

A Transdisciplinary Model of Program Outcomes for Enhanced Evaluation Practice

John LaVelle and Satlaj Dighe

University of Minnesota – Twin Cities

Abstract: *Evaluation is a transdiscipline with a focus upon asking and answering important questions about programs, policies, and interventions. A unified taxonomy of evaluation-specific program outcomes would be a helpful tool to evaluators, program designers, implementers, and policymakers, but one has not yet been proposed and validated in an evaluation context. This study builds from a grounded analysis of 125 programs and over 850 individual program outcomes, from which was developed a taxonomy of nine outcomes specific to evaluating programs and interventions: attitude, affect, behavior, cognition, status, relationship, biological, environmental, and economic. The article defines key terms and directionality and suggests specific fields that provide a helpful lens for understanding programs and improving evaluation practice.*

Keywords: *evaluation, measurement, outcomes, research on evaluation, taxonomy*

Résumé : *L'évaluation est une discipline transversale visant à poser des questions importantes sur les programmes, les politiques et les interventions, et à répondre à ces questions. Une taxonomie uniformisée des résultats de programmes, pour l'évaluation, constituerait un outil utile pour les évaluateurs et évaluatrices, les concepteurs et conceptrices de programmes, les personnes chargées de la mise en œuvre et les responsables des politiques, mais une telle classification n'a pas encore été proposée et validée dans un contexte d'évaluation. La présente étude est fondée sur une analyse de 125 programmes et de plus de 850 résultats de programme précis, à partir desquels on a élaboré une taxonomie de neuf résultats précis pour les programmes et les interventions d'évaluation, qui décrivent des changements au niveau des attitudes, de l'affect, des comportements, de la cognition, du statut, des relations, ainsi que des changements biologiques, environnementaux et économiques. L'article définit les termes clés et la direction, en plus de suggérer des domaines précis qui peuvent offrir une optique intéressante pour comprendre les programmes et améliorer la pratique de l'évaluation.*

Mots clés : *évaluation, mesure, résultats, recherche sur l'évaluation, taxonomie*

Corresponding author: John LaVelle, Department of Organizational Leadership, Policy, and Development, College of Education and Human Development, 323 Burton Hall, 178 Pillsbury Drive SE, Minneapolis, MN 55455; jlavelle@umn.edu

Programs are a set of systematic and purposeful activities that should lead to a set of implicit or explicit outcomes that are definable and measurable through credible inquiry methods (Yarborough, Shula, Hopson, & Cautthers, 2011). The terms “outcome” and “outcomes” are ambiguous, however, and evaluation scholars have not defined the term well in the published literature. Moreover, contemporary evaluation lacks a unified framework for conceptualizing and organizing program outcomes, which can lead to inefficient and ineffective evaluation practice. In this article, we describe the utility of a data-based taxonomy of program outcomes for evaluation practice, and how we developed a taxonomy to organize and define individual, group, and systems-level outcomes.

UNDERSTANDING PROGRAMS, POLICIES, AND INTERVENTIONS

Because programs are complex, working to understand them can be a challenging undertaking (CDC, 1999; Funnell & Rogers, 2011). Evaluators have different visual and written tools at their disposal to frame, clarify, contextualize, explain, and build consensus on what the program is and what impact(s) it is supposed to have on program participants. These tools help describe the *theory* of the program. Funnell and Rogers (2011) compiled a list of terms that are related with program theory, including reasoning, causal maps, results chain, theory of action, theory of change, and more. Other discussions of theory (Chen, 1990, 2014; Donaldson & Lipsey, 2006) have worked to distinguish the kinds of theory that can influence evaluator and stakeholder understanding of programs, broadly partitioning them into program/stakeholder theories and social science theories.

Program and stakeholder theories

Stakeholder and program theories originate from stakeholders and are often represented through logic models and other visual tools (Donaldson, 2007; Funnell & Rogers, 2011; Patton, 2008; Russ-Eft & Preskill, 2009). Broadly, the evaluator facilitates one or more conversations with stakeholders and helps them articulate the program’s assumptions, resources, activities, outcomes, outputs, and external factors that might reasonably influence the success of the program (Chen, 1990, 2014). In the best circumstances, the stakeholders learn from the evaluation process to better understand their program, clarify misperceptions, and build consensus (Patton, 2008; Russ-Eft & Preskill, 2009). Evaluators generally use this theory as the basis for further conversations about key evaluation questions, evaluation design, and other processes intrinsic to evaluation practice (CDC, 1999; Russ-Eft & Preskill, 2009).

Stakeholder-driven theory is valuable because it is grounded in the program itself and contextualized in the program’s current environment. It is also imbued with face validity, which refers to the degree to which a data-collection tool appears to represent the construct it was intended to measure (Shadish, Cook, & Campbell, 2002), giving it credibility with stakeholders. In essence, the theory makes sense to the stakeholders and evaluator. However, an overreliance on

stakeholder-driven program theory can lead to challenges in evaluation practice due to the plausible and practical gap between the program's resources and activities, and the kinds of outcomes that stakeholders and evaluators believe they will achieve. That is, although stakeholders believe their program will achieve a particular outcome, that alone does not necessarily make it plausible or realistic (Chen, 1990; Weiss, 1997).

Social science theories

An additional tool that evaluators find helpful in practice is social and educational science theory. These theories are often used to describe, predict, classify, or guide inquiry about the values of an individual or group of people (Mark, 2008), and each theory is rooted in one or more disciplines dedicated to understanding human functioning and behavior in a number of contexts. There are too many social science disciplines to list here, though the ones that seem to have the most direct relevance to evaluation practice are drawn from psychology, sociology, political science, anthropology, public health, education, and economics. Social science theories are also valuable because they study and describe the impact of individual-level and program-level moderating variables (e.g., participant sex, gender, sexual identity, race, ethnicity, class, socioeconomic status, disability status, and age; program dosage, intensity, or duration of program activity, etc.). They can lead to a more nuanced, different interpretation of the data than a simple analysis might show (Figure 1).

Last, social science theory is valuable in evaluation practice because it builds from strong construct validity (Shadish et al., 2002). A limitation, however, is that social science theory often seeks to study the average effect of a variable over

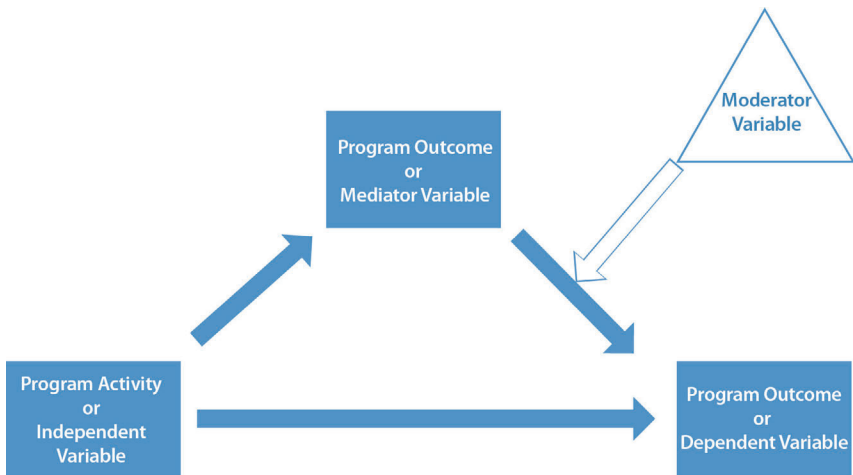


Figure 1. Mediator and moderator variables in evaluation

large groups, rather than focusing on the specific effects of a particular program contextualized in a sociopolitical and economic environment.

Integrating program and social science theory: Program outcomes

Inspired by [Chen's \(1990, 2014\)](#) work, [Donaldson \(2007\)](#) suggested a process for utilizing social science theory as a tool for framing programs and interventions that he termed “theory-driven evaluation science.” The approach garnered support as well as critique from evaluators who believed the approach was too difficult to implement in practice ([Stufflebeam, 2001](#); [Stufflebeam & Shinkfield, 2007](#)). We are intrigued by the idea of integrating existing social science theory as a tool for understanding programs, policies, and interventions, and we believe there is value in stitching together stakeholder-driven theory with social and educational science theory.

Examples on *how* to weave the theories together are relatively rare, although [LaVelle \(2019\)](#) gives a practical example of how to use social science theory to reframe and reorder program outcomes and to perform a plausibility check on the alignment between program activities and outcomes. LaVelle recommends building from the theory-driven approach to evaluation ([Chen, 2014](#); [Donaldson, 2007](#)) into a more holistic evaluation approach that purposefully integrates the best aspects of basic research and theory to identify and operationally define program outcomes, better conduct plausibility checks on the reasonableness of particular outcomes, and guide in the refinement and contextualization of data-collection tools. However, even this example failed to address a fundamental area where stakeholder-driven theory and social science and educational theory could begin to meet: the definition of program outcomes.

DEFINING PROGRAM OUTCOMES

Evaluators often find themselves helping stakeholders articulate and illustrate their program's theory of change as well as developing reasonable definitions of the desired outcomes. A limitation in contemporary evaluation practice is that there does not seem to be a unified definition of what makes something an outcome (see [Table 1](#)) or a taxonomy of program outcomes and their characteristics. Such a taxonomy could be very helpful for both evaluators and program developers/implementers, as well as for other engineers of policies and interventions. The specific definitions of the outcomes would need to change based on context and evaluand type, but the broad categories would ideally transcend context.

We conducted a preliminary search in major evaluation textbooks, reference guides, and peer-reviewed journals that publish articles on evaluation scholarship and practice (e.g., *American Journal of Evaluation*; *Canadian Journal of Program Evaluation*; *Evaluation: An International Journal*; *Evaluation and Program Planning*). When found, the definitions and descriptions of what makes something an outcome are illustrative but ephemeral; they often fail to provide in-depth elaboration of their descriptions or to give more than one or two concrete examples.

Table 1. Summary of “outcome” definitions and descriptions

Source	Definition
Scriven(1994). <i>Evaluation Thesaurus</i> , 4th ed.	“Outcomes are usually the post-treatment effects; but there are often effects during treatment, for example, enjoyment of a teaching style, which we sometimes (casually) call process. In general, we should distinguish: immediate outcomes, end of treatment outcomes, and long-term outcomes to be discovered by follow-ups” (p. 250).
Weiss (1998). <i>Evaluation: Methods for studying programs and policies</i> , 2 nd ed.	“the end results of a program; may be positive or negative” (p. 334).
Mathison (2005). <i>Encyclopedia of evaluation</i> .	“Outcomes are changes, results, and impacts that may be short or long term; proximal or distal; primary or secondary; intended or unintended; positive or negative; and singular, multiple, or hierarchical. Outcomes are enduring changes, in contrast to outputs. Evaluations, especially summative evaluations, measure outcomes at the individual level (changes in knowledge, skills, and attitudes), organizational level (changes in policies, practices, capacity), community level (changes in employment rates, school achievement, recycling), and the policy or governmental level (changes in law, regulations, sources of funding)” (p. 287).
Stufflebeam & Shinkfield (2007). <i>Evaluation theory, models, and applications</i> .	“Outcome need is a level of achievement of outcome in a particular area to fulfill a defensible purpose, such as preparing students for higher education” (p.11).
Patton (2008). <i>Utilization focused evaluation</i> , 4 th ed.	“the choice of language varies under different evaluation approaches ... what is important is not the phrase but that there be a clear statement of the targeted change in circumstances, status, level of functioning, behavior, attitude, knowledge, or skills” (p. 244).
Fitzpatrick, Sanders, & Worthen (2011). <i>Program evaluation: Alternative approaches and practical guidelines</i> , 4 th ed.	“Immediate, intermediate, long-term, and ultimate outcomes: the longitudinal goals for participant change (development)” (p. 160).

Some of the descriptions include information about the level of analysis for outcomes (e.g., Mathison, 2005), an emphasis on development and/or change (e.g., Fitzpatrick et al., 2011), or when the outcomes might be observed (e.g., Scriven, 1994). The descriptions by Patton (2008) and Mathison (2005) were the most

elaborate but were buried in the body of the text, and the small amount of space dedicated to the topic could be easily glossed over by the reader. Reading these descriptions, one might get the impression that evaluators and non-evaluators alike already share a common understanding of what constitutes an outcome. The text around the definitions did, however, include language about short-, medium-, and longer-term outcomes to give the reader a sense of proximity to the program activities, but with the exception of [Patton \(2008\)](#) did not describe outcomes in any greater depth. This look at the literature suggests that previous authors might have made assumptions that stakeholders and evaluators alike would intuit what made something an outcome without further elaboration. We are concerned that such an assumption could lead to misunderstandings between evaluators and stakeholders, as well as within the evaluation profession itself.

Strong alignment between anticipated outcomes and their operational definitions is imperative if evaluators are to create or adapt appropriate data-collection tools and analysis strategies, but strong operational definitions can be challenging to locate and synthesize ([Donaldson, 2007](#); [Donaldson & Lipsey, 2006](#)). However, evaluators would be wise to remember that many fields of inquiry exist to understand and predict different aspects of human life; it is therefore possible and probable that at least some of the outcomes articulated by program stakeholders have already been operationally defined in the scholarly literature.

Before that alignment can happen, it is necessary to address specific empirical questions about program outcomes, which might lead to the development of a flexible taxonomy of program outcomes. By exploring the following questions, we hope both to guide evaluation practice and to help raise evaluators' awareness of resources that could be valuable in contemporary evaluation practice:

1. What kinds of outcome can be identified through an inductive analysis of program logic models?
2. To what degree can the desired outcomes be defined at the individual, group, or systems level of analysis?
3. To what degree do the identified outcomes align with constructs currently identified in educational and social sciences?

METHOD

Because a taxonomy of program outcomes does not currently exist, the researchers used an inductive approach to address the research questions. We used a four-stage approach to this study: data compilation, development of codes, refinement of codes, and application to new data sources. This approach is based on the method advocated by [Miles and Huberman \(1994\)](#). We began by creating a comprehensive list of evaluations we had been involved with during our time in professional practice, either as an evaluator or a supervisor of evaluators ($N = 75$), and by compiling the logic models, flowcharts, logframes, and other visual representations of program operations and anticipated outcomes. We then conducted

a content analysis of the program logic models themselves (Miles & Huberman, 1994), which were taken to represent the most important and salient outcomes for the stakeholders (Russ-Eft & Preskill, 2009). In cases where the program logic model's outcomes were not immediately understandable, we used *in vivo* codes to analyze the program description section of the evaluation plan or report.

We augmented this list with outcomes identified in 50 additional program evaluations described in the *American Journal of Evaluation*, *Evaluation and Program Planning*, the *Canadian Journal of Program Evaluation*, and the grey literature developed by governmental organizations, non-governmental organizations, foundations, and nonprofits. This process resulted in a list of more than 850 program-level outcomes, which we categorized into individual-level outcomes (e.g., “improved student performance on an exam”), group-level outcomes (e.g., “improved school test scores”), and systems-level outcomes (e.g., “improved educational policy for the district”).

Next, consistent with grounded research practice (Denzin & Lincoln, 1994; Miles & Huberman, 1994) we randomly selected 10 logic models to analyze independently. These independent analyses formed the foundation of within-team discussions about the types and characteristics of the outcomes revealed by the *in vivo* coding. Drawing from consensus methodology (Waggoner, Carline, & Durning, 2016), we use two rounds of discussion to be sure we were coding the outcomes consistently and interpreting them similarly. These discussions coalesced into an initial taxonomy that included fourteen categories of outcomes that we subdivided into individual, group, and systems-level outcomes. The initial taxonomy comprised the following groups: attitude, affect, behavior, cognition, knowledge, money, time, environment, relationship, status, access, policy, health (physical), and health (mental).

We repeated the multi-discussion approach to apply the taxonomy to the remaining 115 logic models and refine the model (Waggoner et al., 2016), allowing us to refine and merge several categories of codes. Cognition and knowledge were merged into cognition (thought); Money and time were merged into economics (resources); and we removed both access and policy. We removed health (physical) and health (mental) because of the conceptual overlap they share with status and other outcome categories, such as biology, attitude, and environment. We added biology to represent biological and chemical change due to programs and interventions.

Access and policy were re-conceptualized as program activities (e.g., one creates access *to* something, or one creates policy *to do* something), both of which can lead to other individual or group-level outcomes such as attitude, affect, behavior, status, and more. Written policies can be considered outputs of evaluation activities—as opposed to enacted policies, which are an activity—and access might also be considered as *creating the potential for something*, suggesting that access might be a unit of measurement related to a later outcome (e.g., increase in something).

To validate the taxonomy, we used it to code an additional 10 logic models developed by colleagues working in different contexts of evaluation practice. The

Table 2. Example outcomes organized by major category

Example of program outcome	Level
Outcome Category: Affect	
Students communicate feelings and needs effectively	Individual
Students enjoy reading in the 4th grade and in the future years	Individual
Improved mood, reduced stress, and stimulation of creativity in the participants attending nature therapy session	Individual
Participants feel that they have a supportive cohort of fellow students and staff who want their success	Group
Outcome Category: Attitude	
Nature Therapy course participants form positive attitude towards nature and integrative nature therapy practices	Individual
Employees of XXX organization believe that the career center is beneficial for them	Individual
Students feel more confident in their ability to communicate	Individual
Increase in educational aspirations of students	Individual
More open and welcoming approach in school staff and teachers toward students from migrant and ethnic communities	System
Outcome Category: Behavior	
Diabetes patients form habits of healthy eating	Individual
Improved adherence to diabetes treatment	Individual
Increase in school engagement for students	Individual
Youth avoid risky sexual practices	Individual
Community interventions to prevent intimate partner violence	Group
Fathers are included in family and parenting activities	Group
The county engages in succession planning and strategic hiring from within	Group
Decrease in the abuse of over-the-counter and prescription pain-relief medications	System
Outcome Category: Cognition & knowledge	
Increase in knowledge of causes, symptoms, and medical care options for HIV	Individual
Increased participants' knowledge in nature therapy	Individual
Increased participants' knowledge about local environment and natural resources	Individual
Teachers learn strategies to build relationships with their students	Individual
Improved skills and knowledge of men and women in appropriate and sustainable crop, animal, and aquaculture management practices	Individual
A developed set of skills for recognizing and preventing intimate partner violence	Individual
Increased knowledge and awareness of processes related to immigration status and school enrollment procedures in the immigrant community in XXX city	Group
Increased knowledge and capacity in XXX school district to support students with disabilities to successfully graduate from high school	Systems

(Continued)

Table 2. Continued

Example of program outcome	Level
Outcome Category: Relational	
Staff has a positive and professional relationship with residents (of XXX residence complex) and their colleagues	Individual
Participants leave the program with a strong relationship with a faculty mentor and faculty support for their goals	Individual
Teacher-student relationships support accomplishment in both grades and social-emotional growth	Group
Outcome Category: Economic	
Support for consumers vulnerable to short-term food-price changes	Individual
Availability of credit to individual farmers participating in the XXX program	Individual
Increased job opportunities for the program participants	Individual
Increase in the financial savings of the households participating in XXX program	Group
Increased market access for coffee-bean harvesters in XXX region	Group
Enhanced food security for smallholder farmers and their households in XXX region of XXX country	System
Gender-equitable control of productive assets and resources (in communities participating in XXX agricultural reform program)	System
Outcome Category: Biological & Health	
Decreased levels of blood pressure in patients	Individual
Increased levels of physical activity among the residents of XXX neighborhood	Individual
Reduced intake of salt, sugar, and fats by students in XXX school	Individual
Increase in the purchase of healthy and organic food and beverage products by schools	Group
Decrease in the incidence of cholera in the residents of XXX city	Group
Decreased neonatal mortality and morbidity in XXX district in XXX country	System
Increase in the national life-expectancy rates for XXX country	System
Outcome Category: Environmental	
Increase in the number of people using recycling waste bins and facilities	Individual
Decrease in the amount of waste to be disposed of at landfills due to sludge deposition	Group
Increased community access to safe drinking water	Group
Reduced net greenhouse gas emissions from agriculture, forests, and other forms of land use	System
Outcome Category: Status	
Successful transition of high-school students with disabilities to institutions of higher education	Individual
Increase in number of athletes from XXX university who are physically and mentally fit	Individual
Increased graduation rates for students from XXX community	Group
Higher number of elected women representatives in local, state, and national offices	System
Decreased level of homelessness in XXX city	System

validation process failed to generate new outcome categories, leading us to believe we had reached data saturation, resulting in a valid and trustworthy taxonomy of program outcomes based on the available data (Miles & Huberman, 1994).

RESULTS

The majority of outcome descriptions included an active verb referring to the directionality of the outcome. This supports the Joint Commission's definition of a program and solidifies the proposition that programs exist to induce change or to maintain the status quo at the individual or group level (Yarbrough et al., 2011). Program outcomes were described primarily in terms that were synonymous with "increase," "decrease," or "maintain." We describe these verbs as increasing the amount of the outcome, maintaining a particular level of performance, and decreasing the amount of the outcome criterion.

Some outcomes were described in terms of "improved efficiency," which we took to mean attaining the same or better outcome with a reduced amount of input resources utilized or effort expended. Other programs described their desired outcomes in terms of "resilience," which we took to mean an individual, group, or community resisting change or "bouncing back" to a pre-intervention level of performance on specific outcomes following an intervention or event (Pooley & Cohen, 2010).

Taxonomy of outcomes

The results of the study indicate that the outcomes of the programs we analyzed can be categorized into one of nine categories (see Table 2). That is, every program and intervention we analyzed was trying to incur or prevent change across Attitude, Affect, Behavior, Biology, Cognition, Economic, Environment, Relationship, or Status. Although the specific meanings of each outcome category must be contextualized for the program and the evaluator, we offer the following broad descriptions:

- Attitude: Likes or dislikes that influence our behavior toward a person or a thing (Allport, 1935). Researchers believe that attitudes are the summative product of different processes, including emotions, thoughts, and behaviors (Kalat, 2005).
- Affect/emotion: Affect is any experience of emotion (APA, 2018a), and emotion is a reaction pattern through which an individual experiences an event (APA, 2018c). Both result from an interplay of physiology, actions/behavior, cognition, and phenomenology (feelings). Examples of affect and emotion include happiness, joy, sadness, fear, anger, stress, and more (Carlson, 2005; Kalat, 2005).
- Behavior: An overt, observable act; one's conduct toward the self or toward another, in response to internal or external stimuli (APA, 2018b)
- Cognition: The processes of thinking, gaining knowledge, and dealing with knowledge (Carlson, 2005; Kalat, 2005).
- Status: The perceived quality of someone or something in relation to others (Podolny, 1993); higher status as desirable, lower status as undesir-

able, with the sociometric challenges to go along with it (Katz, 1953); the relative social, professional, or other standing of someone or something. The term can be used to describe a physical or mental health state and shares some overlap with the Biology outcome category.

- Relationship: The way in which two or more concepts, objects, or people are connected, or the state of being connected.
- Biological: the organic and inorganic components of life, including physiological, biological, and chemical interactions, as well as how these respond to environmental, behavioural, and emotional stimuli (Martin & Hine, 2008).
- Environment: The physical and psychological surroundings or conditions in which a person, animal, or plant lives or operates. Also includes the social environment, which is the physical surroundings, social relationships, and cultural milieu within which groups of people function and interact (Barnett & Casper, 2001).
- Economic: Economics works under several working descriptions, including the generation of wealth, understanding scarcity, and the decisions that drive behaviors (Backhouse & Medema, 2009). These terms suggest a focus on the presence of, absence of, or behavior toward resources. In program terms, economics is contextualized as resources generated as a result of program activities, including time, money, and other resource-specific outcomes.

DISCUSSION AND NEXT STEPS

The taxonomy of program outcomes is helpful because it provides a starting point for evaluators so that they can consider the kinds of outcome that might be reasonable given the program activities. It is limited, however, because although the outcomes are largely mutually exclusive, some of the outcomes are the result of interactions between other outcomes (e.g., attitude is composed of emotion, thought, and behavior). The broad domain of social and behavioural sciences can offer additional clarification and helpful suggestions for program conceptualization and outcome measurement. As we developed the outcome taxonomy, many of the desired outcomes seemed to fall into the broad domains of psychology, sociology, political science, anthropology, public health, education, and economics. Each of these disciplines has its own body of knowledge, theoretical content, operational definitions, and preferred techniques for measurement, and we found that they provide a viable starting point for discussing and measuring program outcomes. The interconnected nature of evaluation can be a great asset as evaluators work in different contexts, but it can also be a limitation if evaluators are not aware of or do not keep updated on the major developments in other disciplines.

The results of this study are somewhat consistent with Rockwell and Bennett's (2004) work on conceptualizing program development and performance.

Rockwell and Bennett base their model in [Suchman \(1967\)](#) and [Kirkpatrick's \(1987\)](#) four-stage model for evaluating training programs, though the utility of Kirkpatrick's model has been challenged ([Yardley & Dornan, 2011](#)). Based in extension education, Rockwell and Bennett's model proposes nine outcome categories: environmental conditions, practices, knowledge, attitudes, skills, aspirations, reactions, participation, and resources. By contrast, the current study drew from evaluations across many contexts and included several outcomes not addressed in Rockwell and Bennett's work, such as relationships, status, and a broad contextualization of economic outcomes. A future study might further explore the operational definitions included in each outcome taxonomy to explore the degree to which they are context-bound or span across contexts.

We hope this study serves as a launching point for other scholarship. For example, a challenge in evaluation practice is the timing of anticipated program outcomes. Future research in this area could explore when the outcomes in this taxonomy might be expected to be measurable and could contrast the results with existing theory such as the theory of planned behavior ([Ajzen, 1985, 1991](#)), the health belief model ([Janz & Becker, 1984](#); [Sleet, Trifiletti, Gielen, & Simons-Morten, 2006](#)), and the trans-theoretical model of intrapersonal change ([Prochaska & DiClemente, 1986](#)).

Limitations of this study include the fact that it focused primarily on the outcomes of programs and interventions, and that a sizable proportion of the primary data came from evaluations conducted and supervised by the authors. It is possible that a replication study might yield slightly different results, as might a similar study done with a focus on policy outcomes. In addition, the primary data for this study did not lend themselves well to the identification and analysis of more macro-level/systems-level outcomes, such as the kinds of changes that might be realized through systems-level programs and policies. We hope that other scholars can expand this work through a purposeful focus on systems-level programs in both domestic and international contexts.

A further limitation is that the logic models and their associated outcomes were generated primarily in a North American context, and none of the programs focused explicitly on the interests and needs of Native and Indigenous peoples. Feedback from Julian King and other evaluators abroad raise the possibility of including "culture" as a viable outcome in some circumstances (King, personal communication, 2018). That is, some programs and policies exist in the world focused on restoring and/or maintaining Indigenous cultures. We agree, and it seems that a reasonable argument can be made that although the term "culture" might include attitudes, cognitions, behaviors, and other outcomes already listed in the taxonomy, the culture construct is greater than the sum of those parts and might be considered separately (King, personal communication, 2018). We hope evaluators and scholars working in these contexts can contribute to our understanding of culture as an outcome by expanding on our model.

In the end, we hope that this taxonomy is a useful tool for helping evaluators understand and frame program outcomes and for giving us a common

language to help our stakeholders describe exactly how their programs, policies, and interventions are contributing to the kinds of outcome they are trying to achieve.

REFERENCES

- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11–39). Berlin, Germany: Springer-Verlag.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-t](https://doi.org/10.1016/0749-5978(91)90020-t)
- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 798–844). Worcester, MA: Clark University Press.
- American Psychological Association (APA). (2018a). Affect. In *APA Dictionary of Psychology*. Retrieved from <https://dictionary.apa.org/affect>
- American Psychological Association (APA). (2018b). Behavior. In *APA Dictionary of Psychology*. Retrieved from <https://dictionary.apa.org/behavior>
- American Psychological Association (APA). (2018c). Emotion. In *APA Dictionary of Psychology*. Retrieved from <https://dictionary.apa.org/emotion>
- Backhouse, R. E., & Medema, S. G. (2009). Retrospectives: On the definition of economics. *Journal of Economic Perspectives*, 23(1), 221–233. <https://doi.org/10.1257/jep.23.1.221>
- Barnett, E., & Casper, M. (2001). A definition of “social environment.” *American Journal of Public Health*, 91(3), 465. <https://doi.org/10.2105/ajph.91.3.465a>. Medline:11249033
- Carlson, N. R. (2005). *Foundations of physiological psychology* (6th ed.). Boston, MA: Pearson.
- Centers for Disease Control and Prevention (CDC). (1999). *A framework for program evaluation*. Retrieved from <https://www.cdc.gov/eval/framework/index.htm>
- Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Chen, H. T. (2014). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective*. Thousand Oaks, CA: Sage.
- Denzin, N., & Lincoln, Y. S. (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York, NY: Taylor & Francis.
- Donaldson, S. I., & Lipsey, M. W. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The handbook of evaluation: Policies, programs, and practices* (pp. 56–75). Thousand Oaks, CA: Sage.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Upper Saddle River, NJ: Pearson.
- Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass.

- Janz, N. K., & Becker, M. H. (1984). The Health Belief Model: A decade later. *Health Education & Behavior*, 11(1), 1–47. <https://doi.org/10.1177/109019818401100101>. Medline:6392204
- Kalat, J. W. (2005). *An introduction to psychology* (7th ed.). Belmont, CA: Thomson Wadsworth.
- Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1), 39–43. <https://doi.org/10.1007/bf02289026>
- Kirkpatrick, D. L. (1987). Evaluation. In R. L. Craig (Ed.), *Training and development handbook* (3rd ed.). New York, NY: McGraw-Hill.
- LaVelle, J. M. (2019). Teaching the science of theory-driven evaluation. Manuscript submitted for publication.
- Mark, M. (2008). Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation. In N. L. Smith & P. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–134). New York, NY: The Guilford Press.
- Martin, E., & Hine, R. (2008). Biology. In *A dictionary of biology* (6th ed.). Oxford, England: Oxford University Press.
- Mathison, S. (2005). Outcome. In S. Mathison (Ed.), *Encyclopedia of evaluation* (p. 287). Thousand Oaks, CA: Sage.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2008). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage.
- Podolny, J. M. (1993). A status-based model of market competition. *American Journal of Sociology*, 98(4), 829–872. <https://doi.org/10.1086/230091>
- Pooley, J. A., & Cohen, L. (2010). Resilience: A definition in context. *The Australian Community Psychologist*, 22(1), 30–37.
- Prochaska, J. O., & DiClemente, C. C. (1986). Toward a comprehensive model of change. In W. R. Miller & N. Heather (Eds.), *Treating addictive behaviors: Processes of change* (pp. 3–27). Berlin, Germany: Springer.
- Rockwell, K., & Bennett, C. (2004). *Targeting outcomes of programs: A hierarchy for targeting outcomes and evaluating their achievement*. Lincoln, NE: University of Nebraska – Lincoln. Retrieved from <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1047&context=aglecfacpub>
- Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations a systematic approach: A systematic approach to enhancing learning, performance, and change*. New York, NY: Basic Books.
- Scriven, M. (1994). *Evaluation thesaurus* (4th ed.). Thousand Oaks, CA: Sage.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.
- Sleet, D. A., Trifiletti, L. B., Gielen, A. C., & Simons-Morten, B. (2006). Individual-level behavior change models: Applications to injury problems. In A. C. Gielen, D. A. Sleet, & R. J. DiClemente (Eds.), *Injury and violence prevention: Behavioral science theories, methods, and applications* (pp. 19–40). San Francisco, CA: Jossey-Bass.
- Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation*, 2001(89), 7–98. <https://doi.org/10.1002/ev.3>

- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Jossey-Bass.
- Suchman, E. A. (1967). *Evaluation research*. New York, NY: Russell Sage Foundation.
- Waggoner, J., Carline, J. D., & Durning, S. J. (2016). Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. *Academic Medicine*, 91(5), 663–668. <https://doi.org/10.1097/acm.0000000000001092>. Medline:26796090
- Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, 1997(76), 41–55. <https://doi.org/10.1002/ev.1086>
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies*. Upper Saddle River, NJ: Prentice Hall.
- Yarbrough, D. B., Shula, L. M., Hopson, R. K., & Cauthers, F. A. (2011). *Joint committee on standards on educational evaluation: The program evaluation standards: A guide for evaluators and users*. Thousand Oaks, CA: Sage.
- Yardley, S., & Dornan, T. (2012). Kirkpatrick's levels and education "evidence." *Medical Education*, 46(1), 97–106. <https://doi.org/10.1111/j.1365-2923.2011.04076.x>. Medline:22150201

AUTHOR INFORMATION

John LaVelle is an assistant professor in evaluation studies at the University of Minnesota. His research examines how universities prepare evaluators for applied work, and he has written on the intersection of job markets and university programs, techniques for recruiting evaluators, evaluator competencies, the psychology of evaluators, and the ways in which evaluators can use social science theory to inform their practice. He was awarded the American Evaluation Association's Marcia Guttentag Award in 2019.

Satlaj Dighe is a doctoral candidate in the evaluation studies program at the University of Minnesota. As an evaluator, her work is focused on understanding the impact of special education programs on the academic success of ethnic minority students with disabilities. Her research interests lie in the areas of evaluation capacity building in the Global South, culturally rooted evaluation, and contextualized health evaluations.