed to be cautious with the conclusion since our analysis was based on a small sample of 26 respondents.

Next steps

The results from the survey will inform future deliverables for this SAM project, including the development of an evaluation training manual and a web-based tool to help professionals select stormwater behavior change programs. For more information on the survey or to request access to de-identified survey data, please contact Prof. Joe Cook (joe.cook@wsu.edu).

Note that the full 91-page Appendix is provided in the version of this deliverable posted to the SAM site: https://www.ezview.wa.gov/Portals/_1962/Documents/SAM/D2.4%20FinalSurveyReportFeb272022.pdf



Prioritizing stormwater pollutant risks: an annotated bibliography

Anand D. Jayakaran, Ph.D., P.E.
Washington State University

September 6, 2022



This annotated bibliography is part of a larger project on behavior change campaigns funded by the Stormwater Action Monitoring (SAM) program (funded by permittees and administered by the WA Department of Ecology).”

1 Introduction

In three steps, one can think of the entire ‘theory of change¹’ for education and outreach (E&O) programs to protect receiving water quality. First, an E&O/behavior change program leads to an observable change in behavior of households or firms. Second, those household and business behavior changes reduce pollutant loading to streams, rivers, and estuaries. Third, those reduced pollutant loads improve water quality in the environment. These three steps are predicated on a methodology to infer where and what in a watershed to focus. Historically, the implementation of the Clean Water Act (1972) and its regulatory framework, the NPDES² program, included 303d listings of impaired water bodies and meeting the requirements of subsequent TMDL³ formulations. Within Washington State, the Municipal NPDES Permit applies to a jurisdiction’s MS4⁴ that discharge to state and federal water bodies. The requirement to meet the E&O portion of the Permit applies regardless of whether a water body has a 303d listing or an assigned TMDL.

The NPDES framework, however, is arguably restricted by the pollutants that are being measured and how various states choose to interpret it. An ACWA⁵ report from 2018 (Berckes et al., 2019) on responses from state representatives implementing 303(d) / TMDL programs suggested that many states focused primarily on bacteria and nutrients - with bacteria considered as ‘relatively *easy* work to accomplish based on established methods of calculation.’ Another key finding from that report was that in some states, economic value was used to prioritize 303(d) programs, while human health was used as a priority metric in other states. The report notes that efforts to value clean water from the perspective of biota or an environmental justice lens were universally lacking.

This annotated bibliography provides an overview of stormwater pollutant prioritization by summarizing several critical studies from the last 15 years. The annotated bibliography is based on peer-reviewed studies that focus on the issues at a national level. However, we believe those studies’ outcomes are relevant to Washington state jurisdictions. This document is

¹A theory of change is a methodology that informs the understanding and explaining of how change takes place and how specific interventions can lead to desired outcomes and goals.

²National Pollutant Discharge Elimination System

³Total Maximum Daily Load

⁴municipal separate stormwater systems or public stormwater system

⁵Association of Clean Water Administrators

not intended as a substitute for reading the studies summarised here or others. Instead, the intention is to create a road map of significant works that have emerged in the last few decades. In addition, a short review section of some studies that outline the connection between environmental justice and prioritizing pollutants is presented as a second part of the annotated bibliography. This additional review section illustrates the need to ensure that under-resourced communities that have long borne the brunt of stormwater pollution must factor into any pollutant mitigation plan.

2 Prioritizing Pollutants

2.1 2007 - Prioritizing pollutant risk - a European framework

Eriksson, E., Baun, A., Scholes, L., Ledin, A., Ahlman, S., Revitt, M., Noutsopoulos, C., and Mikkelsen, P. S. (2007). Selected stormwater priority pollutants—a European perspective. *Science of the Total Environment*, 383(1-3):41–51

Eriksson et al. (2007) developed a framework to prioritize stormwater pollutants for a European 5th Framework Project named DayWater. The goal of this framework was to inform a decision support system for developing stormwater source control measures to effect sustainable stormwater management. They used a version of the Chemical Hazard Identification and Assessment Tool (CHIAT), which outlines five steps to identify relevant priority pollutants. The five steps are:

1. Source characterization - over 650 organic compounds, 30 metals, and inorganic trace elements were identified
2. Recipient exposure targets and criteria identification - surface water was designated as the recipient, and aquatic organisms and humans (secondary) were identified as exposure targets.
3. Hazard identification - physicochemical properties of the identified pollutants and their environmental fate were categorized. A focus was placed on PAHs⁶, pesticides, and XOCs⁷.
4. Hazard assessment - this step was excluded per the authors' explanation that hazard assessment was not the focus of this work, but pollutant identification was.

⁶Polycyclic Aromatic Hydrocarbons

⁷Xenobiotic Organic Compounds

5. Stakeholder involvement - three stakeholder meetings were held, each building on outcomes of the previous meeting. The meetings identified specific pollutants, grouped and selected representative pollutants, and finally selected water quality analytes, including metals and organic compounds.

Ultimately, 25 priority pollutants were selected, including 6 water quality parameters (BOD⁸, COD⁹, SS¹⁰, N¹¹, P¹², pH), 7 metals, 3 PAHs, 4 herbicides, and 5 miscellaneous organic compounds. Eriksson et al. (2007) state that their list of priority pollutants was intended for use as a framework for risk and hazard assessments, a basis for comparing stormwater BMPs¹³, and to inform stormwater monitoring programs.

2.2 2012 - Prioritizing the risk posed by stormwater pollutant sources

Lundy, L., Ellis, J. B., and Revitt, D. M. (2012). Risk prioritisation of stormwater pollutant sources. *Water Research*, 46(20):6589–6600

The study by Lundy et al. (2012) sets up another framework for prioritizing pollutant risk - grouped by land use and land cover. They propose evaluating pollutant risk in terms of estimates of the likelihood of occurrence and the severity of its impact. They analyzed data from other studies on loading rates and sources of various pollutants, combined with matrices that assessed the degrees of likelihood of: A) occurrence in stormwater, and B) levels of consequence posed by that pollutant. This work from a decade ago identified road surfaces as the primary source of pollutants and likely a critical space for intervention in the urban ecosystem. They recognized the need for mitigation practices before road runoff enters roadside ditches or piped drainage networks. Of the four pollutants considered by Lundy et al. (2012) - TSS¹⁴, BOD, cadmium, and lead - TSS was considered the pollutant that posed the most significant risk, followed by cadmium. BOD posed the lowest risk to downstream receiving waters. The paper concludes by recognizing the difficulties posed with removing TSS from roadway surfaces, suggesting

⁸biological oxygen demand

⁹chemical oxygen demand

¹⁰suspended sediments

¹¹nitrogen

¹²phosphorous

¹³Best Management Practices

¹⁴total suspended sediments

the use of structural stormwater BMPs - such as infiltration trenches, infiltration basins, and sub-surface flow constructed wetlands - between the road surface and before discharge into receiving waters.

2.3 2016 - Prioritizing heavy metals risk

Ma, Y., Egodawatta, P., McGree, J., Liu, A., and Goonetilleke, A. (2016). Human health risk assessment of heavy metals in urban stormwater. *Science of the Total Environment*, 557:764–772

Ma et al. (2016) looked specifically at the risk that heavy metals in urban stormwater posed to human health. They developed a Hazard Index (HI) derived from traffic and land use metrics because traffic and land use were presumed to be the greatest sources of heavy metals in an urban landscape. The authors state that in 2016, there was no reliable methodology available to quantify the risk posed by heavy metals emanating from these two sources. The study involved vacuuming street and roof dust samples, conducting laboratory analyses of those samples, and assessing the risk to humans through three pathways (stormwater ingestion as drinking water, ingestion of stormwater while swimming, and dermal contact.) Using these methods, they developed a model for heavy metal build-up based on daily traffic volume and land use (as covariates), and a human health risk model based on a hazard index related to total heavy metals and fine solids heavy metals. In order of decreasing risk, they showed the following heavy metals posed risk to human health: *chromium* > *manganese* > *lead* > *aluminum* > *iron* > *cadmium* > *zinc* > *copper* > *nickel*. The authors showed that individual heavy metals do not pose as much risk as a mixture of multiple heavy metals, noting that even low concentrations of chromium, manganese, and lead¹⁵ are extremely toxic so more importance should be paid to a hazard index as opposed to actual concentration values. They also showed that traffic volume was the most significant driver of health risks associated with heavy metals in stormwater.

2.4 2017 - Prioritizing toxic metals and PAH risk

Ma, Y., Liu, A., Egodawatta, P., McGree, J., and Goonetilleke, A. (2017). Assessment and management of human health risk from toxic metals and polycyclic aromatic hydrocarbons in urban stormwater arising from anthro-

¹⁵USEPA state that there is no safe level for lead exposure

pogenic activities and traffic congestion. *Science of the Total Environment*, 579:202–211

In another study by Ma et al. (2017), the build-up of 9 toxic metals (aluminum, cadmium, chromium, copper, iron, lead, manganese, nickel, and zinc) and 15 PAHs on road surfaces was evaluated similarly to their previous work (Ma et al., 2016). Once again, a risk model was developed using daily traffic volume statistics and land use covariates. They found that traffic volume and land use were not significant enough to explain hazard indices. Therefore, they developed another model that includes traffic congestion and additional metrics of anthropogenic activity, such as motor vehicle-related businesses, retail, education, hospitality, catering, and commercial offices. It should be noted that high congestion leaves more time for cars to deposit pollutants. Their results showed that the highest risk levels to human health in terms of metals and PAHs were from stormwater emanating from industrial sites. The risk from metals and PAHs in stormwater emanating from commercial and residential sites depended on anthropogenic activities and traffic congestion.

2.5 2018/2020 - Oregon pollutant toxicity ranking database

Danielsen, A. (2018). Oregon Pollutant Toxicity Ranking Database A Tool for Supporting Risk Assessment of Oregon’s Water Quality. *mathesis*, Portland State University, Portland, OR

As part of a master’s thesis project, a toxics ranking database was developed by (Danielsen, 2018) to inform public education and outreach efforts in Oregon. The database was designed to help target specific behaviors that would lead to ‘quantifiable change’, specifically with regard to metals, pesticides, and consumer product chemicals. The database characterizes pollutant distribution, source, uses, transport, and fate. The ranking was based on toxicity to humans, fish, invertebrates, and amphibians. The primary objective of the database was to develop a toxicology scale that assigned a numerical value to risk level. It should be noted that this database saw further development beyond what is described in the 2018 thesis. In the thesis document, an actual ranking of pollutants is not presented - instead, the methodology and possible future uses of the database are presented. For example, the determination of which chemicals and pollutants to include was based on a November 2017 forum where a group of scientists and stakeholders reviewed survey data to determine which pollutants were of most

concern across the state. The scientists were also asked a series of questions that were used to inform pollutant ranking. I reached out to the author and collaborator to obtain a copy of the latest version of the database in 2021 (Danielsen and Handaly, 2020). The authors of the database acknowledge that more work is needed, but their work is an excellent framework for developing future education and outreach programs. Based on the database's risk ratings for stormwater - heavy metals pose the highest risk, with zinc in metal roofs and car tires with the highest risk score. A suite of pesticides follows zinc in priority, followed by various plastic ingredients and breakdown products - Phthalates, PVC, Bisphenol, and microplastics.

2.6 2019 - The need to consider exposure and toxicity when assessing pollutant mixtures

Altenburger, R., Brack, W., Burgess, R. M., Busch, W., Escher, B. I., Focks, A., Hewitt, L. M., Jacobsen, B. N., de Alda, M. L., Ait-Aissa, S., et al. (2019). Future water quality monitoring: improving the balance between exposure and toxicity assessments of real-world pollutant mixtures. *Environmental Sciences Europe*, 31(1):1–17

Altenburger et al. (2019) speak to the need for determining causal relationships between pollutant mixtures and toxic biological endpoints. They also make the case that while monitoring chemicals in aqueous environments is important, biological responses measured through the development of bioassays provide a holistic picture of chemical burden. They propose developing a metric called Toxic Units (TU), where TU is calculated as the ratio of environmental concentrations to the concentration toxic to a specific aquatic species. In place of disconnected environmental assessments of specific pollutants, they proposed more comprehensive assessments using a line of evidence approach that accounts for chemical occurrence, bioanalytical data to establish concentration-effect relationships, in situ functional responses, and field surveys that characterize the population and community structure. In conclusion, they hypothesize that such an approach will yield better water quality assessments leading to better allocations of resources to tackle the sources of the water quality impairments.

2.7 2019 - Stormwater as a source of mixed contaminants

Masoner, J. R., Kolpin, D. W., Cozzarelli, I. M., Barber, L. B., Burden, D. S., Foreman, W. T., Forshay, K. J., Furlong, E. T., Groves, J. F., Hladik,

M. L., et al. (2019). Urban stormwater: An overlooked pathway of extensive mixed contaminants to surface and groundwaters in the United States. *Environmental Science & Technology*, 53(17):10070–10081

A study by Masoner et al. (2019) evaluated stormwater runoff for 50 events from 21 urban locations across the United States, analyzing stormwater samples for 438 organic chemicals and 62 inorganic ones. Samples were collected in constructed conveyance infrastructure comprising concrete culverts, canals, and open dirt ditches. They also collected and evaluated catchment and stormwater distribution characteristics as additional dependent factors, with runoff from roofs to road surfaces. They found that 215 of the 438 organic chemicals analyzed were detected in their stormwater samples, with 69 of those organic samples detected in over half the samples. Pesticides were the most frequently measured group of organic contaminants. They showed that many of the same chemicals detected in their stormwater samples were also seen in another study (Bradley et al., 2017) that assessed streams impacted by agriculture and development. In fact, there was a greater detection frequency and concentration of neonicotinoid insecticide in urban stormwater compared to another study (Hladik et al., 2014) on agricultural streams in the Midwestern U.S. The authors showed that organic chemical loads from some runoff events were similar to daily treated effluent loads from wastewater treatment plants (WWTP). They conclude that when compared to WWTP effluent, untreated urban stormwater contributes higher loads of PAHs, pesticides, and PCBs; similar loads of household and industrial chemicals and non-prescription pharmaceuticals; and smaller loads of prescription pharmaceuticals, biogenic hormones, and plant/animal sterols.

2.8 2020 - Need for re-examining pollutant sources in stormwater

Müller, A., Österlund, H., Marsalek, J., and Viklander, M. (2020). The pollution conveyed by urban runoff: A review of sources. *Science of the Total Environment*, 709:136125

Müller et al. (2020) focus their work on the sources of pollutants in stormwater as an organizing framework, eschewing the traditional classification of pollutants by their physicochemical properties. Instead, they synthesized existing studies examining emerging and well-established sources of stormwater pollutants. Sources of pollutants that were examined were:

1. Atmospheric deposition.
2. Drainage surfaces - roads and paved surfaces; building materials and surfaces; green areas like parks, lawns, urban forests, and sports facilities.
3. Anthropogenic activities - vehicular sources, road and construction activities, littering, illicit dumping, gardening, and pets/wildlife.
4. Urban drainage - GSI¹⁶, materials used to make pipes, and cross-connections between wastewater and stormwater networks.

They identify atmospheric deposition, vehicles and roadways, and metal cladding around buildings as the primary sources of pollution. They also note that with improvements in manufacturing and processing of materials, a lot of the historical data on pollutant loading rates are obsolete and should be considered ‘historical data’; they caution against reliance upon these data. They conclude by stating that with further development of new consumer materials, new unsampled pollutants are likely to emerge.

2.9 2021 - Tire wear breakdown products in stormwater

Tian, Z., Zhao, H., Peter, K. T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., et al. (2021). A ubiquitous tire rubber-derived chemical induces acute mortality in coho salmon. *Science*, 371(6525):185–189

In this critical study by Tian et al. (2021), a toxicant from the breakdown of vehicle tires was identified as the chief agent of Urban Mortality Runoff Syndrome - the syndrome describing pre-spawn mortality of Pacific Northwest coho salmon in urban creeks of the Puget Sound region. The toxicant identified by Tian et al. (2021) is a toxic quinone transformation product of N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine (6PPD) called 6PPD-quinone (6PPD-q). 6PPD-q is formed by the breakdown of 6PPD, where 6PPD is a widely used compound incorporated into tire rubber to prevent tire rubber from oxidizing and breaking down when exposed to the elements - particularly ozone. Tian et al. (2021) estimated that 6PPD-q loadings to streams near multilane highways range from 0.2 to 3.5 $\mu\text{g}/\text{L}$ per storm event, with LC50¹⁷ for coho salmon estimated to be 0.095 $\mu\text{g}/\text{L}$. With the

¹⁶green stormwater infrastructure

¹⁷LC50 - a measure of toxicity. In this case, the concentration of 6PPD-q in water needed to kill 50% of group of coho salmon after a single exposure.

publication of the Tian et al. (2021) paper, it appears the most toxic pollutants in stormwater, in terms of prevalence and exposure, emanate from vehicles and transportation networks.

An important outcome of this work is that the characterization of stormwater pollution for known priority pollutants is insufficient to explain even the most acute stormwater problems. With the inherent complexity of stormwater, what other problems could we be missing by focusing on monitoring a small group of known contaminants (PAHs, metals, nutrients, conventionals, bacteria)?

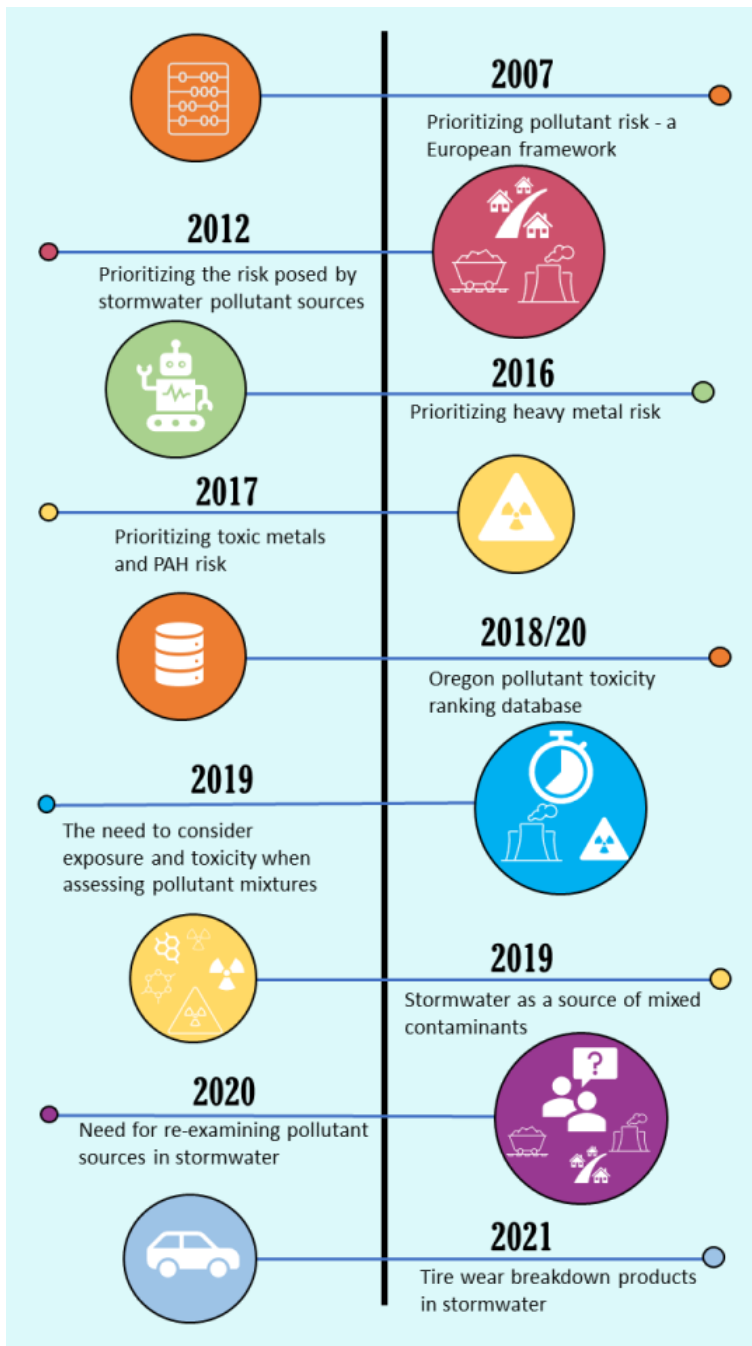


Figure 1: A timeline of recent pollutant prioritization based on nine studies summarized in this document.

3 Prioritizing Environmental Justice

The roots of the environmental justice movement in the US can be traced to the Civil Rights movement but are directly associated with protests in 1982 related to the dumping of PCB¹⁸-contaminated soils in a Black farming community in rural North Carolina (Lehtinen, 2009). The Office of Environmental Justice, established in 1992, operates within the US EPA with the mandate to examine every federal regulation regarding its implications for environmental justice issues.

The HEAL¹⁹ Act passed by the Washington state legislature in 2021 defines environmental justice as “*the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, rules, and policies. Environmental justice includes addressing disproportionate environmental and health impacts in all laws, rules, and policies with environmental impacts by prioritizing vulnerable populations and overburdened communities, the equitable distribution of resources and benefits, and eliminating harm.*” All seven state government agencies covered by the HEAL Act are required to carry out environmental justice assessments “*when making decisions and to assist the agency with the equitable distribution of environmental benefits, the reduction of environmental harms, and the identification and reduction of environmental and health disparities*” - see RCW:70A.02.060.

Waller et al. (1997) break down environmental justice studies to three basic questions:

- Are members of a particular subpopulation subject to disproportionately high exposure?
- Are they experiencing a disproportionate number of adverse outcomes?
- Is their risk of particular outcomes unduly increased by the exposure?

Evaluating the equitable distribution of the burdens associated with environmental pollution across socioeconomic demographics is a complex and evolving area of work. Waller et al. (1997) proposed a risk-exposure model that evaluates the preponderance of a pollutant in an area and the dose a human will receive of that pollutant, describing those phenomena as ex-

¹⁸ polychlorinated biphenyls

¹⁹ Healthy Environment for All (SB 5141)

posure inequity²⁰ and risk injustice²¹, respectively. They used a Bayesian framework to quantify the uncertainty in both exposure and response variables.

Spatial models to map toxic exposure and environmental justice were used within an early geographic information system (GIS) by Bevc et al. (2007). They sought to move past the standard proximity-exposure model where proximity to polluted sites was equivalent to exposure to a specific pollutant. Instead, Bevc et al. (2007) expanded their focus to include health, demographic, and biophysical²² data to develop models of potential mental and physical health. They also point out the strong connection between environmental pollution and human health, an issue many environmental justice scholars of that time overlooked.

Zartarian et al. (2011) developed the Community-Focused Exposure and Risk Screening Tool (C-FERST), a GIS tool to assess community-level exposure and risk to environmental pollutants. In Washington State, the best mapping tool that shows the cumulative risk of environmental pollution at a neighborhood level is described by Min et al. (2019) and is called the Washington Environmental Health Disparities Map. The tool is available at <https://fortress.wa.gov/doh/wtn/WTNIBL/> and combines 19 community health indicators, demographic data, diesel emissions exposure, and hazardous waste proximity.

The Western Washington Municipal Stormwater Permit requires basin planning to determine current water quality, model future water quality, and develop strategies to meet water quality standards. Phase-1 Permittees were required to do so in the 2013-2019 Permit, while Phase-2 Permittees are in the 2019-2024 Permit cycle. Phase-1 Permittees are required to include overburdened communities in the action plan for water quality attainment.

The Permit's requirement to *include* under-resourced communities is in itself a large area of work and study. A study by Hoover et al. (2021) of 119 GSI projects in several large cities in the US found that community engagement tended to be mostly passive, driven by complaints or individual relationships. They suggest a more active engagement strategy with the community is achievable by prioritizing needs identified by a community and creating frameworks for conflict resolution between jurisdiction and community. So that community engagement is a central focus of any environmental mitigation strategy or effort, a community must be engaged early Hoover

²⁰refers to differences in exposure distributions

²¹refers to differences in adverse outcomes due to exposure inequity

²²geological, hydrological, and meteorological characteristics

et al. (2021) and authentically and provided compensation or resources to facilitate engagement Black et al. (2013). We should also remember that every community is heterogeneous and should not be treated as a monolith.

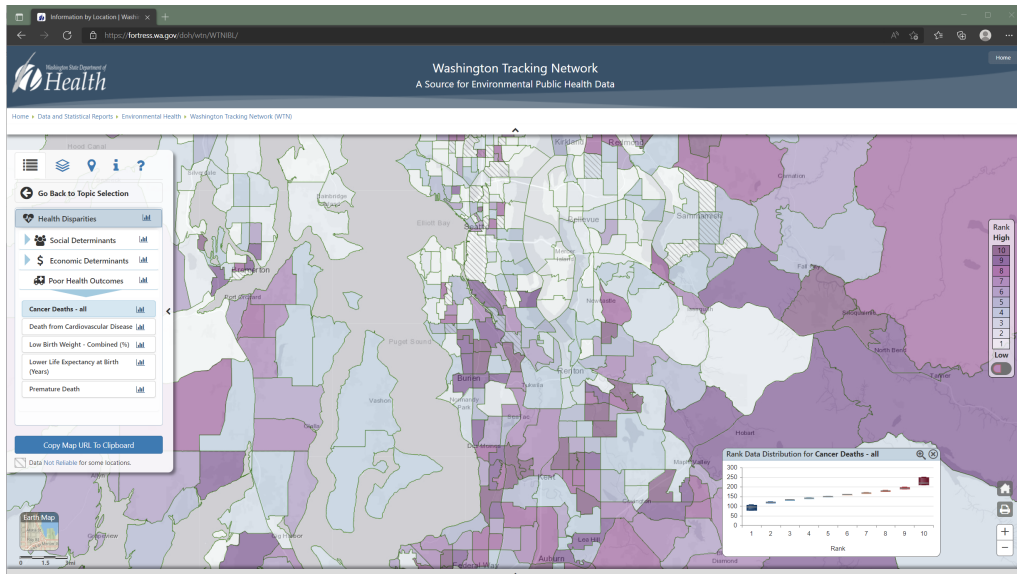


Figure 2: An interactive mapping tool that ranks the cumulative risk from environmental factors faced by Washington neighborhoods - <https://fortress.wa.gov/doh/wtn/WTNIBL/>

4 Conclusions

The HEAL Act requires multiple state agencies in Washington to address environmental justice in the state intentionally. The recent Municipal Stormwater Permit in Washington requires jurisdictions to conduct basin-level planning and include under-resourced communities. State-of-the-art mapping tools offer insight into cumulative exposure risk at the neighborhood scale and help inform environmental policy at multiple scales. Based on the studies summarized here, it is evident that perceptions of pollutant risk have evolved over the last several decades. In recent years, stormwater runoff from roadways and the pollutant mixtures they transport are of particular worry. Using the Health Disparities map in conjunction with basin level

maps to identify roadway sources of stormwater that impact under-resourced communities and targeting those communities for intervention, outreach, education, and incentives could be highly impactful.

From a pollutant risk perspective, engaging health professionals (Venkataraman et al., 2019; Kondo et al., 2015) and community (Jayakaran et al., 2021) are potent avenues for determining what pollutants might be posing the most significant risk. From just the nine studies outlined here, it is clear that pollutants must be viewed as mixtures, and vehicles and roadways are among the most potent contributors of pollutants to stormwater.

- Hoover, F.-A., Meerow, S., Grabowski, Z. J., and McPhearson, T. (2021). Environmental justice implications of siting criteria in urban green infrastructure planning. *Journal of Environmental Policy & Planning*, 23(5):665–682.
- Jayakaran, A. D., Rhodes, E., and Vogel, J. (2021). Stormwater management at the lot level: engaging homeowners and business owners to adopt green stormwater infrastructure. In *Oxford Research Encyclopedia of Environmental Science*.
- Kondo, M. C., Low, S. C., Henning, J., and Branas, C. C. (2015). The impact of green stormwater infrastructure installation on surrounding health and safety. *American Journal of Public Health*, 105(3):e114–e121. PMID: 25602887.
- Lehtinen, A. A. (2009). Environmental Justice. In Kitchin, R. and Thrift, N., editors, *International Encyclopedia of Human Geography*, pages 535–539. Elsevier, Oxford.
- Lundy, L., Ellis, J. B., and Revitt, D. M. (2012). Risk prioritisation of stormwater pollutant sources. *Water Research*, 46(20):6589–6600.
- Ma, Y., Egodawatta, P., McGree, J., Liu, A., and Goonetilleke, A. (2016). Human health risk assessment of heavy metals in urban stormwater. *Science of the Total Environment*, 557:764–772.
- Ma, Y., Liu, A., Egodawatta, P., McGree, J., and Goonetilleke, A. (2017). Assessment and management of human health risk from toxic metals and polycyclic aromatic hydrocarbons in urban stormwater arising from anthropogenic activities and traffic congestion. *Science of the Total Environment*, 579:202–211.
- Masoner, J. R., Kolpin, D. W., Cozzarelli, I. M., Barber, L. B., Burden, D. S., Foreman, W. T., Forshay, K. J., Furlong, E. T., Groves, J. F., Hladik, M. L., et al. (2019). Urban stormwater: An overlooked pathway of extensive mixed contaminants to surface and groundwaters in the United States. *Environmental Science & Technology*, 53(17):10070–10081.
- Min, E., Gruen, D., Banerjee, D., Echeverria, T., Frelander, L., Schmeltz, M., Saganić, E., Piazza, M., Galaviz, V. E., Yost, M., et al. (2019). The Washington State Environmental Health Disparities Map: Development of a community-responsive cumulative impacts assessment tool. *International Journal of Environmental Research and Public Health*, 16(22):4470.

References

- Altenburger, R., Brack, W., Burgess, R. M., Busch, W., Escher, B. I., Focks, A., Hewitt, L. M., Jacobsen, B. N., de Alda, M. L., Ait-Aissa, S., et al. (2019). Future water quality monitoring: improving the balance between exposure and toxicity assessments of real-world pollutant mixtures. *Environmental Sciences Europe*, 31(1):1–17.
- Berckes, J., Iott, T., and Gonzalez, J. (2019). 303(d) and TMDLs: State of the States. Summary report, Association of Clean Water Administrators, Washington DC.
- Bevc, C. A., Marshall, B. K., and Picou, J. S. (2007). Environmental justice and toxic exposure: Toward a spatial model of physical health and psychological well-being. *Social Science Research*, 36(1):48–67.
- Black, K. Z., Hardy, C. Y., De Marco, M., Ammerman, A. S., Corbie-Smith, G., Council, B., Ellis, D., Eng, E., Harris, B., Jackson, M., et al. (2013). Beyond incentives for involvement to compensation for consultants: increasing equity in cbpr approaches. *Progress in Community Health Partnerships*, 7(3):263.
- Bradley, P. M., Journey, C. A., Romanok, K. M., Barber, L. B., Buxton, H. T., Foreman, W. T., Furlong, E. T., Glassmeyer, S. T., Hladik, M. L., Iwanowicz, L. R., et al. (2017). Expanded target-chemical analysis reveals extensive mixed-organic-contaminant exposure in US streams. *Environmental Science & Technology*, 51(9):4792–4802.
- Danielsen, A. (2018). Oregon Pollutant Toxicity Ranking Database A Tool for Supporting Risk Assessment of Oregon’s Water Quality. thesis, Portland State University, Portland, OR.
- Danielsen, A. and Handaly, K. (2020). Oregon Pollutant Ranking List - Database File. UnPublished Database.
- Eriksson, E., Baun, A., Scholes, L., Ledin, A., Ahlman, S., Revitt, M., Noutsopoulos, C., and Mikkelsen, P. S. (2007). Selected stormwater priority pollutants—a European perspective. *Science of the Total Environment*, 383(1-3):41–51.
- Hladik, M. L., Kolpin, D. W., and Kuivila, K. M. (2014). Widespread occurrence of neonicotinoid insecticides in streams in a high corn and soybean producing region, USA. *Environmental Pollution*, 193:189–196.

- Müller, A., Österlund, H., Marsalek, J., and Viklander, M. (2020). The pollution conveyed by urban runoff: A review of sources. *Science of the Total Environment*, 709:136125.
- Tian, Z., Zhao, H., Peter, K. T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., et al. (2021). A ubiquitous tire rubber-derived chemical induces acute mortality in coho salmon. *Science*, 371(6525):185–189.
- Venkataramanan, V., Packman, A. I., Peters, D. R., Lopez, D., McCuskey, D. J., McDonald, R. I., Miller, W. M., and Young, S. L. (2019). A systematic review of the human health and social well-being outcomes of green infrastructure for stormwater and flood management. *Journal of environmental management*, 246:868–880.
- Waller, L. A., Louis, T. A., and Carlin, B. P. (1997). Bayes methods for combining disease and exposure data in assessing environmental justice. *Environmental and Ecological Statistics*, 4(4):267–281.
- Zartarian, V. G., Schultz, B. D., Barzyk, T. M., Smuts, M., Hammond, D. M., Medina-Vera, M., and Geller, A. M. (2011). The Environmental Protection Agency’s Community-Focused Exposure and Risk Screening Tool (C-FERST) and its potential use for environmental justice efforts. *American journal of public health*, 101(S1):S286–S294.



WASHINGTON STATE UNIVERSITY
EXTENSION

A Synthesis and Annotated Bibliography on Stormwater Behavior Change Campaigns

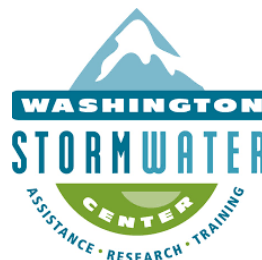
Wisnu Sugiarto and Joseph Cook

School of Economic Sciences

Washington State University

FINAL REPORT

June 21, 2022



Please cite as “Sugiarto, W. and J. Cook. “A Synthesis and Annotated Bibliography of Stormwater Behavior Change Campaigns”. Washington State University Stormwater Center. June 2022.

1. Introduction

This report describes the results of a comprehensive review of peer-reviewed studies and “gray literature” reports that have evaluated behavior change campaigns in water quality or stormwater. It is part of a larger project on these campaigns for the [Stormwater Action Monitoring](#) (SAM) program, funded by permittees and administered by the Washington Department of Ecology. The title of the larger project is “Evaluating the effectiveness of stormwater education and outreach: permittee guidance for addressing challenges through behavior change”, though the project team and our Technical Advisory Committee agreed to narrow our focus to behavior change. By “behavior change”, we mean programs that seek to get people to change a concrete behavior, such as picking up pet waste or reducing fertilizer use, rather than other “education and outreach” programs that seek to communicate information or change attitudes or programs that provide stewardship and volunteer opportunities without clear quantitative measures.

Our charge from SAM for Task 3 was to “conduct a broad literature review of education and outreach (E&O) programs, organized by stormwater issue...”, including a “3-10 pg. draft review of the literature (both “gray” and peer-reviewed) on the effectiveness of E&O programs, organized by stormwater issue, including studies from within and outside Washington State.” The report was also to “evaluate the pros and cons and research quality (e.g. evaluation of methods and data) of each existing E&O evaluation study.”

This report identifies, evaluates, and summarizes existing studies on behavior change campaigns associated with water quality or stormwater management. The audience is behavior change professionals who work in the stormwater and water quality space. This document is organized as follows. First, we discuss our approach in reviewing the literature. We adopt a comprehensive, objective, and reproducible search strategy to capture all relevant studies. Second, we discuss our findings and identify the common trends through descriptive statistics. Third, we highlight evaluation methods that have been used in assessing the effectiveness of behavior change programs. We supplement the evaluation strategy discussions with methodological concerns to identify gaps that can be used to improve future evaluations. Finally, we present a table summary of all collected articles, which is available as an appendix. The articles are also searchable and downloadable on www.waterbehaviorchange.org.

2. Review Criteria

The first guiding principle in conducting our review was relevance. We defined relevance as program reports and published academic articles that evaluated measurable and direct behavior change programs for water quality or stormwater management purposes. Behavior change studies could be “quantitative” or “qualitative” in nature. “Direct” behavior change programs reach their target audience directly, so we excluded studies that focused only on creating awareness or providing stewardship opportunities. Although an awareness campaign can induce behavior change, evaluations of these programs typically focus on changes in awareness or attitudes. Similarly, stewardship programs ask volunteers in the community to help with water quality or stormwater projects (e.g. storm drain monitoring, tree planting), though the aim is again to build awareness. We also excluded programs which focused on surveying residents’ perceptions of water quality, assessing people’s knowledge on stormwater issues, and/or reporting only homeowners’ intent to change. Similarly, there are a number of case studies that discuss the implementation of a social marketing approach in water quality but that had very little or no information on how the program was evaluated. We excluded studies that did not discuss evaluation at all, or only briefly in passing. We included studies that explained how an evaluation was conducted, even if the report provided no detailed information on the evaluation results or the evaluation attempts did not yield any results.

The second principle guiding the search was breadth. We employed multiple approaches to collect articles on behavior change programs related to stormwater management. First, we searched for published academic articles and publicly-available reports on several keyword databases such as [Google Scholar](#), [GreenFile \(EBSCO\)](#), [Academic Search Complete](#), and [Science.gov](#). Second, we found two databases that already contained a collection of stormwater program-related surveys and evaluations. The NPS Outreach Toolbox on the US EPA’s website has a collection of [Surveys & Evaluations](#) from around the United States. In fact, many of our collected reports came from the US EPA’s site. We also included case studies found on the [Tools of Change](#) and [CBSM websites](#). These were primarily community programs from across North America, and they covered other environmental issues beyond water pollution. We also collected program reports from our Washington key informant interviews. Finally, we contacted stormwater professionals or environmental managers at every U.S. state agency, including the U.S. territories. The list of all state agency contacts could be found on the United States Environmental Protection Agency (EPA) page for [stormwater contacts](#). Some of the listed contacts were incorrect, had retired, or were no longer working in the position. Thus, we searched online to obtain the updated contact information.

Finally, we aimed for our literature searches to be replicable. We searched Google Scholar, GreenFile (EBSCO), Academic Search Complete, and Science.gov with the following sets of keywords: 1) “stormwater” “behavior change” AND “education and outreach”, 2) “stormwater” “behavior change”, and 3) “stormwater” AND “behavior change”. This yielded a large number of results, though many were broken links, duplicative studies or were not relevant. We scraped the data, which were the large number of search outputs, and conducted a blind-review process. Each author independently reviewed each collected title and determined if it was relevant to our literature review study. We discussed discrepancies in coding and agreed on which to proceed with. We next independently reviewed the study abstracts. Some abstracts clearly fit the relevance criteria, but not all studies or reports can be assessed from their abstracts. We reviewed all studies and that both authors coded for inclusion based on the abstracts. This blind-review process allowed us to systematically collect search results that meet the relevance criteria and eliminate the ones that do not.

3. Data recorded

We systematically recorded several pieces of information from each collected study and report so that we could provide descriptive summary statistics and evaluate research quality in a consistent and transparent way. The data recorded also allowed us to compile the spreadsheet database (Appendix #). From each study, we recorded the study year, location, targeted behavior, and target audience. We recorded whether the program specifically targeted under-represented groups, such as non-native English speakers or racial minority groups. We abstracted information on the implementation of the program, including whether the implementers used educational posters, distributed brochures and posters, conducted home visits, used financial incentives (including subsidized purchases or free services like soil testing) or held online or in-person workshops or demonstrations.

To help us classify evaluations, we created a series of dummy (0/1) variables that captured each of the following categories, each of which is described in more detail below: a) well-defined **target audience**, b) well-defined **target behavior**; c) validated or pretested data collection **instruments**; d) evaluation included a **comparison group**; e) evaluation includes **pre-intervention data**; f) evaluation uses **observational data** (rather than self-reported data); g) Evaluation measures **long-term** change (≥ 1 year); h) evaluation discusses possible **selection bias** in uptake among target audience; l) evaluation includes **water quality measurements**. Every study was examined based on the 9 categories. Each category would

yield 1 point if the study met the criteria for that category. For example, category (a) was assigned 1 point if the study had a well-defined target audience. Otherwise, it would be assigned 0 points. Thus, the maximum score each study could get was 9. Although the summative score does not distinguish between cases where the information is missing (e.g. no information on pretesting the instruments is discussed) versus cases where sufficient information is reported to allow us to definitely classify the study as “0” (e.g. the study reports that its outcome measure is self-reported behavior), this data is recorded separately in the database. We group studies with scores below 4 as “*fair*”, studies with scores between 4 and 6 as “*good*”, and scores of 7-9 as “*exemplary*”.

Target audience, target behavior: All the major social marketing frameworks share a focus on clearly identifying a specific, end state target behavior and a specific target audience. Although this is largely about implementation rather than evaluation, we include them in our summative measure of study quality as a rough indicator of how well the practitioner applied social marketing ideas. Each is separately worth one point.

Validated instruments: Many studies rely on surveys to collect information on self-reported behavior or attitudes. Good survey practice dictates that survey questions are tested or validated in advance with focus groups or pretests or both. Evaluators can also re-use survey questions used by others that have already been tested.

Baseline data: Collecting pre- and post-intervention data is useful to attribute behavior changes to the program. If a campaign only employed a post-intervention evaluation, it would not have any comparison data to measure the effect of an intervention on program participants. For example, suppose there was a workshop on educating homeowners on substitutes for pesticides, and participants were only asked to complete a post-event survey. The collected survey results suggested that 75% of workshop participants purchased pesticide substitutes after attending the workshop. While the workshop had been helpful in informing participants about the danger of pesticides, it is possible that some of the participants had already planned on using alternative options in the first place. In this case, the workshop primarily served as a refresher course to reaffirm participants to move away from using pesticides, and one cannot conclude that the campaign increased substitute purchases by 75% percentage points. **Comparison or control groups:** Having a control or comparison group helps to establish that the program caused the desired behavior change by isolating its effect from other possible explanations for changes in behavior.

It provides a comparison between people or areas (within the defined target audience) that were exposed to the social marketing campaign and people or areas (again within the defined audience) that were not exposed. As an example, suppose jurisdiction A implemented a city-wide campaign targeted at new pet owners to encourage them to pick up and dispose of pet waste properly. To assess the campaign's effectiveness, behavior among new pet owners in jurisdiction B could in theory also be observed. Alternatively, half of new pet owners in jurisdiction A might be randomly assigned to receive the campaign materials (e.g. a "scoop the poop" sticker and free dog bag dispenser), but follow up surveys on self-reported behavior would be sent to both groups. Note that in both cases changes in behavior in the comparison group help control for other factors that might have changed behavior over the same time period. Finding the resources to identify and survey control groups may, however, be challenging.

Observational vs. Self-reported data: While most studies rely on surveys to ask households to self-report behaviors, this approach suffers from well-known reporting biases. This is particularly true when respondents believe there is a "right" or "socially-appropriate" answer¹; they might be ashamed of answering otherwise and there is little cost to being untruthful. This is clearly a concern in many water quality or stormwater-related behaviors (e.g. picking up pet waste, not dumping chemicals down the drain). Social marketing approaches may even attempt to build a social norm, so they in effect train respondents what the "right" or socially appropriate answer is. Good survey techniques can help ameliorate but not eliminate social desirability bias. Observational studies are very often not possible or practical (i.e. observing whether someone dumps chemicals down their kitchen drain), but have been done in programs for pet waste, dumpster maintenance and others.

Selection bias: Here we record whether the study monitored or discussed differences in uptake *among the target audience*. The main concern is that bias in who takes up the campaign can skew predictions about how a program may scale or expand beyond in the same jurisdiction or a different one.

Example: Suppose a city runs a natural yard care workshop for seniors. Among seniors, the only people who attend are those who are already very environmental-minded (perhaps as measured

¹ For more information, including techniques to reduce social desirability bias, see https://en.wikipedia.org/wiki/Social-desirability_bias

by a survey). Twenty people participate and the team evaluates their lawn care self-reported behaviors before and after the workshop with a survey. Suppose that the workshop is associated with increases in the target behavior, and the team asks for funding to run another 5 workshops. But the next 5 workshops may have less or no effect, and it may not even be possible to get anyone to attend: the program already harvested the “low-hanging fruit”.

Selection bias affects validity. The same program as the example with the same target audience in a different city may not find the same effect if they had fewer environmentally-minded seniors. As in control groups, it can be challenging to find resources to observe the characteristics of *non-participants* (or the community at large) to make these sorts of comparisons, though evaluators can sometimes use existing recent community-wide surveys (for things like environmental attitudes) or census data (for gender, income, age, etc.).

Long-term: Behavior changes induced by a social marketing campaign may be short-lived, so we distinguish between evaluations that looked only at short-term effects versus those that attempted more longer-term evaluations. We (somewhat arbitrarily) chose one year (or longer) as our cutoff for “long-term”.

Water quality information: Our primary focus is on the question of whether a behavior change campaign changed behavior, not whether the campaign actually improved water quality. Even campaigns that successfully change behavior may not cause detectable changes in water quality, depending on the hydrology and pollutant loading. The frequency and timing of water quality monitoring (which could be run by another department or agency) may not sync with the timing of the campaign. Nevertheless, we wanted to highlight the few studies in the literature that did attempt to measure water quality outcomes.

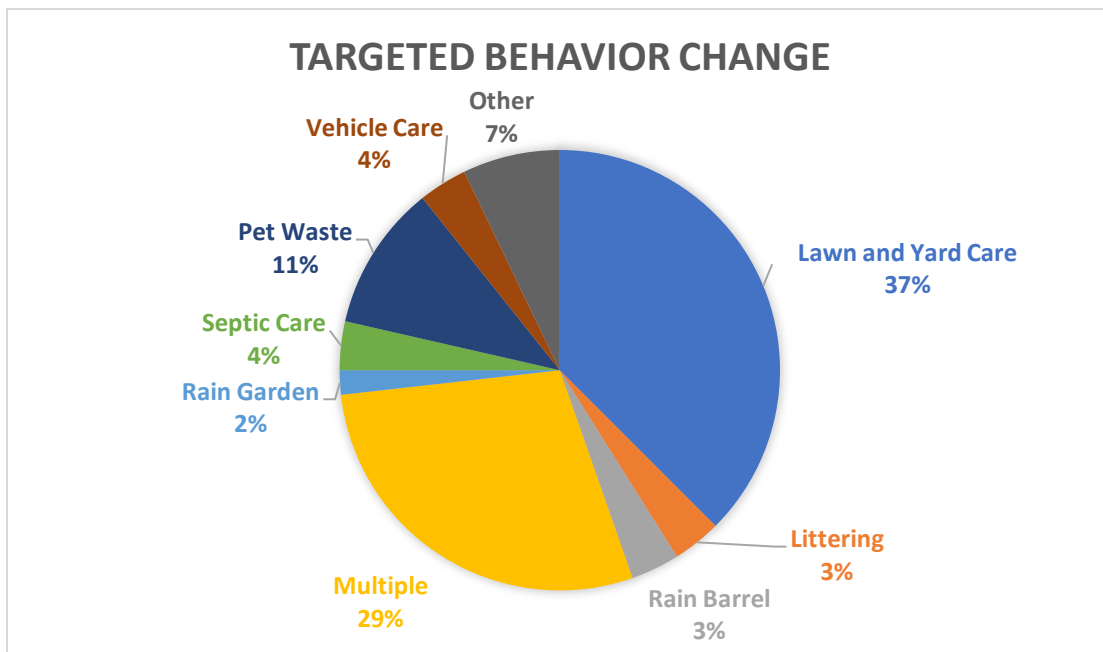
4. Descriptive Statistics on Studies

In total, we judged 56 documents to be relevant for review according to the criteria described in section 2. Fourteen were published in academic journals and the remaining 42 were program reports. One of the documents contained a report of eight case studies that discuss the impact of regulations and public education programs on reducing use of pesticides on residential property. In this case, information from the eight case studies were collected but the document containing the case studies was not counted to avoid redundancy. The oldest document was published in 1979 and the newest reports were released in 2021. Many municipalities involved an external consulting company to implement a behavior change program or conduct a program evaluation. Not surprisingly, there are also municipalities that involve a

university extension or research organization. For example, several Municipal Separate Storm Sewer Systems (MS4s) in Washington partnered with Washington State University Extension.

We categorized target behaviors into nine categories: 1) lawn and yard care practices, 2) rain barrel adoptions, 3) rain garden installations, 4) pet waste pick up and disposal, 5) litter pick up and disposal, 6) septic care, 7) vehicle care, 8) other, and 9) multiple (Figure 1). The lawn and yard care category included many practices, such as reducing the use of pesticides and fertilizers, using more environmental-friendly landscaping products, and recycling grass clippings. The vehicle care category included car washing, fixing oil leaks, and any programs targeting vehicle owners such as reduction in vehicle miles driven. The “other” category would also include a wide range of programs, such as care for street trees, proper disposal of used cooking oil and hazardous household wastes, and proper disposal of wastewater from carpet cleaning. The “multiple” category indicated programs that aim to change more than one behavior. For example, a campaign to maintain the health of the Nippersink Creek Watershed in Illinois wanted people to maintain septic systems properly and use phosphorus-free fertilizers.

Figure 1: Proportions of Targeted Behavior Change

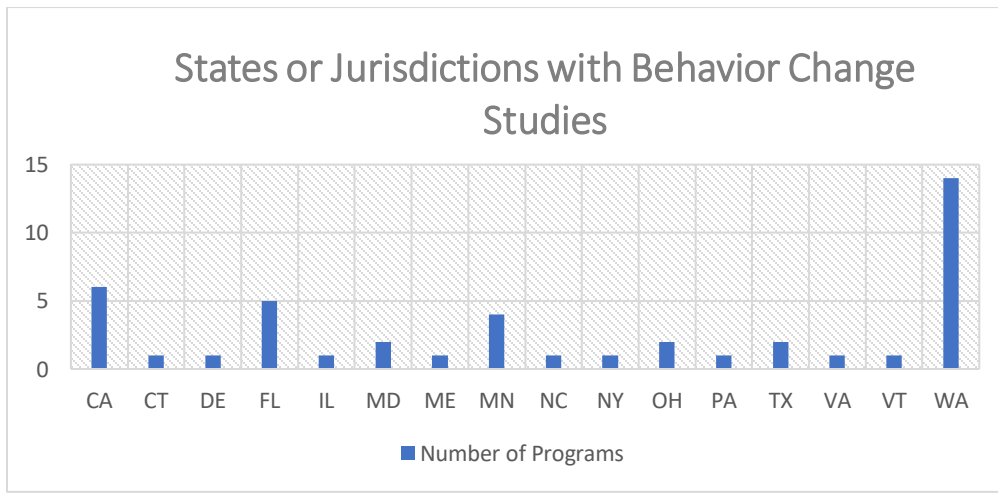


We found that programs related to lawn and yard care to be the most common; 37% (21 of 56) of the collected literature exclusively focused on lawn care practices to reduce the impact of stormwater runoff. Programs that employed multiple targeted behaviors were the next most common, with 16 studies. For example, the Think About Personal Pollution (TAPP) campaign in Tallahassee, Florida, aimed to change people’s behavior through switching to non-phosphorus fertilizers and picking up pet waste. Within these

16 studies, behaviors related to lawn care (29%) and pet waste disposal (17%) were the most common. Other relatively more common studies in the multiple group targeted rain garden or rain barrel adoption. The next most common type of exclusive behavior targeted (11% or 6 of 56) was picking up pet waste, followed by 4% (2 studies) concerned exclusively with septic care and another 4% (2 studies) focused on issues associated with vehicle care.

Geographically, many reports came from organizations in areas near estuaries or lakes, such as municipalities in the Puget Sound (Washington), California, Minnesota, and states surrounding the Chesapeake Bay, such as Maryland and Delaware (see Table 1). We also located several studies from outside of the United States, including programs in Australia, Canada, and Denmark, but note that Figure 2 reports the distribution of collected behavior change studies in US states only. Programs may actually be more common in these areas because of the importance of large, shared waterbodies, politics, larger cities or because many early social marketing studies originated in the Pacific Northwest. It may also be that jurisdictions in those areas were more likely to publicly report their findings. This also reflects differences in regulatory environments and the number of regulated cities.

Figure 2: Distribution of Collected Behavior Change Studies in the United States



Only two of the studies examined interactions with regulations or fines. For example, the small towns of Hudson, St. Lazare, and Notre Dame de L’ille (in Montreal, Quebec) regulated the use of pesticides on residential property (Kassirer and Wolnik, 2004). On May 6th 1991, Hudson became the first municipality in Canada to pass a by-law regulating the use of pesticides on residential property. They employed a

variety of techniques to help residents comply, including the use of home patrols to answer questions and provide advice on alternatives to pesticides. Offenders were warned through warning letters with a threat of escalating fines. The regulation enforcement was also supplemented with the municipalities' efforts to build pride in being pesticide-free areas. For example, a pesticide-free campaign logo was used on all city paperwork for several years in Notre Dame and St. Lazare. The authors found that only communities that supported a regulation with education or made a community agreement were successful in reducing pesticides by a large amount (51-90%). E&O programs without regulations were more popular, but none achieved more than a "low" (10-24%) reduction during the study period

An example of a community agreement is the Drastrup project in Denmark. This pesticide reduction project was managed by the Aalborg municipal government to prevent groundwater contamination and meet European Union drinking water standards. There is a groundwater recharge area near Aalborg, Frejlev, which had a population of 2,000 people. As part of the agreement, land was purchased and farmers had to relocate outside the catchment area. Residents were provided with information about the consequences of pesticide contamination through meetings, website, local media, and newsletter.

The second study involving the use of a regulation also took place in Canada. The Halifax Regional Municipality was the first Canada's larger communities to introduce a pesticide by-law. The implementation was phased in over two years, between April 2001 and April 2003. The municipality contracted Clean Nova Scotia, a local non-profit organization, to process applications for pesticide permits through home visits. These home visits allowed program staff to educate property owners on sustainable maintenance practices

5. Program Evaluation Assessment

In this section, we identify trends, common practices, and gaps in existing evaluations. The purpose is to inform how behavior change evaluations might be improved, not to criticize individual studies, authors, or jurisdictions. We recognize that implementing staff operate under real time and money constraints and vary in their expertise with social marketing or program evaluation.

All (100%) of our curated and evaluated studies (56) for this project had at least one well-defined target audience and at least one well-defined target behavior. Program managers were clear with who they wanted to persuade and what behavior they wanted to change. In almost half (26 of 56) of the collected studies, staff validated their data collection instruments. Common validation practices included pre-testing surveys before dissemination and incorporating focus groups to supplement survey findings. For example, the Mobile Contractor Illicit Discharge Education & Outreach program led by the City of

Wenatchee, Washington, had a pilot test in the City of Pullman, Washington, to validate their survey instrument.

Table 1. Summary of evaluation quality measures

Measure	Percent of studies (n=56)
Well-defined target audience	100%
Well-defined target behavior	100%
Validated instruments	46%
Collected baseline/pre-intervention data	63%
Collected data on control or comparison group	13%
Use observed rather than self-reported data	26%
Address possible selection bias	48%
Measure behavior change after 1 year	36%
Measure water quality	18%

In contrast, we found that very few (7 of 56, or 13%) mentioned the use of a control group. Sixty-three percent of studies (35 of 56) included baseline data to measure changes in behavior caused by the campaign. Approximately 73% (41 of 56) measured behavior solely as self-reported in surveys administered by mail, telephone, or the web surveys. The remainder used observations or sales data to measure changes before and after an intervention. (For example, please see “Community Program #2: Halifax Regional Municipality, Nova Scotia” and “Community Program #5: Chesapeake Bay, Pennsylvania (Harrisburg)” in the bibliography table.) Observation was more common in published journal articles: over half of them included observed behavior.

Thirty six percent (20 of 56) of studies measured “long-term” changes (one year or longer). A few studies were clearly and intentionally interested in measuring the long-term effects of on-going programs. For example, Snohomish County Natural Yard Care Education team deployed a long-term post-outreach survey that was administered five years after receiving education. Approximately half (27 of 56) of studies monitored or discussed potential issues associated with selection bias. These studies did not necessarily analytically correct for any selection bias concern, but at least acknowledged the issue. We found 18% (10 of 56) of studies included some types of water quality measurement. These were relatively more common among the published journal articles, where 36% (5 of 14) included water quality measurement.

Considering the nine categories that we used to help us with measuring program evaluation quality, our review suggests that more than half (37 of 56, or 66%) of the collected behavior change studies are in

the “good” category, scoring between 4 and 6. One quarter (14 of 56, or 25%) of studies would be ranked as “fair”, scoring less than 4. No study achieved a “perfect” score of 9, but 9% (5 of 56) of studies would be categorized as “exemplary”, scoring 7 or 8.

Do these programs “work”? Evaluation results from multiple studies report that behavior change campaigns work in influencing people’s behavior to adopt stormwater runoff best management practices (BMPs). Among the collected studies, 63% (35 of 56) collected baseline data and provided some information on their campaign results. Nine of the 35 studies also collected observational or monitoring data. Outcomes of the behavior change campaigns are generally positive with 57% (18) of the 35 studies reporting successes in persuading people to adopt more environmental-friendly behavior. Forty three percent (15 of 35) of the studies reported what could be described as mixed results, where campaigns had positive and negative outcomes or where positive outcomes were attenuated in some way. For example, Thurston County Washington conducted a long-term evaluation in 2020 of their 2014 behavior change campaign, the GoGreen lawn care program. The evaluation found that the percentage of respondents who said that they used a weed-and-feed product fell from 65% at baseline to 12% in the 6-12 months after training. The reduction in weed-and-feed behavior remained in a follow-up survey six years post-training, though it showed some degree of backsliding: 38% used weed-and-feed. Nonetheless, the study provides evidence that behavior change campaigns can result in sustainable or long-term behavior change.

We would like to highlight one of the “exemplary” studies. In “Education Campaigns to Reduce Stormwater Pollution in Commercial Areas: Do They Work?” (Taylor, A. et al., 2007), the authors discuss the results of a littering reduction campaign in a small commercial area in Melbourne, Australia. To monitor the changes in awareness, attitude, and self-reported behavior of the priority audience associated with littering prevention, the evaluation team used face-to-face survey instruments before, during, and after the campaign. To evaluate changes in people’s actual behavior with respect to littering and littering prevention, the evaluation team observed littering before, during, and after the campaign. They also monitored litter loads entering stormwater by using side entry pit litter baskets at the intervention and control sites to capture pollutants before, during, and after the campaign. The evaluation incorporated self-reported and observed information, used a control group that included an explanation why it was chosen, and monitored the presence of pollutants in the water. In addition, the monitoring process was conducted by staff members who were trained in collecting data. The training, along with a validated rating instrument (the Clean Communities Assessment Tool) and in-depth personal interviews, were helpful for validating the data collection process. They also performed data analysis using

different styles of evaluation to obtain a more comprehensive understanding of key findings and to verify self-reported information.

Conclusions

We conclude by providing several recommendations from observing the trends, common practices, and gaps in existing evaluations. First, we recommend including a control group for assessment purposes. The current evaluation practices suggest that there are significantly more evaluations that collect baseline data than include a control group. This finding is not surprising given that jurisdictions face budget constraints. Nonetheless, this gap jeopardizes the validity of a campaign's results. The use of a control group helps isolate the campaign's effect from other possible explanations for changes in behavior. This is also important for a repeated behavior change campaign that is only based on pre- and post-treatment data from the previous evaluation. Decision makers may decide to repeat a program because of its high success rate in the past, but may not know the success of the campaign was actually due to another, unobserved factor. Likewise, a campaign that may be thought to have been unsuccessful in changing behavior may in fact have suffered from an unobserved common factor.

Second, we recommend addressing issues of selection bias. While half of the collected studies address or acknowledge such issue, we think more evaluations can meet this criterion especially because it is relatively low cost to at least discuss the uptake among the target audience. For example, program evaluators can compare the demographic of program participants with the general residents' demographic of the jurisdiction by reviewing census data. We recognize that social marketing typically encourages practitioners to have a clearly defined and narrowed audience, but an evaluation of the uptake among the target audience is important for scalability or program expansion.

7.5 WEBSITE DECISION-SUPPORT TOOL - SURVEY RESULTS (D4.5)

Task 4.4 Report on survey of jurisdictions testing website

June 21, 2022

Joe Cook

The goal of this task was to “ask jurisdictions who participated in the Task 2 survey for additional feedback on how to improve the website tool and report findings in a 1-3 page memo.” The current version of the website is at www.waterbehaviorchange.org.

Recall that the Task 2 survey used a “snowball” sampling approach where we asked contacts to help us publicize the survey and distribute the survey link to maximize participation. We therefore do not have a list of all jurisdictions who participated in the survey. We are also holding back from contacting those that did leave contact information so that we can contact them again once at the close of the project to advertise the entire project’s work, not just the website.

Instead, we collected feedback on the website from our Technical Advisory Committee and from jurisdictions and organization on Ecology’s listserv for WA permittees. We constructed a simple Qualtrics survey to enter feedback and took comments from early May (the TAC) through mid-June. In all, 22 people provided us with feedback on the website.

We asked users “overall, as currently designed, how useful do you think the website will be to behavior change practitioners in the State?”. Twenty-one percent thought it would be “very” or “extremely” useful, and an additional 58% thought it would be “moderately” useful. We also asked “How likely is it that you would refer another staff person who works in behavior change in stormwater to the website?” Seventy-two percent said it was “somewhat” or “very” likely.

We asked for feedback on the campaign search mechanism, the resources provided, the campaign selection tool and other open-ended feedback.

- Many identified glitches with the search function (returning odd results) which has since been fixed. Users also flagged several typos which have been corrected.
- Several mentioned the re-designed Chesapeake Behavior Change site as a model. We are in the process of changing the way search results are displayed. Rather than displaying just an icon with the study title that links to the pdf, clicking on results will pull up a dedicated page for each campaign that lists the various components of the campaign, as the Chesapeake site does.
- Many users asked for more (or improved) text on the home page explaining what they might use the site for and explaining how the campaigns featured on the map were selected. The pie charts were confusing to many, since they show no data unless one hovers her mouse over them. Some also felt background

images related to behaviors would be more appropriate than an image of the Snake River near Pullman. We will make these changes.

- Two commenters felt that the site should be a repository for social marketing materials that were not evaluations of campaigns. For example, one commenter wants the site to include more early-stage studies, for example on audience research or barriers analysis, and for the site to more generally include resources on best practices for conducting a social marketing campaign. The site is currently focused on providing data only on evaluations, since this data was collected as part of our Task 3 report summarizing what we know about the effectiveness of social marketing and behavior change campaigns in stormwater and water quality. Including formative research could be included in future work, but we feel this is currently out of our scope. We do, however, plan to link directly and prominently to the relevant section of the Chesapeake site on steps for implementing a behavior change campaign. This site does an excellent job and there is no need to duplicate efforts.
- Few users found the “campaign selection” tool helpful. Many were in fact very confused about its purpose, thinking that it was actually more like another search function: the site would spit out examples of campaigns for them to compare and choose from, like buying a tent on REI. We could address this with more (and improved) introductory text to help explain the purpose. Some had comments and suggestions about specific data fields in this exercise, and many wanted the site to give them a ranking or answer about which campaign was best. We discussed this latter point with our TAC, who felt it was a mistake to believe we could substitute our judgments. But the larger comments were about whether they would ever use it. To many, it seemed a waste of time – if I have all this information, why am I typing it into a website? If the purpose was to have us fill out information for our own use, two users suggested, why not just have a downloadable Excel sheet that they could fill in out locally on their own computer, rather than filling the information on the site? We believe this is a useful suggestion, and will discuss with our TAC whether a downloadable sheet might in fact be more useful than revising this section of the website.
- Several users had suggestions for additional resources to include, and for re-labeling and re-grouping some of the resources we already provide. We will implement these suggestions.

Next steps

The team has already began incorporating some of this feedback and will continue to make substantial redesigns and refinements to the site over the summer. It will remain functional and we will push major site updates as they become ready. We anticipate needing

to make only minor design changes by project close in the fall of 2022. From that point forward, support will be needed primarily to add or edit content to the site.

7.6 SUSTAINING THE WEBSITE DECISION-SUPPORT TOOL (D4.5)

Task 4.5 Sustaining the Behavior Change Website

June 22, 2022

Joe Cook

The goal of this deliverable is to “Develop a 1-3 pg. plan that would identify funding needs to maintain the website and add additional studies nationwide in a systematic and quality-controlled manner.” The current version of the website is at www.waterbehaviorchange.org. Below we outline a) what we project the annual workload to be to maintain the website, b) the estimated cost of sustaining this workload, and c) a position description that contains the key technical skills needed and that could be used to recruit a worker. Note that the project PI Joe Cook has a 37% Extension appointment (with a focus on stormwater and green stormwater infrastructure), so the tasks in maintaining the website that can be fulfilled by him can be covered as part of his regular Extension duties. Cook is familiar with how to maintain the codebase.

Annual Workload

Site design will be finalized using funds from this SAM project. We anticipate that any subsequent efforts towards the website will be comprised of a) marketing, b) quality-controlled study additions, c) and addition of new resources.

Marketing can be done on as-needed basis by Cook to groups such as Pacific Northwest Social Marketing Association, the National Municipal Stormwater Alliance, and other stormwater- or social-marketing focused professional groups. It may involve writing 1 – 2 blogposts for the Washington Stormwater Center’s listserv. We estimate this workload at approximately 2 days (16 hours) per year.

It is hard to anticipate how many new behavior change campaigns in stormwater or water quality will need to be added to the site on an on-going basis. This is also related to a website design decision that we will need to be finalized with our TAC: should the site focus only on evaluations of full campaigns (as the site currently does), or should it include earlier stage formative research studies examining barriers or doing audience research? If the focus is on evaluations only, we might anticipate 1-2 new studies being added per month. We estimate that it would take 2-3 hours per new study. This involves a) reading and coding the study, b) possibly contacting the study author’s for clarifications, and c) filling information into the code template and pushing it to the website. This implies approximately 4-6 hours per month, or 9 days per year. In the short-run, this could be accomplished by Cook. The workload would increase substantially if we included formative research studies, and would imply a large “start-up” cost to find and code existing formative studies. The workload would also increase if, for example, the Dept of Ecology required permittees to post their permit-required evaluations to the site (which we would welcome).

The addition of new resources that come available (new papers, web resources, courses) requires only updating the codebase of the website. We estimate this to take approximately one hour, and might occur once per quarter.

In total, then, we estimate the annual to maintain and slowly grow the website to be 2 days per year (marketing) + 9 days per year (additions) and 0.25 days per year (resources), or 11.25 days per year. Assuming 250 days per year, this is approximately $11.25/250 = 5\%$ FTE.

Estimated Cost

As noted above, Cook has been trained on how to update the codebase of the site and can perform the tasks above as part of his Extension appointment. Should the workload increase or Cook no longer be able to dedicate his time to the site, one option would be to hire an undergraduate with the technical skills (described below). The student would also familiarize themselves with social marketing and our coding definitions, and would be supervised by Cook or another stormwater- or behavior-change focused specialist to maintain quality. An upper range for an undergraduate (which might be needed to attract her to the position) is approximately \$20 per hour in 2022. Assuming she worked only on the study addition tasks, this would imply approximately 9 days x 8 hrs/day = 72 hours per year, or \$1,440.

Another option would be to hire a dedicated web developer, though this would need to be in conjunction with another project since the needs for this site are much less than 1 FTE. We estimate that junior web developers in lower-cost living areas like Pullman and Spokane range from \$70-85,000 per year out of school. Taking the upper limit and adding 34% for fringe benefits, this would imply approximately $\$113,900 * 0.05 \text{ FTE} = \$5,695$ per year.

Position Description

We constructed the following position description to help staff identify the skills needed to maintain the codebase of the website:

“Required languages to be familiar with Javascript/Typescript, React, and html/CSS. Some experience in web development or design recommended. Knowledge of libraries such MUI or Chakra along with the ability to modify library elements would be ideal. Relative competence with the tools Github, Firebase and Excel is expected.”

7.7 BEHAVIOR CHANGE EVALUATION DRAFT REPORT TEMPLATE – PILOT TESTING RESULTS (D5.5)

Section #	Report Template Section Title	General Comments	Comments <i>If you responded No or Maybe, state why and provide suggestions for improving the report template content.</i>	Comment Responses
	Template Instructions (pages ii & iii)	Section should be added or revised	In the paragraph explaining that all content is instructions and should be deleted- emphasize that paragraph- change body text color? Highlight in a way that emphasizes difference between added content an	The paragraph was highlighted to emphasize the content.
1	Executive Summary	Example could be more helpful	This section could include a "Suggested Length" portion- to assist in ensuring the summary does not become too long.	Added text about targetted number of words.
2.4	Behavior Change Campaign Strategy	Maybe example should be added Permit required content could be clearer or more helpful	Potentially add a bullet point for the process for the selection of the campaign materials? (how it was decided to use mailer, radio, etc.) Use a bullet point to address languages materials were produced in- to be addressed	comments added to the section about describing the process for selecting campaign materials and the languages the materials were developed in.
3.2	Evaluation Location(s) and Target Population	Consider revising section to improve clarity of instructions and prompts.	This section reads very similar to section 2.2 (Target Audience) It may be beneficial to note in section the directions for section 2.2 that more detail about the Target Population will be discussed further in later sections. Maybe call out the map with the same format used in 2.1?	The target population is a subgroup of the target audience that is the focus of an evaluation. It is possible that these two groups could be the same and there is a note about how to address this already in the section. Call out for map was updated for consistency with section 2.1.
3.3	Key Project Members: Roles and Responsibilities	Example was somewhat helpful and consider adding example	Potentially helpful to provide the list of potential roles and what those roles might consist of in the table. Wondering if this section should be in part 2	Because Table 3-1 is not required as part of the permit, it was left simple. Section 2.2 is just about the campaign and section 3 is about the evaluation which is the focus of the document so section 3.3 was left in section 3.
4.3.3	Hypothesis Testing	Considering adding example	This section may benefit from an example or a description of some of the statistical terms (Null Hypothesis, Alternative Hypothesis, Confidence Interval(s).	Additional guidance on hypothesis testing is included in the guidance manual.

7.8 BEHAVIOR CHANGE EVALUATION REPORT CHECKLIST & TEMPLATE (D5.6 & 5.7)

Section #	Report Template Section Title	Overview of Section Content	Completed
	Template Instructions (pages ii & iii)	This section provides instructions for using the template and should be deleted before finalizing report.	
	Report Template Information (page iv)	This section provides information about who developed the report template and should be deleted before finalizing	
	Public Information/Authors and Contact Information	This section provides information about the author(s) of the report as well as a weblink to where the final report can be accessed by the public.	
	Table of Contents	This section should be updated before finalizing report.	
1	Executive Summary	After the report is complete, develop a brief non-technical summary about the project.	
2	Behavior Change Campaign	Section 2.0 focuses only on information about the campaign.	
2.1	Behavior Change Campaign Background	Describe the behavior change campaign, where the campaign was implemented, and who was responsible for developing and/or implementing the campaign. <i>Reference Chapter 2 of the Evaluation Guidance Manual for more information about Community Based Social Marketing and Social Marketing approaches.</i>	
2.2	Target Audience	Describe the audience that the campaign was designed to target including who the target audience is and how the audience was selected.	
2.3	Target Behavior and BMPs	Describe the target behavior that the behavior change campaign aims to have the target audience adopt and the BMP that was implemented to effect behavior change.	
2.4	Behavior Change Campaign Strategy	Describe how the behavior change campaign was developed and what materials were used to advertise the campaign.	
3	Evaluation Plan Overview	Section 3.0 focuses on information about the evaluation.	
3.1	Evaluation Goals and Description	Describe the goal(s) for the behavior change campaign evaluation and how the goals were accomplished.	
3.2	Evaluation Location(s) and Target Population	Describe the locations where the evaluation was conducted (test site and if applicable, control sites), what the demographics of the target population were, and why the target population was selected for the evaluation. <i>Reference Chapter 3 of the Evaluation Guidance Manual for information about selecting a sample size.</i>	
3.3	Key Project Members: Roles and Responsibilities	Identify key team members of the project team, decision-makers, and/or stakeholders.	
3.4	Evaluation Schedule and Budget	Provide a schedule of when the evaluation tasks occurred along with the evaluation budget	
4	Evaluation Methods	Describe the data collected, the instruments used to collect data, and the methods used to evaluate changes in the target audiences understanding and adoption of the target behavior. <i>Chapters 4, 5, and 6 of the Evaluation Guidance Manual provide guidance for conducting the evaluation.</i>	
4.1	Data Overview	Describe what data was collected to meet the study objectives such as: type of data including whether the data was pre, post, or control, source of data, frequency of collection, number of samples collected, location of sampling,	
4.2	Instruments Used to Measure Change	Describe the different instrument(s) that were used during the evaluation to measure change, including how they were designed, validated, and the process during which they were used.	
4.2.1	Instrument Design	Describe how each instrument was designed.	
4.2.2	Instrument Validation	Describe the process that was employed to validate the instruments	
4.2.3	Data Collection Protocol	Define the procedures used to collect the various types of data collected from each instrument.	
4.3	Data Analysis Methods	Describe the process and methods that were used to analyze the data and address the campaign goals outlined in	
4.3.1	Qualitative Data Analysis Methods	Describe the process and methods that will be used to analyze qualitative data.	
4.3.2	Quantitative Data Analysis Methods	Describe the data analysis method(s) used for all types of quantitative data.	

D5.6 Report Template Checklist

4.3.3	Hypothesis Testing	If hypothesis testing was conducted, describe the testing method here.	
5	Results & Discussion	Section 5.0 describes the results from the analysis, whether the behavior change campaign was effective, as well as any challenges that occurred during the study that may have affected the results.	
5.1	Changes to Understanding and Adoption of Target Behavior	Describe whether there was a change in understanding and adoption of the target behavior.	
5.2	Challenges Identified During the Evaluation	Describe what challenges may have impacted the evaluation and what steps could be taken in the future to reduce the impact of these challenges.	
6	Future Action Recommendations	Describe the future action recommendations based on the results of the evaluation	
7	Glossary	This section the definition of key terms that are relevant to developing the report and should be updated before finalizing report.	
8	References	This section includes references relevant to developing the report template and should be replaced with reference relevant to the evaluation that is the focus of the final report.	
9	Appendices	It is optional to include raw data and other supporting documents (e.g., campaign materials and instruments used to evaluate change) in the report appendix however, this information should be available upon request. If these items are included in the report, the Appendix sections noted are suggested for including the information.	
9.1	Evaluation Schedule		
9.2	Campaign Materials		
9.3	Instruments		
9.4	Raw Data		

PUBLIC EDUCATION & OUTREACH

BEHAVIOR CHANGE CAMPAIGN EVALUATION & REPORT

CONDUCTED TO MEET NPDES MS4 PERMIT REQUIREMENTS

Phase I WWA S5.C.11.a.vi Phase II WWA S5.C.2.a.ii.(e)

Phase II EWA S5.B.1.b.

STUDY TITLE

INSERT PROJECT RELATED IMAGE HERE

Date Report Completed

Prepared For:

Jurisdiction

Department

Client Address Line 1

Client Address Line 2

Client Contact - Phone Number

Prepared By:

Organization

Consultant Address Line 1

Consultant Address Line 2

Phone Number

TEMPLATE INSTRUCTIONS

This template was developed to assist permittees in meeting the Public Education and Outreach (E&O) section of the Municipal Separate Stormwater System (MS4) Permit (MS4 Permit) requirements for evaluation and reporting on the understanding and adoption of a targeted behavior. Specifically, the following MS4 Permit sections: Western Washington (WWA) Phase I S5.C.11.a.vi-vii and WWA Phase II S5.C.2.a.ii.(e)-(f). As discussed in Section 1.1 of the Evaluation Guidance Manual, Eastern Washington (EWA) Phase II MS4 Permit requirements for Education and Outreach are different than the WWA requirements. Since the EWA MS4 Permit deadline for evaluating and reporting on E&O programs has passed guidance specific to the EWA MS4 Permit was not included in this document however this resource may still be useful for EWA Permittees to meet future E&O MS4 Permit requirements. Chapter 1 of the Evaluation Guidance Manual provides more discussion about the E&O MS4 Permit requirements relevant to this document.

The use of this template is not required to meet MS4 Permit requirements. The template was developed as part of a SAM Project to streamline report writing by identifying what information is required by the MS4 Permit. Suggestions for content and the basic information Ecology would like included (shown at the bottom of this page) are also included because they are common evaluation steps that can help permittees meet their MS4 Permit requirements. For these sections, if a particular study does not have information to populate the section, then either note “Not applicable” or “Relevant information was not collected”.

Items highlighted in this template are as follows:

- **Blue text** in a green box refers to items that are required by the MS4 Permits along with the specific permit section denoted in [brackets].
- Text shown in **Bold** are defined in the glossary of this document.
- **Red** text denotes the beginning of an example.

Except for the section headers, all text in this document are instructions or suggestions for content. Before finalizing this document, these items should be deleted and replaced with the information that is relevant to the study that is the focus of this report. In addition, pages ii to iv and the glossary should also be deleted.

An Evaluation Guidance Manual was developed as a companion to this document. The Manual provides Permittees with additional guidance regarding common evaluation methods and additional resources. References to the Manual are included throughout this template where additional guidance can be found.

Ecology has offered the following as information that may be helpful to include in your report to provide additional context above the basic permit reporting requirements in order to tell a more complete story of your process.

- Discussion regarding how Permittees went through social marketing practices such as **Community Based Social Marketing (CBSM)** or similar approach.
- An overview of the **campaign/strategy** that was developed or expanded as well as how the **campaign/strategy** was implemented.
- A description of what is known about the **water quality problem** that the campaign was developed to improve.
- Identify the **target behavior** and **target audience** marketed in the campaign.
- Describe what is known about the **barriers** and **motivators** of the **targeted behavior**.

- Discussion regarding the methods used to evaluate changes in the **target audiences'** understanding and **adoption** of the **targeted behavior** as well as a summary of the results.
- Describe recommended changes to make the **campaign** more effective as well as strategies and the planned process to achieve these results.
- It is optional to include raw data and other supporting documents (e.g., campaign materials and instruments used to evaluate change) in the report appendix however, this information should be identified/described in the report and available upon request.

REPORT TEMPLATE INFORMATION

This report template was developed using funds from the Stormwater Action Monitoring (SAM) by the authors noted below. The authors were advised and supported by a technical advisory group (TAC) made up for Phase I and Phase II Permittees from Eastern and Western Washington as well as Ecology Staff. Electronic copies of this document can be accessed at the following weblink: [add weblink](#). When the template is used for a specific study, this section should be deleted from the report.

AUTHORS

Aimee Navickis-Brasch, PhD, PE
Evergreen StormH2O

MaKenna Lindberg
Osborn Consulting, Inc.

Francesca White, PE
Evergreen StormH2O

Nicole Chen, EIT
Port of Seattle

TECHNICAL ADVISORY GROUP

Andrea Jedel, Ecology

Aislin Gallagher, Kitsap County

Ann Marie Pearce, Thurston County

Susan McCleary, City of Olympia

Katherine Straus, STORM Representative

Peggy Campbell, Snohomish County

Cammy Mills, Kitsap County and STORM Representative

Amanda Mars, Ecology

Eric Lambert, Clark County

Brian Morgenroth, City of Walla Walla

Jessica Shaw, City of Wenatchee

Laura Haren, City of Kent

Trey George, City of Spokane

PUBLICATION INFORMATION

The template user should insert information here about where the final report will be stored and accessible to the public. Include a [weblink](#) and/or contact information.

AUTHORS AND CONTACT INFORMATION

Insert author and contact information here:

Name

Organization

Title

Address

City, State, Zip Code

Email Address

Phone Number(s)

TABLE OF CONTENTS

1.0	EXECUTIVE SUMMARY	1
2.0	BEHAVIOR CHANGE CAMPAIGN.....	2
2.1	BEHAVIOR CHANGE CAMPAIGN BACKGROUND	2
2.2	TARGET AUDIENCE.....	3
2.3	TARGET BEHAVIOR AND BMPS	3
2.4	BEHAVIOR CHANGE CAMPAIGN STRATEGY	4
3.0	EVALUATION PLAN OVERVIEW.....	5
3.1	EVALUATION GOALS AND DESCRIPTION	5
3.2	EVALUATION LOCATION(S) AND TARGET POPULATION	5
3.3	KEY PROJECT MEMBERS: ROLES AND RESPONSIBILITIES	6
3.4	EVALUATION SCHEDULE AND BUDGET.....	6
4.0	EVALUATION METHODS.....	8
4.1	DATA OVERVIEW	8
4.2	INSTRUMENTS USED TO MEASURE CHANGE	8
4.2.1	<i>Instrument Design.....</i>	9
4.2.2	<i>Instrument Validation</i>	9
4.2.3	<i>Data Collection Protocol.....</i>	9
4.3	DATA ANALYSIS METHODS	10
4.3.1	<i>Qualitative Data Analysis Methods</i>	10
4.3.2	<i>Quantitative Data Analysis Methods</i>	10
4.3.3	<i>Hypothesis Testing.....</i>	10
5.0	RESULTS & DISCUSSION	11
5.1	CHANGES TO UNDERSTANDING AND ADOPTION OF TARGET BEHAVIOR.....	11
5.2	CHALLENGES IDENTIFIED DURING THE EVALUATION	12
6.0	FUTURE ACTION RECOMMENDATIONS	13
7.0	GLOSSARY	15
8.0	REFERENCES	19
9.0	APPENDICES	20
9.1	EVALUATION SCHEDULE.....	21
9.2	CAMPAIGN MATERIALS	22
9.3	INSTRUMENTS	23
9.4	RAW DATA	24

1.0 EXECUTIVE SUMMARY

The **executive summary** is a brief (300-500 word) non-technical summary of the project that is typically written for a more general audience and includes the “key” elements of the report. This may include the following:

- Describe the **behavior change campaign** and the **strategy** developed.
- Identify the target audience and the target behavior
- Note where the evaluation was conducted (location) and duration of the evaluation.
- Describe the evaluation goal(s) and objectives as well as how those objectives were accomplished.
- Note the type of **instruments** (i.e., survey, interviews, observations, etc.) used to measure changes in the **target audience’s** understanding and **adoption** of **behavior change**.
- Note the type of data collected (i.e., **pre, post, or control data**) and how much data was collected (i.e., **sample size**).
- Provide an overview of the evaluation results regarding changes in **behavior understanding** and **adoption** of the targeted behavior.
- Recommended changes to the **campaign** to be more effective.
- Provide an overview of the implementation plan for the campaign post evaluation.

2.0 BEHAVIOR CHANGE CAMPAIGN

This section only focuses on the campaign, information about the evaluation plan should be included in Section 4.

2.1 Behavior Change Campaign Background

This section describes the **behavior change campaign**. After reading this section, the reader should understand *what* factors were considered while developing the **behavior change campaign**, *where* the **campaign** was implemented, and *who* was responsible for developing and/or implementing the **campaign**. Suggestions for content include the following items:

[WWA Phase I: S5.C.11.a.iv.] & [WWA Phase II: S5.C.2.a.ii.(c).]

Describe the Community-Based Social Marketing (CBSM) or Social Marketing (SM) strategy that was developed and briefly describe the strategy.

- Describe how this was determined to be the best strategy, and how the strategy addressed the **barriers** and **motivators** that prevented and/or encouraged participation from the **target audience**.
- Provide an overview of the behavior change campaign including how the campaign was implemented and tailored to the community.
- Note whether the campaign was existing or new.
- If the campaign was existing, note how it was more effectively implemented or expanded for a new target audience and/or BMP.
- Introduce the Permittee(s) and/or organization that developed and implemented the **campaign**.
- Note where the campaign was implemented and briefly describe how the location(s) was selected in relation to the **campaign**. Provide maps, photos and/or diagrams to identify boundaries of these locations (Figure 2-1).
- More information about the CBSM and SM approach are detailed in Chapter 2 of the Evaluation Guidance Manual.

[WWA Phase I: S5.C.11.a.] & [WWA Phase II: S5.C.2.a.]

Identify what is known about the **water quality problem** that the campaign is designed to target and explain how the **water quality problem** was identified as a high priority.

***FOR EXAMPLE**, water quality data or TMDL information was used to identify a pollutant of concern in receiving water bodies. Alternatively, observations may also be used to identify a water quality problem: a Permittee’s maintenance crew may observe an increase of fats, oils, and grease in the stormwater system that discharges to receiving waters. This may drive the Permittee to select employees of fast-food restaurants as the target audience and the selected subject area may be proper dumpster management.*

- Explain how changes in the target audience’s behavior could reduce the water quality problem.

INSERT MAP OF AREA WHERE THE CAMPAIGN WAS IMPLEMENTED

FIGURE 2-1 AREA(S) WHERE THE CAMPAIGN WAS IMPLEMENTED

2.2 Target Audience

This section introduces the audience that the campaign was designed to target. After reading this section, the reader should understand *who* the **target audience** is and *how* the audience was selected. Suggestions for content include the following items:

[WWA Phase I: S5.C.11.a.] & [WWA Phase II: S5.C.2.a.]
Identify and describe the **target audience's** characteristics and demographics.
List any languages (besides English) used by the **target audience**.

- Describe community land uses where the **target audience** is located related to the **campaign**.
For example: residential, industrial, or commercial land uses.
- Describe other characteristics of the **target audience** as it relates to the MS4 Permit such as whether the audience is homeowners, teachers, school aged children, overburdened communities, businesses, engineers, construction contractors, developers, development review staff, land use planners, etc.
- If the **campaign** was designed to target multiple populations, briefly describe each population as they relate to the **campaign**.

[WWA Phase I: S5.C.11.a.] & [WWA Phase II: S5.C.2.a.] Describe how the **target audience** was selected as a high priority group that contributes to the **water quality problem**.

- Identify what is known about the size of the **target audience**.
- Discuss how **audience research** (if any) was conducted that helped inform the **campaign strategy**.
- Describe what was known about the **target audience's** relevant stormwater perceptions before the evaluation including the basis or source determining their perceptions. *For example, note if their perceptions were determined from information collected during focus groups, public comments, survey results, etc.*

2.3 Target Behavior and BMPs

This section will introduce readers to the **target behavior** that the **behavior change campaign** aims to have the **target audience** adopt and the **BMP** that was implemented to effect behavior change. *AN EXAMPLE of a BMP compared to a target behavior is if the target behavior is to stop illicit discharges, a spill kit would be the BMP.* This section introduces *how* target behaviors and BMPs were identified and described. Suggestions for **target behavior** content include the following items:

[WWA Phase I: S5.C.11.a.ii.] & [WWA Phase II: S5.C.2.a.ii.]
Describe the preferred **BMPs** and the **target behavior** intended for the **target audience** to adopt in accordance with the **campaign**.
Reference Figure 1-1 in the Evaluation Guidance Manual for a list of potential BMPs.

- Describe how the **target behavior** was identified and why it was selected.
- Describe the key **barriers** and **motivators** for the **target audience** that inhibited or strengthened **adoption** of the **target behavior**. Describe how the **barriers** and **motivators** were identified.
- If a literature search was conducted to identify **barriers** and/or **motivators** of the **target behavior**, briefly summarize those findings here including citations for the sources and include references in Section 8.0.

2.4 Behavior Change Campaign Strategy

This section will help readers understand *how* the **behavior change campaign** was developed and *what materials* were used to advertise the **campaign**. After reading this section, the reader should understand the strategy to implement the **behavior change campaign**. Suggestions for content include the following items:

- Describe any **materials** (e.g., fliers, postcards, brochures, stickers, spill kits, radio ad script, etc.) associated with the **campaign** and describe the process for selecting these materials. Providing a copy of the **materials** in Appendix 9.2 is optional, however, the documents should be available upon request from Ecology.
- If the **materials** were included in the appendix, reference that appendix section here.
- Explain how the **materials** were used to support the **campaign strategy**.
- Describe how the **campaign** was implemented including the type of media (e.g., radio, newspaper, social media, commercials, etc.) used to disseminate the **materials/message**.
- Note if the **behavior change campaign** materials were developed in more than one language.

[WWA Phase I: S5.C.11.a.] & [WWA Phase II: S5.C.2.a.]

Identify if any of the **behavior change campaign materials** were produced in more than one language.

3.0 EVALUATION PLAN OVERVIEW

This section focuses on the evaluation that was conducted to measure changes in understanding and **adoption** of the **targeted behaviors**.

3.1 Evaluation Goals and Description

After reading this section, readers should understand *what* the goal(s) were for the **behavior change campaign** evaluation and how the goals were accomplished. Suggestions for content include the following items:

- Define the evaluation goal(s) (i.e., the reason(s) the evaluation was conducted). **EXAMPLE(S)** of a goal relevant to MS4 Permit Requirements includes:
 - Measure the target populations changes in the understanding and adoption of targeted behaviors resulting from the implementation of the strategy.
 - Recommend changes to the campaign to be more effective
 - Use evaluation results to continue to direct effective methods and implementation of the ongoing behavior change program.
- Generally, describe how the evaluation goals were accomplished (*a more detailed discussion of the evaluation methods should be included in Section 4.0*). Suggestions for content include the following items:
 - Provide a short description of how the evaluation was conducted
 - If the evaluation was conducted in multiple phases, provide a short description of each phase including if **pilot testing** was conducted or if this study is considered a **pilot test**.
 - Provide a list of the evaluation objectives. An objective is a measurable statement that includes an action verb that defines how the project goal(s) will be accomplished. Once all the objectives are complete, the study goal should be achieved. **EXAMPLES** include:
 - Develop and implement a behavior change campaign
 - Developed and disseminate a survey to the target population
 - Analyze data collected to determine if there is change between the pre and post (or control) data sets.
 - Developed a final report and implementation plan (this document)

3.2 Evaluation Location(s) and Target Population

After reading this section, readers should understand the locations *where* the evaluation was conducted (test site and if applicable, control sites), *what* the demographics of the **target population** were, and *why* the **target population** was selected. If these details are the same as the **campaign** as described in Section 2.2, state that here and reference that section rather than repeat the information here. Suggestions for content include the following items:

- Explain if there are any characteristics unique to the **target population** that are different than the **target audience**. **FOR EXAMPLE:** *restaurant employees within city limits may be the target audience for a behavior change campaign, while the target population for the evaluation may be more specifically fast-food restaurant employees along a highly trafficked road.*

- Identify the approximate **target population** size and describe why this **sample size** was selected. If possible, describe how the sample size is representative of the entire **target audience**. A more detailed explanation about how to determine **sample size** is described in Chapter 3 of the Evaluation Guidance Manual.

[WWA Phase I S5.C.11.a.] & [WWA Phase II S5.C.2.a.]
 Based on the target audience’s demographic, the Permittee shall consider delivering its selected messages in languages other than English, as appropriate for the target audience.

- Note if the target population speaks languages other than English and if there are any differences in the language preference between the target audience and the target population.
- Note if there are any known media preference by the target population
- Identify the locations where the evaluation occurred.
- Note where the evaluation was implemented and briefly describe how the location(s) was selected. Provide maps, photos and/or diagrams to identify boundaries of these locations (Figure 3-3).



FIGURE 3-3 EVALUATION AREA

3.3 Key Project Members: Roles and Responsibilities

This section identifies key team members of the project team, decision-makers, and/or stakeholders (i.e., material distributor, data collector, data verifier, auditor, research assistant, QA/QC lead, etc.). Suggestions for this section include:

- List the key team members’ names and roles who were involved in developing and implementing the **campaign** as well as conducting the evaluation. A table format is preferred. See Table 3-1 for an example of how to list key project team members and their roles and responsibilities.

EXAMPLE

TABLE 3-1 KEY PROJECT TEAM MEMBERS: ROLES & RESPONSIBILITIES

KEY TEAM MEMBERS	ROLE
Name, Entity	Project Manager

3.4 Evaluation Schedule and Budget

After reading this section, readers should understand the schedule of when evaluation tasks occurred along with the evaluation budget.

Suggestions for the schedule content include:

- Describe when data collection occurred and for how long. Include information about when data collection occurred in relation to when the campaign was implemented.

- The schedule may include details of the tasks, the expected start and end dates of each task, deliverables, and deliverable deadlines that took place throughout the **campaign**. Documenting these details in table format is preferred. An example is provided in Table 3-2.

EXAMPLE**TABLE 3-2: PROPOSED BEHAVIOR CHANGE CAMPAIGN & EVALUATION TIMELINE**

TASK AND DELIVERABLES	2023											
	J	F	M	A	M	J	J	A	S	O	N	D
Task 1. Name	X	X	X	X	X	X	X	X	X	X	X	X
Task 2. Name	X	X	X									
Task 3. Name				X	X	X						
Task 4. Name							X	X	X			
Task 5. Name										X	X	X

If a budget is included, suggestions for the content include:

- Organize the budget into a table and separate the budget by tasks and subtasks that took place. A table format is preferred. See Table 4-3 for an example.
- Include items such as labor for preparing and validating the **instruments**, collecting data, special training needs, distributing educational **materials**, and any specialized contracting contributed (data **validation** and verification, **social marketing** firms, or other specialized services).

EXAMPLE**TABLE 4-3: SUMMARY OF FEES BY TASK**

TASK #	TASK TITLE	FEES
1	Project Management	\$\$
2	Task 2	

4.0 EVALUATION METHODS

This section focuses on the data collected, the **instruments** used to collect data, and the methods used to evaluate changes in the **target audiences** understanding and adoption of the target behavior. Discussion regarding the results should be included in Section 5.0. Guidance for conducting the evaluation is located in Chapters 4, 5, and 6 of the Evaluation Guidance Manual. *Note: the level of detail included in this section is based on recommendations from a SAM Study and exceeds what is needed to meet Ecology’s annual report requirements.*

4.1 Data Overview

After reading this section, the reader should understand what data was collected to meet the study objectives including items such as: type of data including whether the data was pre, post, or control, source of data, frequency of collection, number of samples collected, location of sampling, purpose of data, etc. A table format is preferred and Table 4-1 provides an example of how to organize the information.

EXAMPLE

TABLE 4-1: DATA NEEDED TO MEET EACH OBJECTIVE

DATA TYPE & INSTRUMENT USED TO COLLECT DATA	SAMPLE SIZE PLANNED	SAMPLE SIZE ACTUAL	HOW DATA WAS COLLECTED	PURPOSE
List of Businesses and Contact information	50	45	Google Maps & Bing Maps cross referencing	Identifies the target population of the evaluation; contact information will be used to schedule site visits
Baseline Observation Inspection	45	40	In-person site visit to collect data	Rule out social desirability bias
Baseline Survey Data	35	32	In-person site visit	Identify target audience barriers
Follow-Up Survey Data	35	28	In-person site visit	Measure adoption of targeted behaviors
Follow-Up Observational Inspection	45	42	In-person site visit to collect data	Rule out social desirability bias; help determine if behaviors have been adopted, ultimately proving the effectiveness of the evaluation

4.2 Instruments Used to Measure Change

After reading this section, readers should understand the different **instrument(s)** that were used during the evaluation to measure change, including how they were designed, validated, and the process during which they were used. In the context of E&O studies, **instruments** are a measurement device (i.e., a survey, test, observation log, interview questions, etc.) used to collect data which is analyzed to measure changes in the **target audiences** understanding and **adoption of target behaviors**. A more detailed discussion regarding

how to select **instruments** to measure **behavior change** is described in Chapter 4 of the Evaluation Guidance Manual.

4.2.1 Instrument Design

This section describes how each **instrument** was designed. For a more detailed description about selecting and designing **instruments**, see Chapter 4 in the Evaluation Guidance Manual. Suggestions for content include the following items:

- Describe the **instruments** used to measure change including how they were designed and developed to measure change in the understanding and adoption of the target behavior. Including a copy of the **instruments** as an appendix is optional, however, the documents should be available upon request from Ecology.
- Note why these **instruments** (e.g., survey, interview questions, observational checklist, etc.) were selected for the target audience.

4.2.2 Instrument Validation

This section describes the process that was employed to validate the **instruments**. **Validation** is the process to verify the **instrument** measures what it was intended to measure and produces consistent results. Suggestions for content below are specific to the **validation** method used, include only information that applies to the methods used on the project. If **instruments** were not validated for the project, that should be noted here. A more detailed discussion regarding **validation** methods is described in Chapter 4 of the Evaluation Guidance Manual.

- Note if **instruments** used were from similar studies that have already been validated. Include the name of the other studies, citations and references to these documents, and the **validation** methods that were used.
- Note if the **instruments** were **field tested** using focus groups, **pilot testing**, etc. and if so, describe how the results from this testing were used to refine the **instrument(s)** before it was implemented for the project.

4.2.3 Data Collection Protocol

This section should define the procedures used to collect the various types of data collected from each **instrument**. Defining these procedures and following them consistently minimizes errors and supports the integrity of the collected data. **Standard operating procedures (SOPs)** are the procedures that define specifically how to conduct an activity. **SOPs** should provide sufficient detail such that the activity is repeatable and can be reproduced by an individual (i.e. third party) unfamiliar with the evaluation.

SOPs for Behavior Change Campaign Evaluations may include:

- How the **instruments** were disseminated to the **target population** (i.e., mailer, email, web-based polls, social media, list-serve, one-on-one or group interview, etc.).
- How surveys were conducted including a summary of the instructions provided to the participant before they took the survey.
- How interviews were conducted including a summary of the instructions provided to the participant before starting an interview; how the participant's questions were addressed during the interview; and how **prompting** and/or **priming** participants was addressed.
- If the **SOPs** are detailed and long, consider putting a brief overview of the **SOPs** and including the actual **SOPs** in the Appendix.

4.3 Data Analysis Methods

After reading this section, readers should understand the process and methods that were used to analyze the data and address the **campaign** goals outlined in Section 3.1. There are different methods that can be used based on type of data (**qualitative**, **quantitative**, etc.) and the complexity of the analysis. Assistance for selecting data analysis methods is provided in Chapter 5 of the Evaluation Guidance Manual.

4.3.1 Qualitative Data Analysis Methods

This section describes the process and methods that will be used to analyze **qualitative** data. This typically includes (Schutt, 2011):

- Organizing the data into spreadsheets and categorizing the data into **codes** and **themes**
- Note if a **peer debriefing** process was used to validate the **codes** and **themes**.

4.3.2 Quantitative Data Analysis Methods

Describe the data analysis method(s) used for all types of **quantitative** data. Suggestions for content include:

- Summary of methods including equations that were used to analyze the data
- Summary of methods used to compare data sets
- A description regarding how multiple-choice, yes/no questions, or scaled survey responses were converted to numerical values such as a **Likert scale**

***EXAMPLE:** Basic statistics was used to calculate the percent change before and after the campaign for multiple choice and yes/no questions. This included dividing the total responses by the yes and no responses.*

4.3.3 Hypothesis Testing

Hypothesis testing is a statistical analysis used to test predictions (hypotheses) about the outcome of an evaluation. This analysis determines whether the difference between two data sets (pre and post data) are meaningful (significant) by determining the odds that the results happened by chance. Additional information about how to perform **hypothesis testing** is described in Chapter 6 of the Evaluation Guidance Manual. *If hypothesis testing was conducted, suggestions for content to include in this section are as follows:*

- Describe the testing method conducted and explain why this method was selected.
- Identify the data sets that were included in the hypothesis testing.
- List the null and alternative hypothesis for each data set with respect to whether a change was measured between the data sets.
- Note the selected confidence interval(s) the testing is based upon.

5.0 RESULTS & DISCUSSION

After reading this section, the reader should understand the results from the analysis, whether the behavior change campaign was effective, as well as any challenges that occurred during the study that may have affected the results.

5.1 Changes to Understanding and Adoption of Target Behavior

This subsection describes whether there was a change in understanding and **adoption** of the **target behavior**. Suggestions for content include the following items:

[WWA Phase I S5.C.11.a.vi.a.] & [WWA Phase II S5.C.2.a.ii.e.1.]
Evaluate and report on the changes in understanding and adoption of targeted behaviors resulting from the implementation of the strategy.

- Provide a brief description regarding whether the **target populations** understanding of the **target behavior** increased, decreased, or stayed the same. Explain your reasoning.
- Provide a brief description regarding whether the **target population's adoption** of the **target behavior** increased, decreased, or stayed the same. Explain your reasoning.
- For **quantitative** data, discuss the pre and post **campaign** data, and/or **control population** data as well as the differences between the data sets.
- For **qualitative** data, provide a summary of the common **themes** found in the pre and post data and discuss how the **themes** changed between the **pre** and **post data**, and/or **control data**.
- Where possible, summarize results into figures, tables, charts, etc.
- Include a copy of the **raw data** in tables in an appendix (optional).
- If hypothesis testing was used to measure the change in understanding or adoption of the targeted behavior, state if there was a **statistically significant difference** between the pre and post data or control data. A more detailed description about conducting a statistical analysis is described in Chapter 6 of the Evaluation Guidance Manual.

***EXAMPLE** of the hypothesis results in a table format is shown in Table 5-1.*

- Provide an explanation if the result of the statistical analysis makes sense and why
***EXAMPLE:** The results of the statistical analysis indicated a decrease in commercial car wash use. The decrease in **adoption** appears to be a market trend, caused by Covid-19 impacts, that has been observed by commercial car washes throughout the nation.*

TABLE 5-1: HYPOTHESIS TESTING RESULTS

TOPIC FOR COMPARISON	P VALUE	STATISTICALLY SIGNIFICANT DIFFERENCE? (P-VALUE < 0.05)
Commercial Car Wash BMP Adoption	0.010	Yes ¹
Residential Car Wash BMPs Adoption	0.153	No

5.2 Challenges Identified During the Evaluation

This subsection will allow readers to understand *what* challenges may have impacted the evaluation and *what* steps could be taken in the future to reduce the impact of these challenges. Suggestions for content include the following items:

- Identify and discuss challenges that potentially impacted the **campaign, instruments**, schedule, or budget. Documenting the challenges in a table format is recommended.
EXAMPLE of challenges identified in a table format shown in Table 5-2.
- Describe steps (mitigation approach) taken to address the challenges identified during the evaluation and note how or if they were resolved.

TABLE 5-2: CHALLENGES AND MITIGATION APPROACH

CHALLENGES	MITIGATION APPROACH
Limited availability of staff to conduct the study	Other jurisdictions assisted with field work and other duties
Target audience unwillingness to complete the survey	Incentives was offered upon the completion of the survey

6.0 FUTURE ACTION RECOMMENDATIONS

After reading this section, the reader should understand *what* future action recommendations are based on the results of the evaluation. Based on the recommendations, suggestions for content include the following:

[WWA Phase I S5.C.11.a.iii.] & [WWA Phase II S5.C.2.ii.b.]

Evaluate and report on any changes to the campaign in order to be more effective; describe the strategies and process to achieve the results.

- Document lessons learned and recommendations for which **campaign's strategy** and schedule will more effectively be implemented based on discussion from Section 5.0.
- Discuss how the lessons learned can be applied to future studies.
 - Consider recommendations that:
 - Increase E&O material outreach to the **target audience**. If any E&O **materials** were translated into a different language other than English, consider strategies to reach that demographic.
 - Increase **campaign** participation
 - Improve campaign instruments
 - Conduct a literature search to identify recommendations and strategies to achieve the recommendations
 - **EXAMPLE** of a recommendation: *Sending E&O materials with utility bill to increase E&O program reach to the target audience.*

[WWA Phase I S5.C.11.a.vii.] & [WWA Phase II S5.C.2.f.]

Use the results of the **campaign** evaluation to direct effective methods for implementation and E&O resources effectively.

- Describe the strategies and process to implement changes to the **campaign** to be more effective.
 - If the desired results were not achieved state how this will be resolved in the future.
EXAMPLE: *There was no or only a small measurable difference between the pre and post data that indicated a change in the target audiences understanding and adoption of targeted behaviors. The strategy will be re-evaluated through pilot testing and revised based on the results.*
 - **EXAMPLE:** *There was a low response rate from the community who first languages was other than English. For future evaluations, the Permittee will strive to increase the response rate by hosting focus groups with these communities to identify alternatives methods for disseminating campaign materials.*
 - Develop and describe an implementation plan to execute changes to the **behavior change program**. The plan should include what the recommendations are, how they were identified, when the recommendations will be implemented, and what is needed to implement the recommendations, such as budget and staff.
 - Documenting the implementation plan in a table format is preferred.

***EXAMPLE** of an implementation plan is shown in Table 6-1. The text in the table is example entries and should be replaced by the lead or participating entity with their preferred recommendations.*

TABLE 6-1: IMPLEMENTATION PLAN

RECOMMENDATIONS	BUDGET	BUDGET SOURCE	STAFF	IMPLEMENTATION TIME FRAME
Focus on face-to-face distribution of behavior change campaign materials at the State Fair	\$100	Jurisdiction	Stormwater Staff	State Fair 2023
Revise behavior change campaign material to increase interaction with posts and education material.	N/A	N/A	Webmaster/Public Relations staff	January 2023

7.0 GLOSSARY

Adoption: Also referred to as behavior adoption. *See Section 2, 5, and 7.* The action (or inaction) of participating in the target behavior; adoption levels are measurable using evaluation instruments and analysis.

Alternative hypothesis: *Section 6.* A prediction that is accepted if the null hypothesis is rejected through hypothesis testing.

Audience research: Investigating the target audience's characteristics, habits, pre-existing beliefs, etc. prior to the behavior change campaign taking place that will divulge useful information to improve the campaign effectiveness.

Barriers: Any physical, spiritual, or societal obstacles preventing a target audience from adopting a target behavior.

Basic statistics: Quick methods of analyzing the difference between data sets that often does not require the use of complex statistical software. Evaluating the differences between measures of central tendency (mean, median, and mode) of two or more data sets are common examples of basic statistics. For example, calculating the average pre and post data response to compare differences.

Behavior change: Similar to behavior adoption, behavior change is the act of adopting new BMPs in lieu of not participating in them.

Behavior change campaign: A portion of the E&O program that is directed to a target audience to change a specific behavior that negatively impacts water quality of the downstream watershed.

Behavior understanding: The target audiences understanding of why preferred BMPs should be adopted including the negative implications of not adopting the BMPs.

Best management practices (BMPs): The schedules of activities, prohibitions of practices, maintenance procedures, and structural and/or managerial practices approved by Ecology that, when used singly or in combination, prevent or reduce the release of pollutants and other adverse impacts to waters of Washington State.

Codes: To analyze open ended data, it is grouped by like-items, and then like-items are given a name that is referred to as a code. Codes are used as a way to organize data to quickly pull assumptions about a group of similar data. Coded data is then organized into themes. Reference the Themes definition for more details.

Campaign: References to the campaign are referring to the behavior change campaign. See behavior change campaign.

Community-based social marketing (CBSM): Community-based social marketing (CBSM) blends social marketing with social and environmental psychology to develop tools that foster lasting behavior change. (McKenzie-Mohr, 2011)

Control data: This data is what was collected from the control population, who has not had any exposure of the behavior change campaign, that provides comparable variable to provide information of the target audiences understanding and adoption of the targeted behaviors. Control data is often collected when it is not possible to collect pre (baseline) or post (follow-up) data. See control populations for additional information.

Control populations: The sample group that produces control data resulting from no changing variables and can be compared against the target population data which has had changing variables, meaning they have not been exposed to the behavior change campaign. The control population typically has similar characteristics as the target population but may be in a different geographic region. For example, if an existing behavior change campaign is the focus on the evaluation it would not be possible to collect baseline data from the target audience. Instead, baseline data is collected from a control population and compared to the post data collected from the target population to measure the change in understanding and adoption of targeted behaviors.

Education and outreach (E&O) program: The overall MS4 Permit requirement to provide stormwater education and measure a behavioral change campaign to a selected target audience, as to provide stewardship opportunities to the general public.

Executive summary: A non-technical summary of the project that is typically written for a more general audience and includes the “key” elements of the report.

Field testing: One or more methods of validating the behavior change campaign methods and instruments should be conducted before broad implementation of the overall behavior change campaign or evaluation. Some examples of field testing include pilot testing or peer review.

Hypothesis testing: A statistical analysis used to test predictions (hypotheses) about the outcome of an evaluation. This analysis determines whether the difference between two data sets (pre and post data) are meaningful (significant) by determining the odds that the results happened by chance.

Instruments: A measurement device (i.e., a survey, test, observation log, interview questions, etc.) used to measure changes in the target audiences understanding and adoption of target behaviors.

Likert Scale: often represented as descriptive answer options that can be easily quantified and analyzed: “Always, Frequently, Sometimes, Never” is one example of a Likert scale.

Materials: Physical items that are used to encourage participation or seen as educational tools used for the behavior change campaign.

Motivators: Incentives for the target audience to take part in the target behavior. These may be tangible items distributed from participating jurisdictions or monetary, socially desirable, or esteem-boosting incentives.

Null hypothesis: A prediction that is assumed to be true unless there is strong evidence against it which is determined through hypothesis testing.

Peer debrief: A technique used in qualitative research where staff who are familiar with the topic that is being evaluated meet to review and discuss the coding that were developed by the researcher for the purpose of validating (through agreement) the coding.

Pilot testing: A quality control method to validate the data collection methods, instruments, and/or campaign on a selected group of the target audience during a practice-round before broadly implementing the campaign and conducting the follow up evaluation. The group who took part in the pilot test will not be used for the target population during an evaluation for the same campaign.

Post data: Also commonly referred to as *follow-up data*. This data is what was collected after the behavior change campaign was implemented that provides information of the target audiences understanding and adoption of the targeted behaviors.

Pre data: Also commonly referred to as *baseline data*. This data is what was collected before the behavior change campaign was implemented that provides information of the target audiences preexisting understanding and adoption of targeted behaviors.

Priming: Also known as prompting. This is a method of intentionally or unintentionally leading the respondent to the desired response through advanced training on the desired response before asking the question.

Prompting: Also known as priming. This is a method of intentionally or unintentionally leading the respondent to the desired response through coerced or particularly worded questioning.

Qualitative: Data that is presented through descriptions or words and is not representative of mathematically applicable values (Takona, 2002).

Quantitative: Numerically significant data values of which mathematical operations of addition, subtraction, multiplication, and division can apply without needed conversion (Takona, 2002).

Raw data: Collected data that has not gone through any flagging, coding, or analysis.

Regional collaboration: Collective jurisdictions and/or agencies combining efforts to achieve MS4 Permit requirements.

Sample size: The portion of the target audience that data will be collected from (see target population). The sample size is set to gather a representative understanding of the target audience as a whole. This size is determined using several methodologies described in Chapter 3 of the Evaluation Guidance Manual.

Social marketing (SM): A process that applies marketing principles and techniques to create, communicate, and deliver value in order to influence target audience behaviors that benefit society (public health, safety, the environment, and communities) as well as the target audience (Lee & Kotler, 2011).

Standard operating procedures (SOPs): The procedures that define specifically how to conduct an activity such as an interview or how surveys were deployed; SOPs should provide sufficient detail such that the activity is repeatable and can be reproduced by an individual (i.e. third party) unfamiliar with the evaluation.

Statistically significant difference: A detectable change between data sets that can be calculated based on a predetermined confidence interval.

Strategy: A method for targeting the preferred behaviors and discouraging the undesirable behaviors.

Study area: The geographical location that behavior change measurements took place.

Target audience: The group that the behavior change campaign is directed towards who participate in the behavior that impacts stormwater quality.

Target behavior: The behavior aimed for the target audience to adopt by implementing a BMP which could improve a water quality issue. See BMP.

Target population: A subgroup of the target audience that includes all members of the ideal sample size. This subgroup of the target audience should of a size that represents the understanding and behaviors of the target audience as a whole.

Themes: Coded data can be grouped together based on the interpreted meaning to narrow down open-ended and/or qualitative data into select categories or themes. For example, barriers which may be filtered into common codes which were organized into primary themes may include: convenience, cost, space, or lack of knowledge.

Validation: The process to verify the instrument measures what it was intended to measure and produces stable results (Guba, 1981).

Water quality problem: Stormwater related pollutant(s) that is triggering the need for the behavior change campaign. This is identified by previous data or observations known to contribute to poor water quality in water bodies.

8.0 REFERENCES

Guba, E. G. (1981). Criteria for Assessing the Trustworthiness of Naturalistic Inquiries. *ERIC/ECTJ Annual Review Paper*, 29(2), 75-91.

Lee, N. R., & Kotler, P. (2011). *Social Marketing: Influencing Behaviors for Good*. SAGE Publications.

McKenzie-Mohr, D. (2011). Introduction: Fostering Sustainable Behavior. *Fostering Sustainable Behavior-Community Based Social Marketing.*, 1-11.

QuestionPro. (2022). *Quantitative Data: Definition, Types, Analysis and Examples*. Retrieved from QuestionPro Survey Software.

Takona, J. (2002). *Educational Research: Principles and Practice*. Lincoln, Nebraska: Writers Club Press.

9.0 APPENDICES

9.1 Evaluation Schedule

Provide a range of dates for when the tasks and subtasks for developing, implementing, and evaluating the **campaign**. A table format is preferred.

EXAMPLE

TABLE 9-1 EVALUATION TIMELINE

TASK & DELIVERABLE	START DATE	COMPLETION DATE
Task 1:		
Subtask		

9.2 Campaign Materials

It is optional to include supporting documents such as campaign materials in the appendix however, this information should be identified/described in the report and available upon request. Delete this section if these documents are not included.

9.3 Instruments

It is optional to include supporting documents such as instruments used to evaluate change in the report appendix however, this information should be identified/described in the report and available upon request. Delete this section if these documents are not included.

9.4 Raw Data

It is optional to include raw data in the report appendix however, this information should be identified/described in the report and available upon request. Delete this section if these documents are not included.

Evaluating the Effectiveness of Stormwater Education and Outreach: Permittee Guidance for Addressing Challenges through Behavior Change

Evaluation Guidance Manual

January 12, 2023

Prepared by:

Evergreen StormH2O
PO Box 19812
Spokane, WA 99228



ACKNOWLEDGEMENTS

This Manual was made possible by the Stormwater Action Monitoring (SAM) program, funded by Permittees and administered by the Washington State Department of Ecology (Ecology). The Manual is one of several deliverables developed for the following project: *Evaluating the Effectiveness of Stormwater Education and Outreach: Permittee Guidance for Addressing Challenges through Behavior Change*. The Manual authors are noted below. Special thanks to McKenna Lindberg, Osborn Consulting, Inc., and Nicole Chen, EIT, Port of Seattle, for their supporting contributions to the Manual development.

Aimee Navickis-Brasch, PhD, PE
Evergreen StormH2O

Francesca White, PE
Evergreen StormH2O

TABLE OF CONTENTS

1.0	INTRODUCTION	1-1
1.1	MANUAL PURPOSE AND BACKGROUND	1-1
1.2	MANUAL ORGANIZATION	1-2
1.3	ADDITIONAL RESOURCES	1-3
1.3.1	<i>Behavior Change Campaign Evaluation Report Template.....</i>	<i>1-3</i>
1.3.2	<i>Website – Tools and Resources for Behavior Change Programs</i>	<i>1-3</i>
1.4	SOCIAL MARKETING AND COMMUNITY-BASED SOCIAL MARKETING RESOURCES	1-3
1.4.1	<i>Professional Organizations:</i>	<i>1-3</i>
1.4.2	<i>Practices and Methods Papers:</i>	<i>1-4</i>
1.4.3	<i>Informational Resources:</i>	<i>1-4</i>
1.4.4	<i>Informational Resources Created and Maintained by For-Profit Subject Matter Specialists/Practitioners:</i>	<i>1-5</i>
2.0	SAMPLE SIZE SELECTION	2-1
2.1	CHAPTER OVERVIEW	2-1
2.2	TARGET AUDIENCE VS TARGET POPULATION	2-1
2.3	FACTORS INFLUENCING SAMPLE SIZE.....	2-2
2.3.1	<i>Level of Precision</i>	<i>2-2</i>
2.3.2	<i>Confidence Level and Interval</i>	<i>2-2</i>
2.3.3	<i>Degree of Variability</i>	<i>2-3</i>
2.4	STRATEGIES FOR SELECTING SAMPLE SIZE.....	2-3
2.4.1	<i>Census for Small Populations.....</i>	<i>2-3</i>
2.4.2	<i>Sample Size of Similar Study.....</i>	<i>2-3</i>
2.4.3	<i>Published Tables.....</i>	<i>2-4</i>
2.4.4	<i>Formulas.....</i>	<i>2-5</i>
2.4.5	<i>Website Calculators.....</i>	<i>2-5</i>
2.5	TARGETED SAMPLE SIZE VS ACTUAL SAMPLE SIZE.....	2-6
2.6	RANDOM SAMPLE COLLECTION	2-7
3.0	EVALUATION INSTRUMENTS	3-1
3.1	CHAPTER OVERVIEW	3-1
3.2	INSTRUMENT TYPES AND SELECTION CONSIDERATIONS.....	3-1
3.2.1	<i>Surveys.....</i>	<i>3-2</i>
3.2.2	<i>Interviews.....</i>	<i>3-3</i>
3.2.3	<i>Focus groups</i>	<i>3-3</i>
3.2.4	<i>Observations.....</i>	<i>3-4</i>
3.2.5	<i>Photos</i>	<i>3-4</i>
3.2.6	<i>Drawings</i>	<i>3-5</i>
3.3	CONSIDERATIONS FOR DESIGNING INSTRUMENTS.....	3-5
3.3.1	<i>Survey and Interview Questions.....</i>	<i>3-6</i>
3.3.2	<i>Survey Design Resources.....</i>	<i>3-7</i>
3.3.3	<i>Social Desirability</i>	<i>3-7</i>
3.3.4	<i>Target Audience Research.....</i>	<i>3-8</i>
3.3.5	<i>Use an Existing Instrument.....</i>	<i>3-8</i>
3.3.6	<i>Multiple Instruments</i>	<i>3-8</i>
3.3.7	<i>Checklists for Collecting Observational Data</i>	<i>3-9</i>
3.4	VALIDATING INSTRUMENTS.....	3-10
4.0	DATA TYPES	4-1
4.1	CHAPTER OVERVIEW	4-1

4.2	DATA MANAGEMENT.....	4-1
4.3	QUALITATIVE.....	4-2
4.4	QUANTITATIVE	4-3
4.5	CODING QUALITATIVE DATA.....	4-3
4.6	CONVERTING QUALITATIVE DATA TO QUANTITATIVE DATA	4-4
5.0	DATA ANALYSIS.....	5-1
5.1	CHAPTER OVERVIEW	5-1
5.2	DESCRIPTIVE STATISTICS.....	5-1
5.2.1	<i>Mean</i>	5-1
5.2.2	<i>Median</i>	5-2
5.2.3	<i>Mode</i>	5-2
5.2.4	<i>Standard Deviation</i>	5-3
5.2.5	<i>Range</i>	5-4
5.3	EVALUATING RESULTS	5-4
5.3.1	<i>Comparing Results</i>	5-4
5.3.2	<i>Hypothesis Testing</i>	5-5
5.4	SOFTWARE OPTIONS	5-8
6.0	REFERENCES	6-1

1.0 Introduction

1.1 Manual Purpose and Background

In Washington State, discharges from Municipal Separate Storm Sewer Systems (MS4s) are regulated under a combined National Pollutant Discharge Elimination System (NPDES) and State Waste Discharge General Permit (MS4 Permit). The MS4 Permits require Permittees to implement a Stormwater Management Program (SWMP) that includes an Education and Outreach (E&O) Program that is designed to meet specific goals. These goals vary depending on whether the permit is issued to a Western Washington (WWA) Phase I or Phase II jurisdiction or an Eastern Washington (EWA) Phase II jurisdiction. **Table 1-1** Provides a summary of the goals based on the different types of permits.

Table 1-1 MS4 E&O Program Design

Education and Outreach in MS4 Permit	WWA Phase I	WWA Phase II	EWA Phase II
Build general awareness about methods to address and reduce stormwater runoff.	✓	✓	
Effect behavior change to reduce or eliminate behaviors and practices that cause or contribute to adverse stormwater impacts.	✓	✓	
Create stewardship opportunities that encourage community engagement in addressing the impacts from stormwater runoff.	✓	✓	
Educate target audiences about the impacts of stormwater discharges to water bodies and the steps to take to reduce pollutants in stormwater.			✓

In the 2019–2024 WWA and EWA MS4 Permits, all three included requirements for evaluating and reporting on behavior change campaigns (WWA) and E&O programs (EWA) for changes in the understanding and adoption of target behaviors. The specific requirements vary between the WWA and EWA Permits, as shown in **Figure 1-1** and **Figure 1-2**. Most notably, the WWA Permits explicitly state to “*follow social marketing practices and methods, similar to community based social marketing (CBSM) and develop a campaign that is tailored to the community, including development of a program evaluation plan. Each Permittee shall: develop a strategy and schedule (a) to more effectively implement the existing campaign or; (b) to expand the existing campaign to a new target audience or BMP; or (c) for a new target audience and BMP behavior change campaign*” [WWA Phase I S5.C.11.a.iv. and Phase II S5.C.2.a.ii.(c)]. In contrast, the EWA Permit does not mention social marketing or CBSM [EWA Phase II S5.B.1.]. However, in all three permits, Permittees have similar requirements for evaluating and reporting on the understanding and adoption of targeted behaviors.

The specific WWA Permit language [Phase I S5.C.11.a.vi. and Phase II S5.C.2.a.ii.(e)] is as follows:

No later than March 31, 2024, evaluate and report on:

- 1. The changes in understanding and adoption of targeted behaviors resulting from the implementation of the strategy; and*

2. *Any planned or recommended changes to the campaign in order to be more effective; describe the strategies and process to achieve the results.*

The purpose of this Manual is to assist Permittees with meeting the evaluation requirement, which is due on March 31, 2024, for WWA and was due on December 31, 2021, for EWA. While the EWA MS4 Permit deadline passed before this document was developed, this resource may still be useful for EWA Permittees to meet future E&O MS4 Permit requirements.

This Manual describes professionally recommended approaches and concepts that can be used to conduct an evaluation of the behavior change campaign component of the E&O requirements; however, it is not required that this Manual be used to meet MS4 Permit requirements. It was designed to provide guidance for any size project. Some projects may use information from every chapter to conduct a successful evaluation, and some projects may not need that much detail to conduct a successful evaluation. **For example**, small projects that use photograph comparisons or observational data will likely use very little math and may not find Chapter 5 useful, whereas larger projects that collect survey data may find Chapter 5 useful to analyze their data and support their results. While this Manual provides resources about social marketing and CBSM (Social Marketing and Community-Based Social Marketing Resources [Section 1.4]), it does not provide guidance for developing a behavior change campaign.

Note: MS4 Permit language has been included in this document to provide the reader with context for why this Manual was developed. There are slight variations between the WWA MS4 Phase I and Phase II Permits, and the MS4 Permits are updated and reissued every five years. Please refer to the current version of the MS4 Permit that applies to your jurisdiction for exact permit language.

1.2 Manual Organization

The following provides information about the content of each chapter in this document, and each chapter has examples of how to apply the approaches and concepts described. Chapters 2 through 5 are in the order typically followed when an evaluation plan is developed. Ideally, an evaluation plan should be developed prior to implementing a campaign so that all the data can be developing during the campaign.

- **Chapter 1 Introduction** (this chapter) introduces the reader to the Manual and describes relevant permit requirements and the Manual's purpose and organization. It also provides additional resources that were developed as part of this project, as well as information about social marketing and CBSM resources.
- **Chapter 2 Sample Size Selection** provides an overview of common methods for selecting a minimum sample size to evaluate the campaign.
- **Chapter 3 Evaluation Instruments** provides tools (surveys, observational checklists, etc.) used to measure the target audience's change in the understanding and adoption of the targeted behavior campaign for the E&O program. It also provides information about different evaluation instruments, including suggestions for selecting and validating instruments.
- **Chapter 4 Data Types** introduces the different types of data and provides guidance for organizing both qualitative and quantitative data in preparation for data analysis.
- **Chapter 5 Analysis Methods** provides an overview of common data analysis methods for both qualitative and quantitative data. Discussion about the values of hypothesis testing is also included.

1.3 Additional Resources

As part of the SAM project, two additional tools were developed to assist Permittees with meeting their E&O requirements for evaluating behavior change campaigns. These resources include a report template and a website. The following provides additional details about these resources.

1.3.1 Behavior Change Campaign Evaluation Report Template

A report template was developed to support Permittees in meeting their E&O MS4 Permit requirements for reporting on the evaluation of understanding and adoption of targeted behaviors. It is recommended that the report template be used in tandem with this Manual. The report template was developed to streamline report writing by identifying what information is required by the MS4 Permits and providing suggestions for content, which includes informal suggestions from Ecology as basic information to include in a report. The suggestions for content are included because they are common steps in the evaluation process and provide a more complete story of this process. However, all that is required to be submitted to Ecology is what is written in the MS4 Permits, which were written to provide Permittees with flexibility for reporting their process and results.

1.3.2 Website – Tools and Resources for Behavior Change Programs

The website waterbehaviorchange.org was created to provide tools and resources for behavior change programs. It is recommended that the website be used as a companion to this document to help jurisdictions assess the effectiveness of existing campaigns around the country. The website compiles every known evaluation of a behavior change campaign in stormwater or water quality. The site provides details on campaign implementation and evaluation and rates the research quality of the evaluation.

1.4 Social Marketing and Community-Based Social Marketing Resources

Mention of social marketing and community-based social marketing (CBSM) first appeared in the 2019–2024 WWA MS4 Permits as methods for developing behavior change campaigns. Social marketing has been around since the early 1970s (Social Marketing Services, Inc., 2008) and is used to promote behavior change that improves public health and prevents injuries. Lee & Kotler define social marketing as “*a process that applies marketing principles and techniques to create, communicate, and deliver value in order to influence target audience behaviors that benefit society (public health, safety, the environment, and communities) as well as the target audience*” (Lee & Kotler, 2011). CBSM blends social marketing with social sciences (social and environmental psychology) and draws from the concept that sustainable behavior change is most effectively achieved when it involves direct contact with people and initiatives delivered at the community level (McKenzie-Mohr, 2011).

Providing guidance for developing a behavior change campaign is beyond the scope of this Manual. Instead, resources below provide more information about social marketing and CBSM. Additional resources provided by Technical Advisory Committee (TAC) members are included in this section. These resources are organized into the following categories: professional organization, practices and methods papers, informational resources (not tied to a for-profit entity), and informational resources that are created and maintained by for-profit subject-matter specialists/practitioners.

1.4.1 Professional Organizations:

- **Pacific Northwest Social Marketing**, <https://pnsma.org/> – Professional organization made up of members of the social marketing community across the Pacific Northwest
 - Learning forums and events
 - SPARKS Conference (annual)
 - Resources
- **Social Marketing Association of North America (SMANA)**, <https://smana.org/> – U.S.-based social marketing association serving Central and North America
 - Listserv
 - Webinars, events, training opportunities
 - Conferences
 - [Resources & Guiding Principles](#)
 - [Social Marketing Quarterly](#) (peer-reviewed journal) for members
- **International Social Marketing Association (iSMA)**, <https://isocialmarketing.org/>
 - Webinars, events, trainings, and news

1.4.2 *Practices and Methods Papers:*

- **Consensus Definition of Social Marketing**, <http://smana.org/wp-content/uploads/2017/04/iSMA-Consensus-definition-of-Social-Marketing-Oct-2013.pdf>
- **Global Consensus on Social Marketing Principles, Concepts and Techniques**, <http://smana.org/wp-content/uploads/2017/04/ESMA-AASM-SMANA-endorsed-Consensus-Principles-and-concepts-paper.pdf>
- **Social Marketing Evidence of Effectiveness, 2018**, <http://smana.org/wp-content/uploads/2018/11/Final-List-of-Key-Social-Marketing-Evidence-of-Effectiveness-citations-Nov-2018.pdf>
- **DRAFT Social Marketing Statement of Ethics** (issued for consultation in September 2022, with end of consultation February 15, 2023), https://docs.google.com/document/d/1pCpW15DPyL5a5-D9Ip2x1-gOA_Kr-znK/edit

1.4.3 *Informational Resources:*

- **Social Marketing: Messaging for Behavior Change**, <https://www.epa.gov/system/files/documents/2022-11/EPA%20Social%20Marketing%20Training-%2010.25.22%20FINAL.pdf>
150-page module developed for U.S. Environmental Protection Agency
- **Creating Messages that Drive Behavior Change**, <https://www.epa.gov/recyclingstrategy/creating-messages-drive-behavior-change>, U.S. Environmental Protection Agency
 - What is Social Marketing?
 - Learn How to Create a Social Marketing Program (recorded presentations, presentation slides, guide, and worksheet)

- **Getting Your Feet Wet with Social Marketing**, <https://fyi.extension.wisc.edu/wateroutreach/files/2015/12/GettingYourFeetWet1.pdf>, A Social Marketing Guide for Watershed Programs, Jack Wilbur, Utah Dept of Agriculture & Food – Designed to walk watershed groups, municipalities, etc., through the process of developing and implementing a watershed outreach campaign, including stormwater focus.
- **Chesapeake Behavior Change** <https://www.chesapeakebehaviorchange.org/>
Serves as a repository to publish behavior change campaigns and view other organizations' campaigns to encourage collaboration and learning opportunities.

1.4.4 Informational Resources Created and Maintained by For-Profit Subject Matter Specialists/Practitioners:

- **Social Marketing Primer and Step-by-Step Guide and Workbook**, <https://cplusc.com/social-marketing-workbook/>, Downloadable PDFs created by C+C
 - **Social Marketing Primer**, <https://cplusc.com/wp-content/uploads/2022/10/CC-Social-Marketing-Primer.pdf>, [linked on EPA's site]
 - **Planning for Effective Social Marketing Campaigns: A Step-by-Step Guide and Workbook**, <https://cplusc.com/wp-content/uploads/2021/09/CC-Social-Marketing-Workbook.pdf>
- **Tools of Change**, <https://toolsofchange.com/en/home/>, – Collection of behavior change, social marketing, and community-based social marketing case studies
 - Planning guide
 - Case studies
 - Topic resources
 - Webinars and workshops
- **Community-Based Social Marketing**, <https://cbsm.com/> – CBSM founder, Doug McKenzie-Mohr's site includes:
 - Upcoming training opportunities listed on site
 - Resource links to articles, reports, cases, forums, and colleagues
 - Publications: *Fostering Sustainable Behavior* (Third Edition), Doug McKenzie-Mohr, PhD
- **Social Marketing Services, Inc.**, <https://www.socialmarketingservice.com/> – Nancy Lee, author of 15 books on social marketing. This website site includes links to:
 - Planning worksheets (free, [downloadable PDF](#) provides step-by-step guide through the process of creating a social marketing plan)
 - Journal articles
 - Upcoming learning opportunities are listed on iSMA, SMANA, and Nancy's website.
 - Publications co-authored by Nancy Lee and Philip Kotler:
 - Success in Social Marketing: 100 Case Studies from Around the Globe (2022)
 - Social Marketing: Changing Behaviors for Social Good (Sixth Edition, 2023)
 - Social Marketing to Protect the Environment: What Works (2011)
 - Marketing in the Public Sector: A Roadmap for Improved Performance (2006)

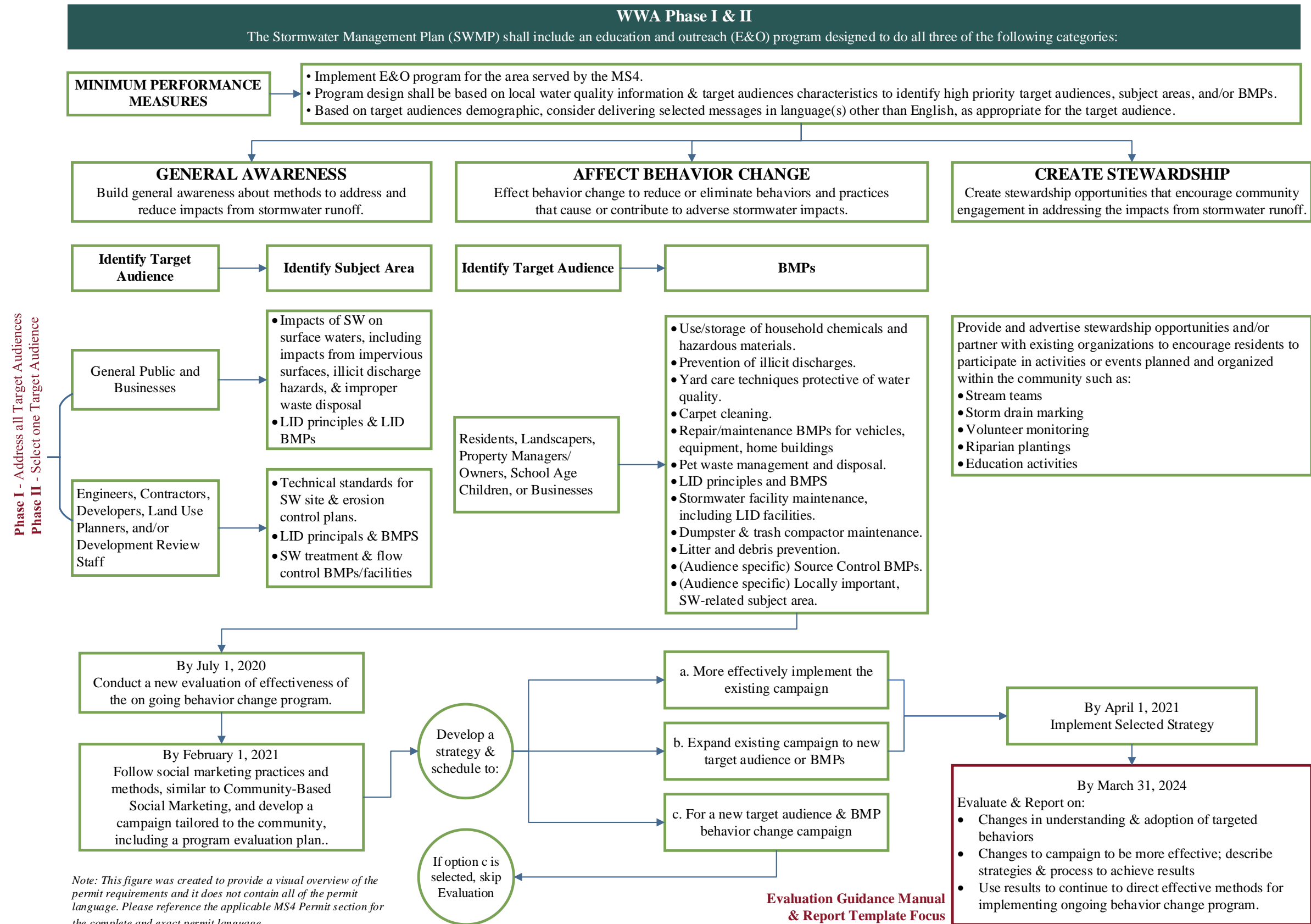


Figure 1-1 WWA Phase I & II Illustration of MS4 E&O Program Requirements

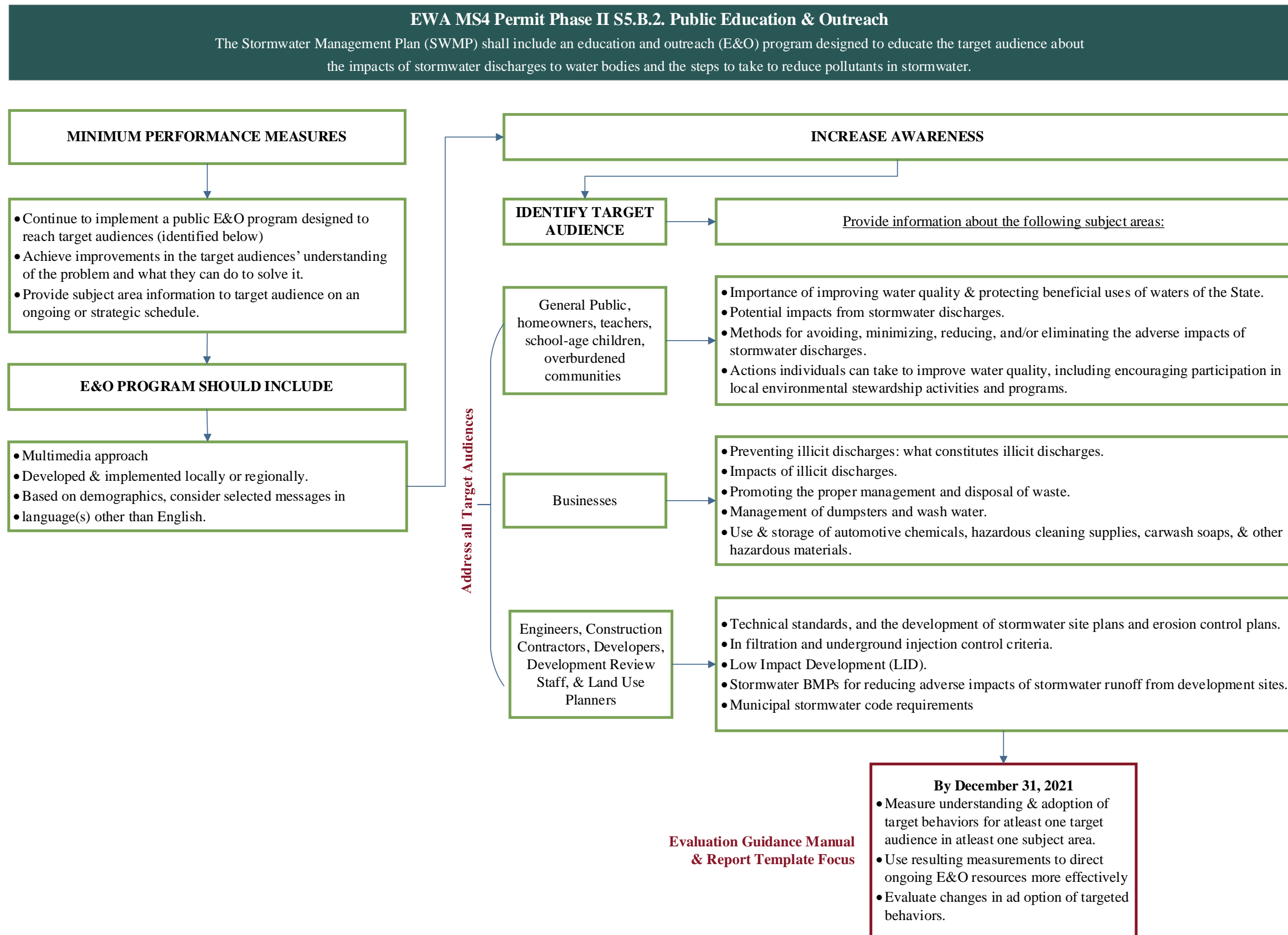


Figure 1-2 Illustration of EWA MS4 Permit E&O Requirements

2.0 Sample Size Selection

2.1 Chapter Overview

Careful consideration and selection of an appropriate sample size is an important early step in the evaluation planning process. This is because sample size can influence the study design, including the choice of evaluation instruments (Chapter 3) and analysis methods (Chapter 5). In addition, it is important to determine how much data (sample size) is needed from the target audience. The purpose of this chapter is to provide guidance for selecting and justifying a minimum sample size. Specific items included in this chapter are as follows:

- The difference between a target audience and target population (Section 2.2)
- Factors that can influence the sample size selection process (Section 2.3)
- Common strategies for selecting a sample size (Section 2.4)
- How differences between the target sample size and actual samples collected can influence a study approach (Section 2.5); specifically, instrument selection (Chapter 3) and data analysis methods (Chapter 5)
- Discussion about random sample collection (2.6)

2.2 Target Audience vs Target Population

For most evaluations, it is not feasible to collect data from the entire **target audience** (also known as priority audience)¹. Instead, a subgroup of the target audience is studied that is ideally representative (has similar characteristics) of the target audience. This subgroup is commonly referred to as the “target population” (Takona, 2002). **Example 1** illustrates the difference between the target population and target audience. For this example, behavior change materials would be distributed to the entire target audience, but evaluation data would be collected only from the target population.

Example 1: Several jurisdictions jointly developed a behavior change campaign that targets all Washington State drivers. As part of their MS4 Permit requirements, the jurisdictions have decided to conduct a study to evaluate the campaign. They have decided the sample size for the study will be drivers

¹ Many social marketing and CBSM professionals have moved away from the term “target audience” because it is perceived negatively by populations who have been “targeted” in an adverse way in the past. Instead, many now use the term “priority audience”. Throughout this manual we continue to use the term target audience because a) it is the term currently used in Permits, b) it is the term that other documents and resources developed for this project use, and c) it is the term readers will encounter in most of the existing printed and online resources provided for choosing, designing, and evaluating behavior change campaigns.

from a large city within the state. In this example, drivers from the city would be the target population, while State of Washington drivers would be the target audience.

Audience segmentation is a common approach in social marketing that is conducted to break down the target audience into smaller subgroups based on common interests or characteristics. The target population may end up being the same subgroup as the segmented audience subgroup; however, there are differences between these subgroups. The target population is the size (number of participants needed) of the sample needed to be representative of the target audience, and the size is identified to determine how much data should be collected, whereas audience segmentation is done for the purpose of developing a more tailored campaign. Common ways target audiences are segmented include age, gender, geography, income, habits, etc. (Wilbur, 2006). **Example 2** provides an example of evaluating a program that used audience segmentation. Section 1.4 contains resources with more information about audience segmentation. If it is not feasible to collect data for the entire segmented audience, follow the procedures described in Section 2.4 to select a sample size, and the subgroup included in the evaluation would be considered the target population.

Example 2: A jurisdiction developed a behavior change campaign that would target restaurants and focuses on proper BMPs for disposing of restaurant fats, oils, and grease (F.O.G.). Since the characteristics of restaurants vary substantially (e.g., food trucks, fine dining, fast food) and may influence how the restaurant disposes of F.O.G., the jurisdiction decided to segment the audience to 20 fast-food restaurants (of the total 100 restaurants within city limits). Materials were provided to only fast-food restaurants and evaluation data was only collected from this subgroup.

2.3 Factors Influencing Sample Size

This section outlines how factors can influence sample size. These factors are important to consider, as they directly relate to different types of analysis methods as well as the accuracy of the study results (Israel, 1992a). For example, certain types of statistical analysis require a minimum number of samples or an equal number of participants within the control group and the target population of an evaluation (McKenzie-Mohr, 2011). Additional discussion regarding how sample size can influence instrument selection and the type of analysis methods selected for a study is found in Chapters 3 and 5, respectively.

2.3.1 Level of Precision

Level of precision (e) defines a range in which the true value of the target population is estimated. The range is typically stated as a percentage, such as $e = \pm 5$ percent (Olejnik, 2016; Israel, 1992b; Israel, 1992a). For example, if the results indicate that 75% of the target population adopted a new behavior, and a $\pm 5\%$ level of precision was used to select the sample size, then we can assume that 70–80% of the target population have adopted the behavior.

2.3.2 Confidence Level and Interval

Confidence level helps to quantify whether a result is likely due to chance or a factor of interest. In the context of this document, a factor of interest would be a variable such as the campaign strategy that is or is not influencing behavior change. A confidence interval is selected by the researcher, which is used to determine when results are statistically significant (reference Chapter 5, Hypothesis Testing, for additional information about statistical significance). A typical confidence interval (α) is $\alpha = 0.05$, meaning there is a 95% confidence level that the result is real instead of being due to chance. Conversely, there is a 5% chance of concluding that a relationship exists between a variable studied, even though no

relationship exists in the target population. For sample size selection, larger sample sizes are typically associated with higher confidence levels (Olejnik, 2016; Israel, 1992b; Israel, 1992a).

2.3.3 Degree of Variability

The degree of variability refers to differences in the characteristics of the target population that may influence the study results. If the target population is more heterogeneous (has different characteristics), a larger sample size is required to achieve a greater level of precision. This heterogeneity can be captured in the “standard deviation” of a variable around its average – larger standard deviations indicate more heterogeneity and larger sample sizes needed (see Section 5.2.4 for more details). The more homogenous (similar characteristics) the target population, the smaller the sample size is needed (Olejnik, 2016; Israel, 1992b; Israel, 1992a). **For example**, suppose the target audience for a behavior change campaign is segmented to only fast-food restaurant managers. Because this business type has more similar characteristics, fast-food managers would be considered a more homogenous population.

2.4 Strategies for Selecting Sample Size

There are many different strategies for selecting a sample size (that is, selecting the size of the target population). Five simple methods are described in this section, and **Table 2-1** provides an overview of these strategies along with the recommended applications for when to use these methods. More details about each method, along with examples for applying the method, are included in the following subsections.

Table 2-1 Overview of Strategies for Selecting Sample Size

Sample Size Selection Method	Method	Recommended Applications
Census for Small Populations	The entire target audience is used as the sample size	Studies with small target populations (less than 200)
Sample Size of Similar Study	Sample size is selected based on sample sizes from similar studies	Any size study
Published Tables	Published tables are used to select sample size	Studies that have identified a target level of precision, confidence level, and variability
Formulas	Equations are used to calculate sample size	Studies that have identified a target level of precision, confidence level, and variability
Website Calculators	An online calculator is used to calculate sample size	Studies that have identified a target confidence level and interval

2.4.1 Census for Small Populations

This approach proposes to collect data on the entire target audience and is best suited for evaluations where the target audience is small (less than 200). For this strategy, the target audience would also be the target population. The benefit of this method is that it is simple and provides data on the entire target audience, which would eliminate consideration for many of the factors described in Section 2.3. The disadvantage to this method is that it may be cost prohibitive for large target audiences (Israel, 1992a; Israel, 1992b). In **Example 2**, if data was collected from all 20 of the fast-food restaurants, that would be an example of using the census method to select sample size.

2.4.2 Sample Size of Similar Study

This method proposes to use the same sample size used in a similar study. The advantage of this method is that it is easy to determine a sample size. The disadvantage is that if errors were made selecting the sample size in the similar study, there is a risk of repeating the same errors in your own study. With this method, it is important to review the procedures the researchers used to select their sample size before applying the same size on a study, to make sure there was justifiable reasoning for the sample size (Israel, 1992a; Israel, 1992b). **For example**, conduct a review of literature of evaluations that are similar to your own study and compare the size of the target audience to the sample size (target population) the researchers selected. (The waterbehaviorchange.org website described in Section 1.3.2 contains many articles focused on behavior change evaluations that could be reviewed.) Then calculate the average sample size from the articles reviewed to identify a “typical” sample size for your study.

2.4.3 Published Tables

Published tables can be used to select a sample size, such as the one shown in **Table 2-2**. Before determining whether this method is appropriate for your project, it is important to consider the Section 2.3 factors influencing sample size (Israel, 1992a; Israel, 1992b). The advantage to this method is that it is easy to use. The disadvantage is that the factors influencing sample size may not be known by the researcher, such as level of precision and confidence level. If these variables are not known, common values used in similar evaluations could be used. **For example**, the target audience is 1,000 people for a behavior change campaign focused on owners of private stormwater facilities. Assuming a 95% confidence level and $\pm 5\%$ level of precision (common values), the sample size (target population) is 286. This was determined by finding the target audience size in the first column and the level of precision in the third column ($\pm 5\%$).

Table 2-2 Sample Size Where Confidence Level is 95%

Size of Target Audience	Target Population Sample Size (n) for Level of Precision (e) of:			
	$\pm 3\%$	$\pm 5\%$	$\pm 7\%$	$\pm 10\%$
100	a	81	67	51
125	a	96	78	56
150	a	110	86	61
175	a	122	94	64
200	a	134	101	67
225	a	144	107	70
250	a	154	112	72
275	a	163	117	74
300	a	172	121	76
325	a	180	125	77
350	a	187	129	78
375	a	194	132	80
400	a	201	135	81
425	a	207	138	82
450	a	212	140	82
500	a	222	145	83
600	a	240	152	86
700	a	255	158	88
800	a	267	163	89

Size of Target Audience	Target Population Sample Size (n) for Level of Precision (e) of:			
	±3%	±5%	±7%	±10%
900	a	277	166	90
1,000	a	286	169	91
2,000	714	333	185	95
3,000	811	353	191	97
4,000	870	364	194	98
5,000	909	370	196	98
6,000	938	375	197	98
7,000	959	378	198	99
8,000	976	381	199	99
9,000	989	383	200	99
10,000	1,000	385	200	99
15,000	1,034	390	201	99
20,000	1,053	392	204	100
25,000	1,064	394	204	100
50,000	1,087	397	204	100
100,000	1,099	398	204	100
>100,000	1,111	400	204	100

Table reproduced from the following citation (Israel, 1992a).

- a. The entire population should be sampled.

2.4.4 Formulas

There are many formulas that can be used to calculate sample size. Equation 1 was included in this Manual because it is simple. The sample size (target population) is calculated based on the target audience size, selected level of precision, and an assumed confidence level of 95%. The advantage of this method is that it can be used to calculate the sample size (target population) for different levels of precision or target audiences' sizes. The disadvantage is that the level of precision may not be known. As described in Section 2.4.3, the level of precision may be assumed. **For example**, the target audience is 35,000 and a ±6% level of precision was selected by the researcher. Using **Error! Reference source not found.**, $N=35,000$ and $e=0.06$, the target sample size (n) is 276. *Note: Equation 1 was also used to calculate the values in Table 2-2.*

$$n = \frac{N}{1+Ne^2}$$

Equation 1

Where:

- n = Sample size (target population size)
 N = Target audience size
 e = Selected level of precision

2.4.5 Website Calculators

There are many websites that have calculators that use formulas to calculate the sample size. A sample size calculator recommended by a TAC member is from the Creative Research Systems website (<https://www.surveysystem.com/sscalc.htm>). The advantage to this method is that it is easy to use. The

disadvantage is that the Section 2.3 factors influencing sample size may not be known by the researcher, such as confidence level and interval. If these variables are not known, common values used in similar evaluations could be used. **For example**, the target audience is 100,000 people and you want to determine the number of people to survey (target population) to collect a representative sample of the target audience. Using the sample size calculator on the Creative Research Systems website and assuming a 95% confidence level and 0.05 confidence interval (common values used on evaluations), the sample size (target population) is 383. **Figure 2-1** provides a screen shot of the calculator and the results. *Note: The reference to the target population in Figure 2-1 is referred to as the target audience in this Manual.*

Sample Size Calculator

This Sample Size Calculator is presented as a public service of Creative Research Systems [survey software](#). You can use it to determine how many people you need to interview in order to get results that reflect the target population as precisely as needed. You can also find the level of precision you have in an existing sample.

Before using the sample size calculator, there are two terms that you need to know. These are: **confidence interval** and **confidence level**. If you are not familiar with these terms, [click here](#). To learn more about the factors that affect the size of confidence intervals, [click here](#).

Enter your choices in a calculator below to find the sample size you need or the confidence interval you have. Leave the Population box blank, if the population is very large or unknown.

Determine Sample Size

Confidence Level: 95% 99%

Confidence Interval:

Population:

Sample size needed:

Figure 2-1 Example of a Website Sample Size Calculator

2.5 Targeted Sample Size vs Actual Sample Size

The sample size selected for a study reflects the number of responses or data collected from the target population. This targeted sample size is typically not the same as the number of surveys mailed or interviews planned. **For example**, Giacalone, et al., implemented a telephone survey to collect information on public perception, knowledge, behaviors, and willingness to get involved in improved stormwater management. Surveys were sent to 1.5 million people located in five different cities and, of those, only 13.4% were willing to complete the survey (Giacalone, Mobley, Sawyer, Witte, & Eidson, 2010). As such, the amount of data planned to be collected (number of surveys mailed, interviews planned, etc.) may need to be increased to compensate for nonresponses; even then, the targeted sample size may not be achieved. Because of this, we recommend starting off an evaluation with the desired target population sample size and then, after the study is complete, comparing the actual sample size to the targeted sample size to estimate the representativeness of the results. **For example**, if your target audience is 100,000 people and a 3% level of precision was selected, based on **Table 2-2**, the target sample size would be 1,099. If after all data has been collected from 100 people, then the final report could describe the results based on a level of precision of $\pm 10\%$ for this sample size. This was determined by locating the target audience size in the first column (100,000) and the target population size in the last column (level of precision $\pm 10\%$).

2.6 Random Sample Collection

With the exception of collecting data on every member of the target audience (Sec 2.4.1), statisticians advise that where possible you randomly choose which person, business or location to collect data from. By choosing randomly, you maximize the chances that what you learn from the random sample will be representative of the larger group. Without randomization, your evaluation is at risk of “selection bias” and the findings can be misleading.

For example, suppose a pet waste program was evaluated by having a staff member stand near a city-provided station that provided poop bags and a garbage bin. By interviewing only the people who used the station, the researchers inadvertently selected for precisely the people who had successfully made the behavior change. In this case, they would overestimate how successful their campaign had been. By randomly selecting dog walkers along various paths or sidewalks, the researchers would gain a better understanding of the effectiveness of the campaign. Although this example has very obvious selection bias, it can affect evaluations in ways that are often hard to diagnose or anticipate.

Randomization can seem daunting and may not be feasible in some cases. Two simple strategies often suffice. First, if a complete list of the target audience is already available, then one can use a random number generator in Excel (the function RAND) to draw a random sample. **For example**, if the target audience is 100 people and the target population size was determined 51 (Table 2-2), the RAND function could be used to generate a random number for each of the 100, sort the rows by this number, and interview the first 51. Second, where a full list of the target population is unavailable, one can use simple rules like interviewing every fourth dog walker that passes a particular spot, or every seventh house on a street. One can also use an old-fashioned coin toss to determine whether data is collected or not.

3.0 Evaluation Instruments

3.1 Chapter Overview

The purpose of this chapter is to provide an overview of the different types of evaluation instruments (referred to as instruments from this point forward), considerations for selecting and designing instruments, and suggestions for validating instruments. In the context of a behavior change campaign evaluation, an instrument is a measurement device (a survey, interview questions, an observation log, etc.) used to collect data that can be used to assess changes in the target population's understanding and adoption of a targeted behavior. The instruments covered in this chapter include surveys, interviews, focus groups, observations, photos, and drawings. The measurement occurs by comparing data collected using an instrument both before and after implementing a campaign, or by comparing data collected from a control group not exposed to the campaign to data collected from the target population after they were exposed to the campaign.

Instruments fall into two broad categories: researcher-administered and participant-completed. They are distinguished by those the researcher administers versus those that are completed by the participants (target population).

- An example of a researcher-administered instrument would be an observation log completed while observing the target population's behavior.
- An example of a participant-completed instrument is a survey questionnaire that the target population completes following specific instructions (Biddix, 2022).

Instruments may also be classified by the type of data they collect qualitative (i.e., open-ended questions) or quantitative (i.e., multiple-choice surveys). Additional discussion about data types is included in Chapter 4.

3.2 Instrument Types and Selection Considerations

This section provides an overview of the different types of instruments along with considerations for selecting an instrument. Instruments should be selected and developed prior to implementing a campaign to ensure the right data is collected during the campaign evaluation. It is important to note that seldom is only one instrument appropriate for a study. Further, there are typically trade-offs to selecting one instrument over another (Takona, 2002). **For example**, collecting observational data allows the researcher to document participants' actual behavior, whereas a survey is completed by participants who would self-report their behavior.

Research indicates that observational data is typically more accurate than surveys because it documents actual behavior (Kimberlin & Winterstein, 2008; Grove & Fisk, 1992). Because behavior is self-reported by participants in surveys, surveys are subject to social desirability bias where the participant answers the question in a manner, they believe is favorable by others. This can result in over-reporting good behavior or under-reporting undesirable behavior (Grimm, 2010). The trade-off is that, while observational data is typically more accurate than surveys, it is also typically more expensive to collect and analyze compared to survey data, particularly if the survey is administered in an electronic format. That being said, it is not always feasible or appropriate to collect observational data, and every evaluation regardless of the instrument provides value.

An overview of each instrument covered in this chapter is described in **Table 3-1**, followed by a more detailed description of the instrument. Considerations for designing instruments are described in Section 3.3.

Table 3-1 Overview of Instruments and Selection Consideration

Instrument	Description	Considerations for Selection
Surveys	A questionnaire that is typically sent to the target population (participant) who completes the questionnaire following specific instructions. Questions may be closed- or open-ended.	<ul style="list-style-type: none"> • Any sample size • Low response rate • Less expensive compared to other instruments for large sample sizes
Interviews	An interactive form of data collection that involves an interviewer reading prepared questions to participants and recording their answers. Questions are typically open-ended.	<ul style="list-style-type: none"> • More suitable for a smaller sample size • Higher response rate • Time-consuming to collect and analyze data
Focus Groups	A small gathering of people who discuss a specific subject under the guidance of a moderator to better understand the target population's perceptions and collect their feedback. Questions are typically open-ended.	<ul style="list-style-type: none"> • More suitable for a smaller sample size • Inexpensive • Typically requires more than one instrument to conduct a complete evaluation
Observations	Data is collected by the researcher observing and documenting the target population's actual behavior. A predeveloped checklist is typically used to collect data.	<ul style="list-style-type: none"> • More accurate data because actual behavior is documented • Larger sample sizes can be time-consuming and expensive • May not reveal as much about understanding a targeted behavior
Photos	A camera is used to record any changes in behavior before and after the behavior change campaign takes place.	<ul style="list-style-type: none"> • Any sample size • Best for documenting inanimate objects such as dumpsters • May increase data management to track where photos were taken and when
Drawings	Drawings developed by the target populations before and after participating in an educational program are used to assess changes in understanding and perceptions.	<ul style="list-style-type: none"> • Any sample size but typically better suited for smaller sample sizes • Best suited when the target population is school age (K–6) • Time-consuming to analyze data

3.2.1 Surveys

A survey is a process of collecting data that typically involves sending a combination of questions (questionnaire) to the participants who provide responses to the questions. The questions may be closed- (e.g., multiple-choice, yes/no) or open-ended (Kumar, 2011). The goal of a survey is to learn more about the target population: specifically, about their understanding and adoption of a targeted behavior. Surveys can be administered to any sample size and are typically more cost effective for larger audiences compared to other instruments. Response rates for surveys tend to be lower compared to other

instruments. Common methods used to distribute surveys include mail, email, and web links posted on social media or mailed with utility bills.

Because the survey is completed by the participant, the information reported may not be completely accurate due to social desirability bias, discussed above (Grimm, 2010). Careful survey design can minimize or reduce the effects of social desirability bias, as described in Section 3.3.3.

3.2.2 Interviews

An interview is an interactive method of collecting data that typically involves an interviewer reading prepared questions to a participant and then recording the participant's response. The questions are typically open-ended but may also include closed-ended questions. Responses to interview questions typically provide more depth than survey responses because the research can ask more probing questions that provide insights to the participant's responses (Kumar, 2011; Wilbur, 2006). Interviews are more time-consuming to conduct and analyze the data. Consequently, they are better suited for a smaller sample size. In addition, interview response rates tend to be higher than survey response rates (Nehe, 2021). Some evaluations may include both surveys and interviews: the survey is used to learn about the target population and interviews are conducted on a subset of the target population to gain clarification and additional insight on their survey responses. In this case, the interview questionnaire is typically developed based on the survey responses. Interviews are normally conducted face to face, over the phone, or via online video conference.

Similar to surveys, because the interview is completed by the participant, the information reported may not be completely accurate due to social desirability bias (Grimm, 2010). Careful interview and questionnaire design can minimize or reduce the effects of social desirability bias, as described in Section 3.3.3.

3.2.3 Focus groups

A focus group is a small gathering of people in an interactive setting where they discuss a specific subject under the guidance of a moderator or the researcher. The researcher will raise specific questions or issues to stimulate discussion among the focus group participants, and the information collected by the research is used to understand the target population's perceptions and collect their feedback (Kumar, 2011). Typically, data collected from focus groups are used to develop other instruments or aspects of a campaign (Wilbur, 2006). For behavior change campaigns and evaluations, focus groups can be used to develop a better understanding of how the audience perceives a target behavior and provide an opportunity for the audience to discuss in detail their regular behaviors and barriers that prevent them from changing their behavior. Focus groups may also be used to collect feedback on behavior change campaign materials before they are implemented (McKenzie-Mohr, 2011). The sample size for a focus group is usually small, ranging from 6 to 10 people who gather in the same room or an online video conference. Like interviews, focus groups are also subject to social desirability bias. Bias can be reduced by explaining why the participants were chosen to participate in the focus group, what the researcher is wanting to understand about their perceptions or behaviors, and how that information will be used (McKenzie-Mohr, 2011).

Similar to surveys and interviews, because the responses in focus groups are provided by the participant, the information reported may not be completely accurate due to social desirability bias. In a group setting, this can be particularly challenging, especially if one person reports good behavior; the remaining participants may feel uncomfortable providing honest responses about undesirable behavior (Grimm,

2010). Careful design of a focus group outline can reduce social desirability bias, as described in Section 3.3.3.

3.2.4 Observations

Observational data is collected by the researcher observing and documenting the target population's actual behavior. Because actual behavior is being documented (as opposed to self-reporting with surveys), the data is typically more accurate (Kimberlin & Winterstein, 2008; Grove & Fisk, 1992). Observations should be naturalistic and recorded as discretely as possible to be sure authentic behavior is being observed about an individual or a group. Otherwise, if the individual is aware they are being watched, they may act differently than they would if they did not know that what they were doing was being recorded (Kumar, 2011). Prior to collecting observational data, the researcher will develop a checklist that is used to record the anticipated behavior and describe relevant observations. This instrument is well suited for documenting behavior that can be observed in a public place, but it may not reveal as much about the target population's understanding of a targeted behavior. Changes in behavior are measured by comparing the data collected before and after a behavior campaign is implemented. This method works well with any sample size; however, the larger the sample size, the more expensive and time-consuming the data collection and analysis process becomes.

An indirect way to observe and understand changes in the target audience's behavior is to measure outputs. **For example**, if the campaign focused on pet waste disposal and included adding bag dispensers at parks, the number of bags removed from the dispenser could be counted. Because of uncertainty regarding why bags were taken from the dispensers, and not actually observing whether they were used to pick up pet waste, this type of information is typically supplementary to other data that is being collected. An example would be a project with a limited budget for collecting observational data may use results from counting outputs to further support their observational data results. Or these measurements could be collected over a long term to assess whether there is an increase in the demand for bags, which may indicate an increase in understanding and changes in behavior. If this method is used, it will be important to also consider the community growth where the bags are located to determine whether the demand is just increasing with the increase in population size.

3.2.5 Photos

Photographs may also be used to collect data about behavior change. A camera is used to observe and record any changes in behavior before and after the behavior change campaign takes place. The behavior change is measured by comparing the before and after photos side by side. Photographs typically work best with inanimate objects, such as dumpsters, and provide proof of behaviors such as dumpster management. Generally, it is best to ask permission before taking photos of private property. Notifying the property owner that the picture will be taken may introduce issues with the property owners/workers attempting to change their behavior before the researcher arrives to take the photo, so it is recommended that permission be sought at the time of the site visit and right before the photo is taken. Try to take a photo in the exact spot for both the before and after photo so the same views are compared. Typically, a checklist is developed to document and compare the observations.

Example: A campaign focused on dumpster management with a target audience of automobile repair shops within the city. Photos were taken to understand the businesses' dumpster management practices twice at each business during the evaluation: once before the campaign was implemented to collect baseline data, and again after the campaign was implemented to collect follow-up data. Photos were taken of the dumpsters and oil containers, as well as the locations of the nearest storm drains. The site visits

took place at unannounced times, photos were taken from the same locations, and then the checklists were completed in the office. **Figure 3-1** provides an example of a before and after photo.



Figure 3-1 Dumpster Photo Before (left) and After (right) E&O Campaign

3.2.6 Drawings

Drawings can be used to evaluate changes in the target population’s understanding, perceptions, and adoption of behaviors. This method is best suited for younger school-age children (K-6) and can be time-consuming to prepare for as well as collect and analyze data. Typically, the target population is asked to develop a drawing related to a specific topic (baseline data). Then they participate in an educational program and afterwards they are asked to draw the same thing (follow-up data). The two drawings are then compared to assess changes as a result of participating in the educational program. A checklist of relevant items in the drawings is then developed and used to identify what items are present in the drawings before and after the educational program (Xu, Read, Sim, & McManus, 2009; Miele, 2014). Like items are then grouped and coded into themes (more discussion on coding qualitative data is in Section 4.5), and differences in the two drawings related to each theme are calculated to measure change.

3.3 Considerations for Designing Instruments

This section focuses on things to consider when designing and developing instruments. When selecting an instrument for a particular evaluation, a critical consideration is whether the instrument is reliable and valid.

- A reliable instrument will collect similar data made on the same participants if the study is repeated.
- A valid instrument will measure what it is intended to measure (Takona, 2002).

Reliable and valid instruments are developed by selecting an instrument that is appropriate for a particular study, carefully designing the instrument to answer the study questions, and validating the instrument before it is used. Sections 3.2 and 3.4 provide considerations for selecting and validating instruments.

When designing instruments, it is important to collect data that will help answer the specific questions needed to conduct an evaluation (McKenzie-Mohr, 2011). Instruments should be designed to measure both the understanding and the behaviors of the target population. Instruments can easily become lengthy and complex while under development. Keeping the contents clear and brief is important for answering the question that the evaluation is intended to answer and prevents the instrument from managing data that is not needed. It may be helpful to assess the instruments after they are developed and remove questions that provide data that is interesting to know but will not drive your decision making relative to the evaluation goals.

3.3.1 *Survey and Interview Questions*

Considerations for developing survey and interview questions:

- When developing questions, only collect the required details needed to measure changes in behavior and understanding.
- Participants should be clear on what the questions are asking. Consider shaping questions in terms that a person with no stormwater background knowledge would understand. For public surveys, using a 5th grade reading level is suggested.
- Each survey should include instructions regarding the purpose for the survey, the jurisdiction(s) involved, and instructions for completing the survey. Instructions should be listed clearly and noticeably just before the first questions is asked.
- Most surveys should be designed to have participants complete the survey in 10 minutes or less. Any more time than this and the participants' attention-span declines and they are less likely to complete the survey.
- Questions may be either closed-ended or open-ended. Some key points and differences for each type of question are as follows:
 - Closed-ended questions offer limited options for responses, such as questions that have multiple-choice or yes/no response options. Closed-ended questions also include response options on the Likert Scale (see Section 4.4 for more details). Closed-ended questions are easier to analyze and can minimize misinterpretation of participants' responses, which can be an issue with open-ended questions.
 - Open-ended questions give participants the opportunity to answer in their own words and are typically designed to elicit more information that can be provided with closed-ended questions. However, it is time-consuming to analyze open-ended responses (see Section 4.5) and it typically takes the participant more time to answer the questions compared to closed-ended questions.
- Designing good survey or interview questions involves selecting the questions needed to meet the evaluation goals and evaluating the questions to make sure they are clear and answer the questions intended. The following four questions can be used to evaluate survey questions (Fowler, Jr., 1984):
 1. Is this a question that can be asked exactly as written?
 2. Is this a question that will mean the same thing to everyone?
 3. Is this a question that people can answer?
 4. Is this a question that people will be willing to answer?

- After the survey or interview questions have been developed, follow the suggestions in Section 3.4 to validate the instrument, which will improve the quality of the instrument and the data collected using the instrument.

3.3.2 *Survey Design Resources*

A TAC member provided the following information, which includes free resources that may also be helpful for survey design.

- **NOAA Coastal Management – Introduction to Survey Design and Delivery**
 - <https://coast.noaa.gov/digitalcoast/training/survey-design.html>
 - NOAA also offers training on survey design, and the pdf from its course is available for free download.
- **Survey Monkey™ – 10 Best Practices for Creating Effective Survey**
 - <https://www.surveymonkey.com/mp/survey-guidelines/>
 - Survey Monkey™ provides a simple starting place for survey design.
- **Harvard Questionnaire Design Tip Sheet**
 - <https://psr.iq.harvard.edu/book/questionnaire-design-tip-sheet>
- **Survey Fundamentals – A Guide to Designing and Implementing Surveys**
 - <https://osteopathic-medicine.uw.edu/docs/getting-started-research/survey-fundamentals.pdf>

3.3.3 *Social Desirability*

Social desirability bias is a type of response bias where survey respondents answer questions in a manner they believe will be viewed favorably by others. It can take the form of over-reporting "good behavior" or under-reporting "bad" or undesirable behavior (Grimm, 2010). This section provides suggestions for surveys, interviews, or focus groups that can minimize or reduce the effects of social desirability bias (Ipsos, M. O. R. I., Autumn 2012).

- Carefully consider the following:
 - **How research is introduced.** Avoid priming participants to respond in a more socially acceptable manner, or let them know it is okay to admit undesirable behavior.
 - **How questions are worded can encourage respondents to answer truthfully.** This might include presenting statements other people have made during an interview and then asking the respondent to provide a response that is closest to their own views. This sends a message that there are a range of "acceptable" responses, rather than one "right" one. Include statements in the survey or interview instructions indicating that "there are no right or wrong answers," to help reduce concerns participants might have about being judged for their responses.
 - **Ask participants what they do (or would do), not just what they think.** Research has indicated there is a disparity between self-reported opinions vs self-reported actions or a willingness to change actions. Asking participants what they do or would do is typically less affected by social desirability bias.
- Use multiple types of instruments and sources to collect and cross-check data to assist with understanding and interpreting the responses. For example, if a survey is released to the public

about car wash wastewater management, consider also collecting observational data, and even trends in washing cars at commercial car washes, and using the combination of results to support the survey findings.

- Ask the same question with the response option in reverse order. For example, if a question asks, “how likely are you to adopt a behavior” and the response options are “extremely likely, likely, neutral, unlikely, extremely unlikely,” ask the same question at the end of the survey but put the response options in reverse order: “extremely unlikely, unlikely, neutral, likely extremely likely.” If the participant answers, “extremely unlikely” and “extremely likely” to the same question, this may be an indication that their responses are not valid, and consideration should be given to excluding the data from their survey in the final dataset (Hopper, 2013). This same approach can also be done with reverse-wording the question but leaving the response options in the same order for both questions. For more information on this topic, consult the following resource: <https://www.formpl.us/blog/how-to-get-the-truth-on-surveys-why-respondents-lie>

3.3.4 Target Audience Research

Both social marketing and community-based social marketing (CBSM) recommend conducting target audience research to better understand existing behaviors and barriers that inhibit individuals from engaging in preferred behaviors. This information is also important for developing study instruments that will be used as part of the evaluation. Two ways to collect information about the target audience are noted below. Additional social marketing and CBSM resources are located in Section 1.4.

- Conduct a literature review to identify the target audience and/or determine what is known about the habits and demographics of a target audience relevant to a behavior change campaign. The CBSM website, <https://cbsm.com/>, is an option for reviewing environmentally-related case studies (Mckenzie-Mohr and Associates, 2005-2022). Web search engines or Google Scholar are also good resources for looking up published reports or more scholarly articles. Many of these options are free and readily available.
- Conduct surveys or focus groups to better understand the target audiences’ current behaviors and barriers to behavior change. This information is commonly used to develop behavior change campaigns, and the information collected can also be used to develop survey or interview questions.

3.3.5 Use an Existing Instrument

Instruments developed from other studies can be reused, which may eliminate the need to develop a new instrument. However, each study is unique, and an existing instrument will likely need to be adapted to the new study. Reference the waterbehaviorchange.org website for articles that may contain examples of instruments that have already been developed.

3.3.6 Multiple Instruments

It can be beneficial to use more than one instrument for data collection. One reason to do so is that it may be necessary to answer all the evaluation questions. For example, observational data may be used to understand whether changes in behavior occurred, but additional data may need to be collected (e.g., using a survey or interview questions) to determine whether the target audience’s understanding of the behavior has changed or why they changed their behavior. When considering using more than one instrument for data collection, assign each instrument a purpose for what data will be collected, and identify how that data will answer the evaluation questions to confirm the additional data is needed. An

additional benefit of using more than one instrument for data collection is it can improve the validity of the results, particularly if the results are similar from each instrument. The disadvantage of using more than one instrument is it may cost more time, money, and resources than using only one instrument.

3.3.7 Checklists for Collecting Observational Data

The purpose of the checklist is to reduce the time needed to record the observations and analyze the data. Checklists are typically developed by conducting research (literature reviews, focus groups, etc.) to understand the anticipated behavior and barriers of the target population and then pilot testing (see Section 3.4) to validate the checklist before it is used in a study. In addition, developing a checklist into a standardized form before starting to collect data will minimize errors in the process of collecting, recording, and analyzing errors (Radhakirishna, 2012). **Figure 3-2** provides an example of a checklist.

<p>Observational Data Form #_____</p> <p>Name of the Inspector: _____</p> <p>Inspection Date and Time: _____</p> <p>Jurisdiction the inspector works for: _____</p> <p>Location of the observed evidence of residential car wash (neighborhood, street name, etc.):</p> <p>_____</p> <p>_____</p> <p>Were any of the following car wash practices observed?</p> <p><input type="checkbox"/> Vehicle washed on pervious surface (grass, dirt, or gravel) and wash water not entering street</p> <p><input type="checkbox"/> Vehicle washed on impervious surface</p> <p><input type="checkbox"/> Washing of the engine, undercarriage, mounted equipment, or tires</p> <p><input type="checkbox"/> Objects used to divert car wash wastewater away from storm drain to permeable surface</p> <p><input type="checkbox"/> Other, please specify</p> <p>Is there evidence of car wash wastewater entering the storm drain?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p>Please provide a brief description of what you observed (for example: no barriers used to prevent wash water from entering storm drain, barriers used to prevent wash water from entering storm drain but is not effective, etc.)</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p>
--

Figure 3-2 Checklist from a Car Wash Wastewater Management Evaluation

3.4 Validating Instruments

After an instrument has been developed, the next step is to validate the instrument. Validation is a process used to verify that the instrument measures what it was intended to measure and produces stable results (Guba, 1981). Three common methods for validating instruments include:

- Peer Debriefing – Distribute the instrument to a group of your peers and have each of them review/use the instrument. Then have the group meet to debrief on their assessment of the instrument. This will include discussion regarding whether (a) there is more than one way to interpret a question or instructions, (b) the terminology seems clear for a diverse audience, or the terms should be revised because they may not be understood by the general public, or (c) whether the questions or instructions should be revised to improve clarity. The instruments are then revised until the group mutually agrees.
- Field testing instruments before broad implementation – This may include using focus groups or pilot testing the instruments before they are implemented for a study. (Focus groups are described in more detail in Section 3.2.3.) Pilot testing would include implementing the instruments with small subgroups of the target population or separate control groups. Data collected from pilot testing is then used to update the instruments before they are used as part of the evaluation. Data collected from pilot testing would not be included as part of data collected from the actual evaluation.
- Use established instruments from similar studies that have already been validated. Refer to Section 3.3.5 for more information on this topic.

4.0 Data Types

4.1 Chapter Overview

Once data has been collected using the instruments listed in Chapter 3, data will need to be prepared for analysis. Depending on the data collection instrument used, either qualitative or quantitative data will be produced. Qualitative values consist of descriptions, whereas quantitative data can typically be measured or counted and has numerical values. This chapter focuses on data management, providing an overview of qualitative and quantitative data types, and guidance for coding qualitative data as well as converting qualitative data to quantitative data.

4.2 Data Management

Data management is the organization, storage, and preservation (or archiving) of data collected during the evaluation. It is the everyday management of the data during the data collection and analysis phases of a project and is an important step to reduce the potential for errors (Radhakirishna, 2012). Proper data management also ensures that, should an unanticipated change in key team members take place, the project can be more easily continued by the new team member. It is generally recommended that a plan be developed prior to data collection that outlines how data will be managed.

The remainder of this section focuses on data organization. To accurately measure if there was a change in behavior, data will need to be collected both before (baseline) and after (post or follow-up data) a campaign is implemented. Baseline data provides information about the target population before they are exposed to the campaign, and follow-up data provides information about the target population after they are exposed to a campaign. Baseline and follow-up data are then compared to evaluate changes in the target population's understanding and adoption of a behavior. Both forms of data should be collected in the exact same manner so that, ideally, the only changing variable in the study was that the campaign took place. If it is not possible to collect baseline data, data collected from a control group may be used instead. Control groups are not exposed to the campaign materials and should have characteristics similar to the target population. The same instruments used on the target population should be used to collect data from the control group. In addition, Sugiarto and Cook (2022), recommend collecting a combination of baseline and follow-up data as well as data from a control group because it is considered the gold standard in high-quality evaluations: it eliminates many of the factors which could lead you to mistaken conclusions. However, collecting additional data from control groups uses limited resources and may not be feasible in many cases (Sugiarto & Cook, 2022).

Suggestions for organizing data are as follows and illustrated in **Figure 4-1**:

- Excel[®] or a similar program is recommended for organizing data.
- Organize data first by the evaluation instrument used to collect the data and then separate data into baseline, follow-up, or controlled responses. Depending on the amount of data collected, it may be easiest to separate each set of data (e.g., baseline or controlled and follow-up) into different worksheets.
- Put each question or item from the evaluation instrument into the column header and arrange responses from each participant into separate rows in the same column. This will allow for an easier comparison of data.

- If data is coded into themes (Section 4.5) or responses are converted to numerical values (Section 4.6), then themes or values can easily be added to the adjacent column in the same row.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1	Participant #	Location Where Data was collected	Data Collector Initials	Question 1	Questions 2	Question 3
2	1	South Park	ABC	Participant 1 response to Question 1	Participant 1 response to Question 2	Participant 1 response to Question 3
3	2	South Park	ABC	Participant 2 response to Question 1	Participant 2 response to Question 2	Participant 2 response to Question 3
4	3	South Park	ABC	Participant 3 response to Question 1	Participant 3 response to Question 2	Participant 3 response to Question 3
5	4	North Park	DEF	Participant 4 response to Question 1	Participant 4 response to Question 2	Participant 4 response to Question 3
6	5	North Park	DEF	Participant 5 response to Question 1	Participant 5 response to Question 2	Participant 5 response to Question 3
7						

Figure 4-1 Example of Survey Data Organization

4.3 Qualitative

Qualitative data is descriptive data (non-numerical) that can be placed into categories (Creswell, 2013). Qualitative data generally refers to text, such as open-ended responses to survey, interview, or focus group questions, but also includes data collected from observations, photos, and pictures. Quantitative and qualitative data provide different information and are often used together to develop a better understanding of the target population (Austalian Bureau of Statistics, n.d.). Collecting and analyzing qualitative data can provide insights into quantitative results. **For Example:**

- **Quantitative Data:** Observational data collected of people walking their dogs found that 40% do not pick up their dogs' poop.
- **Qualitative Data:** During interviews with dog owners, the top reason why they do not pick up poop is because they forgot to bring a bag with them.

Nominal and *ordinal* data are two forms of qualitative data. Nominal data groups variables into categories that are purely descriptive and do not have any numerical value. Some examples of nominal data include sex, religion, or race. Nominal data may be collected through asking questions such as open-ended questions or answering questions that have a given list of multiple-choice or yes/no response options. Observations recorded from pictures, site visits, or drawings are also considered nominal data. These observations should be recorded through text to express the description of the resulting behavior, whereas ordinal data groups variables into ordered categories, which has a natural order or rank based on some hierarchal scale such as high to low (Kumar, 2011). An example of ordinal data is a survey question that asks how much a person agrees with a statement and the response options include statements such as “Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree.”

Qualitative data can be collected using several different evaluation instruments, as described in Chapter 3, including questionnaires, interviews, observations, pictures, or drawings. After collecting qualitative data,

that data is commonly coded (Section 4.5) and/or converted to quantitative data (Section 4.6) to make data analysis simpler. Additional discussion about data analysis is included in Chapter 5.

4.4 Quantitative

Quantitative data has a numerical value that expresses a certain quantity, amount, or range. **For example**, if a survey question asks how often a person washes their car each year, the response would be a numerical value and is considered quantitative data. Numerical data can be represented in many ways, including percentages, proportions, or rates of change. Interval and ratio data are forms of quantitative data that represent positions along continuous number lines rather than categories like qualitative data. Quantitative data is also amenable to statistical analysis (Kumar, 2011).

Intervals represent values that have a defined numerical scale where the order of the variables is known as well as the difference between the variables; however, the zero point is arbitrary. Examples of interval data include credit scores and SAT scores. In both cases, it is not possible to get a zero score. Likert Scales are another example of interval data that is often used to give quantitative value to qualitative data. A Likert Scale is similar to ordinal data in which a survey question asks how much a person agrees with a statement and the response options include statements such as “Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree.” The difference is that after data is collected, the response options are converted to a numerical value for data analysis (Kumar, 2011). For example, “Strongly Agree” responses are converted to a 5, “Agree” responses are converted to a 4, and so on. With a Likert Scale, zero has no real meaning.

Ratio data is like interval data in that the numerical distance between points is consistent and can be measured. The difference is that the zero point reflects an absolute zero, unlike interval data in which zero is arbitrary (Kumar, 2011). Examples of ratio data include weight, height, length, area, length of time, or duration. For each of these, zero is possible (e.g., zero duration or weight).

Quantitative data can be used to conduct statistical analyses, including calculating the average response and hypothesis testing, as described in Chapter 5. Analyses such as averages or percent change use quantified values to calculate behavior change and change in understanding. The number of data points, whether that be a quantitative response or the number of times an observation was recorded, can be analyzed as quantitative data. Qualitative data can also be converted to quantitative to simplify data analysis, as described in Section 4.6.

4.5 Coding Qualitative Data

The possibilities of open-ended responses and observations are limitless, with some responses similar but not exactly the same. Similar responses can be coded and grouped together into themes (Gibbs, 2008). Coding is the process of labeling and organizing qualitative data to identify different themes that make it easier to interpret the evaluation findings. The process begins with assigning labels to words or phrases from the target population’s responses. These labels represent the important themes in the response, and labeling data makes it easier to group similar responses (Insights, n.d.). **For example**, a survey question asks the target population why they have not adopted a particular behavior. **Table 4-1** provides an example of responses that were labeled (e.g., the text highlighted in yellow below) and then similar codes were grouped into themes.

Table 4-1 Example Themes Identified from Responses

Example Responses	Theme – Definition
<ul style="list-style-type: none"> It is easier to do what I do now It would take too much time 	Convenience – it is more convenient or takes less time to keep their current behavior
<ul style="list-style-type: none"> It costs more than what I do now The recommended products are expensive 	Cost – it costs less to keep the same behavior, or the recommended behavior is perceived as expensive
<ul style="list-style-type: none"> I was not aware that what I am doing has a negative impact on water quality There is no stormwater in neighborhood 	Unaware – unaware of the impact of their behavior or of stormwater

Things to consider when coding responses:

- A response code may fit into more than one theme, which is acceptable. Record all themes that code responses belong to. All codes will count as one response toward the overall total number of responses for a theme. **For example**, a survey question asks the target population why they have not adopted a particular behavior and they provide the following response: *it is easier to do what I do now, and the recommended products are too expensive*. Based on the themes shown in Table 4-1, there are two themes in this response: convenience (easier to do what I do now) and cost (recommended products are too expensive). Once the data coding is complete, then the number of responses for each theme would be counted and the response in this example would count toward both convenience and cost.
- Clearly define each theme (shown in the second column of Table 4-1) and use the definition consistently in coding all responses. These definitions will also make it easier to code if more than one person is reviewing the data.
- Themes should be the same between baseline and follow-up data review. This will make it easier to compare the differences between the data sets and determine whether there is a measurable difference.

Determining which responses (labels) belong to which theme is based on the interpretation of the person reviewing the data. To confirm the validity of the coding, consider conducting a peer review to backcheck coding methods. This may include a peer who works in the same field and is familiar with the subject of the study. After all the data has been coded by the initial data reviewer, the peer will review a portion of the data to determine whether they agree with the themes identified by the initial data review for different responses. Then the initial data reviewer will meet with their peer to discuss and compare their results until they mutually agree on the interpretation of the coding, which may include changing how some responses were coded or adding additional codes if needed.

4.6 Converting Qualitative Data to Quantitative Data

Qualitative data can be converted to quantitative data to simplify data analysis. **For example**, counting the number of responses to each multiple-choice response option for each question. This information could then be used to calculate the percentage of responses to each multiple-choice option, which makes it easier to compare baseline and follow-up data. Alternatively, multiple-choice or yes/no responses can be converted to a numerical scale similar to using the Likert Scale, as described in Section 4.4. The response options would be assigned using an even incremental range of values. If applicable, the options

would be scored considering the relativity to the desired answer, with the desired response receiving the highest value and the most undesirable the lowest value. The scoring could be assigned respectively as 5, 4, 3, 2, and 1. It could also be scored respectively as 2, 1, 0, -1, -2. The zero score works well for neutral responses; then, when the average is calculated, it is easier to compare baseline and follow-up results as more positive or negative responses. The actual values of these numbers do not matter as long as the difference between the scores is consistently equal. Converting responses to a numerical scale is particularly important if hypothesis testing is used to demonstrate whether there is a statistical difference between two data sets. Hypothesis testing is discussed in Section 5.3.2.

Example: A campaign focused on disposal of F.O.G and mop wash water with a target audience of restaurants within the city. Survey responses were collected as baseline and follow-up data, both before and after the implementation of the campaign. Researchers took the survey responses and grouped them together into themes, assigning a numeric qualitative value to each of these themes. Because each question posed different responses, the codes vary for each question. A small sample of the questions and their paired coded responses are shown in **Table 4-2**. After the data has been coded, the results can be analyzed following the methods in Section 5.2. If hypothesis testing is conducted, the numerical scales for each response code would be compared for each question to determine whether there is a statistically significant difference between the baseline and follow-up data. An example of hypothesis testing using this data is included in Section 5.3.2.

Table 4-2 Example Converting Qualitative Data to Quantitative Data

Question	Numerical Scale and Response Codes
What are the impacts of F.O.G./wash water if they reach the storm system?	1 – Does not understand the harm 2 – Knows it is bad but unsure why 3 – Fully understands the impacts of F.O.G. and wash water if they reach the storm system
How are employees educated on F.O.G./wash water disposal?	1 – Trained on proper disposal when hired 2 – Video training once when hired 3 – None/some employees are educated on this topic
Are specific employees trained to inspect and clean the grease traps/interceptors?	1 – Managers 2 – Cooks 3 – No specific employees 4 – Majority of employees 5 – Maintenance or external company

5.0 Data Analysis

5.1 Chapter Overview

After data has been organized and qualitative data has been coded and/or converted to quantitative data (Chapter 4), the next step is to analyze and compare the data. Section 5.2 provides guidance for using basic statistics to calculate and describe the central tendency and variance in data sets. Section 5.3 provides guidance for comparing the data sets to assist with determining whether there is a difference between the baseline and follow-up data. Examples for applying the different methods are also included in this chapter. Section 5.4 provides a list of software options that can be used to perform the data analysis in this section.

5.2 Descriptive Statistics

Descriptive statistics summarize information about a data set that can be broken down into measures of the central tendency or variability. *Central tendency* describes the center position of a data set, and measures of central tendency include the mean, median, and mode. *Variability* describes the spread of a data set, and measures of variability include standard deviation and range. This section provides guidance and examples for calculating central tendency and spread.

5.2.1 Mean

The mean reports the average value of a given data set. This is the most used measurement of central tendency. The mean is calculated by summing all the variables in a data set and dividing by the total number of variables in the data set, as shown in Equation 2. The mean is typically used for normally distributed data, which typically has a low number of outliers. Examples for calculating the mean are included at the end of this subsection. *If the mean is calculated using a program such as Excel™, the formula is =Average(cell1, cell2, ...).*

$$\bar{x} = \frac{\sum x_i}{n} \times 100\% \quad \text{Equation 2}$$

Where:

- \bar{x} = average or mean
- $\sum x$ = sum of the variables in the data set
- n = number of variables in the data set

EXAMPLE: A multiple-choice question was used on a survey and 100 people (n=100) responded to the question. There were four response options labeled as A, B, C, and D. The total number of responses to each option was as follows: A (10), B (25), C (15), and D (50). The average percentage of responses to each option can be calculated using Equation 2. An example calculation for option A is as follows. If the analysis is repeated for each response option, the results would be: 10% responded to option A, 25% to option B, 15% to option C, and 50% to option D.

$$\bar{x} = \frac{10}{100} = 0.1 * 100\% = 10\%$$

EXAMPLE: A survey question asked people to indicate how much they agree with a statement, and the response options included “Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree.” 100 people responded to the question and the number of responses provided for each response option is shown in **Table 5-1**. The data can be analyzed by determining the average percentage of responses to each response option using the method described in the previous example or by determining the average overall score, as shown in Table 5-1. For the average score, the response options are first converted to a Likert Scale, as described in Section 5.6. Then the number of responses to each option are multiplied by the corresponding Likert Scale value to determine a score for each response, and the total response score is summed. Finally, the average score is determined using Equation 2. The results can then be described as follows: the responses indicate an average score of 3.7, meaning that on average the target population’s response is between *agree* and *neutral* to the survey question statement. This type of analysis makes it easier to compare results between baseline and follow-up data.

Table 5-1 Mean Example Data for Likert Scale

Response Option	Number of Responses	Likert Scale	Response Score
Strongly Agree	25	5	125
Agree	40	4	160
Neutral	20	3	60
Disagree	10	2	20
Strongly Disagree	5	1	5
Total (Σx_i) =			370

$$\bar{x} = \frac{370}{100} = 3.70$$

5.2.2 Median

The median is the middle value of an ordered data set and is not affected by outliers. The median would work best for responses to open-ended questions that have been sorted into a range of numerical values (Section 4.6). The median is determined by listing all numbers in ascending order and then locating the middle number. *If the median is calculated using a program such as Excel™, the formula is =median(cell1, cell2, ...).*

EXAMPLE: A survey asked how often the target population participated in a specific behavior, such as washing their car each year. Nine people responded to the question and their responses in ascending order were: 0, 1, 2, 3, **4**, 4, 6, 6, 24. The median, or middle number, in this data set is 4. The average value could also be reported for this data set. However, the disadvantage of using the average is, if the data has large outliers (such as 24 times per year), it will strongly influence the average, making the median a better representation of the middle value.

5.2.3 Mode

Mode is the value that appears most frequently in a data set. A data set may have one mode, more than one mode, or no mode at all. Mode is most useful when describing categorical data such as qualitative data that have coded into themes, as described in Section 4.5, and the themes have been a numerical value (Section 4.6). *If the mode is calculated using a program such as Excel™, the formula is =mode(cell1, cell2, ...).*

EXAMPLE: Table 4-1 provided an example for coding responses from open-ended questions. After the data was coded into themes, it was found that the responses included 10 about convenience, 5 about cost, and 2 about being unaware. The mode, or most frequently reported response, would be convenience.

5.2.4 Standard Deviation

The standard deviation is a measure of the average distance of the individual data points from the mean. It is the most used method for describing the variability or spread of a data set. Data sets that have multiple values similar to the mean will have a lower standard deviation, whereas data sets with multiple values that are spread out (i.e., much larger or smaller than the average value) will have a larger standard deviation. Equation 3 is used to calculate standard deviation, and an example calculation is included at the end of this subsection. *If the standard deviation is calculated using a program such as Excel™, the formula is =stdev(cell1, cell2,...).*

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \quad \text{Equation 3}$$

Where:

s = standard deviation which is also denoted as σ

EXAMPLE: If the data from the median example was used to calculate the average number of times the target population washed their car each year, it would be helpful to also report the standard deviation to indicate how spread out the responses were. Table 5-2 provides an example for a sample size of nine respondents ($n=9$). For this example, the average number of times the target population reported washing their cars was 5.56 times per year, with a standard deviation of 7.21 times per year. The results are typically reported as 5.56 ± 7.21 , and these results mean that the target population provided a wide range of responses. If the standard deviation were smaller, such as ± 0.21 , that would mean the target population provided similar responses.

Table 5-2 Example Standard Deviation Calculation

Reported Frequency	$x - \bar{x}$	$(x - \bar{x})^2$
0	-5.56	30.86
1	-4.56	20.75
2	-3.56	12.64
3	-2.56	6.53
4	-1.56	2.42
4	-1.56	2.42
6	0.44	0.20
6	0.44	0.20
24	18.44	340.20
$\bar{x} = 5.56$		$\sum (x - \bar{x})^2 = 416.22$

$$s = \sqrt{\frac{416.22}{9 - 1}} = 7.21$$

5.2.5 Range

Range is the simplest technique for reporting. Range is the differences between the highest and lowest values in a data set. Extreme data points (or outliers) will increase the range. Removing outliers from a data set will decrease the range.

EXAMPLE: Using the data from **Table 5-2**, the range would be the highest reported car washing frequency (24) minus the lowest reported frequency (0), which equals 24. The results are typically reported with the average, such as: the average number times the target population reported washing their cars was 5.56 times per year and the range of responses was 24 times per year.

5.3 Evaluating Results

Once the data from each data set has been analyzed as described in Section 5.2, the results from the baseline and follow-up (or control group) will need to be compared to determine whether there is a change in the understanding and adoption of a targeted behavior. This section describes two methods for comparing results. Section 5.3.1 describes a simplified method and Section 5.3.2 describes methods for conducting hypothesis testing. The advantages and disadvantages of both methods are also described.

5.3.1 Comparing Results

The simplest method to evaluate the results is to compare the baseline results (collected before a campaign was implemented) to the follow-up results (collected after a campaign was implemented) and note the differences (changes) in the results. If it is not possible to collect baseline data, data collected from a control group may be used instead (Section 4.2). This method works best when the same instruments are used to collect baseline and follow-up data and when there is a limited amount of data to work with. The advantage of this method is that it is simple and provides a fast evaluation of the two data sets. The disadvantage of this method is that it does not consider the confidence level of the result regarding whether the result is due to chance or a factor of interest, as described in Section 5.3.2.

EXAMPLE: A campaign focused on F.O.G. and wash water management was evaluated using interview questions. One of the questions the target population was asked was, “*What are the impacts of F.O.G./wash water if they reach the storm system?*” A total of 20 fast-food restaurant managers were interviewed before the campaign was implemented (baseline data) and again after the campaign was implemented (follow-up data). **Table 5-3** provides a summary of the number of responses received to each response code along with the change or difference between the baseline and follow-up data. The results indicate that there are fewer managers that do not understand the harm (four fewer managers) or know it is bad but are unsure why (one fewer managers), and there are more managers that fully understand the impacts of F.O.G. and wash water if they reach the storm system (five more managers). These results indicate that after the campaign there was an increased understanding of the impacts of F.O.G. and wash water.

Table 5-3 Example of Simple Method for Comparing Results from Coded Responses

Numerical Scale and Response Codes	Number of Baseline Responses	Number of Follow-up Responses	Change
1 – Does not understand the harm	8	4	-4
2 – Knows it is bad but unsure why	7	6	-1
3 – Fully understands the impacts of F.O.G. and wash water if they reach the storm system	5	10	+5

EXAMPLE: The same survey was used to collect baseline and follow-up data about a target population’s willingness to change their behavior, and the responses were converted to a Likert Scale. The average response to each question is reported in **Table 5-4**, along with the change in the average response from the baseline survey to the follow-up survey. For the Likert Scale, a response of 5 indicates they are very likely to change their behavior, and a response of 1 means they are not likely to change their behavior. Based on the change in the results to all the survey questions, there is an overall increase in the target audience’s willingness to change their behavior after being exposed to the campaign.

Table 5-4 Example of Simple Method for Comparing Results for a Likert Scale

Survey Question #	Baseline Results Average	Follow-up Results Average	Change
1	4.0	4.5	+0.5
2	3.5	3.6	+0.1
3	3.75	3.5	-0.25
4	3.25	3.5	+0.25
5	3.0	3.5	+0.5

5.3.2 Hypothesis Testing

The purpose of hypothesis testing is to determine whether there is a statistically significant difference between the two data sets (i.e., baseline data and follow-up data) based on assumptions (see null and alternative hypothesis below). The advantage of using hypothesis testing is that it provides a confidence level regarding whether the result is likely due to chance or the factor of interest. In the context of this document, a factor of interest would be a variable such as the campaign strategy. Hypothesis testing is not required by the MS4 Permits to evaluate differences in data sets; however, it is nearly always a feature of high-quality evaluations. Hypothesis testing works best with large data sets with a minimum of twelve (n=12) samples in each data set needed to conduct testing.

Hypothesis testing starts with defining a null hypothesis and an alternative hypothesis, which are assumptions about the results. Both terms are described below along with relevant hypothesis examples.

- **Null Hypothesis (H_0)** – there is no significant difference between the two data sets. **For example**, a null hypothesis would mean there is no change in understanding or adoption of the target behaviors between the baseline data and the follow-up data.
- **Alternative Hypothesis (H_A)** – there is a significant difference between the two data sets. **For example**, an alternative hypothesis would mean there is a change in understanding or adoption of targeted behaviors.

There are several different methods used to test the hypothesis, and the appropriate method is based on the type of data. Two common methods used in educational research are: (1) Mann-Whitney or Wilcoxon Rank Sum Test, which is for non-normally distributed data and compares the medians between two sets of data, and (2) Paired T-Test, which is for normally distributed data and compares the averages/means and standard deviations of two data sets². Statistical software is commonly used to determine whether data is normally or non-normally distributed and to conduct hypothesis testing. For a Paired T-Test, Excel®TM can be used to conduct the analysis. Reference Section 5.4 for additional software options.

Hypothesis testing is used to confirm or reject the null hypothesis. The confidence interval is selected by the researcher and is used to describe the likelihood that the true value lies within the data set, meaning that results accurately represent the target population's response. A typical confidence interval is $\alpha = 0.05$, meaning there is a 95% confidence level that the result is real instead of being due to chance or error. Conversely, there is a 5% chance of concluding that a relationship exists, even though no relationship exists in the target population (Olejnik, 2016; Israel, 1992b; Israel, 1992a). A researcher may choose to adjust their confidence interval or level to describe the results. For example, results where $\alpha = 0.05$ or less could be considered statistically significant, and α between 0.051 and 0.10 could be considered moderately significant $\alpha > 0.10$ considered insignificant. Most researchers agree that less than a 90% confidence interval is not a robust statistical association.

EXAMPLE: A survey was implemented to determine how likely the target population is to change their behavior. The response options were “Likely, Neutral, Unlikely.” The responses were converted to the following Likert Scale: 3 – Likely, 2 – Neutral, and 1 – Unlikely. A summary of the baseline and follow-up responses is shown in **Table 5-5**. The data was then input into a statistical software program called MiniTabsTM. A 95% confidence level and a confidence interval of $\alpha = 0.05$ were selected along with the null hypothesis noted above (no difference between data sets) and input into the software. Next, the data was evaluated to determine whether the data is normally distributed using a Normality Test, and the results are shown in **Figure 5-1**. Since the p-value reported from normality testing is greater than our selected $\alpha = 0.05$, and the data points do not follow a straight line, the data is considered non-normally distributed, so the Mann-Whitney or Wilcoxon Rank Sum Test was used to conduct the hypothesis

² Paired t-tests depend on the assumption that the data is normally-distributed. In other words, that if one plotted the data, they would resemble a bell curve.

testing. The results of the analysis are shown in **Figure 5-2**. Since the reported p-value is less than 0.05, which indicates that there is a statistically significant difference between the base line data. Based on these results we would reject the null hypothesis and accept the alternative hypothesis (there is a significant difference between the two data sets). Next, the baseline and follow-up data need to be compared to assess whether the differences between the data sets indicate that the target population is more or less likely to change their behavior. This can be done by summing and comparing the response codes. Considering that the sum of values in **Table 5-5** increased from the baseline to follow-up data and that Likert Scale responses with higher scores show an increase in the willingness to change behavior, these results suggest that, as a result of the campaign being implemented, the target population is more likely to change their behavior.

Table 5-5 Summary of Survey Responses

Participant #	Baseline Response	Follow-up Response
1	1	2
2	1	2
3	2	3
4	1	2
5	2	2
6	1	2
7	3	3
8	2	2
9	1	1
10	1	2
11	3	3
12	1	1
13	2	3
14	1	2
15	1	2
16	3	3
17	1	2
18	2	2
Sum =	29	39
Average =	1.61	2.17

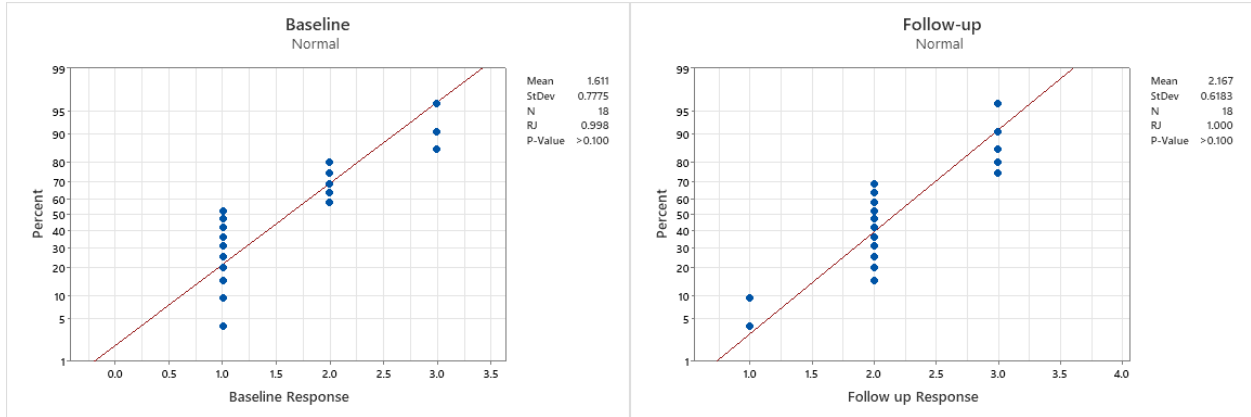


Figure 5-1 Results from Normality Testing

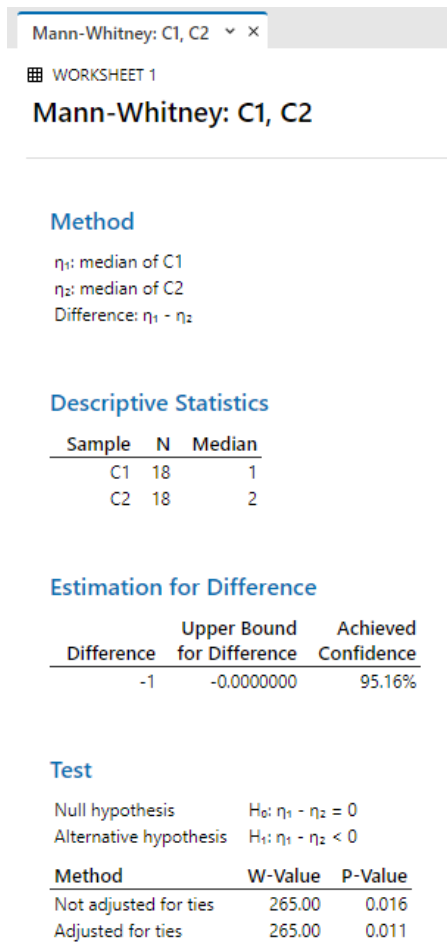


Figure 5-2 Summary of Hypothesis Testing Results

5.4 Software Options

Data analysis is typically done with some type of software. For the descriptive analysis methods, software such as Excel™ is typically sufficient to perform the calculations. For hypothesis testing, more robust software may be needed, and **Table 5-4** provides a summary of options.

Table 5-4 Summary of Statistical Analysis Software Options

Software	Types of data	Additional Description	Cost	Current Version Released	Manufactured/Developed By	Link to Website
Excel			Free version; \$139.99 without Microsoft Office 365	Microsoft Excel 2019	Microsoft	
Minitab	t tests; one and two proportions; normality test; chi-square; equivalence tests	Offers government pricing on implementation, training, & maintenance	Starts at (\$1,400/user)/yr	20.1.3 (January 2021)	Minitab Inc.	Data Analysis, Statistical & Process Improvement Tools Minitab
Statistical Package for the Social Sciences (SPSS)	t-tests, ANOVA, z-tests, confidence intervals, proportions, non-parametric tests, etc.		Starts at (\$99.00/user)/month	27.0.1.0 (November 2020)	IBM corporation	SPSS Statistics - Overview IBM
Statistical Analysis System (SAS)		Advertises to benefit a number of industries (public sector being one)	Must contact for pricing	9.4M7 (August 2020)	SAS Institute North Carolina, USA	Data Management Software SAS
R	ANOVA; t-tests; "linear and generalized linear models, nonlinear regression models, time series analysis, classical parametric and nonparametric tests, clustering and smoothing"	A programming language used for statistical computing and graphics (charts, graphs, etc.); Base for Rstudio software	Free	4.0.4 (February 2021)	Ross Ihaka & Robert Gentleman from R core team	R: The R Project for Statistical Computing (r-project.org)
Rstudio	vectors; lists; matrices; arrays; factors; data frames	Uses the R language to develop statistical programs; Provides further functionality for R	Starts at \$995/yr	Rstudio 1.4 (January 2021)	Founded by J.J. Allaire	RStudio Open source & professional software for data science teams - RStudio
Stata			Starts at (\$765/user)/yr	Stata 16.1 (February 2020)	StataCorp	Stata: Software for Statistics and Data Science
Web: G-Power	t tests; F tests; χ^2 tests; z tests; ANOVA (one-way & multi-way); chi-square tests; some exact tests	Compute data and graphics	Free	3.1.9.7 for Windows (March 2020); 3.1.9.6 for Mac (February 2020)	Heinrich-Heine-Universität Dusseldorf (HHU) – German company	Universität Düsseldorf: gpower (hhu.de)
Web: Sample Power	t tests; ANOVA; McNemar's Z test; Cox; test odds	Web-based calculator	Free		SPSS	Power and Sample Size Calculators HyLown
Web: StatPages.net		Statistical search engine	Free			

6.0 References

(n.d.).

Australian Bureau of Statistics. (n.d.). *Statistical Language - Quantitative and Qualitative Data*. Retrieved from What are quantitative and qualitative data?: <https://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Statistical+Language+-+quantitative+and+qualitative+data#:~:text=What%20are%20quantitative%20and%20qualitative,symbol%2C%20or%20a%20number%20code>.

Biddix, P. J. (2022, December 1). *Instrument, Validity, Reliability. Uncomplicated Reviews of Educational Research Methods*. Retrieved from Research Rundowns: <https://researchrundowns.com/quantitative-methods/instrument-validity-reliability/>

Creswell, J. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Thousand Oaks, California: SAGE.

Fowler, Jr., F. J. (1984). *Survey Research Methods*. Beverly Hills, CA: Sage.

Giacalone, K., Mobley, C., Sawyer, C., Witte, J., & Eidson, G. (2010). Survey Says: Implications of a Public Perception Survey on Stormwater Education Programming. *Journal of Contemporary Water Research & Education*, 92-102.

Gibbs, G. R. (2008). *Analysing Qualitative Data*. Sage Publications Inc.

Grimm, P. (2010). Social Desirability Bias. *Wiley International Encyclopedia of Marketing*.

Grove, S. J., & Fisk, R. P. (1992). Observational data collection methods for services marketing: An overview. *Journal of the Academy of Marketing Science*, 217-224.

Guba, E. G. (1981). Criteria for Assessing the Trustworthiness of Naturalistic Inquires. *ECTJ*, 75-91.

Hopper, J. (2013, January 30). *Versta Research*. Retrieved from Tips on “Reverse Wording” Survey Questions: <https://verstaresearch.com/blog/tips-on-reverse-wording-survey-questions/>

Insights. (n.d.). *Learn to code qualitative data*. Retrieved from Coding Qualitative Data: How to Code Qualitative Research: <https://getthematic.com/insights/coding-qualitative-data/>

Ipsos, M. O. R. I. (Autumn 2012). Four ways to get the truth out of respondents. . *The social research newsleTTer from ipsos mori scoTland*.

Israel, G. D. (1992a, November). Determining Sample Size. *University of Florida Cooperative Extension Service*.

Israel, G. D. (1992b, October). Sampling the Evidence of Extension Program Impact. *University of Florida Cooperative Extension Service*.

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 2276-2284.

Kumar, R. (2011). *Research Methodology: a step by step guide for beginners*. Thousand Oaks, California: Sage Publications.

Lee, N. R., & Kotler, P. (2011). *Social Marketing: Influencing Behaviors for Good*. SAGE Publications.

- Mckenzie-Mohr and Associates. (2005-2022). *Community Based Social Marketing*. Retrieved from Community Based Social Marketing: CBSM.org
- McKenzie-Mohr, D. (2011). Fostering Sustainable Behavior. *Fostering Sustainable Behavior-Community Based Social Marketing*.
- Miele, E. (2014). Using the draw-a-scientist test for inquiry and evaluation. *Journal of College Science Teaching*, 43(4), 36-40.
- Nehe, J. (2021). Survey Versus Interviews: Comparing Data Collection Tools for Exploratory Research. *The Qualitative Report*, 541-554.
- Olejnik, S. F. (2016, October 28). Planning Educational Research: Determining the Necessary Sample Size. *The Journal of Experimental Education*, 40-48.
- Radhakirishna, R. (2012). Ensuring Data Quality in Extension Research and Evaluation Studies. *Journal of Extension* 50(3), 3.
- Social Marketing Services, Inc. (2008). *Social Marketing Services, Inc.*
- Sugiarto, W., & Cook, J. (2022). *A Synthesis and Annotated Bibliography on Stormwater Behavior Change Campaigns* . Pullman, WA: Washington State University Stormwater Center.
- Takona, J. (2002). *Educational Research: Principles and Practice*. Lincoln, Nebraska: Writers Club Press.
- Wilbur, J. (2006). *Getting Your Feet Wet with Social Marketing: A Social Marketing Guide For Watershed Programs*. Retrieved from Extension University of Wisconsin-Madison: <https://fyi.extension.wisc.edu/wateroutreach/files/2015/12/GettingYourFeetWet1.pdf>
- Xu, D., Read, J., Sim, G., & McManus, B. (2009). Experience it, draw it, rate it: capture children's experiences with their drawings. *Proceedings of the 8th international conference on interaction design and children*, 226-270.