GUIDANCE FOR EVALUATING TRAFFIC SAFETY CULTURE STRATEGIES

Final Report

prepared for

CALIFORNIA	IOWA	MISSISSIPPI
CONNECTICUT	KANSAS	MONTANA
GEORGIA	KENTUCKY	NEVADA
IDAHO	LOUISIANA	TEXAS
ILLINOIS	MICHIGAN	UTAH
INDIANA	MINNESOTA	VERMONT
		WASHINGTON

in cooperation with

THE U.S. DEPARTMENT OF TRANSPORTATION
FEDERAL HIGHWAY ADMINISTRATION

January 2021

prepared by

Eric Austin, Ph.D.
Jay Otto, M.S.
Kelly Green, M.P.A.

Nicholas Ward, Ph.D.
Holly Watson
Katherine Dively, M.S.

Center for Health and Safety Culture
Montana State University
Bozeman, MT

EVALUATION

R E S E A R C H   P R O G R A M S

MDT★

Guidance for Evaluating Traffic Safety Culture Strategies

Prepared by:

Eric Austin, Ph.D., Jay Otto M.S., Kelly Green, M.P.A.,

Nicholas Ward, Ph.D., Holly Watson, and Katie Dively M.S.

Center for Health and Safety Culture

Montana State University

Bozeman, MT

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA/MT-21-001/8882-309-14	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Guidance for Evaluating Traffic Safety Culture Strategies	5. Report Date January 2021
	6. Performing Organization Code

7. Author(s) Eric Austin, Ph.D., (https://orcid.org/0000-0001-7742-9065), Jay Otto M.S. (http://orcid.org/0000-0003-0975-0249), Kelly Green M.P.A. (http://orcid.org/0000-0002-6846-0447), Holly Watson (https://orcid.org/0000-0002-9290-6381), Nicholas J. Ward, Ph.D. (http://orcid.org/0000-0002-2792-8082), and Katie Dively M.S. (http://orcid.org/0000-0001-7850-2367)	8. Performing Organization Report No.

9. Performing Organization Name and Address Center for Health and Safety Culture Montana State University P.O. Box 170548, Bozeman, MT 59717-0548	10. Work Unit No.
	11. Contract or Grant No. 08882-309-14

12. Sponsoring Agency Name and Address Research Programs Montana Department of Transportation (SPR) http://dx.doi.org/10.13039/100009209 2701 Prospect Avenue, PO Box 201001 Other sponsoring agencies: California DOT, Connecticut DOT, Idaho Transportation Department, Illinois DOT, Indiana DOT, Iowa DOT, Louisiana DOT and Development, Mississippi DOT, Nevada DOT, Texas DOT, Utah DOT, Vermont Agency of Transportation and Washington State DOT	13. Type of Report and Period Covered Final Report (January 2019 – August 2020)
	14. Sponsoring Agency Code 5401

16. Abstract

This project conducted a literature review of current practices in the evaluation of traffic safety culture strategies. This review focused on transportation safety literature but also extended to evaluating safety culture in other public health sectors. A description of the literature was provided as a report, which was also converted to a journal submission. A separate resource document was created to provide traffic safety stakeholders with guidance about the steps and conditions that are necessary for the evaluation of traffic safety culture strategies.

17. Key Words Traffic Safety, Traffic Safety Culture, Strategy Evaluation, Evidence-based Decisions, Evaluative Thinking	18. Distribution Statement No restrictions.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 100	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Disclaimer

This document is disseminated under the sponsorship of the Montana Department of Transportation (MDT) and the United States Department of Transportation (USDOT) in the interest of information exchange. The State of Montana and the United States assume no liability for the use or misuse of its contents.

The contents of this document reflect the views of the authors, who are solely responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the views or official policies of MDT or the USDOT.

The State of Montana and the United States do not endorse products of manufacturers.

This document does not constitute a standard, specification, policy or regulation.

Alternative Format Statement

MDT attempts to provide accommodations for any known disability that may interfere with a person participating in any service, program, or activity of the Department. Alternative accessible formats of this information will be provided upon request. For further information, call 406/444.7693, TTY 800/335.7592, or Montana Relay at 711.

Table of Contents

List of Figures

List of Tables

1   INTRODUCTION

1.1   Traffic Safety

In recognition of the importance of traffic safety as a public health issue, many states have adopted a vision of zero traffic fatalities and serious injuries. Although this is a bold vision, it is also a necessary vision, because no traffic fatality or serious injury is ever acceptable.

To reach this vision, innovative strategies are needed such as those that can change traffic safety culture. Indeed, "creating a positive safety culture"[1] has been named as one of the pillar strategies for the Road to Zero plan to end roadway deaths. But, as with any traffic safety strategy, evaluation is essential to ensure it is effective.

Reaching the vision of zero traffic fatalities and serious injuries requires that all our strategies be effective. To determine their effectiveness, it is important that these strategies are evaluated.

1.2   Traffic Safety Culture

Traffic safety culture can be defined as the beliefs shared by a group of road users or stakeholders that influence their behaviors that impact traffic safety.[2] This definition of culture sets up a relationship between beliefs and behaviors. Specifically, when individuals have certain beliefs, they are more likely to engage in certain behaviors. For example, if people believe that it is safe to have hands-free cell phone conversations while driving, they are more likely to engage in this risky driving behavior.

There are several basic types of beliefs including:

- Expectations about the consequences of behavior (e.g., "If I drive after using cannabis, I am more likely to cause a crash.")

- Perceptions about how common a behavior is (e.g., "I believe most people speed.")

- Perceptions about how acceptable or expected a behavior is (e.g., "My spouse expects me to use a seat belt.")

- Perceptions about an individual's ability to perform the behavior (e.g., "I am comfortable not answering my cell phone while driving.")

Traffic safety culture strategies focus on changing beliefs like these. When these beliefs change, people's behaviors are likely to change, and this change in behavior is more likely to be sustained.

---

[1] Road to Zero: A Plan to Eliminate Roadway Deaths, National Safety Council. Retrieved from: https://www.nsc.org/road-safety/get-involved/road-to-zero
[2] Traffic Safety Culture Primer. Retrieved from: https://www.mdt.mt.gov/other/webdata/external/research/docs/research_proj/tsc/TSC_PRIMER/PRIMER.pdf

In contrast, a speed bump is a physical way of changing behavior. People tend not to speed over speed bumps but will resume their speed when the speed bumps are no longer present. A speed bump does not change underlying beliefs about speeding and therefore does not result in sustained behavior change.

Traffic safety culture strategies use specific experiences designed to change beliefs. For example, workplace traffic safety training is a specific experience designed to change a worker's beliefs about specific driving practices. The training might discuss the increased risk for crashing while talking on a cell phone when driving. The training could provide information about how most employees do not drive while using a cell phone and that leadership, management, and supervisors expect drivers not to use their cell phones while driving. A workplace policy prohibiting cell phone use while driving could be reviewed. Supervisors could meet with employees and discuss how work procedures will take place without using cell phones while driving. By growing healthy beliefs among workers, the likelihood of risky driving is reduced. As fewer drivers engage in risky driving (e.g., distracted driving), fewer crashes will occur, and traffic safety will improve. This process is summarized in Figure 1.

| A TSC Strategy is Implemented (e.g., training) | Beliefs Change | Behavior Changes | Traffic Safety Improves (e.g., fewer crashes) |

Figure 1. Diagram of how a traffic safety culture strategy leads to improved traffic safety

Understanding how a traffic safety culture strategy leads to improving traffic safety is important when considering evaluating a traffic safety culture strategy. There are many potential problems that could result in a traffic safety culture strategy being ineffective.

Using the same workplace training example shared previously, imagine what would happen if only 10% of the workers were trained. Only training 10% of the workers would significantly reduce the likelihood that beliefs across the workforce would change, thus reducing the likelihood that behaviors across the workforce would change, thus reducing the likelihood that crashes would be reduced.

Suppose everyone took part in the training, but the training was poorly implemented and did not change people's beliefs. If beliefs did not change, it would be unlikely that behaviors would change, and traffic safety would not improve.

Suppose everyone took part in the training, and the training changed beliefs, but it changed the wrong beliefs – beliefs that did not matter or did not influence the behavior. Behavior would not change, and traffic safety would not improve.

Understanding how a traffic safety culture strategy leads to improving traffic safety will inform how a traffic safety culture strategy should be evaluated. Specifically, the evaluation might capture what percentage of the workers took part in the training, to what degree the training changed beliefs, how much subsequent risky driving behaviors changed, and whether crashes were reduced.

1.3   Problem Statement

To reduce the number of traffic crashes and resulting injuries and fatalities, traffic safety agencies are developing and implementing new intervention strategies aimed at changing traffic safety culture. However, efforts to systematically evaluate these new programs are not advancing as rapidly as the strategies themselves. Barriers to evaluating traffic safety culture strategies include a lack of suitable evaluation designs and/or a lack of agreement about the design elements of an evaluation appropriate for these strategies.

1.4   Project Purpose

The purpose of this project was to conduct a literature review of the current "state of the science" in evaluating traffic safety culture strategies. To begin, staff reviewed the literature for examples of evaluation methods applied to traffic safety culture strategies. Due to a lack of examples, this review was extended to other public health domains to learn from evaluations performed by other disciplines. The results of this review were then used to produce three project deliverables: journal article; summary of evaluation guidance (toolkit); and a poster summarizing the key steps for evaluation. In this final report, we describe each of these deliverables. To provide context, we begin by providing an overview of the literature review method and results.

2 LITERATURE REVIEW

2.1 Method

As described in Appendix I, a keyword search was performed using Montana State University's Library "CatSearch" meta-search engine. This search engine combines all the University's research services into one comprehensive search tool. Word and phrase combinations were chosen to capture literature with "traffic safety," "traffic safety culture," "transportation safety," and "transportation safety culture." This search was augmented with a follow-on search of specific search engines including ProQuest Central, Elsevier Science Direct Journals Complete, and Emerald A-Z Current Journals to corroborate the results from CatSearch.

All searches were structured to include only results published in English and did include studies that were conducted both in and outside the United States. While the search strategy did not use any date parameters to restrict the search, only four articles published before 2000 were found.

Perhaps because traffic safety culture is a new concept in traffic safety, this search did not yield results that included the implementation and evaluation of traffic safety culture strategies.

As a result, a new set of word and phrase combinations were selected to search outside traffic and transportation safety domains (most notably, healthcare). We used search terms including "culture change" and "safety culture" as title and subject searches with combinations of added key word searches including "intervention," "evaluation," and others. This final search process resulted in 64 articles to review.

2.2 Results

As discussed in Appendix I, the key finding from this literature review was the absence of any identified journal articles that included the evaluation of a traffic safety culture strategy, perhaps because such strategies are new within transportation safety. However, the literature review did reveal some relevant patterns in the form of evaluations used in other domains of safety culture.

There is a growing trend in a wide range of program settings of all types that emphasizes the importance of gathering the most rigorous data possible to gauge whether programs have the impact their designers intended and if they function as planned. Culture change and safety culture literature have seen similar efforts to gather scientific evidence of effectiveness (Hill, Kolanowski, Milone-Nuzzo, & Yevchak, 2011) and support evidence-based efforts (Petriwskyj, Parker, Wilson, & Gibson, 2016a, 2016b).

The review of the published literature revealed several issues that contribute to a better understanding of the evaluation strategies that have been used to assess efforts to change culture:

- There was no one definition or theory of "culture," which suggests either a lack of consensus about this concept or the presence of domain specific definitions. There were some cases where no definition was provided. Not reporting a definition of culture within

an evaluation is bad practice, because without it, it is not clear what the strategy is intended to change.

- There was no one method for measuring culture. Methods of assessment ranged from inferences based on found artifacts or observed behavior to the indirect assessment of beliefs through surveys and interviews. There can be concerns that inferences about artifacts and behaviors can be biased by the beliefs of the observer. Some forms of survey and interview can be unreliable depending on how the respondents interpret the questions.
- There is no one form of accepted evaluation. The design and analysis of evaluations varied widely depending on the nature of the strategy and the domain of application. This variability may reflect different efforts to minimize threats to the validity of the evaluation results in the (unique) context that the strategy was implemented.
- Increasing the rigor of the design and analysis of an evaluation increases its complexity, which also increases its cost and time commitment. There is a need to balance rigor with a form of evaluation that is still possible and practical to implement (Chen, 2014).
- Important factors that impact rigor include
  - the choice of proper (matched), accessible comparison groups,
  - the development of reliable and valid measures, and
  - the administration of those measures with a sufficiently large sample to support the desired analyses.

Admittedly, setting up the necessary level of rigor in the evaluation of strategies can pose challenges to program managers, especially those with limited resources, staffing, and ability.

3 JOURNAL ARTICLE

Appendix II includes a copy of the journal article based on the literature review that will be submitted for publication. Below is the title and abstract for that article:

Assessing the Impact of Culture: A Systematic Analysis of Culture Interventions and Evaluations in Different Organizational Settings

Abstract: Over the last twenty years, transportation agencies have increasingly added culture-based approaches to the existing education, engineering, and enforcement strategies being used as a means of reducing traffic related injuries and fatalities. Despite this increased interest, there have been comparatively few evaluations of interventions designed to enhance traffic safety culture. At the same time, many other organization types have adopted culture-based strategies either to improve safety or to enhance other elements of organizational performance. In aggregate, the evaluations of culture-focused interventions across a range of settings offer an untapped body of information about the models of culture being leveraged to affect change, the intervention strategies used to impact culture, the impacts of these strategies, and more. This article presents the results of a systematic analysis of evaluations of culture-focused interventions across a variety of settings and seeks to identify patterns that could be useful to both researchers and practitioners. The findings of the study suggest that there are areas of substantial consensus regarding the nature and features of culture and the potential effectiveness of culture-based programs. At the same time, the findings also suggest that more conceptual and empirical work is warranted to further refine our understanding of culture and its functions and to build deeper understanding of how to leverage culture effectively to support health and safety efforts.

4   SUMMARY GUIDANCE

Appendix III contains the toolkit created that provides guidance on evaluating traffic safety culture strategies. This guidance is based on the importance of evidence and the adoption of an evaluative-thinking mindset.

4.1   Evidence-Based Decisions

Reaching the vision of zero traffic fatalities and serious injuries will require that we apply limited resources to innovative strategies that are effective and sustainable. Evidence is needed to inform these decisions because evidence-based decision making is the foundation of responsible investment and management of any public health program.

> *Evidence is extremely important for researchers, practitioners, and policy makers charged with the task of making decisions around the funding and implementation of public health programs* (Puddy & Wilkins, 2011, p. 3).

4.2   Evaluative Thinking

> *"Evaluative thinking is a cognitive process in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves skills such as identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking and making informed decisions in preparation for action"* (Archibald, 2013).

To be effective, evidence-based decision making requires us to think as an evaluator. Evaluative thinking is a form of problem solving that extends beyond the collection of evidence to include learning lessons from that evidence, then integrating this knowledge into processes that make our strategies *more* effective in the future. In short, "evaluative thinking is learning for change" (Bennett & Jessani, 2011, p. 24).

Because reaching the Vision Zero goal will challenge us to be more effective, evaluative thinking is a key to our success. To this end, evaluative thinking is something we must share and pursue together. But, to be effective with evaluative thinking, there is a need to have a basic understanding of evaluation design and its implications for the validity of the produced evidence.

To help support this understanding, this project examined the published literature about evaluation methods used in transportation and other public health sectors. The primary goal of this review was to explore common practices in the evaluation of traffic safety culture strategies. As a result, number of key steps in the evaluation process were summarized.

4.3   Evaluation Steps

Given the variability in evaluation methods reported in the literature, it was impossible to specify one best practice for evaluating traffic safety culture strategies. However, it was possible to summarize general guidance for such evaluations based on recommended practices by the

Centers for Disease Control and Prevention (U.S. Department of Health and Human Services Centers for Disease Control and Prevention, 2011):

1. **Identify, Recruit, and Engage Stakeholders.** Stakeholders include people responsible for the strategy (e.g., funders, contractors, etc.), people affected by the strategy (e.g., general population, workplaces, etc.), and those who will use the evaluation results. Early participation by stakeholders is necessary to identify questions and concerns and support access to quality data to ensure an effective evaluation. Those affected by the strategy should be included to measure exposure to the strategy and help identify unintended consequences including potential harms.

    A key purpose of stakeholder involvement is to specify "standards" for effectiveness. What does an effective evaluation mean for this strategy in this context? How does each stakeholder define and envision success? What outcomes are important to the needs of each stakeholder? Should the evaluation bolster a sense that the strategy caused the change in outcomes or is it OK just to assess change? It is important to understand these distinct perspectives to align expectations about potential interpretations of the evaluation results.

2. **Describe the Strategy.** Before starting an evaluation, it is necessary to agree on a detailed description of the strategy, including the conditions necessary for its implementation: "a comprehensive [strategy] description clarifies all the components and intended outcomes of the [strategy], thus helping you focus your evaluation on the most central and important questions."[3] Understanding how the strategy causes a change in outcomes (and later positive impact to traffic safety) is critical to designing an evaluation. This understanding will inform potential process measures (e.g., how many people experienced the strategy), intermediate outcome measures (e.g., which beliefs and behaviors to examine for change), and impact measures (e.g., crash types) as well as provide insights as to how much time the strategy will take to cause changes. Practitioners can reach out to the strategy developer and ask for the "theory of change" for the strategy (the theory of change lays out the science behind how a strategy has been shown to cause the expected outcomes). Additionally, a practitioner could need a contractor implementing a strategy to articulate how the strategy causes the expected outcomes.

3. **Identify Data Measures and Comparisons to Be Performed.** In this step, the stakeholders identify the reliable and valid data that measure the process, outcome, and impact of the strategy. The sources and methods to collect these data are also named. A comprehensive plan is developed for the evaluation, which includes the type(s) of

---

[3] https://www.cdc.gov/eval/guide/step2

planned comparison(s). As discussed previously, carefully consider the type of comparison, because it affects the ability to draw conclusions about the strategy.

4. **Make Meaning.** This step involves analyzing the data and interpreting the results. Considerations addressed in the first step can inform efforts in this step. Too often, evaluations are reduced to one simple question: "Did the strategy work?" Often the answer is: "Yes and no." Making meaning of the evaluation should allow for greater learning to inform how to make the strategy more effective. More questions to ask include:

    a. Was the strategy implemented as intended? Why or why not? What could be done better next time?

    b. Did the strategy reach the intended audience? Why or why not? What could be done better next time?

    c. Did the strategy result in the intermediate outcomes (e.g., in beliefs and behaviors) expected? Why or why not? What could be done better next time?

    d. Did the strategy result in the desired impact? Why or why not? What could be done better next time?

The evaluation may not inform all these questions. Asking these questions may guide how future evaluations may be changed to answer more questions. The intent is to use the evaluation results to improve effectiveness over time by enhancing learning.

5. **Accumulate and Share Wisdom (e.g., lessons learned)**. A single evaluation, if explored and discussed by stakeholders, can generate many lessons that can inform future actions. These lessons are often much more valuable than simply answering the question "Did the strategy work?" Stakeholders should give time to review and discuss the evaluation results and gather lessons to share with other stakeholders.

An evaluation can have greater impact if the lessons learned reach a variety of audiences that need the information to make decisions about strategies, planning, funding, etc. To be accessible and usable, lessons should use language familiar to stakeholders.

It is also important to accumulate lessons learned and evidence for a strategy over time, because a single evaluation may not be enough to truly understand how best to implement a strategy or to convince stakeholders to continue support for the strategy.

5 GUIDANCE POSTER

Here, we provide a screen shot of the poster created for this project (Figure 2). This poster focused on evaluative-thinking and the key steps in an evaluation.

Figure 2. Screen shot of TRB poster based on guidance.

6 CONCLUSIONS

As with many forms of traffic safety strategies – there is still insufficient evidence for the effectiveness of traffic safety culture strategies. In part, this is because too few traffic safety culture strategies have been implemented and evaluated.

Vision Zero is bold but necessary because no traffic fatality is ever acceptable. To reach this vision, we need effective strategies. This will include the use of innovative strategies to change traffic safety culture. These strategies focus on changing shared beliefs that influence our choices to behave either safely or recklessly. For such strategies to become more widely used, we need more evidence that they are effective and sustainable. Evidence-based decisions depend on reliable and valid data from well-designed evaluations to measure the process, outcome, and impact of strategies.

It is beneficial to us and the communities we serve to think in evaluation terms, especially when we have limited resources to achieve the Vision Zero. Evaluative thinking is a form of problem solving involved in designing, selecting, and distributing resources for traffic safety programs. It seeks credible evidence to provide answers about the effectiveness and sustainability of traffic safety programs.

One role we can all share is to be proponents for quality evidence and the need for credible evaluations. To that end, Table 1 offers some talking points to help with discussions about the importance of evaluative thinking.

Table 1. Promoting Evaluative Thinking

Many traffic safety practitioners and stakeholders already engage in forms of evaluative thinking. Discussing the value of evaluative thinking within traffic safety will grow its importance. Here are talking points to foster conversations about the importance of evaluative thinking.

1. Fatal crashes and serious injuries have a significant impact on public health.

2. Zero traffic fatalities and serious injuries is the only acceptable goal.

3. To be successful in reaching this goal, we must learn to use innovative strategies and grow evidence of their effectiveness.

4. Evaluations inform which strategies are effective and generate knowledge about how to make strategies more effective and sustainable.

5. Traffic safety practitioners can seek opportunities to include process, outcome, and impact evaluations in the projects they implement, manage, and fund.

6. Effective evaluations require quality data and proper comparisons.

7. Evaluations should include engaging stakeholders, developing careful descriptions of strategies, and finding quality data and proper comparisons.

8. Traffic safety practitioners can create opportunities to review and discuss evaluation results with stakeholders to gather lessons learned and show opportunities for improvement in future efforts.

9. More consistent and rigorous evaluations will accelerate learning and effectiveness of strategies in improving traffic safety.

10. Investing in training to help staff become more familiar with evaluation design and contracting with evaluators will improve the effectiveness of strategies and traffic safety.

7   REFERENCES

Archibald, T. (2013). Evaluative Thinking. *Free Range Evaluation,* WordPress. Retrieved from: https://tgarchibald.wordpress.com/2013/11/11/18/

Bennett, G. and N. Jessani (2011), The Knowledge Translation Toolkit, Bridging the Know–Do Gap: A Resource for Researchers, International Development Research Centre, http://ajpponline.org/resources/downloads/04-TheKnowledgeTranslationToolkit.pdf.

Chen, H. T. (2014). *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective. 2nd ed.* Thousand Oaks, CA: SAGE.

Hill, N. L., Kolanowski, A. M., Milone-Nuzzo, P., & Yevchak, A. (2011). Culture Change Models and Resident Health Outcomes in Long-Term Care. *Journal of Nursing Scholarship: An Official Publication of Sigma Theta Tau International Honor Society of Nursing 43* (1).

Petriwskyj, A., Parker, D., Wilson, C. B., & Gibson, A. (2016a). Evaluation of Subscription-Based Culture Change Models in Care Settings: Findings From a Systematic Review. *The Gerontologist 56* (4): e46–e62.

Petriwskyj, A., Parker, D., Wilson, C. B., & Gibson, A. (2016b). What Health and Aged Care Culture Change Models Mean for Residents and Their Families: A Systematic Review. *The Gerontologist 56* (2): e12–e20.

Puddy, R. W. & Wilkins, N. (2011). Understanding Evidence Part 1: Best Available Research Evidence. A Guide to the Continuum of Evidence of Effectiveness. Atlanta, GA: Centers for Disease Control and Prevention.

U.S. Department of Health and Human Services Centers for Disease Control and Prevention. (2011). Office of the Director, Office of Strategy and Innovation. Introduction to program evaluation for public health programs: A self-study guide. Atlanta, GA: Centers for Disease Control and Prevention. Retrieved from: https://www.cdc.gov/eval/guide/cdcevalmanual.pdf

APPENDIX I – LITERATURE REVIEW

Guidance for Evaluating Traffic Safety Culture Strategies

Task 1: Literature Review

Prepared by:

Eric Austin, Ph.D., Kelly Green, M.P.A., Jay Otto M.S., and Nicholas Ward Ph.D.

Center for Health and Safety Culture

Western Transportation Institute

College of Engineering

Montana State University – Bozeman

Disclaimer

This document is disseminated under the sponsorship of the Montana Department of Transportation (MDT) and the United States Department of Transportation (USDOT) in the interest of information exchange. The State of Montana and the United States assume no liability for the use or misuse of its contents.

The contents of this document reflect the views of the authors, who are solely responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the views or official policies of MDT or the USDOT.

The State of Montana and the United States do not endorse products of manufacturers.

This document does not constitute a standard, specification, policy, or regulation.

Alternative Format Statement

MDT attempts to provide accommodations for any known disability that may interfere with a person participating in any service, program, or activity of the Department. Alternative accessible formats of this information will be provided upon request. For further information, call 406/444.7693, TTY 800/335.7592, or Montana Relay at 711.

Table of Contents

1. EXECUTIVE SUMMARY

In recent years, traffic safety agencies have developed and implemented new initiatives aimed at changing both agency and road user culture as a way to reduce the number of injuries and fatalities on public roads and highways. A preliminary review of the Transportation Research Board's Transportation Research International Documentation and Research in Progress databases, as well as other research databases, found several broad lines of research on culture. One line focuses on social and safety culture of communities as it relates to the behaviors of community members who operate cars, ATVs, watercraft, snowmobiles, and other vehicles (Mulder and de Rooy 2018; Hanchrow 2017; Li, Gkritza, and Albrecht 2014). A second body of research studies the organizational culture of transportation agencies and examines culture both as it affects either safety orientated activities or organizations' cultural capacity for innovation and change (Bedford, Egan, and Graham 2017; Brunetto, Xerri, and Nelson 2014). A final area of cultural research explores these questions in maritime, air, and other non-road transportation domains (Fu and Chan 2014; Mearns et al. 2013; López de Castro et al. 2013).

As the use of culture-based safety initiatives has expanded, systematic evaluations of the operations and impacts of these new programs have not advanced as rapidly as the programs themselves. Several authors have noted that road safety campaigns (one type of strategy used to change traffic safety culture) are rarely subjected to a formal and complete evaluation (Robertson et al. 2015; Hoekstra and Wegman 2011). This lack of accessible evaluation data severely restricts the advancement and adoption of effective campaigns because there is (1) no guidance on how to improve campaigns, (2) no evidence to discontinue ineffective campaigns, and (3) no impetuous to advance safety campaign techniques. Both peer-reviewed and professional literature suggests that there is a consistent set of barriers to both conducting evaluations and using the results in instances when evaluations are conducted. Commonly cited barriers include factors such as a lack of time and resources, insufficient knowledge to conduct or use evaluations, and skeptical attitudes among program staff about the process and results of evaluations (Brescianai 2011; Holosko 1996). The General Accountability Office (GAO) reports that less than 40% of the agencies they examined had conducted formal evaluations of their programs. However, 80% of the agencies that conducted evaluations reported multiple benefits from having done so. Thus, rather than provide a hypothetical example of a complete evaluation, we instead reference the European Campaigns and Awareness-Raising Strategies in Traffic Safety (CAST) project, which developed standard tools for evaluating roadway safety campaigns (Vaa et al. 2009) and reporting their effectiveness (Boulanger et al. 2009). Both these tools are supported by a comprehensive guidance manual for designing, implementing, and evaluating roadway safety campaigns (Delhomme et al. 2009). Transportation agencies are advised to review this tool and manual as part of the design, implementation, and evaluation steps of the strategic approach.

The short-term benefits of this project include:

- The research will result in a summary analysis of formative and summative evaluation designs as well as any outcomes identified by the existing studies.
- That comparative assessment will be utilized to develop evaluation process guidance of traffic safety culture strategies for current practitioners based on available best practices.
- The assessment will also provide recommendations to develop better evaluations and ultimately more effective programs.

In the long-term:

- These findings will benefit researchers conducting future evaluations and their ability to craft successful summative and formative designs as well as program managers who either conduct or assess contracted evaluations.
- These findings will lead to more effective strategies as better evaluations allow program managers to make more informed decisions about selecting strategies and program developers to create more effective strategies.

This Task 1 Report reviews the existing literature on the evaluation of traffic safety culture and other culture change initiatives. The purpose of this literature review is to catalog the designs used to conduct those evaluations, and compile the major findings of the evaluations regarding the impacts of the culture change programs evaluated. The research strategy used to conduct this literature review entailed first searching the databases of peer-reviewed publications for English language evaluations of traffic safety culture initiatives, as well as other culture change programs beyond traffic safety. Because the literature within the field of traffic safety that evaluates implemented culture-change initiatives is relatively small, a literature review of published research on the implementation and evaluation of culture change initiatives within organizations more generally and across a variety of disciplines was completed as well. The literature that was identified through an iterative search strategy was then assessed to identify the evaluation methodology used, including data collection and analysis techniques used.

The examination of the literature gathered for this Task 1 literature review revealed evaluation designs including quasi-experimental, single-group, qualitative, and formative approaches. The review also found several systematic or secondary reviews of existing studies. While specific recommendations and guidance for practitioners will be provided in subsequent materials developed for this project, this review of this literature allowed us to examine implemented culture change initiatives over time within other disciplines to inform the possible design and implementation of traffic safety culture change strategies. While the shift toward cultural approaches to safety programming is relatively new to transportation and traffic domains, culture change and cultural interventions have been used in other organization types for nearly three decades. Assessment of what is known about the effectiveness of culture-based strategies in these areas will enable researchers and program staff to begin to determine what can be drawn from these settings and applied or adapted to traffic safety. A systematic assessment of this broader literature poses a different challenge than the review of traffic safety culture research and evaluations. This broader culture literature is so extensive that narrowing the analysis to those studies that are most applicable to traffic safety required refining the search and collection strategies and then the synthesis of those results into a meta-analysis useful to traffic safety researchers and program staff. Nevertheless, this literature review does demonstrate that there are a consistent set of evaluation strategies and designs that seek to balance the need for rigorous assessment of program impacts and the challenges of conducting high-quality research in complex field conditions.

2. INTRODUCTION

In an effort to reduce the number of traffic crashes and resulting injuries and fatalities, traffic safety agencies are developing and implementing new intervention strategies aimed at changing road user culture. While there is an emerging body of literature that provides guidance and shares experience of conducting evaluations of these programs (Lewis et al. 2019a, 2019b), systematic evaluations of the implementation and impacts of these new programs are not advancing as rapidly as the programs themselves. At this point, there are neither well-developed summative/outcome evaluations nor formative/process evaluations of most existing programs. Compounding this lack of systematic evaluation is an underlying lack of consensus about or development of the sorts of evaluation designs capable of yielding results that researchers and program managers can be confident in to support future programming and resource allocation decisions.

In contrast to summative evaluations, formative or process evaluations examine the implementation or operation of a program in order to determine if how a program is organized or implemented impacts its effectiveness. Formative evaluations can also provide useful information for program managers on the process of implementation and operation in the event that the program is expanded or replicated. The focus of these assessments is on design and functional performance of programs, regardless of the causal or logic model used. The review of existing formative research revealed even fewer process evaluations of existing programs than summative evaluations. Similarly, the review found no evidence of existing meta-analyses that would enable more global claims about the operational approaches that have proven effective or not.

To address the need for a better understanding of the availability and applicability of robust summative and formative evaluation designs, and in an effort to build a rich body of outcome and process data, this literature review:

1. Conducts a comprehensive systematic analysis of available evaluations of traffic safety culture initiatives in order to catalog and assess both their designs and findings. This results in a better understanding of the state of the field with respect to what is known about the effectiveness of existing culture-focused interventions and countermeasures and identifies, catalogs, and assesses the evaluation designs including their associated impact indicators and measures.
2. Conducts a parallel examination of what is known about formative and summative designs used to evaluate culture change initiatives in other fields including organization development, community development, and community health. An examination of these related fields yielded additional information about both the effectiveness and rigor of the evaluation designs as well as knowledge generated about the effectiveness and operation of culture change programs in those fields.

As the conceptual development of culture-based safety strategies becomes more refined and agencies move to implement those strategies, there is a need to systematically assess which program models are effective and why, as well as what program models function well within organizations and across wider communities.

One focus of this research is oriented toward summative or outcome evaluations, which are those that assess the effectiveness of programs with respect to their capacity to affect desired outputs and/or outcomes. Where possible, these evaluations endeavor to separate the impacts of the program from other factors that may simultaneously affect the outcomes of the program in

question. Summative evaluations are also critical in calculating the size of the program's effect, particularly in public sector programming where scarce resources need to be allocated to programs with the most substantial influence. An initial review of the existing literature revealed a small number of self-identified summative evaluations, as well as a number of case and single-group studies that assessed the impact of single or stand-alone programs. However, these efforts vary substantially in design, and the review did not uncover any meta-analyses of these empirical assessments that would support any conclusions about the relative effectiveness of different program models.

In contrast to summative evaluations, formative or process evaluations examine the implementation or operation of a program in order to determine if how a program is organized or implemented influences its effectiveness. The focus of these assessments is on design and functional performance of programs regardless of the causal or logic model used. The review of existing formative research revealed even fewer process evaluations of existing programs than summative evaluations. Similarly, the review found little evidence of existing meta-analyses that would support more global claims about the operational approaches that have proven effective or not.

3. MATERIALS AND METHODS

To identify articles for this project, a keyword search was performed using Montana State University's Library "CatSearch" meta-search engine. This search engine consolidates all of the University's research services into one, comprehensive search tool. Word search and phrase combinations initiated broad searches of "traffic safety," "traffic safety culture," "transportation safety," and "transportation safety culture." This search was augmented with a follow-on search of specific search engines including ProQuest Central, Elsevier Science Direct Journals Complete, and Emerald A-Z Current Journals in order to corroborate the results from CatSearch. All searches were structured to include only results published in English, and did include studies that were conducted both in and outside the United States. While the search strategy did not use any date parameters to restrict the search, only four articles published before 2000 were identified. As expected, this search did not yield results in these domains that included the implementation and evaluation of traffic safety culture change initiatives. As a result, new keyword and phrase searches were conducted to expand the search outside traffic and transportation safety domains to capture culture change initiatives that were implemented and evaluated for additional studies to include in this literature review. We used search terms including "culture change" and "safety culture" as title and subject searches with combinations of additional key word searches including "intervention," "evaluation," and others. This search allowed us to focus on culture change initiatives that have been both implemented and evaluated in other disciplines and organizations, which let us focus on the culture change initiative, its implementation, and any results due to the interventions.

4. RESULTS

The search strategy described in the Methods section above resulted in a set of 64 articles, the content of which is summarized in this section. The results below focus on the evaluation strategies and techniques used to assess the effectiveness of culture change initiatives, including those within safety culture, organizational culture, and social culture. The review is organized around the major types of designs found in the literature.[4]

4.1. Quasi-Experimental Designs

The goal of any evaluation effort is to produce the most valid and reliable information about the program or intervention being examined. Generally, in the social sciences, experimental designs that randomly assign subjects into one or more program groups and a control group are considered the most rigorous strategy. Random assignment, which ensures that there are no systematic differences between the two (or more) groups, eliminates the need to conduct pre-tests. However, because of the ethical and logistical difficulties of conducting experimental evaluations in the field, quasi-experimental designs are a common way to mitigate threats to internal validity as much as possible while balancing the ethical and logistical challenges of "real world" evaluations.[5]

The majority of the quasi-experimental evaluations identified in the literature focused on individuals as the unit of analysis and sought to understand the impact of the culture intervention on either members of the organization or recipients of services provided by the organization. Among those evaluations that used individuals as the unit of analysis, the most common data collection strategy was the use of surveys or questionnaires to gather baseline and subsequently post-test data from both the intervention and comparison groups (Caspar, O'Rourke, and Gutman 2009; Gonzalez-Formoso et al. 2019; Xu et al. 2018; Ginsburg et al. 2005). Others administered surveys to both the test and comparison groups before and after the intervention, but also gathered additional information from other stakeholders (Hermer et al. 2017) or secondary sources of performance data (Guzmán et al. 2017).

A second set of evaluations used the organization as the unit of analysis and assessed the degree to which organizational indicators of culture had changed rather than individual measures. Hermer et al. (2018) for example, used program records to examine nursing homes' culture, performance, and decisions regarding the choice to opt into a state-wide program. In another healthcare-focused

---

[4] While there is a great degree of conceptual and terminological standardization among evaluators and within the academic study of evaluation, there is still some variation among authors and evaluators about how specific studies and evaluation strategies are described. Moreover, because field evaluations are ultimately designed in ways that reflect programmatic needs and practical realities, many projects do not perfectly conform to ideal types of evaluation designs and may blend designs and strategies. As a result, the typology of designs here and placement of studies identified in this review of the literature into that typology is based on the framework and outline in Posavac, and an examination of 1) methodological description provided by the study's authors, and 2) an assessment by the researchers of the core or central elements of the study (Posavac 2011). Following that framework, the studies included here are presented in narrative form, but will also be organized into a summary table for the final report.

[5] It is worth noting that, according to Posavac (2011), time-series designs constitute a second form of quasi-experimental approach. Time-series studies collect information on the same measure or measures at multiple points in time – more than just once before and after an intervention – in order to assess the impact of that intervention and exclude other potential variables like a historical event or maturation.

---

evaluation, a team of evaluators conducted a comparison of facilities that adopted a culture change with a propensity score[6] compared to a set of facilities that did not adopt the program (Grabowski et al. 2014). Using existing facility-level performance data, the evaluators analyzed differences between those facilities that had engaged in culture change efforts with a comparable set of facilities that had not done so. Researchers then used a regression analysis to examine whether the culture change efforts impacted facility-level performance. Similarly, evaluators studying a German healthcare intervention collected performance data from 60 general healthcare practices, which were randomly assigned into a test group of 28 practices and a control group of 32 practices prior to an intervention and again 12 months after the practices (Hoffmann et al. 2014). Performance data were analyzed using analysis of variance (ANOVA) in order to identify whether there were performance differences between the test and control groups of practices.

4.2. Single-Group Designs

The availability of and challenges associated with implanting an evaluation design utilizing a comparison group often makes quasi-experimental designs infeasible. Nevertheless, program managers and funders need the best possible information about whether programs generate change in the target audiences. There are several design strategies which examine just the target group to assess the impact of an intervention. Our review of the literature identified three distinct single-group designs within efforts to evaluate culture change initiatives: pre-post test, mixed methods, and post-test only designs.

4.2.1. Single-Group Pre- and Post-Test Designs

Single-group pre-post test evaluation strategies, as the term suggests, gather data from the target group about variables related to the desired outcomes prior to the intervention and again after the intervention has taken place and then compares the results to determine whether or how much change has occurred.

Evaluations of culture change found in this search of the literature fell into two categories. The first focused on assessing changes among staff before and after the culture intervention and included changes in attitudes values and beliefs that were reflective of culture (Marcinkoniene and Kekale 2007; Harvey et al. 2001) or included an examination of how those attitudes and values manifest in knowledge and behavior (Jones et al. 2013).

A second group of evaluations gathered data before and after the intervention, which aimed to understand impacts beyond the staff of the target organization itself. Two related studies examined not only the presence of a changed culture in the target organization, but also the relationship between that change and specific patient outcomes (Meddings et al. 2017; Smith et al. 2017). Others (Lee-Fay et al. 2018; Nielsen 2014) looked at patient outcomes such as depression rates or injury rates but also outputs such as social care that were necessary to achieve outcome objectives.

Various forms of regression analysis, which seek to identify causal relationships among variables, were the most common form of analysis among all of these evaluations (Lee-Fay et al. 2018; Meddings et al. 2017; Smith et al. 2017; Jones et al. 2013). Other evaluations included additional

---

[6] Propensity score matching is a means of controlling for variables other than the intervention that might cause changes.

descriptive forms of analysis that identified the existence of relationships between variables but did not indicate that those relationships were causal (Nielsen 2014; Harvey et al. 2001).

4.2.2. Mixed Methods Designs

For a variety of reasons, pre and post tests may not be possible in those circumstances, mixed methods evaluation designs often provide an alternative approach, which is more feasible and still provides useful, systematic information about program effects. Mixed method designs gather different types of programmatic information from different sources and utilize different data gathering techniques or approaches. Data types and sources can include individual interviews, observations, focus groups, surveys, the collection of secondary data, or others. Typically, mixed methods designs are used as a way of enhancing the robustness of an evaluation or when the validity of any one data source can be enhanced with the addition of parallel but independent data, which corroborates or reinforces the findings from others. In other words, mixed methods evaluations then use a comparative examination of various data types in order to "triangulate" or corroborate the findings of one data source with those of another, independent data source.

The most common form of mixed methods evaluations of culture change efforts utilized surveys or questionnaires as a core source of information and then added additional sources of data including interviews and secondary sources of performance data (Bradley et al. 2018; Jorritsma and Wilderom 2012; Simons et al. 2015), interviews and observations (Curry et al. 2015), or interviews and surveys of secondary stakeholders (Bystydzienski et al. 2017). One other study that relied primarily on survey responses supplemented it with secondary, performance data (Rachele 2012). A final study began with surveys and then utilized distinct interviews with two different stakeholder groups (Barratt-Pugh and Bahn 2015).

Other common evaluation strategies relied on interviews as the primary source of information, coupled with observations (Bowers, Nolet, and Jacobson 2016; Nævestad 2010). In one case, (Cottingham et al. 2008) interviews and observations were supplemented further with other post-intervention, secondary data.

4.2.3. Longitudinal Studies

Longitudinal studies are those that gather data, typically from the same individuals or for the same variable, multiple times, over a set period of time, rather than simply before and after an intervention. Only one culture change evaluation was found that used a longitudinal design (Jarvie et al. 2008). This study gathered performance data every 4-6 months over the course of a 24-month program, as well as data on a secondary program indicator and compared those measures to program outcomes to assess the impact of the intervention.

4.3. Qualitative Designs

Broadly, qualitative designs are considered to be those that do not gather or analyze data that is or can be presented numerically. Qualitative designs are often used when researchers want to develop a richer or more nuanced understanding of factors that either cannot be easily quantified or when quantification loses or obscures something important, as can be the case with psychosocial concepts like culture and associated values, beliefs, and attitudes. Qualitative designs can include the collection of observational data, interviews, text analysis, and others. While qualitative designs can be quite diverse both in the forms of data collected and the types of analysis possible, our examination of the literature identified one common qualitative design, interviews, which were

analyzed using content analysis.[7] Content analysis is a method of examining text or verbal communications information and can be used to analyze other data types including images or videos. Analysis involves systematically reviewing and labeling or coding elements of the data – words, phrases, or passages - that exhibit meaningful content for the study. For example, coding for culture might include expressions of values, attitudes, or beliefs.

The most common qualitative design identified in this review utilized individual interviews, which sought to elicit information about culture-linked beliefs among stakeholder groups (Silvester, Anderson, and Patterson 1999) and different levels of staff regarding their understandings of culture and the impact of culture change initiatives (Wankhade and Brinkman 2014). Several other studies utilized focus group interviews rather than individual interviews. Focus groups have the advantage of enabling a small research team to have open-ended interviews with a larger number of participants, more quickly, and without dramatically increasing the volume of textual data to be analyzed. The potential risk of focus groups is that the social dynamics of a group may preclude some participants from fully sharing their perspectives, or discussions may evolve in such a way as to impede individual or minority perceptions from emerging. However, this can be overcome by using follow-up interviews or integrating focused groups into a mixed-methods design. Focus group-based designs included those that sought to understand the relationship between cultural attitudes and organizational systems and processes (Wankhade and Brinkman 2014) exploring the relationship between culture-change strategies and their impact on a new cultural understanding (Conceição and Altman 2011).

4.4. Systematic Literature Reviews and Meta-Analysis

Systematic reviews are a type of literature review that uses a systematic methodological approach to comprehensively gather literature relevant to a particular research question, rigorously examine, and, finally, synthesize elements of the identified literature that directly relates to the question at hand. A meta-analysis, by contrast, entails the aggregation of data from multiple studies into a single dataset which can then be analyzed and from which broader or more generalizable conclusions might be drawn. Because of the need for high-quality, relatively homogenous data, meta-analyses are relatively uncommon. A total of five studies were identified in this project that fell within the realm of a systematic literature review or meta-analyses. Several of these studies gathered a collection of literature based on the iterative refinement of a set of search terms and strategies, as well as inclusion and exclusion criteria. Once the relevant literature was compiled, these studies analyzed the literature through a pre-selected conceptual lens to assess the quality of the studies and their results (Hill et al. 2011), the degree to which the culture change initiatives comport with an existing organization change model (Johnson et al. 2016), or the initiatives fit with a pre-existing framework of organizational attributes that relate to culture and culture change (Morello et al. 2013). One systematic review did not begin its analysis based on a pre-existing framework but instead used a form of coding similar to content analysis to identify emergent themes (Sammer et al. 2010). Two final studies identified in this research, and conducted by the same research team, used the analytical approached developed by the Joanna Briggs Institute (JBI) to conduct an assessment of both quantitative and qualitative studies in a way that resembled a

---

[7] It should be noted that there are evaluations included in this literature review that gathered qualitative observations and used content analysis as a means of analyzing those data. However, those studies that used observations as a central strategy in concert with other forms of data gathered are included among the mixed methods designs.

more formal meta-analysis (Petriwskyj et al. 2016a, 2016b). The JBI framework and tools lay out a standard set of quality attributes that the Institute argues should be present in all studies that support evidence-based healthcare enabling the assessment of individual studies by an effort to aggregate the findings of those studies.

4.5. Formative Evaluations

Formative evaluations, sometimes described as process evaluations or program monitoring studies, examine a program itself including its activities, functions, procedures, and operations in order to understand how those operations conform with program plans and whether program functions may impact its efficacy. The methods used for formative evaluations are similar to those used in summative evaluations. The most common research strategy among the formative studies found in this research utilized interviews as a central component. Several targeted semi-structured interviews, which asked open-ended questions about a limited set of broad topics or focus areas in an effort to understand how a program functioned with respect to cultural attributes targeted by the intervention (Farokhzadian, Dehghan Nayeri, and Borhani 2018), techniques used to sabotage a culture change initiative (Harris 2002), or to elicit the perspectives and experiences of key groups of stakeholders involved in a culture change intervention (King and Moulton 2013). A second formative strategy relies on formalized or systematic participant observations of the intervention and its impacts (Ward et al. 2018; Frame, Watson, and Thomson 2008). One study uncovered in this research integrated participant observations into an iterative process tied to the implementation of a culture change initiative similar to an action-research project rather than a retrospective analysis of it (Kakabadse and Kakabadse 2002).[8]

4.6. Findings about the Effects of Culture Change Initiatives

The final objective of this study was to assess the findings of the evaluations identified in the search of the literature. The following summary of those findings is organized into two major categories:

- Findings about the impacts of culture change initiatives on program outputs including the attributes or indicators of the target culture, or impacts of the change initiative on program outcomes such patient health outcomes or organizational performance measures.
- Findings about efficacy or importance of the mechanisms used to affect culture change including training and education, system or structural changes, leadership roles, and others.

4.6.1. Output and Outcome Findings

Many of the evaluation studies uncovered in this review focused on the effect the interventions had on the program's outputs, outcomes or in some cases both. The description of those findings in the following sections is broken into outputs or what is produced by the intervention being evaluated, and outcomes, which are indicators or elements of the program's ultimate objectives. For example, an output of a patient service program might include the hygiene behavior of a nurse or other service provider, which is seen as a necessary precursor to and component of improving a program outcome such as improved patient health measured by infection rates. Evaluators

---

[8] Action-research refers to a broad range of strategies which combine data gathering with the use of those data to directly and immediately inform and enhance the project being studied.

sometimes find it easier to measure outputs than outcomes, particularly when outcomes are more abstract or are integrated with a variety of factors, such as outcomes like patient wellbeing. Outputs are often easier to identify and measure, and when part of a program's logic model, they can sometimes be used as a proxy for outcomes.

4.6.1.1. Outputs:

Although the evaluations identified in this research approach the concept of culture and culture change from a variety of different culture concepts and models, there is significant continuity in the indicators of culture, the interventions focused on, and the measurement of evaluations. One of the best examples of an evaluation that identified program outputs as a precursor to outcomes was Jarvie et al.'s (2008) study of healthcare provider behavior which examined hand-hygiene behaviors as an output, which was a central antecedent to infection rates. Several of the evaluations focused on changes in attitudes among the target population as an indicator of change (Nævestad 2010; Cottingham et al. 2008; Jorritsma and Wilderom 2012). While more difficult to study, many models of culture include attention to shared values and norms as a central element, and at least one evaluation included an effort to assess changes at that level (Hermer et al. 2018). As a more practical and accessible indicator of culture change, many of the evaluations concentrated their assessments instead on individual behaviors (Nielsen 2014; Zuschlag, Ranney, and Coplen 2016; Jones et al. 2013), while others chose to focus on collective or organizational behaviors including teamwork (Xu et al. 2018; Nielsen 2014) and communications (Xu et al. 2018). These social variables were the most common organizational behaviors included in the evaluations reporting interventions in specific target organizations (Gonzalez-Formoso et al. 2019; Hoffmann et al. 2014; Xu et al. 2018).

Though not specifically an output of the interventions and programs studied, several evaluations highlighted an important finding related to the relational and behavioral outputs noted above. That is, a number of evaluations found that there were variations in program effects for attitudinal, behavioral, or other outputs, that were expressed. They include, for example, factors like teamwork and communications that take place across different bureaucrat levels, between organizational roles, or across professional silos or specializations within the organization studied (Harvey et al. 2001; Xu et al. 2018; Silvester, Anderson, and Patterson 1999). Finding such varying, and in some instances, unintentional effects of the interventions being evaluated is a valuable piece of information for other program managers to consider when developing culture change interventions and considering how to ensure consistent, optimal results.

4.6.1.2. Outcomes:

Program outcomes are often challenging to measure, in part because they are often somewhat abstract like "patient well-being," or because they are influenced by a variety of factors making it difficult to discern the effect of any one variable. Nevertheless, a number of the evaluations contained in this review included assessments of outcome impacts. Those outcomes included injury rates (Nielsen 2014), error rates (Hoffmann et al. 2014), organization performance measures (Zuschlag, Ranney, and Coplen 2016), patient outcomes including infection rates and psychosocial behaviors (Meddings et al. 2017; Smith et al. 2017; Hermer et al. 2018; Lee-Fay et al. 2018; Bradley et al. 2018; Jarvie et al. 2008).

4.6.2. Mechanisms Utilized for Culture Change

In addition to examining the output and outcome effects of culture change interventions, many of the evaluations in this research also described the mechanisms or components of culture change initiatives. By far the most common element of culture change initiatives is the importance of developing and utilizing collaboration, team-based, or relational approaches. More than a dozen evaluations highlighted some element of teamwork, collaboration, or interaction as an important element of either the culture change intervention or subsequent performance of the organizations studied (Bowers, Nolet, and Jacobson 2016; Farokhzadian, Dehghan Nayeri, and Borhani 2018; Ward et al. 2018; Hoffmann et al. 2014; Jones et al. 2013; Xu et al. 2018; Nielsen 2014; Zuschlag, Ranney, and Coplen 2016; Kakabadse and Kakabadse 2002; Eilers and Camacho 2007; Cottingham et al. 2008; Bradley et al. 2018; Barratt-Pugh and Bahn 2015; King and Moulton 2013). A related finding of several studies noted that the reduced use of top-down authority or command-and-control management approaches was important to both changing the culture and thereby improving performance (Hudson 2007; Marcinkoniene and Kekale 2007). Other evaluations noted the importance of ensuring leadership and management engagement in and support for the change efforts (Ginsburg et al. 2005; Sammer et al. 2010; Farokhzadian, Dehghan Nayeri, and Borhani 2018; Conceição and Altman 2011).

More instrumental mechanisms identified in the evaluations reviewed here included the use and effectiveness of training and educational strategies (Ginsburg et al. 2005; Gonzalez-Formoso et al. 2019; Simons et al. 2015; Bystydzienski et al. 2017; Jorritsma and Wilderom 2012; King and Moulton 2013). Several evaluations also found that ensuring that the interventions were data or evidence driven had a positive impact on their success (Hudson 2007; Eilers and Camacho 2007; King and Moulton 2013). At least two (Nielsen 2014; Simons et al. 2015) drew a connection between training and education and particular organizational attitudes that were important to the effectiveness of the intervention and subsequent organizational performance, namely the development of attitudes that support continuous and double-loop learning.[9] Lastly, consistent with Schein's (2004) model of organizational culture, a number of the evaluations (Farokhzadian, Dehghan Nayeri, and Borhani 2018; Xu et al. 2018; Lee-Fay et al. 2018) found that attention to organizational systems and infrastructure was important as well. Consideration of organizational systems and infrastructure includes elements like ensuring policies and procedures are consistent with the culture being developed. It would also include scrutiny of organizational systems like human resources, planning, budgeting, and the like. The importance of developing systems and infrastructure that is consistent with the culture is also supported by Edwards and Jabs' (2009) findings about the negative impact that rules and bureaucratic systems can have on organizational performance.

4.6.3. Mixed and Negative Results

While the majority of studies identified in this research found that the culture change interventions studied had positive impacts, there were some that had mixed or negative results. Because of the breadth of their reviews, several of the systematic literature reviews found variation in the effectiveness of the culture change studies they examined (Morello et al. 2013; Shier et al. 2014;

---

[9] Double-loop learning is a concept drawn originally from cybernetics, wherein an organization not only routinely collects information about its performance relative to pre-established operating parameters and then adjusts its operations accordingly (which is single-loop learning), but also gathers information about whether those operating parameters themselves are adequate and may adjust the parameters if need be.

Hill et al. 2011). Several evaluations of specific interventions (Simons et al. 2015; Smith et al. 2017; Meddings et al. 2017; Bowers, Nolet, and Jacobson 2016) showed mixed outcomes and at least two (Edwards and Jabs 2009; Wankhade and Brinkman 2014) found negative outcomes, and, in total, this suggests that while culture change initiatives can successfully change culture and have positive effects on both output and outcome measures of the settings where they occur are complex.

4.7. Social-Culture Change Interventions

One last finding of note is that our review of the published literature on evaluations of culture change interventions identified only two studies that could be considered focused on social culture, rather than organizational culture. One (Livingood, Allegrante, and Green 2016) focused on the role of mass communication on shifting social culture relative to tobacco use. A second (Griffiths et al. 2009) focused on the intersection between organization culture and social culture as it relates to tissue donation. The limited number of studies is important to note for programs that aim to enact culture change at a broader social or community level rather than within organizations. It is conceivable that the complexity of social culture by comparison to organizational culture may make changing it and measuring that change more complicated. The fact that there are few evaluations does not help our understanding of these efforts and may suggest that the challenges are significant enough to limit efforts to either achieve change at this level or to assess the impacts of those efforts. The search of the wider literature on traffic safety culture, organizational culture, and social culture returned a significant number of studies that were not evaluations of culture-change interventions but were instead empirical tests of the concepts and relationships. It is possible that there are studies of social culture that fall into this category but were not captured by the search for this research because they were not specifically evaluations. A potentially useful next step for additional or future research would be to look for such research, which may reveal information about broader social or community culture change efforts.

5. CONCLUSIONS

The goal of this stage of the project is to conduct a review of the available evaluation literature assessing culture change interventions focused on traffic safety culture as well as culture change initiatives in other disciplines. The results of this review will provide:

- A catalog of the evaluation designs that have been used to assess the efficacy of culture change efforts, and
- A review of the findings of the evaluations and what those findings reveal about strategies used in and effects of the culture change initiatives.

Subsequent stages of the project will add:
- An assessment of the efficacy, rigor, and applicability of the identified designs, and
- Guidance and recommendations for practitioners to use as they develop and implement evaluations of their own programs.

There is a growing trend in a wide range of program settings of all types that emphasizes the importance of gathering the most rigorous data possible to gauge whether programs have the impact their designers intend and if they function as planned. Culture change and safety culture literatures have seen similar efforts to gather scientific evidence of effectiveness (Hill et al. 2011) and support evidence-based efforts (Petriwskyj et al. 2016a, 2016b).

The review of the published literature revealed a number of issues that contribute to a better understanding of the evaluation strategies that have been used to assess efforts to change culture. First and most broadly, *the studies identified by this research demonstrate Chen's (2014) claim about practical program evaluation effectively balancing rigor and practicality*. The evaluations reviewed here utilize a variety of designs and analytical strategies in an effort to minimize threats to validity. Those threats, including history, maturation, sample selection, and instrumentation, can affect an evaluation's capacity to accurately measure changes to culture and, in some cases, the impact of that culture change on output and outcome objectives. At the same time, enhancing the rigor of a study can also increase the costs, not only monetarily but also in the time and complexity of conducting an effective evaluation. Factors like the selection of appropriate, accessible comparison groups, the development of valid instruments, and the administration of those instruments with a sufficiently large sample to support thorough analysis pose challenges to program managers, especially those with limited resources, staffing, and expertise. The importance of choosing rigorous approaches and tools for analysis is revealed in two related studies (Meddings et al. 2017; Smith et al. 2017). Both found positive changes in the culture of the target entities as well as positive changes in the outcome measures of the associated organizations. On the surface, finding positive change in both the independent variable (culture) and dependent variable (infection rates) is encouraging. However, when examined through the use of inferential analysis (regression), evaluators did not find a causal relationship between the variables.

Another key takeaway is that *the studies identified in this review demonstrated a wide range of research strategies*, both in design and analysis, that enabled researchers and program managers to develop a clearer understanding of their programs' operations and impacts.

With respect to the findings of the evaluations examined for this literature review, several conclusions are notable. While both the outputs and outcomes evaluated by these studies are quite

diverse and reflect the specifics of the programs where the culture change interventions took place, *there is substantial evidence that culture change initiatives can have a positive impact*. The findings of some evaluations are mixed, and negative outcomes indicate some of the complexity and challenges faced by culture-based approaches. However *even studies with negative outcomes do not suggest that there are insurmountable barriers to successfully implementing culture change initiatives*. Similarly, the *evaluations reviewed here demonstrate a diverse set of factors and mechanisms that are potentially important to the success of culture-based initiatives*. Chief among these is the importance of interaction via collaborative and team-based processes, which is to be expected because culture is a shared, social concept. Alternatively, additional crucial elements such as the influence of leadership and management support were explored and considered. The consistency between organizational systems and structures on one hand and culture on the other are not surprising but important, nonetheless.

6. REFERENCES

Barratt-Pugh, Llandis, and Susanne Bahn. 2015. "HR Strategy during Culture Change: Building Change Agency." *Journal of Management and Organization* 21 (6): 741–754.

Bedford, Denise A. D., Dominick Egan, and Hugh T. Graham. 2017. "Role of Culture in Adoption of Enterprise Collaboration Technologies: Case Study." *Transportation Research Record: Journal of the Transportation Research Board* 2646 (1): 1–7. https://doi.org/10.3141/2646-01.

Boulanger, Ankatrien, Stijn Daniels, Marko Divjak, Isabelle Goncalves, Annette Meng, Inger Synnøve Moan, Teti Nathanail, et al. 2009. "Evaluation Tool for Road Safety Campaigns." 2.3. Campaigns and Awareness-Raising Strategies in Traffic Safety. Belgian Road Safety Institute. http://www.roadsafetyknowledgecentre.org.uk/misc/userDownloadProtected.php?context=attachedFile&zone=knowledgeItem&file=66&filenameOverride=Evaluation+Tool.pdf.

Bowers, Barbara, Kimberly Nolet, and Nora Jacobson. 2016. "Sustaining Culture Change: Experiences in the Green House Model.(SPECIAL ISSUE–GREEN HOUSE MODEL OF NURSING HOME CARE)." *Health Services Research* 51 (1).

Bradley, Elizabeth H, Amanda L Brewster, Zahirah Mcnatt, Erika L Linnander, Emily Cherlin, Heather Fosburgh, Henry H Ting, and Leslie A Curry. 2018. "How Guiding Coalitions Promote Positive Culture Change in Hospitals: A Longitudinal Mixed Methods Interventional Study." *BMJ Quality & Safety* 27 (3). http://qhc.bmj.com/content/27/3/218.full.pdf.

Brescianai, Marilee J. 2011. "Identifying Barriers in Implementing Outcomes-Based Assessments Program Review: A Grounded Theory Analysis." *Research and Practice in Assessment* 6: 5–16.

Brunetto, Yvonne, Matthew Xerri, and Silvia Nelson. 2014. "Building a Proactive, Engagement Culture in Asset Management Organizations." *Journal of Management in Engineering* 30 (4): 04014014. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000251.

Bystydzienski, Jill, Nicole Thomas, Samantha Howe, and Anand Desai. 2017. "The Leadership Role of College Deans and Department Chairs in Academic Culture Change." *Studies in Higher Education* 42 (12): 2301–2315.

Caspar, Sienna, Norm O'Rourke, and Gloria M Gutman. 2009. "The Differential Influence of Culture Change Models on Long-Term Care Staff Empowerment and Provision of Individualized Care *." *Canadian Journal on Aging / La Revue Canadienne Du Vieillissement* 28 (2): 165–175.

Chen, Huey T. 2014. *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective*. 2nd ed. Thousand Oaks, CA: SAGE.

Conceição, Simone, and Brian Altman. 2011. "Training and Development Process and Organizational Culture Change." *Organization Development Journal* 29 (1): 33–43.

Cottingham, Ann, Anthony Suchman, Debra Litzelman, Richard Frankel, David Mossbarger, Penelope Williamson, DeWitt Baldwin, and Thomas Inui. 2008. "Enhancing the Informal

Curriculum of a Medical School: A Case Study in Organizational Culture Change." *Journal of General Internal Medicine* 23 (6): 715–722.

Curry, Leslie A, Erika L Linnander, Amanda L Brewster, Henry Ting, Harlan M Krumholz, and Elizabeth H Bradley. 2015. "Organizational Culture Change in U.S. Hospitals: A Mixed Methods Longitudinal Intervention Study." *Implementation Science* 10 (1).

Delhomme, Patricia, Werner De Dobbeleer, Sonja Forward, Anabela Simões, Giannis Adamos, Alain Areal, Julien Chappé, et al. 2009. "Campaigns and Awareness Raising Strategies in Traffic Safety." Public 3.2a. Campaigns and Awareness-Raising Strategies in Traffic Safety.

Edwards, Mark, and Lorelle Beth Jabs. 2009. "When Safety Culture Backfires: Unintended Consequences of Half-Shared Governance in a High Tech Workplace." *The Social Science Journal* 46 (4): 707–723.

Eilers, Angela M., and Armando Camacho. 2007. "School Culture Change in the Making: Leadership Factors That Matter." *Urban Education* 42 (6): 616–637.

Farokhzadian, Jamileh, Nahid Dehghan Nayeri, and Fariba Borhani. 2018. "The Long Way Ahead to Achieve an Effective Patient Safety Culture: Challenges Perceived by Nurses.(Report)." *BMC Health Services Research* 18 (1).

Frame, Joanne, Janice Watson, and Katie Thomson. 2008. "Deploying a Culture Change Programme Management Approach in Support of Information and Communication Technology Developments in Greater Glasgow NHS Board." *Health Informatics Journal* 14 (2).

Fu, Yan-Kai, and Tsung-Lung Chan. 2014. "A Conceptual Evaluation Framework for Organisational Safety Culture: An Empirical Study of Taipei Songshan Airport." *Journal of Air Transport Management* 34 (January): 101–8. https://doi.org/10.1016/j.jairtraman.2013.08.005.

Ginsburg, Liane, Peter G. Norton, Ann Casebeer, and Steven Lewis. 2005. "An Educational Intervention to Enhance Nurse Leaders' Perceptions of Patient Safety Culture." *Health Services Research* 40 (4): 997–1020.

Gonzalez-Formoso, Clara, Ana Claveria, M.J. Fernandez-Dominguez, F.L. Lago-Deibe, Luis Hermida-Rial, Antonio Rial, Francisco Gude-Sampedro, Salvador Pita-Fernandez, and Victoria Martin-Miguel. 2019. "Effectiveness of an Educational Intervention to Improve the Safety Culture in Primary Care: A Randomized Trial.(Clinical Report)." *BMC Family Practice* 20 (1).

Grabowski, David C., A. James O'Malley, Christopher C. Afendulis, Daryl J. Caudry, Amy Elliot, and Sheryl Zimmerman. 2014. "Culture Change and Nursing Home Quality of Care." *The Gerontologist* 54 (Suppl1): S35–S45.

Griffiths, J., M. Verble, S. Falvey, S. Bell, L. Logan, K. Morgan, and F. Wellington. 2009. "Culture Change Initiatives in the Procurement of Organs in the United Kingdom." *Transplantation Proceedings* 41 (5): 1459–1462.

Guzmán, Azucena, Jennifer Wenborn, Tom Swinson, and Martin Orrell. 2017. "Evaluation of the 'Ladder to the Moon, Culture Change Studio Engagement Programme' Staff Training:

Two Quasi-experimental Case Studies." *International Journal of Older People Nursing* 12 (3): n/a–n/a.

Hanchrow, Gregory. 2017. "International Safety Management – Safety Management Systems and the Challenges of Changing a Culture." *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation* 11 (1): 125–31. https://doi.org/10.12716/1001.11.01.15.

Harris, Lloyd C. 2002. "Sabotaging Market-Oriented Culture Change: An Exploration of Resistance Justifications and Approaches." *Journal of Marketing Theory and Practice* 10 (3): 58–74.

Harvey, Joan, Helen Bolam, David Gregory, and George Erdos. 2001. "The Effectiveness of Training to Change Safety Culture and Attitudes within a Highly Regulated Environment.(Statistical Data Included)." *Personnel Review* 30 (5 6).

Hermer, Linda, Natasha S Bryant, Madeline Pucciarello, Carolina Mlynarczyk, and Bridget Zhong. 2017. "Does Comprehensive Culture Change Adoption via the Household Model Enhance Nursing Home Residents' Psychosocial Well-Being?" *Innovation in Aging* 1 (2).

Hermer, Linda, Laci Cornelison, Migette L. Kaup, Judith L. Poey, Robyn Stone, and Gayle Doll. 2018. "The Kansas PEAK 2.0 Program Facilitates the Diffusion of Culture-Change Innovation to Unlikely Adopter." *The Gerontologist* 58 (3): 530–39.

Hill, Nikki L, Ann M Kolanowski, Paula Milone-Nuzzo, and Andrea Yevchak. 2011. "Culture Change Models and Resident Health Outcomes in Long-Term Care." *Journal of Nursing Scholarship : An Official Publication of Sigma Theta Tau International Honor Society of Nursing* 43 (1).

Hoekstra, Tamara, and Fred Wegman. 2011. "Improving the Effectiveness of Road Safety Campaigns: Current and New Practices." *IATSS Research* 34 (2): 80–86. https://doi.org/10.1016/j.iatssr.2011.01.003.

Hoffmann, B, V Müller, J Rochon, M Gondan, B Müller, Z Albay, K Weppler, et al. 2014. "Effects of a Team-Based Assessment and Intervention on Patient Safety Culture in General Practice: An Open Randomised Controlled Trial." *BMJ Quality & Safety* 23 (1). http://qhc.bmj.com/content/23/1/35.full.pdf.

Holosko, Michael J. 1996. "Obstacles to Conducting Program Evaluations: Implications for Public Health Evaluators." *Journal of Health & Social Policy* 8 (1): 91–101. https://doi.org/10.1300/J045v08n01_08.

Hudson, Patrick. 2007. "Implementing a Safety Culture in a Major Multi-National." *Safety Science* 45 (6): 697–722.

Jarvie, Lisa J, Paul D R Johnson, Meryanda E Jodoin, Celene Mcmullan, Roger H C Gregory, Kaye Bellis, Katie Cunnington, Fiona L Wilson, Diana Quin, and Anne-Maree Kelly. 2008. "Significant Reductions in Methicillin-resistant Staphylococcus Aureus Bacteraemia and Clinical Isolates Associated with a Multisite, Hand Hygiene Culture-change Program and Subsequent Successful Statewide Roll-out." *Medical Journal of Australia* 188 (11): 633–640.

Johnson, Anya, Helena Nguyen, Markus Groth, and Karyn Wang. 2016. "Time to Change: A Review of Organisational Culture Change in Health Care Organisations." *Journal of Organizational Effectiveness* 3 (3): 265–288.

Jones, Katherine J, Anne M Skinner, Robin High, and Roni Reiter-Palmon. 2013. "A Theory-Driven, Longitudinal Evaluation of the Impact of Team Training on Safety Culture in 24 Hospitals." *BMJ Quality & Safety* 22 (5). http://qhc.bmj.com/content/22/5/394.full.pdf.

Jorritsma, Petra, and Celeste Wilderom. 2012. "Failed Culture Change Aimed at More Service Provision: A Test of Three Agentic Factors." *Journal of Organizational Change Management* 25 (3): 364–391.

Kakabadse, Andrew, and Nada Kakabadse. 2002. "Making 'Modernising Government Initiatives' Work: Culture Change through Collaborative Inquiry." *Public Administration and Development* 22 (4): 337.

King, Jaime, and Benjamin Moulton. 2013. "Group Health's Participation In A Shared Decision-Making Demonstration Yielded Lessons, Such As Role Of Culture Change." *Health Affairs* 32 (2): 294–302.

Lee-Fay, Low, Shruti Venkatesh, Lindy Clemson, Dafna Merom, Anne-Nicole Casey, and Henry Brodaty. 2018. "Feasibility of LifeFul, a Relationship and Reablement-Focused Culture Change Program in Residential Aged Care." *BMC Geriatrics* 18 (1).

Lewis, I., Elliott, B., Kaye, S., Fleiter, J. & Watson, B. (2019a). In N. Ward, B. Watson & K. Fleming (Eds.), Traffic Safety Culture: Definition, Foundation and Application. Bingley, UK: Emerald Group Publishing Limited. 275-296.

Lewis, I., Forward, S., Elliott, B., Kaye, S., Fleiter, J. & Watson, B. (2019b). Designing and Evaluating Road Safety Advertising Campaigns. In N. Ward, B. Watson & K. Fleming (Eds.), Traffic Safety Culture: Definition, Foundation and Application. Bingley, UK: Emerald Group Publishing Limited. 297-320.

Li, Wanjun, Konstantina Gkritza, and Chris Albrecht. 2014. "The Culture of Distracted Driving: Evidence from a Public Opinion Survey in Iowa." *Transportation Research Part F: Traffic Psychology and Behaviour* 26 (September): 337–47. https://doi.org/10.1016/j.trf.2014.01.002.

Livingood, Jr., William C., John P. Allegrante, and Lawrence W. Green. 2016. "Culture Change from Tobacco Accommodation to Intolerance: Time to Connect the Dots." *Health Education & Behavior* 43 (2): 133–138.

López de Castro, Borja, Francisco J. Gracia, José M. Peiró, Luca Pietrantoni, and Ana Hernández. 2013. "Testing the Validity of the International Atomic Energy Agency (IAEA) Safety Culture Model." *Accident Analysis & Prevention* 60 (November): 231–44. https://doi.org/10.1016/j.aap.2013.08.017.

Marcinkoniene, Romualda, and Tauno Kekale. 2007. "Action Research as Culture Change Tool." *Baltic Journal of Management* 2 (1): 97–109.

Mearns, Kathryn, Barry Kirwan, Tom W. Reader, Jeanette Jackson, Richard Kennedy, and Rachael Gordon. 2013. "Development of a Methodology for Understanding and Enhancing

Safety Culture in Air Traffic Management." *Safety Science* 53 (March): 123–33. https://doi.org/10.1016/j.ssci.2012.09.001.

Meddings, Jennifer, Heidi Reichert, M Todd Greene, Nasia Safdar, Sarah L Krein, Russell N Olmsted, Sam R Watson, Barbara Edson, Mariana Albert Lesher, and Sanjay Saint. 2017. "Evaluation of the Association between Hospital Survey on Patient Safety Culture (HSOPS) Measures and Catheter-Associated Infections: Results of Two National Collaboratives." *BMJ Quality & Safety* 26 (3). http://qhc.bmj.com/content/26/3/226.full.pdf.

Morello, Renata Teresa, Judy A Lowthian, Anna Lucia Barker, Rosemary McGinnes, David Dunt, and Caroline Brand. 2013. "Strategies for Improving Patient Safety Culture in Hospitals: A Systematic Review." *BMJ Quality & Safety* 22 (1): 11–18. https://doi.org/10.1136/bmjqs-2011-000582.

Mulder, Sanne, and Diederik de Rooy. 2018. "Pilot Mental Health, Negative Life Events, and Improving Safety with Peer Support and a Just Culture." *Aerospace Medicine and Human Performance* 89 (1): 41–51. https://doi.org/10.3357/AMHP.4903.2018.

Nævestad, Tor-Olav. 2010. "Evaluating a Safety Culture Campaign: Some Lessons from a Norwegian Case." *Safety Science* 48 (5): 651–59. https://doi.org/10.1016/j.ssci.2010.01.015.

Nielsen, Kent J. 2014. "Improving Safety Culture through the Health and Safety Organization: A Case Study." *Journal of Safety Research* 48: 7–17.

Petriwskyj, Andrea, Deborah Parker, Christine Brown Wilson, and Alexandra Gibson. 2016a. "Evaluation of Subscription-Based Culture Change Models in Care Settings: Findings From a Systematic Review." *The Gerontologist* 56 (4): e46–e62.

———. 2016b. "What Health and Aged Care Culture Change Models Mean for Residents and Their Families: A Systematic Review." *The Gerontologist* 56 (2): e12–e20.

Posavac, Emil J. 2011. *Program Evaluation: Methods and Case Studies*. 8th ed. Upper Saddle River, NJ: Prentice Hall.

Rachele, Jude. 2012. "The Diversity Quality Cycle: Driving Culture Change through Innovative Governance." *AI & SOCIETY* 27 (3): 399–416.

Robertson, Robyn D, Charlotte Pashley, Canadian Automobile Association, and Traffic Injury Research Foundation of Canada. 2015. *Road Safety Campaigns: What the Research Tells Us*. http://epe.lac-bac.gc.ca/100/200/300/traffic_injury_research/road_safety_campaigns/2015_RoadSafety Campaigns_Report_2.pdf.

Sammer, Christine E., Kristine Lykens, Karan P. Singh, Douglas A. Mains, and Nuha A. Lackan. 2010. "What Is Patient Safety Culture? A Review of the Literature: Patient Safety Culture." *Journal of Nursing Scholarship* 42 (2): 156–65. https://doi.org/10.1111/j.1547-5069.2009.01330.x.

Schein, Edgar H. 2004. *Organizational Culture and Leadership*. 3rd ed. San Francisco: Jossey-Bass.

Shier, Victoria, Dmitry Khodyakov, Lauren W. Cohen, Sheryl Zimmerman, and Debra Saliba. 2014. "What Does the Evidence Really Say About Culture Change in Nursing Homes?" *The Gerontologist* 54 (Suppl1): S6–S16.

Silvester, Joanne, Neil Anderson, and Fiona Patterson. 1999. "Organizational Culture Change: An Inter-Group Attributional Analysis." *Journal of Occupational and Organizational Psychology* 72: 1–23.

Simons, Pascale A. M., Ruud Houben, Annemie Vlayen, Johan Hellings, Madelon Pijls-Johannesma, Wim Marneffe, and Dominique Handijck. 2015. "Does Lean Management Improve Patient Safety Culture? An Extensive Evaluation of Safety Culture in a Radiotherapy Institute." *European Journal of Oncology Nursing* 19 (1): 29–37.

Smith, Shawna N, M Todd Greene, Lona Mody, Jane Banaszak-Holl, Laura D Petersen, and Jennifer Meddings. 2017. "Evaluation of the Association between Nursing Home Survey on Patient Safety Culture (NHSOPS) Measures and Catheter-Associated Urinary Tract Infections: Results of a National Collaborative." *BMJ Quality & Safety* 27 (6). http://qhc.bmj.com/content/early/2017/09/26/bmjqs-2017-006610.full.pdf.

Vaa, Truls, Phillips Ross, Adamos Giannis, Areal Alain, Karen Ausserer, Patricia Delhomme, Divjak Marko, et al. 2009. "Effects of Road Safety Campaigns." Public D-1.3. Campaigns and Awareness-Raising Strategies in Traffic Safety. Institut Belge pour la Sécurité Roultiére.

Wankhade, Paresh, and John Brinkman. 2014. "The Negative Consequences of Culture Change Management." *The International Journal of Public Sector Management* 27 (1): 2–25.

Ward, Marie E., Aoife De Brún, Deirdre Beirne, Clare Conway, Una Cunningham, Alan English, John Fitzsimons, et al. 2018. "Using Co-Design to Develop a Collective Leadership Intervention for Healthcare Teams to Improve Safety Culture." *International Journal of Environmental Research and Public Health* 15 (6): 1182. https://doi.org/10.3390/ijerph15061182.

Xu, Xiao Ping, Dong Ning Deng, Yong Hong Gu, Chui Shan Ng, Xiao Cai, Jun Xu, Xin Shi Zhang, Dong Ge Ke, Qian Hui Yu, and Chi Kuen Chan. 2018. "Changing Patient Safety Culture in China: A Case Study of an Experimental Chinese Hospital from a Comparative Perspective.(ORIGINAL RESEARCH)(Case Study)." *Risk Management and Healthcare Policy* 11.

Zuschlag, Michael, Joyce M. Ranney, and Michael Coplen. 2016. "Evaluation of a Safety Culture Intervention for Union Pacific Shows Improved Safety and Safety Culture." *Safety Science* 83 (March): 59–73. https://doi.org/10.1016/j.ssci.2015.10.001.

APPENDIX II – JOURNAL ARTICLE

Assessing the Impact of Culture: A Systematic Analysis of Culture Interventions and Evaluations in Different Organizational Settings

Abstract: Over the last twenty years, transportation agencies have increasingly added culture-based approaches to the existing education, engineering, and enforcement strategies being used as a means of reducing traffic related injuries and fatalities. Despite this increased interest, there have been comparatively few evaluations of interventions designed to enhance traffic safety culture. At the same time, many other organization types have adopted culture-based strategies either to improve safety or to enhance other elements of organizational performance. In aggregate, the evaluations of culture-focused interventions across a range of settings offer an untapped body of information about the models of culture being leveraged to affect change, the intervention strategies used to impact culture, the impacts of these strategies, and more. This article presents the results of a systematic analysis of evaluations of culture-focused interventions across a variety of settings and seeks to identify patterns that could be useful to both researchers and practitioners. The findings of the study suggest that there are areas of substantial consensus regarding the nature and features of culture and the potential effectiveness of culture-based programs. At the same time, the findings also suggest that more conceptual and empirical work is warranted to further refine our understanding of culture and its functions and to build deeper understanding of how to leverage culture effectively to support health and safety efforts.

1. Introduction

Over the last decade, a substantial number of traffic safety agencies have developed and implemented new initiatives that seek to change both agency and road user culture as a way to reduce the number of injuries and fatalities on public roads and highways. As the number and use of culture-based safety initiatives has increased, systematic evaluations of those operations and their impacts have not kept up with the programs themselves. A growing number of researchers have noted that road safety campaigns - one type of strategy used to change traffic safety culture - are rarely subjected to a formal and complete evaluation (Hoekstra & Wegman, 2011; Robertson & Pashley, 2015). This lack of accessible evaluation data severely restricts the advancement and adoption of effective campaigns because there is (1) no guidance on how to improve campaigns, (2) no evidence to discontinue ineffective campaigns, and (3) no impetus to advance safety

campaign techniques. Both peer-reviewed and professional literature suggest that there is a consistent set of barriers to both conducting evaluations and using the results in instances when evaluations are conducted. Commonly cited barriers include factors such as a lack of time and resources, insufficient knowledge to conduct or use evaluations, and skeptical attitudes among program staff about the process and results of evaluations (Brescianai, 2011; Holosko, 1996). The U.S. General Accountability Office (GAO) reports that less than 40% of the federal agencies they examined in the United States had conducted formal evaluations of their programs. However, 80% of the agencies that did conduct evaluations reported multiple benefits from having done so (Government Accountability Office, 2013). There is, however, a more extensive body of literature that evaluates other safety culture and organizational culture interventions outside of traffic safety. As is often the case, disciplinary specificity has prompted researchers and practitioners alike to retain a fairly narrow focus on what is known within their specific discipline and tends to give little attention to what may be gleaned from other fields of study. As a result, there has yet to be a broader examination of culture change initiatives across disciplines and settings. In an effort to close this gap, this study presents the results of a systematic analysis of evaluations conducted on traffic safety culture initiatives, as well as evaluations of safety culture and organizational culture change in other industries and settings.

1.1. The Evolution of Culture Theories and Their Application

Although the concept of culture has become increasingly adopted across a range of academic disciplines, has been utilized in a variety of settings, and has intuitive appeal, it has also been critiqued for being insufficiently clear and precise (Cox & Cox, 1996; Hale, 2000). To provide as much clarity as possible, it is useful to first ground and locate the approach to culture being deployed here. Anthropologist Clifford Geertz was among the first scholars to develop and operationalize a definition and systematic approach to the study of culture. Geertz, in his seminal book *The Interpretation of Cultures*, describes culture as, "a system of inherited conceptions expressed in symbolic forms by means of which men communicate, perpetuate, and develop their knowledge about and attitudes toward life" (Geertz, 1973). Despite the fact that such inherited conceptions reside in the minds of individuals who are a part of any culture, that culture, according to Geertz, is public in that its expression is manifest in patterns of social interaction. Not surprisingly, recognition of the presence and function of cultural attributes such as the communication, development, and perpetuation of knowledge and attitudes quickly moved from anthropology to other disciplines and was recognized in narrower, more specific settings including organizations. Throughout the 1980s and '90s, organization theorists began to explore not just the development and perpetuation of cultural values, beliefs, and their functions in organizations, but how those attributes could affect employee actions and ultimately the performance of organizations as a whole. Organization theorists had long recognized the limitations of both direct supervision and the use of rules and procedures as the sole basis of managing employee performance (see for example the work of Luther Gulick (1937) or Herbert Simon (1976)). The intentional development and management of organizational culture, it seemed, could be used as a means of establishing a set of shared perceptual attributes including values and beliefs that, especially when coincident with the organization's policies, procedures,

mission and objectives, could enhance individual and ultimately organizational effectiveness (see, for example, (Schein, 2004).

1.1.1. Safety culture

As research on organizational culture has evolved, researchers and practitioners alike have refined and developed greater specificity in the application of culture to particular settings and concerns that include the articulation of managerial priorities, the availability and distribution of resources, and the development of policies and procedures that support – or inhibit – consistency with articulated values (Nieva, 2003). Among these more specific areas of focus, safety culture or organizational safety culture (OSC) has emerged as a concept relating more narrowly to the beliefs and values concerning health and safety within an organization and the degree to which those attributes are embodied in practices and expressed in performance (Clarke, 1999). OSC has been used as a contributing element of a wide range of organizational analyses (Cox & Flin, 1998) and intervention initiatives designed to make the workplace less risky (Luria & Rafaeli, 2008). Definitions of organizational as well as safety culture, however, have remained variable and often ambiguous. Among those that move toward operational levels of detail, Reason (2000), for example, argues that safety culture expresses the "ability of individuals or organizations to deal with risks and hazards so as to avoid damage or losses and yet still achieve their goals" (p. 5). More recently, OSC has been described as the "assembly of underlying assumptions, beliefs, values and attitudes shared by members of an organization, which interact with an organization's structures and systems and the broader contextual setting to result in those external, readily-visible, practices that influence safety" (Edwards et al., 2013). Others, including Cox and Cox (1996) and Hale (2000), have picked up on the abstract and conceptual character of safety culture and raised concerns about the clarity, precision, and utility of the concept. Despite, and to some degree in response to these concerns (Havold & Nesset, 2009), researchers and practitioners have continued to extend and expand the use of OSC.

1.1.2. Traffic safety culture

Given the evolution of theories of culture, organizational culture, and now organizational safety culture, it is little surprise that a cultural approach has also made its way into transportation and traffic safety. Among the earliest instances of this trend can be found in the 2007 AAA report, which provided an initial outline (drawing from OSC) of what traffic safety culture is and a call to action for researchers and practitioners in this nascent field (Hedlund, 2007). That initial interest continued to grow in the subsequent years, as is evidenced by the breadth and diversity of efforts described during the 2011 Transportation Research Board sponsored conference in Washington DC (Turnbull, 2011). Efforts to further develop and refine both the concepts and practices associated with traffic safety culture appeared in pages of this journal with the 2014 special edition, perhaps most notably those pieces by Ward and Ozkan (2014) and Edwards, Freeman, Soole, and Watson (2014). Some of the most recent work provides a particularly detailed model and etiology of traffic safety culture and user behavior (Ward et al., 2019) and an extensive description of the attributes or indicators of culture and methodological recommendations for their assessment (Otto et al., 2019).

1.1.3. Problem and research question

There is widespread and growing recognition within and beyond the field of traffic safety that the effectiveness of interventions focused on education, engineering, and/or enforcement have limits and can benefit from the addition of other change strategies. Culture-based approaches have gained increasing attention and are being adopted in a variety of forms and across a wide variety of settings. Nevertheless, there is a great degree of variation in the understandings of culture used to inform these approaches and, correspondingly, disparity in how those interventions are assessed. These variations and disparities pose a challenge to practitioners wanting to adopt culture-based strategies, as they seek to identify practices that are most likely to accomplish their programs' objectives.

In an effort to build more continuity and deeper, shared understandings of culture-based theories, their use, and implications, this study examines what a more systematic analysis of the organizational and safety culture literature reveals about:

- The current models and corresponding uses of cultural theory being utilized across fields.
- The indicators and measures used to assess culture and culture change.
- The strategies and techniques to assess those interventions.
- The outcomes revealed by existing assessments.

2. Material and Methods

The research design used for this study follows the approach described by Kapucu, Hu, and Khosa (2017) in their analysis of network literature in public administration. We describe it here as a systematic analysis in that it utilizes a more rigorous examination of a body of literature than a traditional literature review (Ham-Baloyi & Jordan, 2016), but it does not aggregate and analyze the data gathered by included studies as would be the case with a meta-analysis. Nevertheless, the approach to systematic analysis used here enables researchers to perform a more substantive and sophisticated assessment of patterns and relationships within a body of literature than a traditional literature review. In this study, we seek to develop a better understanding of what we can discern from the existing literature about culture change initiatives, the approaches and culture models used, the effectiveness of those efforts, and how effectiveness is measured in ways that improve the understanding of both researchers and practitioners.

2.1. Data Collection

To maximize the consistency and accessibility of the literature examined in this study, we chose to use peer reviewed journal articles, rather than professional publications or books. To identify articles for the project, a keyword search was performed using Montana State University Library's "CatSearch" meta-search engine. This search engine consolidates access to all of the University's databases including InfoTrac, Academic Search Complete, JSTOR, Lexis-Nexis, and others into one comprehensive search engine. Initially, keyword and phrase combinations were used as the basis of broad searches for "traffic safety," "traffic safety culture," "transportation safety," and "transportation safety culture." This search was augmented with a follow-on search of specific search engines that previously yielded the largest number of results

including ProQuest Central, Elsevier Science Direct Journals Complete, and Emerald A-Z Current Journals to corroborate the results from CatSearch. Lastly, because the focus of this research is to better understand what is known about culture-based approaches to improve traffic safety, our search also included the Transportation Research Board's Transportation Research International Documentation (TRID) and Research in Progress (RIP) databases.

Searches were structured to include only results published in English and included studies that were conducted both in and outside the United States. While the search strategy did not use any date parameters to restrict the search, only four articles published before 2000 were part of the initial search results. As expected, based on a previous, preliminary examination of the literature, this search did not yield results focused on the implementation and evaluation of traffic safety culture change initiatives. As a result, new keyword and phrase searches were conducted to expand the search outside traffic and transportation safety domains to capture published evaluations of other safety culture and organizational culture change initiatives. Search terms including "safety culture" and "culture change" were used for both title and subject searches, in combination with additional key words including "intervention," "evaluation," and other variations. This search allowed us to focus on culture change initiatives that have been both implemented and evaluated in other disciplines and organization types, which let us focus more widely on culture change initiatives, their implementation, and any results attributed to the interventions. The search and review of materials resulted in a final set of 59 articles that were then analyzed. Figure 1 provides a summary flow chart of the search and screening steps used to yield the data set, and is based on the PRISMA flow diagram (Moher et al., 2009).
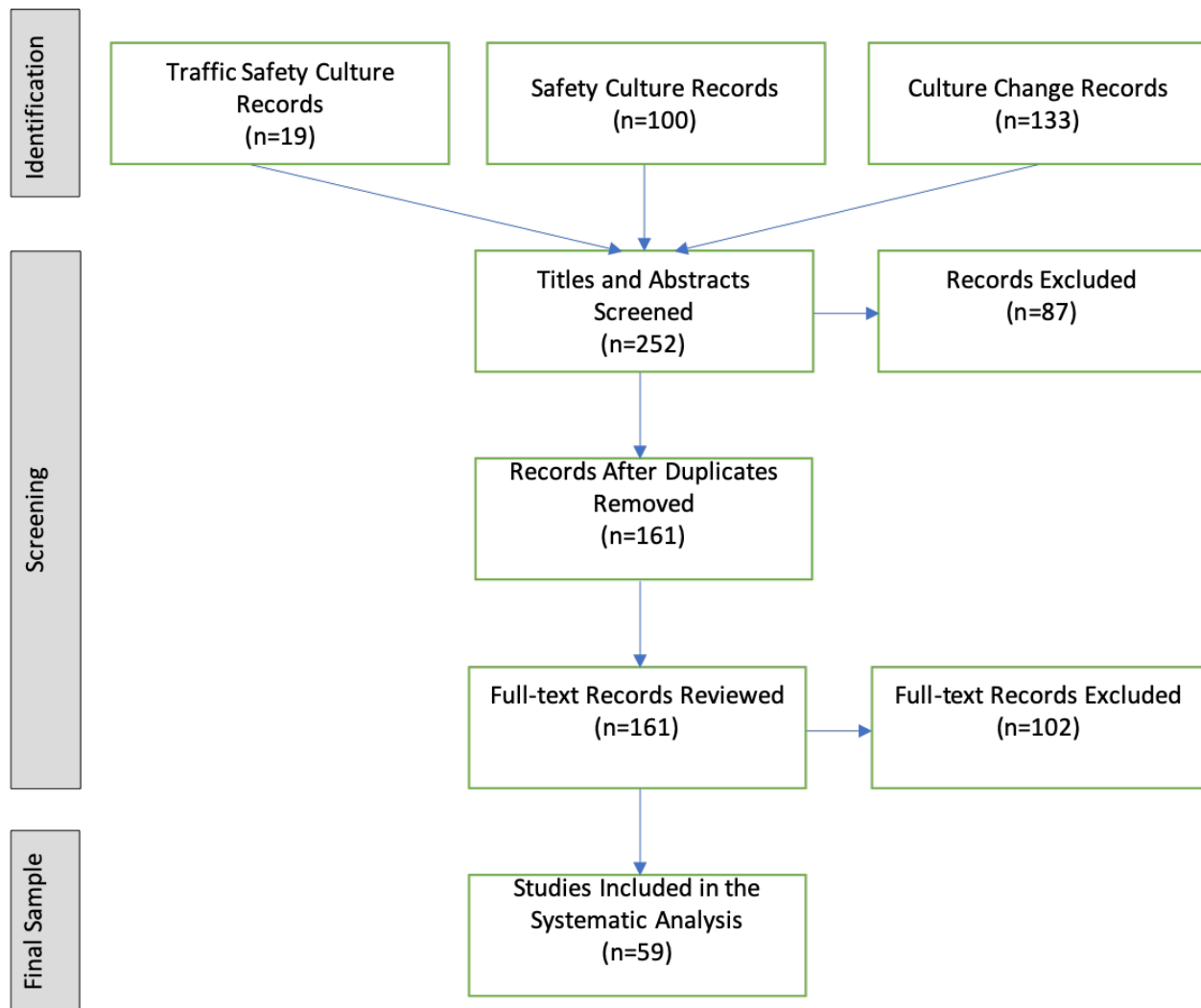
Figure 1 – Search and Screening Process

2.2. Analysis

Once the sample of articles was complete, PDF files for all of the articles identified were imported to the qualitative data management and analysis software NVivo for coding and analysis. The articles were coded using a combination of open and axial coding. Unlike a more typical grounded theory study (Strauss & Corbin, 1990) where no pre-existing theory provides a framework for coding, this study began with a preliminary set of propositional codes that had been developed from existing literatures. A preliminary coding scheme that included features such as research design, analysis, and setting or industry, was augmented with additional, open codes that emerged during the coding process.[10] The coding scheme resulted in a two-level

---

[10] It should be noted that coding for this study focused on items or files, rather than references. For example, when coding for the research design used, our concern was identifying the research design identified in each article,

hierarchy of codes called parent and child nodes. For example, the parent node for research design had a number of child nodes including quasi-experimental, single-group mixed methods, qualitative, etc. This hierarchy of parent and child nodes enabled the researchers to more easily structure the analysis at later stages of the project.

In an effort to enhance the reliability of the coding process, several of the first articles were coded independently by two researchers and then reconciled. That reconciliation was used to ensure clarity and consistency in the coding process going forward. Subsequently, a second selection of articles was chosen at the later stages of the coding process, and the coding completed by one researcher for those articles was again reconciled with that of a second researcher to ensure that the use and understanding of the codes had not diverged throughout the course of the coding process. Finally, the list of references that resulted from the process coding were reviewed for consistency.

The analysis of the coded materials included three main elements. The first involved an examination of the patterns that emerged from the coding process both within and across parent and child nodes. The examination of patterns included consideration of various codes' frequencies, both those that were more frequent and also those that were unique or unusual.

The second element of the analysis involved conducting a series of word frequency queries. These queries allow us to look at how frequently words appeared in the entire data set, but also within categorical subsets of the data. These queries are a means of identifying differences and similarities in focus and emphasis, for example, across industries. That is, by doing a word frequency query within each industry, we are able to get some indication of what is emphasized or prioritized within an industry and how that might vary between industries.

The final element of the analysis involved a series of matrix or cross-tab queries. These queries allow for the identification of patterns in the relationships or intersection of different parent and child nodes. For example, a query that compares the intersection of all the child nodes within the Culture Theory node (i.e. the categories of or cultural theories used by each study) with industry allows us to see if certain industries or sectors tend to use any one particular cultural/theoretical framework in their interventions or analyses by comparison to another industry.

3. Results

The results of our analysis are presented in three categories. The first focuses on the models and theories of culture identified in the literature we analyzed and related patterns that emerged. The second area of analysis focuses on prospective patterns in the literature related to the industry or

---

rather than the number of times each article referred to the research design used. The focus on items or files rather than references within each file allowed us to treat the article as the unit of analysis rather than the concept. Qualitative studies that use this form of content analysis are often more interested in the frequency of references, for example the number of times interviewees mention a topic, because it reveals something about the participants' concerns, perceptions and priorities as indicated by frequency of reference, which can then be tranced into how those concerns appear across the sample of participants as a whole. However, because our interest is on the patterns within the literature, a focus on references risks obscuring those patterns.

organizational type (e.g. health care, education, transportation, etc.) studied. The final area of focus is on the evaluation designs and analytical approaches used by the literature we analyzed.

3.1. Theories and Models of Culture

The first area on which the analysis focused was the theories or models of culture adopted by the initiatives being evaluated in this body of literature and any patterns that emerged regarding the links between the models of culture and:

- Outcome indicators or measures of culture change or of program impact.
- The unit of analysis used by the study and the target of the intervention.
- Impacts of the intervention or change initiative being evaluated.

The working hypothesis with which we began the analysis was that the cultural theories or models used by change agents would fall into categories associated with the major organizational culture theories. Our initial assumption was that if we could identify those theoretical orientations, we could begin to look for patterns that link those orientations to indicators and measures of change, and even impacts of the interventions. Figure 2 below is the "node tree map" that shows the "child" or sub-codes that fell under the overarching "Culture and Other Theory" node. These categories resulted from the process of axial and open coding for culture theories. The size of each box represents the relative frequency of each theory we identified in the literature. As we anticipated, Edgar Schein's organizational culture model was common. However, the two other child nodes for culture, "Culture-Other" and "No Clear or Explicit Model," were created because no other specific model of culture revealed itself in the literature. The "Culture-Other" node was created for those articles that had a specific definition or description of culture, but that could not be traced to a recognized organizational culture source (e.g., Westrum) or other cultural theory source (e.g., Clifford Geertz). The "No Clear or Explicit Model" was created for those articles that note a cultural approach but that do not define, describe, or otherwise articulate a specific understanding of what culture is or how it precisely functions in a way that can be linked to an identifiable cultural theory. The final child node that should be noted in Figure 2 is the "Learning-Systems" node, which was created to capture those articles that explicitly identified systems theory (e.g., the work by W. Edwards Deming or Joseph Juran) and related approaches including learning organizations (e.g., Peter Senge).
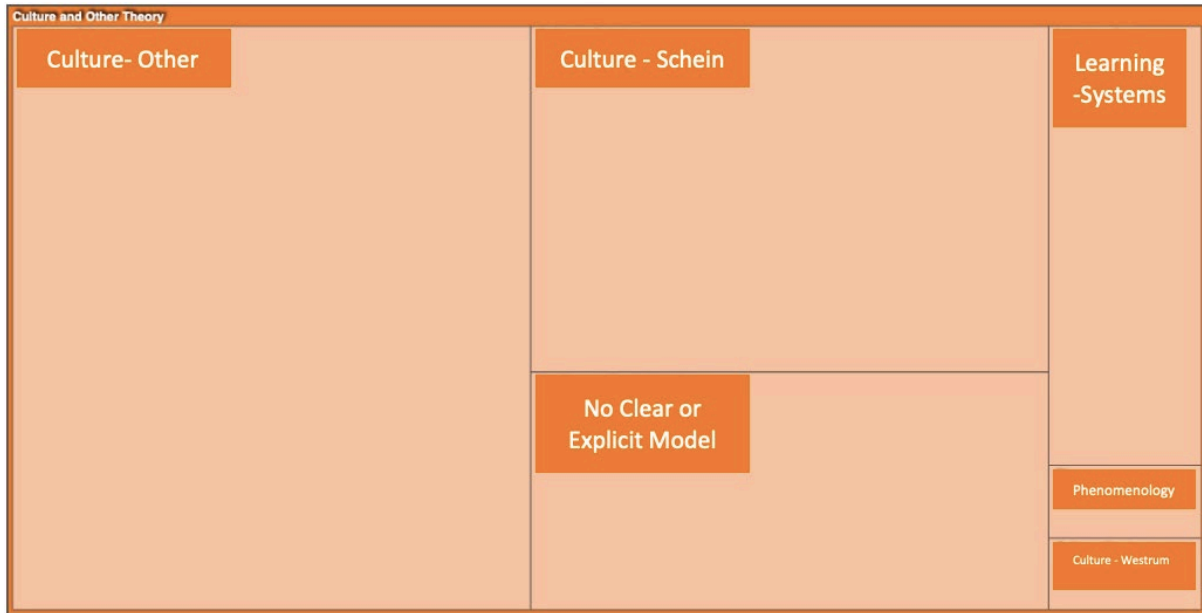
Figure 2 – Culture and Other Theory Node Tree Map

Figure 3 is the node tree that identifies the Outcome Indicators or Measures that were revealed in the analysis of the literature. As was expected, the most common outcome indicators and measures reflected common attributes of most cultural theories, namely that there are particular attitudes/beliefs, artifacts, or structures (e.g., policies, procedures and processes, as well as formal or informal structural forms such as authority or patterns of communication). It's also notable that a substantial number of articles explicitly recognized that particular values, like caring or transparency, were important indicators or measures of culture and culture change. These attributes are, in turn, often linked to culture-based behaviors, or behaviors that express or reflect the attitudes and values of the culture. Not surprisingly, the analysis also identified that organizational performance outcomes, such as patient or student outcomes, were also present in the literature.

Figure 3 – Outcome Indicator-Measure Node Tree Map

In addition to looking at the basic trends in the literature related to the cultural and other theories that inform the interventions and strategies and the outcome indicators and measures used, our analysis also examined the intersection or overlap between different concepts and their codes.[11]
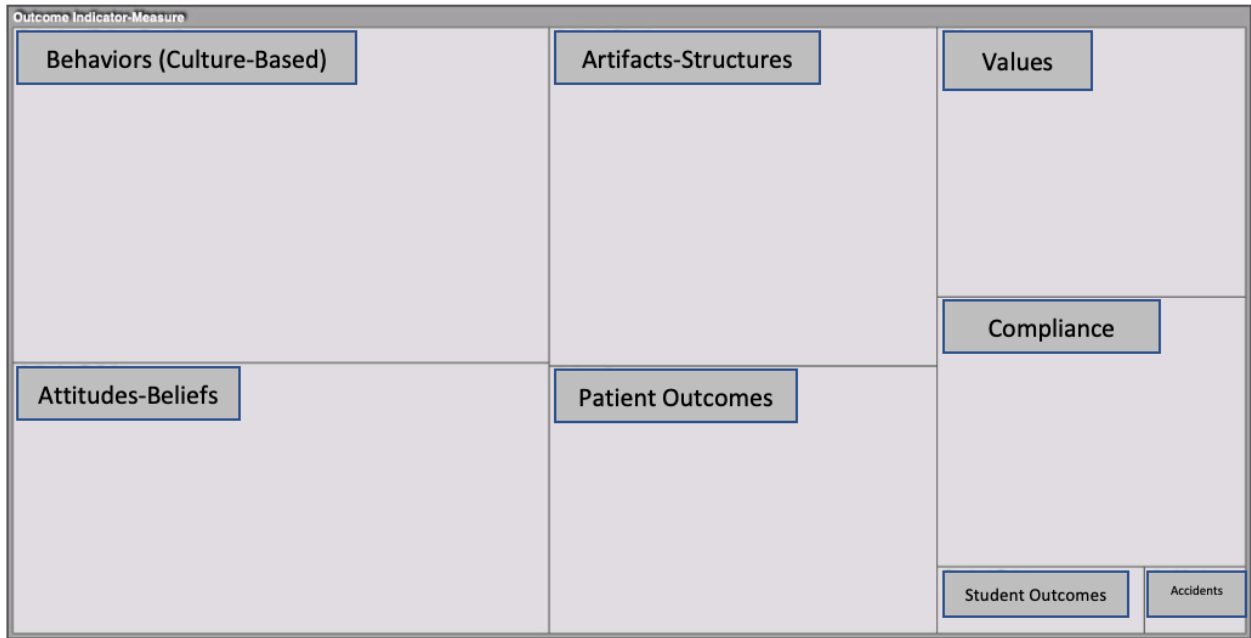
The first of these queries examined the intersections between the cultural or other theory used to inform the intervention and study and the outcome indicator or measure used. Several elements of this query, the results of which are shown in Table 1, are notable. The first is that indicators and measures that are consistently a part of cultural theories, namely attitudes, beliefs, and behaviors are by far the most common indicators and measures used in the studies we compiled, especially among those articles that work either from a cultural model based in Edgar Schein's work or that have some other, specific cultural framework. Similarly, a substantial number of these same articles identify that culture is expressed or embedded in the settings studied, into artifacts, namely things like policy, procedure, and practice or in formal or informal structures such as authority or communications.

Another notable pattern has to do with the commonality of "values" as an indicator in the studies in this sample. Nearly all cultural theories recognize that any given culture is likely to include a fairly specific set of shared values. However, relatively few studies, regardless of the underlying cultural theory identified, have values as an indicator or variable to be changed or augmented as a result of the intervention. One caveat to this comment is that values can overlap with beliefs and particularly attitudes. For example, a safety culture that prompts workers to care for each

---

[11] Matrix Queries in NVivo are similar to cross-tabs used in other quantitative studies. Because of the qualitative approach used here, the output of the matrix queries is presented graphically and in narrative form, rather than using frequencies.

other such that they are more likely to intervene to prevent risky behavior could describe care as a behavior, an attitude, or a value.

Lastly, and somewhat unexpectedly, relatively few studies identified or focused on outcome measures that a culture would purportedly influence (i.e. patient outcomes, student outcomes, or accidents/collisions). Even among those studies of healthcare organizations, which as we will describe further below, were by far the most common industry present in the sample of articles we identified, relatively few studies focused on those ultimate, program outcomes. The vast majority of studies focused primarily on intervening variables, which are components or elements of culture, rather than outcome variables that culture is intended to impact. One other indicator identified through the open coding process, was compliance – i.e., whether the organization, unit, or individuals studied complied with applicable regulatory regimes. In this way, compliance serves as both a convenient proxy for other outcomes, like patient outcomes; it may also be an outcome in and of itself, in that compliance, especially if non-compliance results in sanctions, can easily become an outcome.

	Artifacts-Structures	Attitudes-Beliefs	Behaviors (Culture Based)	Values	Compliance	Patient Outcomes	Student Outcomes	Accidents-Collisions
Culture - leaning-TQM-Systems								
Culture - no clear or explicit model								
Culture - Other								
Culture - Schein								

0	1-4	5-9	10-14	15+

Table 1 – Matrix Query Results: Culture by Outcomes Indicators/Measures

A second matrix query, shown in Table 2, focused on cultural theories within this literature in relation to any outcome effects identified in the study. Column one (Outcome Effects) includes those studies that identify interventions had positive effects, column two (Outcome - Mixed)

includes those studies with mixed or some combination of both positive and negative effects, or no effects were found or described for those studies in column three (Outcome-No Effect). As Table 2 indicates, the vast majority of the studies in our sample found mixed or positive effects. Within studies associated with particular cultural or other organizational theories, those that fell within "Other" node had the largest portion of studies that had mixed and especially positive outcomes. Those studies that didn't articulate or specify an explicit model of culture had the fewest number of studies with positive or mixed results. While it would require more research to identify any causal link between causal theory and program outcomes, these findings suggest the possibility that a lack of a clear cultural theory may result in the lack of a clear causal model, either to establish a culture change intervention or to establish a model for how culture impacts the performance outcomes of the organization. In the absence of a clear causal model, it may be more difficult to craft a program that effectively impacts either culture or organizational performance outcomes.

	Outcome Effects	Outcome - Mixed	Outcome - no effect
Culture - leaning- TQM- Systems			
Culture - no clear or explicit model			
Culture - Other			
Culture - Schein			

0	1-4	5-9	10-14

Table 2 – Matrix Query Results: Culture by Outcome

The final query centered on cultural theories, in this case in relation to the unit of analysis used by the studies in the sample. The results of this query indicate that a large majority of the studies

in the sample, regardless of the underlying cultural or organizational theory, focus on interventions that target the organization as a whole or a specific unit within the organization. A smaller, but still substantial number of studies focused on change industry-wide. Only a small portion of the studies focused on change primarily or solely at the level of the individual.

	Community	Individual	Industry wide	Organization	Unit-Department
Culture -leaning-TQM-Systems					
Culture - no clear or explicit model					
Culture - Other					
Culture - Schein					

0	1-4	5-9	10-14

Table 3 – Matrix Query Results: Culture by Unit of Analysis

3.2. Industries and Sectors

The second area of inquiry and analysis focused on patterns associated with the various industries or sectors of society within which the articles in the sample fell. The node tree in Figure 4 shows the seven sectors or industries from which nearly all of the studies in our sample fell. As we noted in the methods section above, the search strategy used to identify articles to be included in this analysis did seek to find traffic safety culture articles but was otherwise neutral

with respect to targeting industries or sectors. The search terms and strategy focused on organizational culture, safety culture, and culture change, regardless of setting.

By far the largest number of articles were focused on the healthcare industry and healthcare organizations. Education and transportation accounted for a substantially smaller, but still notable, portion of the studies in the sample. Studies focused on organizations in the energy industry and those that we classified as "Private Industry" saw slightly fewer studies than education and transportation. Private Industry, for the purposes of coding articles in this study, included any private sector organization or operation outside of energy or transportation. The final two sectors, Economic Development and Community Development, were created as axial codes, meaning that we anticipated that these would be common areas of focus in the literature, but that was not the case. Our sample of articles included only one article in each of these two areas.
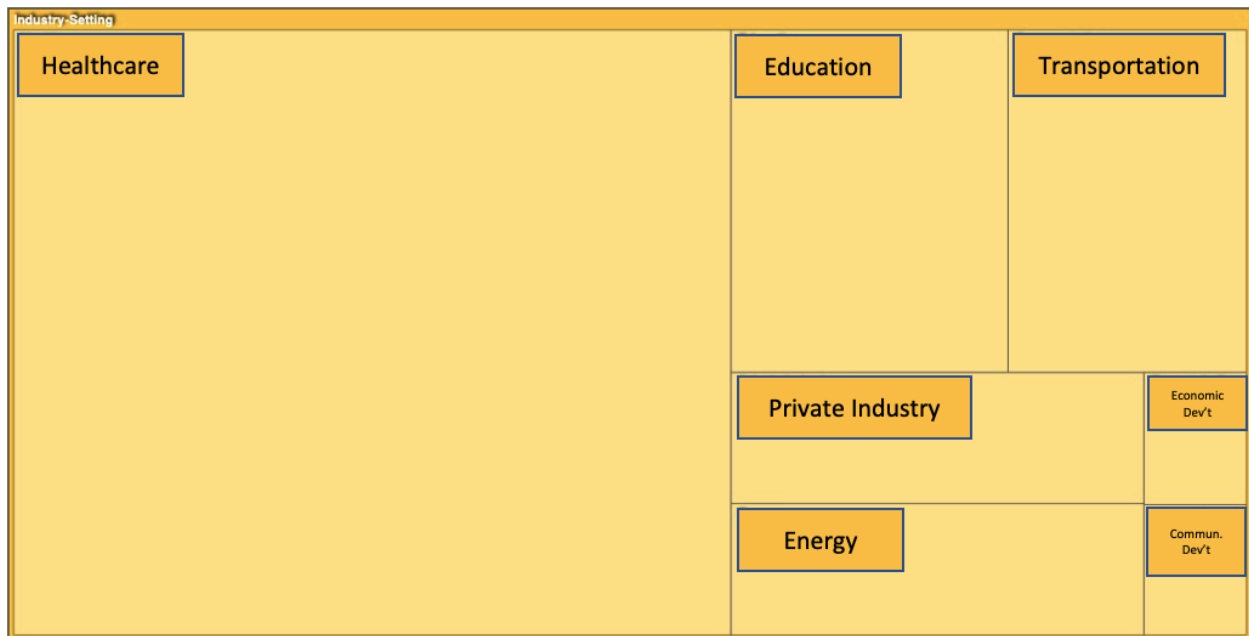
Figure 4 – Industry Node Tree Map

In addition to examining the commonality of different industries in the literature, we also looked for patterns in the relationship between the industries and settings in the sample and other variables or parent nodes in the overarching coding scheme. The first relationship we looked at was between industries or settings and the theories of culture used as a part of the studies and interventions. Table 4 shows the results of this matrix query and indicates that the majority of studies across industries have adopted and articulated some theory of culture. Interestingly, the majority of studies that were coded into the "No Clear or Explicit Model" node came from the healthcare industry. While this would appear to partly reflect the large number of studies from that industry, this result would seem to merit further investigation, particularly in line with the question posed in the previous section about whether the lack of a clear cultural theory leads or is

related to the lack of a clear causal model between change initiatives and culture or between culture or organizational performance outcomes.

	Culture - Other	Culture - Schein	Culture Westrum	Learning-Systems	No Clear or Explicit Model
Industry Healthcare					
Industry-Education					
Industry-Energy					
Industry-Private Industry					
Industry-Transportation					

0	1-4	5-9	10-14	15+

Table 4 – Matrix Query Results: Industry and Culture Model

One further analysis looked at the relationship between industry and other nodes or variables and sought to identify patterns between industry and the outcome indicators or measures used by the studies in each sector. As was true of the examination of the intersection of culture theories and indicators and measures, this analysis reveals a general pattern of culture-based attributes including attitudes, behaviors, and artifacts present across industries. There were some exceptions to this general pattern. For example, none of the studies from private industry focused on behaviors. As was noted earlier, few studies in the sample explicitly identified values as an indicator within their studies, though the analysis by industry suggests that those initiatives and studies that did identify value indicators were in the healthcare and education sectors respectively.

	Accidents-Collisions	Artifacts-Structures	Attitudes-Beliefs	Behaviors (Culture Based)	Compliance	Patient Outcomes	Student Outcomes	Values
Industry Healthcare								
Industry-Education								
Industry-Energy								
Industry-Private Industry								
Industry-Transportation								

0	1-4	5-9	10-14	15+

Table 5 – Matrix Query Results: Industry and Outcome Indicator/Measure

The final analysis conducted with industry or setting being a key focus looked at patterns in the relationship between industry and outcome effects, again coded as positive outcome effects, mixed effects, or no effects. As was the case with the results of the matrix query above looking at the relationship and patterns between culture models and outcome effects, here again we found that the majority of studies across industries found positive, or at the least, mixed outcomes resulting from the interventions evaluated in each article. There were a small number of studies from the healthcare industry that were unable to identify positive effects. Whether the presence of these few negative evaluations is a feature of a tendency toward more robust or critical analysis in the healthcare industry, a feature of the larger number of studies being more likely to have some negative outcomes, the lack of a clear cultural theory or corresponding causal model, or some other or combination of these factors is not clear from these data. However, the pattern of results suggests that culture change interventions and interventions using culture to impact other outcome variables can have positive impacts on culture and ultimately on other organizational performance or outcome measures.

	Outcome Effects	Outcome – Mixed	Outcome – no effect
Industry Healthcare			
Industry-Education			
Industry-Private Industry			
Industry-Transportation			
Industry-Energy			

0	1-4	5-9	10-14

Table 6 – Matrix Query Results: Industry and Outcome Effects

3.3. Research Designs and Implications

The last category of analysis looked at the research designs identified in the articles in our sample generally and within each of the industries or sectors in the sample. Broadly, Figure 5 indicates that, across the entire sample of articles analyzed, qualitative, quasi-experimental,[12] and single-group mixed methods designs are the most frequently used to study specific initiatives or interventions. Typically, mixed methods designs use a combination of quantitative and qualitative approaches, though in some instances studies use a mix of different quantitative designs (e.g. surveys and secondary performance data). The analysis also found a smaller but still substantial number of systematic analyses, meta-analyses, or literature reviews, all of which drew together and examined the existing literature in various ways. In addition to these designs,

---

[12] For the purpose of this study, quasi-experimental designs included only those that utilized a comparison group, and did not include time-series studies as described by Posavac (2011). By comparison, experimental designs also use a comparison group, but randomized assignment into the comparison and experimental groups, and also use a single or double-blind strategy for participants and/or researchers.

the analysis identified a small number of single-group, pretest/post-test studies and formative evaluations. Although they tend to be logistically and programmatically difficult to conduct, the analysis did identify a small number of studies self-described as experimental.

When examining the pattern of designs in relationship to industry, several observations can be made from the results of the matrix query summarized in Table 7. First, it is striking that two of the most common designs within healthcare are quasi-experimental and qualitative. While these two approaches are sometimes considered quite different in terms of their aims and even their rigor, on further consideration it seems reasonable that these distinct strategies are relatively common because they provide different forms of information about organizational performance and thereby enable healthcare organizations to present different information or make different kinds of arguments to distinct audiences. Other notable elements of this query include a larger portion of single-group, mixed methods approaches in education. Although the relative number of studies drawn from the transportation sector is relatively low, it is also notable that those studies fall into just three categories: meta-analysis, single-group mixed methods, and systematic analysis. Although distinct, if the meta-analysis studies and systematic analyses are combined based on the logic that both approaches collect, aggregate, and assess data from across existing studies, then these aggregating approaches are disproportionally common in the transportation sector. This may reflect the fact that the transportation industry has come to utilize culture-based approaches after other sectors have already done so and, as a result, turn to aggregating designs as a way of taking a broader look at what is known to more quickly assimilate relevant learnings into efforts in that industry.
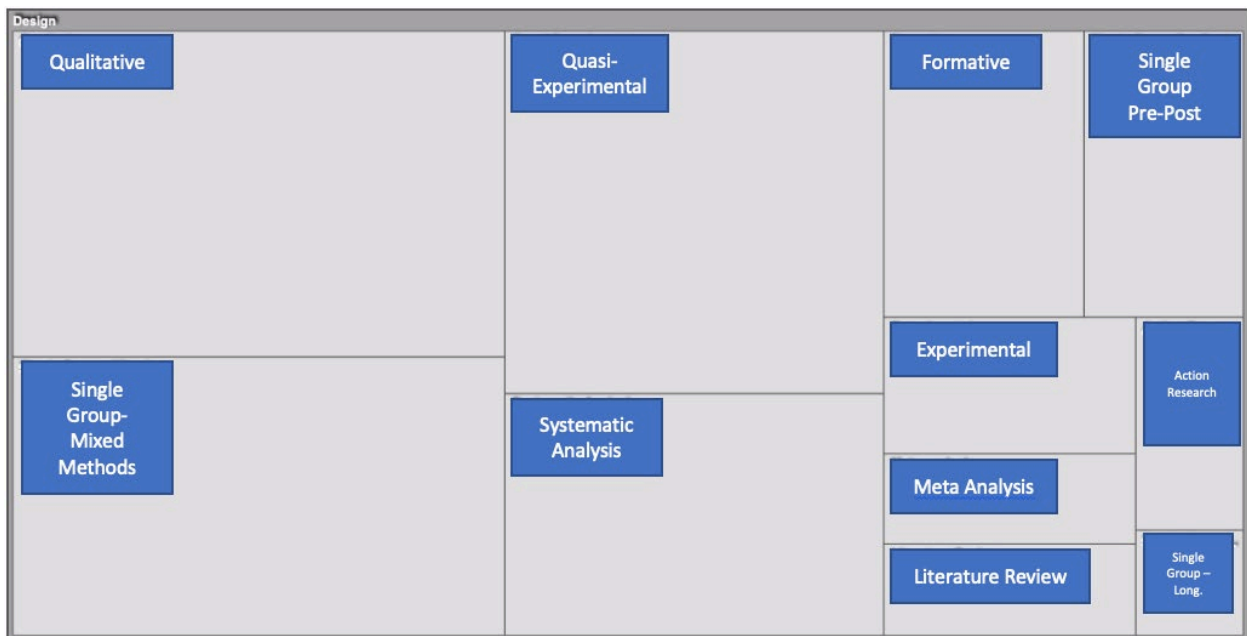
Figure 5 – Design Node Tree Map

	Industry Healthcare	Industry-Education	Industry-Energy	Industry-Private Industry	Industry-Transportation
Action Research					
Experimental					
Formative					
Literature Review					
Meta-analysis					
Qualitative					
Quasi Experimental					
Single Group - Longitudinal					
Single Group - Mixed					
Single Group Pre-Post					
Systematic Analysis					

0	1-4	5-9	10-14

Table 7 – Matrix Query Results: Industry and Research Design

4. Discussion

4.1. Study Limitations

One methodological aspect of the approach used in this study that should be noted again here is that the coding process used was primarily focused on the item – i.e., the article – as the unit of analysis rather than the number of references coded within each article. The focus on coding items rather than all references within each item allowed us to discern patterns across the body of literature that might have been obscured if we had focused on coding every reference in every article. However, this coding strategy does not result in a set of codes that are entirely mutually exclusive. For instance, when coding for the cultural theory used in each study, it is likely that a study falls into only one code category, meaning that it uses only one cultural theory to inform the intervention and evaluation. However, if a study identified two distinct theories, both would be coded and both will appear in the results of relevant queries done for the analysis. As a result, when presenting the results of various matrix queries conducted in the next section, we have chosen not to report the numbers of items generated by the query but instead have presented them in color-coded/shaded categories that represent a range of frequencies. This approach is consistent with the norms of reporting the findings of qualitative research, and because unlike cross-tabs from a quantitative design, the row and column totals from a matrix query may vary slightly from the total number of items in the sample, potentially causing confusion.

A second issue is that despite the efforts used to broaden the scope of this study to include culture interventions in other industries and towards ends beyond safety, the study is still limited to a fairly small sample of evaluations. As a result, it is difficult to draw strong conclusions about the trends and patterns, and their implications for practice, particularly within any one industry or program area. A related limitation is that the search strategy used to build the data set focused almost entirely on peer review articles. It is possible that there are other sources of information about evaluations of culture-based interventions that do not appear in the peer reviewed literature.

4.2. Implications of the Current Study

The patterns and relationships identified through the examination of this literature lead to several observations. The first is that there is wide-spread evidence across the literature indicating that initiatives to influence organizational culture or safety culture more narrowly, or to leverage culture as a means of improving other performance outcomes can be effective. This appears to be true regardless of the specific cultural theory utilized to inform the intervention. There are, however, a surprising number of studies that lack a clear definition of culture, which suggests that there may be a corresponding lack of clarity in the causal models assumed between the interventions used and anticipated culture impacts or between the cultural attributes and their impact on performance outcomes. Additional research would be necessary to disentangle these relationships. Specifically, there is a need for further research to examine how important it is to have a clear theoretical grounding and whether any intervention is likely to be as successful as the next, regardless of whether there is any clear model either articulated by the program staff or embedded by reference by program designers who draw from those ideas only to have them become obscured. Nevertheless, having a clear definition and model of culture, a

correspondingly clear causal model that links elements of a culture change intervention to culture, or one that clearly links culture to target outcome variables would seem to be a benefit to both scholars and practitioners.

Among those studies that do have clearer theoretical grounding and corresponding clarity and specificity about critical variables and the relationship between them, there is a substantial amount of consistency in the indicators and measures across studies, industries, and even designs. Cultural attributes and expressions in the form of attitudes and beliefs, behaviors, and values, as well as the degree to which these attributes are manifested in the structures and practices of the organization – its artifacts – is widespread in this empirical literature. It is unsurprising that specific industries or sectors also develop additional outcome measures relevant to operations in that sector, whether related to compliance requirements that serve as a proxy for target outcomes, or measures of the target outcomes themselves. As practitioners and researchers continue to conduct evaluations of culture-based change initiatives, it may be useful for evaluators and program managers to look beyond their industries as a way to identify new and evolving understandings of culture and its function, as well as the development of evaluation designs and research strategies.

Lastly, there is also some evidence in this literature that there is a tendency toward testing models, be they theoretical, conceptual models, or measurement and design models rather than more practical assessment and sharing of learnings about the efficacy of interventions. This may be a reflection of our sampling strategy being focused on material published in the peer-reviewed rather than professional journals. However, the study of culture-based strategies is largely an applied realm of research. Moreover, because the goal of these interventions is organizational and program performance and, in the case of safety culture, the ultimate outcome is human health and safety, keeping those ends in mind is critical for researchers, practitioners, and citizens alike.

4.3. Recommendations for Future Research

As suggested in the discussion of the study limitations above, one recommendation for further research in this area is to broaden the search for published evaluations in order to identify and examine evaluations that appear in professional publications or that are self-published. Although the diversity of these publications and lack of a centralized search tool make the collection of these evaluations more difficult, the examination of evaluations beyond those in the peer reviewed literature has the potential to substantially expand the sample size. If this broader set of sources could be systematically gathered an analyzed, they could provide more insight into the models of culture being used, indicators and measures of culture, evaluation designs used, and outcomes identified.

A second line of research that will likely prove valuable is an extended qualitative analysis of the interventions used to change an existing culture, or to leverage culture in support of improving safety behavior. Because the structure and function of culture is highly contextual, a qualitative study of culture-change interventions can help to build a deeper, more nuanced and detailed understanding of the intervention strategies used, contextual factors at play, and the impacts of these various factors. This line of research will require shifting the unit of analysis from the

article or report to the reference, in order to better identify and map themes and patterns related to the interventions and their impacts.

Finally, as the number of evaluations of culture-based interventions in traffic safety grows, it will be important to conduct further systematic, or even meta-analyses of these efforts. This will be particularly true for interventions that strive to change the culture of communities of road users rather than organizations.

Acknowledgements

References

Brescianai, M. J. (2011). Identifying Barriers in Implementing Outcomes-Based Assessments Program Review: A Grounded Theory Analysis. *Research and Practice in Assessment*, *6*, 5–16.

Clarke, S. (1999). Perceptions of organisational safety: Implications for the development of safety culture. *Journal of Organizational Behavior*, *20*, 185–198.

Cox, S. J., & Cox, T. (1996). *Safety, Systems and People*. Oxford.

Cox, S. J., & Flin, R. (1998). Safety culture: Philosopher's stone or man of straw? *Work and Stress*, *12*, 189–201.

Edwards, J. R. D., Davey, J., & Armstrong, K. (2013). Returning to the roots of culture: A review and re-conceptualization of safety culture. *Safety Science*, *55*(June), 70–80. https://doi.org/10.1016/j.ssci.2013.01.004

Edwards, J. R. D., Freeman, J., Soole, D., & Watson, B. (2014). A framework for conceptualizing traffic safety culture. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 317–325.

Geertz, C. (1973). *The Interpretation of Culture*. Basic Books.

Government Accountability Office. (2013). *Strategies to Facilitate Agencies' Use of Evaluation in Program Management and Policy Making* (GAO-13-570; Program Evaluation, p. 39). General Accountability Office. https://www.gao.gov/assets/660/655518.pdf

Gulick, L., & Urwick, L. (1937). *Notes on the theory of organization*. Columbia University, Institute of Public Administration.

Hale, A. R. (2000). Culture's Confusions. *Safety Science*, *34*(1–3), 1–14.

Ham-Baloyi, W. ten, & Jordan, P. (2016). Systematic review as a research method in post-graduate nursing education. *Health SA Gesonheid*, *21*, 120–128. https://doi.org/10.1016/j.hsag.2015.08.002

Havold, J. I., & Nesset, E. (2009). From safety culture to safety orientation: Validation and simplification of a safety orientation scale using a sample of seafarers working for Norwegian ship owners. *Safety Science*, *47*(3), 305–326.

Hedlund, J. (2007). *Improving Traffic Safety Culture in the United States: The Journey Forward Summary and synthesis*. AAA, Foundation for Traffic Safety.

Hoekstra, T., & Wegman, F. (2011). Improving the effectiveness of road safety campaigns: Current and new practices. *IATSS Research*, *34*(2), 80–86. https://doi.org/10.1016/j.iatssr.2011.01.003

Holosko, M. J. (1996). Obstacles to Conducting Program Evaluations: Implications for Public Health Evaluators. *Journal of Health & Social Policy*, *8*(1), 91–101. https://doi.org/10.1300/J045v08n01_08

Kapucu, N., Hu, Q., & Khosa, S. (2017). The State of Network Research in Public Administration. *Administration & Society*, *48*(8), 1087–1120.

Luria, G., & Rafaeli, A. (2008). Testing safety commitment in organizations through interpretations of safety artifacts. *Journal of Safety Research*, *39*, 519–528. https://doi.org/doi:10.1016/j.jsr.2008.08.004

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*, *7*(7). https://doi.org/10.1371/journal.pmed1000097

Nieva, V. F. (2003). Safety culture assessment: A tool for improving patient safety in healthcare organizations. *Quality and Safety in Health Care*, *12*(90002), 17ii–1723. https://doi.org/10.1136/qhc.12.suppl_2.ii17

Otto, J., Ward, N. J., & Finley, K. (2019). Guidance for the Measurement and Analysis of Traffic Safety Culture. In *Traffic Safety Culture: Definition, foundation, and application* (pp. 65–91). Emerald Group Publishing Limited.

Posavac, E. J. (2011). *Program Evaluation: Methods and Case Studies* (8th ed.). Prentice Hall.

Reason, J. (2000). Safety paradoxes and safety culture. *Injury Control Safety Promotion*, *7*, 3–14. https://doi.org/10.1076/1566-0974(200003)7:1;1-V;FT003

Robertson, R. D., & Pashley, C. (2015). *Road safety campaigns: What the research tells us*. Canadian Automobile Association, Traffic Injury Research Foundation of Canada. http://epe.lac-bac.gc.ca/100/200/300/traffic_injury_research/road_safety_campaigns/2015_RoadSafetyCampaigns_Report_2.pdf

Schein, E. H. (2004). *Organizational Culture and Leadership* (3rd ed.). Jossey-Bass.

Simon, H. A. (1976). *Administrative behavior: A study of decision-making processes in administrative organization*. Free Press.

Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. SAGE.

Turnbull, K. (2011). *Improving Roadway Safety Programs Through University–Agency Partnerships*. Keck Center of the National Academies.

Ward, N. J., Otto, J., & Finley, K. (2019). Ten Principles of Traffic Safety Culture. In *Traffic Safety Culture: Definition, Foundation and Application* (pp. 21–40). Emerald Group Publishing Limited.

Ward, N. J., & Ozkan, T. (2014). In consideration of traffic safety culture. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 291–292. http://dx.doi.org/10.1016/j.trf.2014.09.004

APPENDIX III – SUMMARY GUIDANCE

Guidance for Evaluating Traffic Safety Culture Strategies

"Traffic Safety Culture is defined as the shared belief system of a group of people, which influences road user behaviors and stakeholder actions that impact traffic safety."[1]

The purpose of this document is to provide guidance to traffic safety practitioners about evaluating traffic safety culture strategies.

Center for Health and Safety Culture
Montana State University
P.O. Box 170548
Bozeman, MT 59717-0548
Phone: 406-994-7873
Fax: 406-994-1697
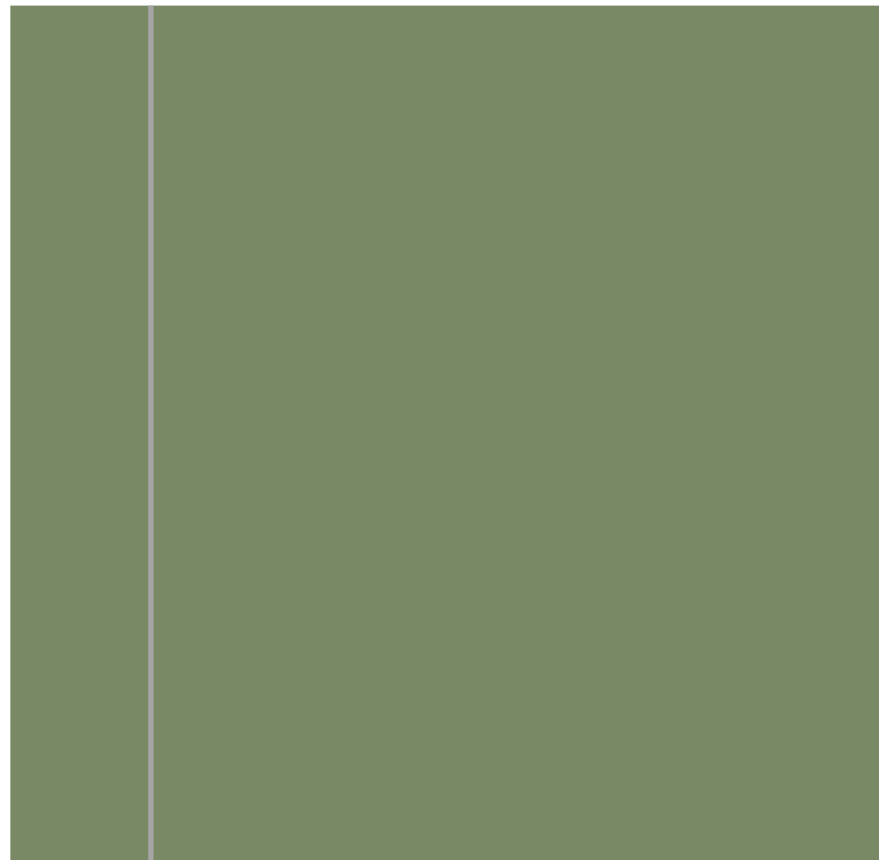www.CHSCulture.org

---

Contents

BACKGROUND RESEARCH ON EVALUATING TRAFFIC SAFETY CULTURE STRATEGY

This guidance document was developed as a component of a project funded by the Partnership for the Transformation of Traffic Safety Culture Transportation Pooled Fund Program lead by the Montana Department of Transportation. An article entitled "Assessing the Impact of Culture: A Systematic Analysis of Culture Interventions and Evaluations in Different Organizational Settings" that was written for this project established the context for this guidance. The following is the abstract of that article.

Over the last twenty years, transportation agencies have increasingly added culture-based approaches to the existing education, engineering, and enforcement strategies being used as a means of reducing traffic related injuries and fatalities. Despite this increased interest, there have been comparatively few evaluations of interventions designed to enhance traffic safety culture. At the same time, many other organization types have adopted culture-based strategies either to improve safety or to enhance other elements of organizational performance. In aggregate, the evaluations of culture-focused interventions across a range of settings offer an untapped body of information about the models of culture being leveraged to affect change, the intervention strategies used to impact culture, the impacts of these strategies, and more. This article presents the results of a systematic analysis of evaluations of culture-focused interventions across a variety of settings and seeks to identify patterns that could be useful to both researchers and practitioners. The findings of the study suggest that there are areas of substantial consensus regarding the nature and features of culture and the potential effectiveness of culture-based programs. At the same time, the findings also suggest that more conceptual and empirical work is warranted to further refine our understanding of culture and its functions and to build deeper understanding of how to leverage culture effectively to support health and safety efforts.
[see https://www.mdt.mt.gov/research/projects/trafficsafety-strategies.shtml]

Overview

The purpose of this document is to provide guidance to traffic safety practitioners about evaluating traffic safety culture strategies. It begins with a description of traffic safety culture strategies and is followed by a summary of evaluation types, components of effective evaluations, and steps to follow to complete an evaluation. It concludes with an evaluation example from an actual project to improve traffic safety culture.

Evaluation is a large, diverse field of scientific research. Clearly, this guidance document cannot cover all there is to know about evaluation. However, it can promote an idea called "evaluative thinking."[2] Evaluative thinking is a problem-solving approach to designing, selecting, and allocating resources to traffic safety strategies. It seeks credible evidence to provide answers about the effectiveness and sustainability of traffic safety strategies.

> *Evaluative thinking is a cognitive process in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of the evidence, that involves skills such as identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective-taking and making informed decisions in preparation for action.[3]*

This guide can help traffic safety practitioners bolster their knowledge about evaluation and include evaluation in their proposal requests and activities – in other words, promote evaluative thinking. This guide is not intended to teach practitioners how to conduct extensive evaluations themselves. Instead, it provides guidance so that more evaluation activities are included in efforts to grow positive traffic safety culture – thereby improving effectiveness of these strategies.

After studying this guide, traffic safety practitioners will be able to:

    A.   Discuss the importance of evaluating traffic safety culture strategies

    B.   Understand types of evaluations and components of effective evaluations

    C.   Ask appropriate questions of evaluation proposals to select effective evaluations

    D.   Better understand and make meaning of completed evaluations

PROMOTING EVALUATIVE THINKING

Many traffic safety practitioners and stakeholders already engage in forms of evaluative thinking. Discussing the value of evaluative thinking within the traffic safety community will grow its importance. Here are talking points to foster conversations about the importance of evaluative thinking.

1. Fatal crashes and serious injuries have a significant impact on public health.

2. Zero traffic fatalities and serious injuries is the only acceptable goal.

3. To be successful in reaching this goal, we must learn to use innovative strategies and grow evidence of their effectiveness.

4. Evaluations inform which strategies are effective and generate knowledge about how to make strategies more effective and sustainable.

5. Traffic safety practitioners can seek opportunities to include process, outcome, and impact evaluations in the projects they implement, manage, and fund.

6. Effective evaluations require quality data and appropriate comparisons.

7. Evaluations should include engaging stakeholders, developing careful descriptions of strategies, and identifying quality data and appropriate comparisons.

8. Traffic safety practitioners can create opportunities to review and discuss evaluation results with stakeholders to gather lessons learned and identify opportunities for improvement in future efforts.

9. More consistent and rigorous evaluations will accelerate learning and effectiveness of strategies in improving traffic safety.

10. Investing in training to help staff become more familiar with evaluation design and contracting with evaluators will improve the effectiveness of strategies and ultimately traffic safety.
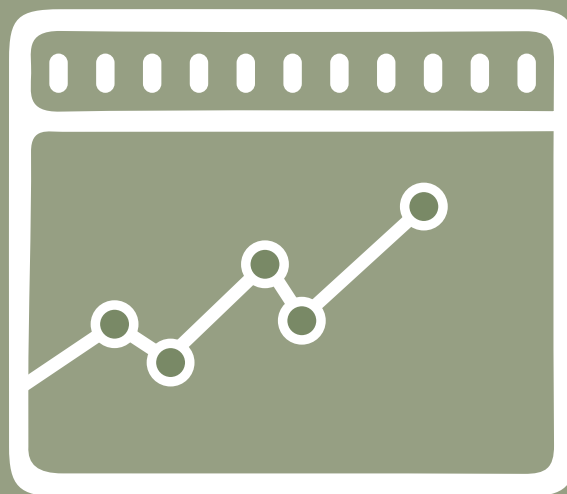
Traffic safety practitioners invest significant time and resources in strategies to improve traffic safety. Everyone wants these resources to be invested in strategies that actually make a difference. Evaluation can ensure these investments are effective.

According to the American Evaluation Association, "evaluation involves assessing the strengths and weaknesses of strategies "to improve their effectiveness."[4] By understanding the strengths and weaknesses of a strategy, those implementing the strategy (including policies and laws) can make adjustments to make the strategy more effective.

Information about if and how a strategy works provides important evidence. This evidence becomes the basis for considering a strategy as "evidence-based." Evidence is critical to making good decisions (e.g., evidence-based decision making).

The Centers for Disease Control and Prevention (CDC) lists several important reasons for evaluating strategies:[5]

- To assess effectiveness and inform good management practices by
  - comparing actual outcomes with intended outcomes,
  - comparing outcomes with those of previous years, and
  - establishing realistic intended outcomes (standards) for future performance.
- To foster sustained improvements in traffic safety by
  - focusing attention on issues important to the effectiveness of the strategy,
  - promoting a strategy by documenting and sharing its effectiveness,
  - recruiting new partners (who want to join in contributing to effective strategies),
  - enhancing the image of the strategy,
  - sustaining or increasing funding,
  - providing direction and informing training for staff and partners to implement the strategy effectively in the future,
  - informing what training and technical assistance is needed to improve effectiveness,
  - informing long-range planning, and
  - justifying the investment of resources by legislators or other stakeholders by showing the strategy is effective.
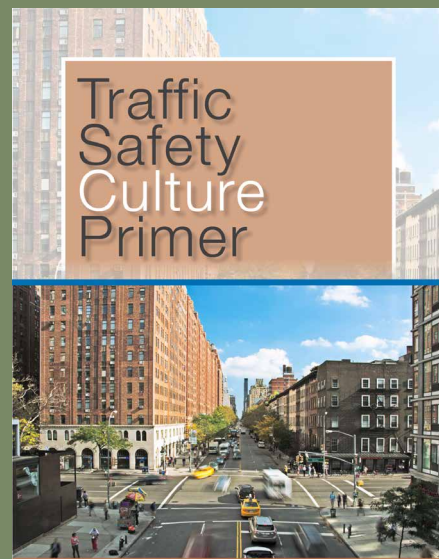
Evaluating Traffic Safety Culture Strategies

Traffic safety culture can be defined as the beliefs shared by a group of road users or stakeholders that influence their behaviors that impact traffic safety. This definition of culture establishes a relationship between beliefs and behaviors. Many of our beliefs come from the culture of the groups we belong to such as family, school, workplace, or community. Specifically, when individuals have certain beliefs, they are more likely to engage in certain behaviors. For example, if people believe that it is safe to have hands-free cell phone conversations while driving, they are more likely to engage in this risky driving behavior.

Understanding how a traffic safety culture strategy leads to improving traffic safety is important when designing an evaluation. Such evaluations need to include evidence not only about changes in behavior, but also changes in beliefs that support those behaviors and the outcomes that result from those behaviors.

LEARNING MORE ABOUT TRAFFIC SAFETY CULTURE

Learn more about traffic safety culture by reading the Traffic Safety Culture Primer or Google "traffic safety culture primer."

Traffic Safety Culture Primer

There are several basic types of beliefs including:

- Expectations about the consequences of behavior (e.g., "If I drive after using cannabis, I am more likely to cause a crash.")

- Perceptions about how common a behavior is (e.g., "I believe most people speed.")

- Perceptions about how acceptable or expected a behavior is (e.g., "My spouse expects me to use a seat belt.")

- Perceptions about an individual's ability to perform the behavior (e.g., "I am comfortable not answering my cell phone while driving.")

Traffic safety culture strategies focus on changing beliefs like these. When these beliefs change, people's behaviors are likely to change, and this change in behavior is more likely to be sustained.

In contrast, a speed bump is a physical way of changing behavior. People tend not to speed over speed bumps but will resume their speed when the speed bumps are no longer present. A speed bump does not change underlying beliefs about speeding and therefore does not result in sustained behavior change.

Traffic safety culture strategies use specific experiences designed to change beliefs. For example, workplace traffic safety training is a specific experience designed to change a worker's beliefs about specific driving practices. The training might discuss the increased risk for crashing while talking on a cell phone when driving. The training could provide information about how most employees do not drive while using a cell phone and that leadership, management, and supervisors expect drivers not to use their cell phones while driving. A workplace policy prohibiting cell phone use while driving could be reviewed. Supervisors could meet with employees and discuss how work procedures will take place without using cell phones while driving. By growing healthy beliefs among workers, the likelihood of risky driving is reduced. As fewer drivers engage in risky driving (e.g., distracted driving), fewer crashes will occur, and traffic safety will improve. This process is summarized in Figure 1.

A TSC Strategy is Implemented (e.g., training) → Beliefs Change → Behavior Changes → Traffic Safety Improves (e.g., fewer crashes)

Figure 1. Diagram of how a traffic safety culture (TSC) strategy leads to improved traffic safety

Understanding how a traffic safety culture strategy leads to improving traffic safety is important when considering evaluating a traffic safety culture strategy. There are many potential problems that could result in a traffic safety culture strategy being ineffective.

Using the same workplace training example shared previously, imagine what would happen if only 10% of the workers were trained. Only training 10% of the workers would significantly reduce the likelihood that beliefs across the workforce would change, thus reducing the likelihood that behaviors across the workforce would change, thus reducing the likelihood that crashes would be reduced.

Suppose everyone participated in the training, but the training was poorly implemented and did not change people's beliefs. If beliefs did not change, it would be unlikely that behaviors would change, and traffic safety would not improve.

Suppose everyone participated in the training, and the training changed beliefs, but it changed the wrong beliefs – beliefs that did not matter or did not influence the behavior. Behavior would not change, and traffic safety would not improve.

Understanding how a traffic safety culture strategy leads to improving traffic safety will inform how a traffic safety culture strategy should be evaluated. Specifically, the evaluation should verify the process of change underlying the strategy. For example, an evaluation might capture what percentage of the workers participated in the training, to what degree the training changed beliefs, how much subsequent risky driving behaviors changed, and whether crashes were reduced. This simple example shows that there may be several ways to evaluate a traffic safety culture strategy.
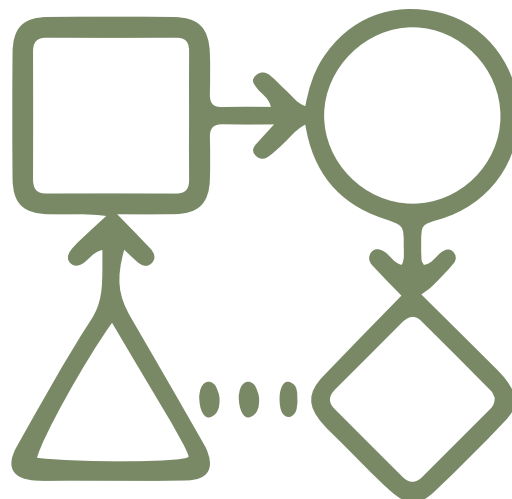
Evaluation Types

Because there are different factors impacting the effectiveness of a strategy, there are different types of evaluation. Each evaluation type provides important information to make a strategy more effective. The CDC summarizes three types of evaluation:[4]

- **Process evaluations** examine the way the strategy was implemented. This represents the first box in Figure 1.
  - Was the strategy (e.g., workplace training) implemented exactly how it was designed? This is also referred to as implementing a strategy with fidelity. For training, this might include assessing the number of sessions the training required, how many sessions were completed, how much of the content was covered, etc.
  - Did the strategy reach a sufficient portion of the population (e.g., percentage of workers) to make a difference?
- **Outcome evaluations** determine if the change the strategy was designed to make occurred. This represents the middle two boxes in Figure 1. An outcome evaluation could assess to what degree the strategy (e.g., workplace training) changed beliefs. An outcome evaluation could also assess to what degree there was a change in behaviors.
- **Impact evaluations** assess the consequences of changes that result in improved public health. This form of evaluation focuses on progress toward a traffic safety goal (such as few fatal crashes), which is represented by the final box in Figure 1. For example, a impact evaluation could assess to what degree there was a reduction in distracted driving related crashes (and injuries) in a workplace.

All three types of evaluation require the gathering of data and then using those data to make comparisons and draw conclusions. Therefore, effective evaluations require using quality data and making appropriate comparisons.
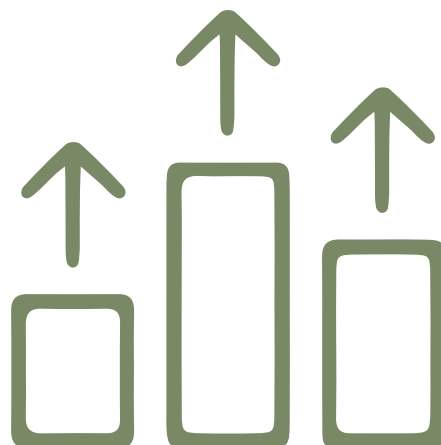
Key Components of Effective Evaluations

The quality of the information (or evidence) provided by an evaluation depends on two key components of the evaluation design: the quality of the data and the way data are compared.

Data Quality

Information about the process, outcomes, and impacts of strategies is assessed from many kinds of data gathered from different sources. These data become the basis for drawing conclusions. Therefore, the quality of the data determines the quality of the conclusions made based on that data.

There are two important aspects of data quality.

1. **Data must be reliable – the data are accurate and have consistent measures.**

    For example, suppose we ask two participants in a workplace training how many people were present at the training. One person says 4, and the other says 10. This is not a reliable measure of how many people attended the training. Another data source needs to be used to provide a reliable measure of how many people attended the training (perhaps a sign-in sheet).

    Reliability can also be compromised on measures that may change over time. For example, suppose we want to know how people feel, on average, over the period of a week. One way to measure this might be to ask people once how they feel "right now." Another way might be to ask people several times a day over several days and average these responses. The second method will mostly likely create a more reliable because people's feelings can fluctuate during a week so asking them at only one point of time may measure an unusual feeling that is not representative of how they usually felt during the week.

2. **Data must be valid – the data truly represent the concepts that are being measured.**

For example, suppose we want to assess whether a strategy changed beliefs about distracted driving. A survey question might ask, "Do you think driving distracted is dangerous - yes or no?" The person might not know what "driving distracted" means, or they might want to answer "sometimes," but that is not an option. The results from asking this question may not be a valid indicator of people's beliefs about distracted driving. A better question might be, "How dangerous is driving while having a conversation on a hands-free cell phone?" with five choices ranging from "not at all dangerous" to "extremely dangerous."

Another challenge is when data do not represent what we think they represent. For example, we compare the number of distracted driving citations between two communities and draw the conclusion that one community has more distracted driving than another community. The data indicate how many citations were written – which may or may not reflect the prevalence of distracted driving in the community. In fact, the number of citations written may be a better indicator of enforcement activity in each community. In this case, the number of citations is not a valid measure of distracted driving.

To help illustrate concerns about reliability and validity of data, Table 1 shows examples of data that could be measured in different types of evaluation.

Table 1. Examples of Reliability and Validity Concerns with Possible Measures Used in Different Types of Evaluation

Evaluation Type	Source and Method for Gathering Data	Concerns About Reliability and Validity	Recommendations
Process			
Was the strategy implemented as designed?	Self-reports by those implementing the strategy indicating if they implemented it as designed	Those implementing the strategy may not know what is required during implementation to assure effectiveness. They may also have a biased opinion regarding the quality of implementation.	Develop checklists to assess implementation and, if possible, use observers to complete the checklists. Require documentation of activities (e.g., publication affidavits for placed media, number of sessions conducted with attendance records, etc.)
Did the strategy reach the intended audience?	Information provided by those implementing the strategy	While those implementing the strategy can perceive they reached most people, there could be gaps.	Collect information directly from people who participated in the strategy. This could include questions on annual workplace surveys (or personnel evaluations) about attendance at safety training or random sample surveys of larger populations asking how often they heard a message or received information.
Outcome			
Did beliefs change as expected by the strategy?	Interviewers use focus groups to assess people's beliefs	The interviewers may have different interviewing styles, which affects responses.

The presence of the interviewer and other participants may create social pressure for respondents to provide only desirable answers. | Use tested or standardized self-report surveys that ensure respondent anonymity. |
| Did behavior change as expected by the change in beliefs? | Citation or arrest data

Data reported by others without using strong methodologies (like workplace supervisors reporting prevalence and frequency of behaviors among direct reports) | Citation or arrest data typically assess law enforcement activity and may not be an indication of underlying behaviors.

People's perceptions of other people's behaviors are typically very inaccurate. | Use tested or standardized self-report surveys that ensure respondent anonymity.

Use observational studies with well-designed methodologies using trained observers. |
| **Impact** | | | |
| Did crashes related to the behavior decrease?

Did fatalities and serious injuries decrease? | Crash databases | Limitations and inconsistencies in how crash data are collected and reported may limit validity and reliability. For example, assessing distraction after a crash may be problematic resulting in underreporting of distraction-related crashes.

Furthermore, assessing impairment from drugs may be limited by lack of valid biological tests. | Understand the limitations of crash databases and use extra care when comparing results across systems that may use different methods. |

Data Comparisons

Once reliable and valid data are collected, comparisons are used to make meaning of the data. There are at least four ways to make comparisons.

- **Benchmark-based evaluations** compare data with a stated reference. These references may be specified by stakeholders or based on previous implementation of the strategy. Examples include:
  - 80% of the employees will agree that not using a seat belt violates company policy.
  - Less than 10% of the population will report driving within two hours of consuming alcohol.
  - There will be fewer than 35 speed-related crash fatalities this year.

  Because benchmark-based evaluations compare a data measure with a set value, this type of comparison does not measure change. Therefore, it is unclear if the strategy resulted in meeting the benchmark or something else (or even if the benchmark was met prior to implementing the strategy). Benchmark-based evaluations cannot claim the strategy caused the change in the outcome or impact.

- **Time-based evaluations** compare data across different time points. The time points could be at the beginning and end of implementing the strategy or could be on a regular basis (like every year). The data between these time points are then compared to assess change. Examples include:
  - Beliefs about impaired driving are compared at the beginning of the first session and at the last session of a class for individuals cited for repeatedly driving under the influence of alcohol.
  - Seat belt use (as measured by observational studies) is compared to previous years.
  - Crash statistics from a specific area over the summer months (i.e., May to September) are compared from year to year.

  It is critical to acknowledge that many other factors besides the strategy can affect outcomes. Time-based evaluations cannot claim the strategy was the cause of the change (e.g., a change in driving behaviors could be due to the economy). Similarly, a lack of change may not necessarily be the consequence of an ineffective strategy; other factors may have changed such as a significant change in the population and age of drivers.

12

- **Place-based comparisons** compare data across different locations, usually within the same time period. To isolate the effect of the strategy, one location (test site) is the place where the strategy is implemented. The other place is similar but does not have the strategy implemented (control site). For example:

  - One county in a state implements a new strategy (test site) while another county with similar characteristics (e.g., road type, population, etc.) does not implement the strategy (control site). Outcome measures (e.g., observed seat belt use or crash data) for the same time period are compared.

  If the two sites are different <u>only</u> in terms of the implemented strategy, then any measured differences between these sites might be attributable to the strategy itself. However, this assertion depends on how comprehensively these sites were matched. It is very difficult to find two sites that match perfectly.

- **Combined time-place evaluations** use a combination of the time-based and place-based evaluation methods by comparing different places at two points of time. The "before" and "after" measures for each place are compared. The changes assessed in each place are then compared against each other.

  - One county in a state implements a new strategy (test site) while another county with similar characteristics (e.g., road type, population, etc.) does not implement the strategy (control site). Several measures (beliefs, behaviors, crash data) are collected in the same way before and after the strategy is implemented. In the "control site" county, none of the measures show any statistically significant changes (as expected, since the strategy was not implemented in this county). There are changes in the measures in the "test site" county.

  This evaluation design has the advantage of permitting multiple comparisons, which can reinforce conclusions about whether the strategy caused the change in measures (this is called "causality"). Notably, if the test site had similar speeding compared to the control site before the strategy was implemented AND speeding reduced at the test site but no changes were measured over the same time period at the control site after the strategy was implemented, we can be more confident in claiming that the strategy caused the change.

  It is important to note that other factors could still explain the change. For example, assume that during the implementation period, a large number of young adult males leave the test county (e.g., for work, join the military, or go to university). Such a change in the population could also cause a change in speeding behaviors in the test site – a change that was not caused by the strategy.

Steps to Plan, Implement, and Make Meaning of an Evaluation

The CDC promotes five core steps for implementing evaluations , which have been expanded upon here:[4]

1. **Identify, Recruit, and Engage Stakeholders.** Stakeholders include people responsible for the strategy (e.g., funders, contractors, etc.), people affected by the strategy (e.g., general population, workplaces, etc.), and those who will use the evaluation results. Early participation by stakeholders is necessary to identify questions and concerns and support access to quality data to ensure an effective evaluation. Those affected by the strategy should be included to measure exposure to the strategy and help identify unintended consequences including potential harms.

> A key purpose of stakeholder involvement is to specify "standards" for effectiveness. What does an effective evaluation mean for this strategy in this context? How does each stakeholder define and envision success? What outcomes are important to the needs of each stakeholder? Should the evaluation bolster a sense that the strategy caused the change in outcomes or is it OK just to assess change? It is important to understand these distinct perspectives to align expectations about potential interpretations of the evaluation results.

2. **Describe the Strategy.** Before starting an evaluation, it is necessary to agree on a detailed description of the strategy, including the conditions necessary for its implementation: "a comprehensive [strategy] description clarifies all the components and intended outcomes of the [strategy], thus helping you focus your evaluation on the most central and important questions."[4] Understanding how the strategy causes a change in outcomes (and subsequent positive impact to traffic safety) is critical to designing an evaluation. This understanding will inform potential process measures (e.g., how many people experienced the strategy), intermediate outcome measures (e.g., which beliefs and behaviors to examine for change), and impact measures (e.g., crash types) as well as provide insights as to how much time the strategy will take to cause changes. Practitioners can reach out to the strategy developer and ask for the "theory of change" for the strategy (the theory of change lays out the science behind how a strategy has been shown to cause the expected outcomes). Additionally, a practitioner could require a contractor implementing a strategy to articulate how the strategy causes the expected outcomes.

14

3. **Identify Data Measures and Comparisons to Be Performed.** In this step, the stakeholders identify the reliable and valid data that measure the process, outcome, and impact of the strategy. The sources and methods to collect these data are also identified. A comprehensive plan is developed for the evaluation, which includes the type(s) of planned comparison(s). As discussed previously, carefully consider the type of comparison, because it affects the ability to draw conclusions about strategy.

4. **Make Meaning.** This step involves analyzing the data and interpreting the results. Considerations addressed in the first step can inform efforts in this step. Too often, evaluations are reduced to one simple question: "Did the strategy work?" Often the answer is: "Yes and no." Making meaning of the evaluation should allow for greater learning to inform how to make the strategy more effective. Additional questions to ask include:

   a. Was the strategy implemented as intended? Why or why not? What could be done better next time?

   b. Did the strategy reach the intended audience? Why or why not? What could be done better next time?

   c. Did the strategy result in the intermediate outcomes (e.g., in beliefs and behaviors) expected? Why or why not? What could be done better next time?

   d. Did the strategy result in the desired impact? Why or why not? What could be done better next time?

   The evaluation may not inform all these questions. Asking these questions may guide how future evaluations may be modified to answer additional questions. The intent is to use the evaluation results to improve effectiveness over time by enhancing learning.

5. **Accumulate and Share Wisdom (e.g., lessons learned)**. A single evaluation, if explored and discussed by stakeholders, can generate many lessons that can inform future actions. These lessons are often much more valuable than simply answering the question "Did the strategy work?" Stakeholders should allocate time to review and discuss the evaluation results and gather lessons to share with other stakeholders.

   An evaluation can have greater impact if the lessons learned reach a variety of audiences that need the information to make decisions about strategies, planning, funding, etc. To be accessible and usable, lessons should use language familiar to all stakeholders.

   It is also important to accumulate lessons learned and evidence for a strategy over time, because a single evaluation may not be enough to truly understand how best to implement a strategy or to convince stakeholders to continue support for the strategy.

Evaluation Example: A Case Study[6]

This section presents a case study of an evaluation of a project completed by the Center for Health and Safety Culture for the Idaho Transportation Department (ITD) to decrease alcohol-impaired driving by encouraging people to intervene and prevent others from driving when impaired. The case study reviews each of the five steps described previously.

Background

In a previous research project, the Center for Health and Safety Culture ("the Center") identified beliefs associated with bystanders speaking up to prevent others from driving after drinking. This research identified potential messages that could be used in a media campaign to increase bystander engagement. The purpose of the project described in this case study was to test these messages using a universal media campaign, engage local stakeholders to use the media to reduce impaired driving, and evaluate the strategy's impact on beliefs and behaviors about bystander engagement as a way to reduce alcohol-related crashes.

Step 1. Identify, Recruit, and Engage Stakeholders

The initial stakeholders included leaders within the Idaho Transportation Department (including the Office of Highway Safety) and the researchers from the Center working on the project. These stakeholders met regularly to discuss the project design that would best meet the purpose of the project. The group decided to engage three communities to implement the strategy and use the remainder of the state as a comparison.

The three communities identified to implement the strategy were Blackfoot, Lewiston, and Twin Falls. These communities were selected because of their geographic distribution across the state, diversity of size, and their high rates of alcohol-impaired driving crashes.

Stakeholders from the three communities were identified and recruited to participate in initial training about the project. Twenty-one individuals from the three communities participated in a two-day training that reviewed the background for the strategy, how the strategy would be evaluated, and potential ways they could support the strategy.

Step 2. Describe the Strategy

The strategy was a media campaign to be augmented with additional supportive materials that could be used by local stakeholders in each of the three communities. The messages for the media campaign were designed to grow specific beliefs associated with bystander engagement including:

- "Most Idaho adults do not drink and drive."

- "Most Idaho adults agree they should try and prevent a stranger from driving after drinking."

- "Most Idaho adults agree they would try and prevent a stranger from driving after drinking."

- "Most Idaho adults agree with strongly enforcing impaired driving laws."

Additionally, media was created to demonstrate what it looked like to actually speak up to prevent impaired driving. These examples were captured in short video messages designed for placement on television. The media campaign was branded "Courageous Voices Create Safe Roads." Media including television and radio ads were developed using this brand and placed in these three communities from late 2013 to late 2014. Supportive materials including a brochure, speaking points, sample presentation, press releases, and a website landing page were also created. A media buyer was contracted to place the media to reach the three communities. The media buyer worked with stakeholders from ITD and the Center to develop a media plan for the media placement.

Step 3. Identify Data Measures and Comparisons To Be Performed

Process measures to assess the placement of the media included

- affidavits from the media buyer on exact placement locations, times, etc.;

- earned media placements in local newspapers (letters to the editor, etc.); and

- distribution of the supportive materials.

Outcome measures to assess changes in beliefs and behaviors included survey responses by adults. Paper surveys were mailed to a random sample of households in each of the three pilot communities as well as across the rest of the state before and after the media campaign. The responses in each sample were compared to reveal change. The three communities acted as the test sites, and the rest of the state acted as the control site.

Alcohol-related crashes in the three communities and the state before and after the campaign were used as a measure of impact.

17

Combined time-place comparisons were made to assess change. The responses in the three test sites were aggregated together. The means measuring each of the beliefs and behaviors of the responses before and after the media campaign were compared. Similar comparisons were made for the sample representing the control site (i.e., the rest of the state).

Step 4. Make Meaning

Process measures indicated that the media were placed as planned by the media buyer. The affidavits showed placements on radio stations, television stations, newspapers, and billboards. However, there was no earned media generated (no letters to the editor, articles written, etc.), and none of the supportive material created was ever distributed in bars or restaurants that serve alcohol.

Outcome measures included responses to surveys. Comparisons of survey responses before and after the media placement showed statistically significant improvements in beliefs addressed in the media messages in the test sites. Specifically, agreement with the belief that most adults think people should try to prevent a stranger from driving after drinking enough alcohol to be impaired and agreement with the statement that "I should try to prevent a stranger..." statistically significantly increased ($p<0.001$ and $p=0.008$, respectively).

Furthermore, the perception that most people would support individuals who chose to prevent a stranger from driving after drinking too much increased ($p<0.001$) as did the perception that most people would try to intervene ($p<0.001$). Other related beliefs increased as well.

Beliefs not addressed in the media messages showed no changes. No changes were seen in responses outside of the three test sites (i.e., the control site) supporting the notion that the media messages caused the changes measured in the test sites.

The surveys revealed no changes (in either the test sites or the control site) in self-reported behaviors about intervening to try and prevent a stranger from driving after having too much to drink, calling 911 to report a potentially impaired driver, or driving within two hours of drinking.

Because these behaviors are rare (most people do not drive after having too much to drink, therefore, few people can intervene), measuring changes in these behaviors can be challenging with small survey samples.

Impact measures included crash data. Crash reports indicated a slight reduction in alcohol-related crashes during the year of the campaign. However, the alcohol-related crash reduction in the pilot communities occurred at a rate similar to the reduction at the state level. Thus, the messages did not appear to have reduced alcohol-related crashes.

Step 5. Accumulate and Share Wisdom

The process measures revealed that there was not enough stakeholder engagement within the three communities to use the supportive media created. There was no known engagement (such as using the supporting media materials, working to change local practices or policies, or engaging specific groups such as schools or community groups) by local stakeholders to support the media strategy in the test sites after participating in the training. An important lesson learned was just how much effort is required to encourage local stakeholders to engage in media strategies.

The outcome measures (based on analyses of the surveys in the test sites and control site) indicated the media strategy changed the targeted beliefs even with such a short implementation period (about 12 months).

Neither changes in behaviors (outcome measure) nor reductions in alcohol-related crashes (impact measure) were found. These results are consistent with previous efforts conducted by the Center for Health and Safety Culture in which behavior change often requires several years of intense messaging and is more likely to occur when supported by other strategies at the local level.

As a result of the evaluation, recommendations were made to improve the effectiveness of the strategy:

- Continue leveraging the existing positive norms at the community level that can provide energy to foster local coalitions to take additional steps to address traffic safety.

- Use highly targeted media to reach those most in a position to act. For example, use the media developed for placement in alcohol retail establishments in future efforts to address impaired driving.

- Invest more in local involvement and leverage the media to engage action and policy at the community level. This may require "seed" funding and/or partnerships with existing entities at the community level. Local stakeholders can use the media as a catalyst to promote family engagement, school or driver education programs, workplace safety programs, enforcement strategies, and local policy change.

- Shift from viewing media campaigns as only a tool for behavior change to viewing campaigns as a catalyst to support local efforts to address traffic safety thus resulting in sustained, long term change in traffic safety culture. While sustained media efforts can impact behavior, augmenting media strategies with local efforts using multiple strategies is more likely to result in greater and sustained change.

The results and recommendations were compiled in a report and presentation. The presentation was shared with key stakeholders including the public board that oversees the Idaho Transportation Department.

19

Conclusion

Reaching zero traffic-related deaths and serious injuries will require new thinking – including evaluative thinking. Evaluative thinking is a problem-solving approach to designing, selecting, and allocating resources for traffic safety strategies. It seeks credible evidence to provide answers about the effectiveness and sustainability of traffic safety strategies.

Traffic safety culture strategies focus on changing beliefs that influence behaviors related to traffic safety. For such strategies to become more widely used, we need more evidence that they are effective and more knowledge about how to implement them effectively.

Traffic safety practitioners can use process, outcome, and impact evaluations to grow evidence and knowledge. For evaluations to be useful, they must use quality data and make appropriate comparisons. Stakeholders should be involved in developing an evaluation. After developing a clear description of the strategy, quality data and appropriate comparisons can be identified for use in the evaluation.

Once evaluation results are gathered and analyzed, stakeholders should make meaning of the results, accumulate wisdom (i.e., lessons learned), and identify opportunities to apply the knowledge in the future.

Glossary

Adapted from an *Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide*[4]

Accountability: The responsibility of managers and staff to provide evidence to stakeholders and funding agencies that a strategy is effective and in conformance with its coverage, service, legal, and fiscal requirements.

Accuracy: The extent to which an evaluation is truthful or valid in what it says about a strategy, project, or material.

Case study: A data collection method that involves in-depth studies of specific cases or projects within a strategy. The method itself is made up of one or more data collection methods (such as interviews and file review).

Causal inference: The logical process used to draw conclusions from evidence concerning what has been produced or "caused" by a strategy. To say that a strategy produced or caused a certain result means that, if the strategy had not been there (or if it had been there in a different form or degree), then the observed result (or level of result) would not have occurred.

Comparison group: A group not exposed to a strategy or treatment. Also referred to as a control group.

Conclusion validity: The ability to generalize the conclusions about an existing strategy to other places, times, or situations. Both internal and external validity issues must be addressed if such conclusions are to be reached.

Confidence level: A statement that the true value of a parameter for a population lays within a specified range of values with a certain level of probability.

Control group: In quasi-experimental designs, a group of subjects who receive all influences except the strategy in exactly the same fashion as the treatment group (the latter called, in some circumstances, the experimental or strategy group). Also referred to as a non- strategy group.

Cost-benefit analysis: An analysis that combines the benefits of a strategy with the costs of the strategy. The benefits and costs are transformed into monetary terms.

Cost-effectiveness analysis: An analysis that combines strategy costs and effects (impacts). However, the impacts do not have to be transformed into monetary benefits or costs.

Cross-sectional data: Data collected at one point in time from various entities.

Data collection method: The way facts about a strategy and its outcomes are amassed. Data collection methods often used in strategy evaluations include literature search, file review, natural observations, surveys, expert opinion, and case studies.

Descriptive statistical analysis: Numbers and tabulations used to summarize and present quantitative information concisely.

Evaluation design: The logical model or conceptual framework used to arrive at conclusions about outcomes.

Evaluation plan: A written document describing the overall approach or design that will be used to guide an evaluation. It includes what will be done, how it will be done, who will do it, when it will be done, why the evaluation is being conducted, and how the findings will likely be used.

Evaluation strategy: The method used to gather evidence about one or more outcomes of a strategy. An evaluation strategy is made up of an evaluation design, a data collection method, and an analysis technique.

Experimental (or randomized) designs: Designs that try to ensure the initial equivalence of one or more control groups to a treatment group by administratively creating the groups through random assignment, thereby ensuring their mathematical equivalence. Examples of experimental or randomized designs are randomized block designs, Latin square designs, fractional designs, and the Solomon four-group.

External validity: The ability to generalize conclusions about a strategy to future or different conditions. Threats to external validity include selection and strategy interaction, setting and strategy interaction, and history and strategy interaction.

Focus group: A group of people selected for their relevance to an evaluation that is engaged by a trained facilitator in a series of discussions designed for sharing insights, ideas, and observations on a topic of concern.

Ideal evaluation design: The conceptual comparison of two or more situations that are identical except that in one case the strategy is operational. Only one group (the treatment group) receives the strategy; the other groups (the control groups) are subject to all pertinent influences except for the operation of the strategy, in exactly the same fashion as the treatment group. Outcomes are measured in exactly the same way for both groups and any differences can be attributed to the strategy.

Implicit design: A design with no formal control group and where measurement is made after exposure to the strategy.

Indicator: A specific, observable, and measurable characteristic or change that shows the progress a strategy is making toward achieving a specified outcome.

Inferential statistical analysis: Statistical analysis using models to confirm relationships among variables of interest or to generalize findings to an overall population.

Interaction effect: The joint net effect of two (or more) variables affecting the outcome of a quasi-experiment.

Internal validity: The ability to assert that a strategy has caused measured results (to a certain degree), in the face of plausible potential alternative explanations. The most common threats to internal validity are history, maturation, mortality, selection bias, regression artifacts, diffusion, and imitation of treatment and testing.

Interviewer bias: The influence of the interviewer on the interviewee. This may result from several factors, including the physical and psychological characteristics of the interviewer, which may affect the interviewees and cause differential responses among them.

Literature search: A data collection method that involves an identification and examination of research reports, published papers, and books.

Longitudinal data: Data collected over a period of time, sometimes involving a stream of data for particular persons or entities over time.

Matching: Dividing the population into "blocks" in terms of one or more variables (other than the strategy) that are expected to have an influence on the impact of the strategy.

Maturation: Changes in the outcomes that are a consequence of time rather than of the strategy, such as participant aging. This is a threat to internal validity.

Measurement validity: A measurement is valid to the extent that it represents what it is intended and presumed to represent. Valid measures have no systematic bias.

Multiple lines of evidence: The use of several independent evaluation strategies to address the same evaluation issue, relying on different data sources, on different analytical methods, or on both.

Natural observation: A data collection method that involves on-site visits to locations where a strategy is operating. It directly assesses the setting of a strategy, its activities, and individuals who participate in the activities.

Non-response bias: Potential skewing because of non-response. The answers from sampling units that do produce information may differ on items of interest from the answers from the sampling units that do not reply.

Non-sampling error: The errors, other than those attributable to sampling, that arise during the course of almost all survey activities (even a complete census), such as respondents' different interpretation of questions, mistakes in processing results, or errors in the sampling frame.

Objective data: Observations that do not involve personal feelings and are based on observable facts. Objective data can be measured quantitatively or qualitatively.

Objectivity: Evidence and conclusions that can be verified by someone other than the original authors.

Outcome evaluation: The systematic collection of information to assess the impact of a strategy, present conclusions about the merit or worth of a strategy and make recommendations about future strategy direction or improvement.

Outcomes: The results of strategy operations or activities; the effects triggered by the strategy. (For example, increased knowledge, changed attitudes or beliefs, reduced risky behaviors, reduced morbidity and mortality.)

Population: The set of units to which the results of a survey apply.

Primary data: Data collected by an evaluation team specifically for the evaluation study.

Probability sampling: The selection of units from a population based on the principle of randomization. Every unit of the population has a calculable (non-zero) probability of being selected.

Process evaluation: The systematic collection of information to document and assess how a strategy was implemented and operates.

Program/Strategy evaluation: The systematic collection of information about the activities, characteristics, and outcomes of strategy to make judgments about the strategy, improve strategy effectiveness, and/or inform decisions about future strategy development.

Program/Strategy goal: A statement of the overall mission or purpose(s) of the strategy.

Qualitative data: Observations that are categorical rather than numerical, and often involve knowledge, attitudes, perceptions, and intentions.

Quantitative data: Observations that are numerical.

Quasi-experimental design: Study structures that use comparison groups to draw causal inferences but do not use randomization to create the treatment and control groups. The treatment group is usually given. The control group is selected to match the treatment group as closely as possible so that inferences on the incremental impacts of the strategy can be made.

Randomization: Use of a probability scheme for choosing a sample. This can be done using random number tables, computers, dice, cards, and so forth.

Reliability: The extent to which a measurement, when repeatedly applied to a given situation consistently produces the same results if the situation does not change between the applications. Reliability can refer to the stability of the measurement over time or to the consistency of the measurement from place to place.

Sample size: The number of units to be sampled.

Sampling error: The error attributed to sampling and measuring a portion of the population rather than carrying out a census under the same general conditions.

Sampling method: The method by which the sampling units are selected (such as systematic or stratified sampling).

Sampling unit: The unit used for sampling. The population should be divisible into a finite number of distinct, non-overlapping units, so that each member of the population belongs to only one sampling unit.

Secondary data: Data collected and recorded by another (usually earlier) person or organization, usually for different purposes than the current evaluation.

Selection and program/strategy interaction: The uncharacteristic responsiveness of strategy participants because they are aware of being in the strategy or being part of a survey. This interaction is a threat to internal and external validity.

Selection bias: When the treatment and control groups involved in the strategy are initially statistically unequal in terms of one or more of the factors of interest. This is a threat to internal validity.

Setting and program/strategy interaction: When the setting of the experimental or pilot project is not typical of the setting envisioned for the full-scale strategy. This interaction is a threat to external validity.

Stakeholders: People or organizations that are invested in the strategy or that are interested in the results of the evaluation or what will be done with results of the evaluation.

Standard: A principle commonly agreed to by experts in the conduct and use of an evaluation for the measure of the value or quality of an evaluation (e.g., accuracy, feasibility, propriety, utility).

Standard deviation: The standard deviation of a set of numerical measurements (on an "interval scale"). It indicates how closely individual measurements cluster around the mean.

Statistical analysis: The manipulation of numerical or categorical data to predict phenomena, to draw conclusions about relationships among variables or to generalize results.

Statistical model: A model that is normally based on previous research and permits transformation of a specific impact measure into another specific impact measure, one specific impact measure into a range of other impact measures, or a range of impact measures into a range of other impact measures.

Statistically significant effects: Effects that are observed and are unlikely to result solely from chance variation. These can be assessed through the use of statistical tests.

Stratified sampling: A probability sampling technique that divides a population into relatively homogeneous layers called strata and selects appropriate samples independently in each of those layers.

Subjective data: Observations that involve personal feelings, attitudes, and perceptions. Subjective data can be measured quantitatively or qualitatively.

Surveys: A data collection method that involves a planned effort to collect needed data from a sample (or a complete census) of the relevant population. The relevant population consists of people or entities affected by the strategy (or of similar people or entities).

Testing bias: Changes observed in a quasi-experiment that may be the result of excessive familiarity with the measuring instrument. This is a potential threat to internal validity.

Theory of change: A theory of change lays out the science behind how a strategy is expected to cause the desired outcomes.

Treatment group: In research design, the group of subjects that receives the strategy. Also referred to as the experimental or strategy group.

Utility: The extent to which an evaluation produces and disseminates reports that inform relevant audiences and have beneficial impact on their work.

References

1   Traffic Safety Culture Primer. Retrienved from https://www.mdt.mt.gov/other/webdata/external/research/docs/research_proj/tsc/TSC_PRIMER/PRIMER.pdf

2   Schwandt, T. A. (2018). Evaluative Thinking as a Collaborative Social Practice: The Case of Boundary Judgment Making. New Directions for Evaluation, 2018(158), 125–137.

3   Archibald, T. (2013). "Evaluative Thinking." Free Range Evaluation, WordPress, Retrieved July 27, 2020 from https://tgarchibald.wordpress.com/2013/11/11/18/.

4   American Evaluation Association. An Evaluation Roadmap for a More Effective Government. Washington, D.C. Retrieved on July 24, 2020 from: https://www.eval.org/evaluationroadmap

5   U.S. Department of Health and Human Services Centers for Disease Control and Prevention. Office of the Director, Office of Strategy and Innovation. Introduction to program evaluation for public health programs: A self-study guide. Atlanta, GA: Centers for Disease Control and Prevention. (2011). Retrieved from: https://www.cdc.gov/eval/guide/cdcevalmanual.pdf

6   Otto, Jay, & McMahill, Annmarie. (2015). Courageous Voices Pilot Study [Final Report]. Idaho Transportation Department.