Spring 2023

A Culturally Responsive Evaluation Lens to Logic Model Design

Ciara Cascharelle Knight

A Culturally Responsive Evaluation Lens to Logic Model Design

Ciara C. Knight

Claremont Graduate University

2023

Approval of the Dissertation Committee

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Ciara C. Knight as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Psychology.

Dr. Tarek Azzam, Chair
Claremont Graduate University
Senior Visiting Fellow
University of California Santa Barbara
Associate Professor of Education

Dr. Wanda Casillas
Claremont Graduate University
Associate Research Professor & Director of
MS in Evaluation and Applied Research (Online)

Dr. Stewart Donaldson
Claremont Graduate University
Professor & Executive Director of
the Claremont Evaluation Center
and the Evaluators' Institute

Dr. Katrina Bledsoe
Abt Associates
Principal Associate of Global Equity,
Diversity, & Inclusion

Abstract

A Culturally Responsive Evaluation Lens to Logic Model Design

By

Ciara C. Knight

Claremont Graduate University: 2023

Culturally responsive evaluation (CRE) is an approach that centers all evaluation processes around the culture of a program's secondary stakeholders. Specifically, this entails ensuring shared meanings in a group through communication. However, minimal connections have been made between CRE and logic model designs. Logic models commonly used by evaluators are data visualization and communication tools designed to aid in effectively communicating a program's theory. Nevertheless, little is understood about the role culture plays in this process.

This multiphase mixed methods study explored the integration of CRE to logic model designs using individualism and collectivism (IC) as a construct for culture and as a basis for tailoring the designs. Specifically, Phase 1 utilized an exploratory sequential design to develop the Logic Model Knowledge Test (LMKT) based upon expert evaluation insight and Amazon Mechanical Turk (MTurk). This addressed how learning from a logic model can be assessed. The LMKT was designed to assess the understanding of logic model components and the content of the program theory. It yielded good validity and reliability and was used in the final phase of the study.

Phase 2 used a quantitative two-group comparison design to answer the question, which measures of IC would be more likely to help inform how people interpret logic models differently? Of three existing measures, Wagner (1995) was identified as a highly reliable

measure that better distinguished IC-oriented people. Accordingly, it was used in the final phase of the study.

Lastly, Phase 3's mixed methods experimental design utilizing the LMKT and Wagner's (1995) IC measure was implemented to help answer six questions regarding whether culturally tailored language, visualizations, or both impacted participants' visual efficiency (VE) and perceived credibility of the program (CP). The findings suggest collectivistic MTurk participants had higher VE across seven culturally tailored logic model conditions. Collectivistic-oriented participants could better quickly and accurately answer the LMKT questions with less mental effort across all logic model conditions than individualistic-oriented participants. Yet, these conditions did not impact IC MTurk participants' CP. Participants rated the program theory as credible across the logic model conditions regardless of their IC orientation.

The findings provided partial empirical support for integrating CRE in logic model designs. Moreover, they added to the validity of the IC construct and Wagner's (1995) measure. The findings also had valuable implications for evaluation practice through measures that could be used to better understand stakeholder cultural values aligned with communication styles and their knowledge of logic models. Finally, these same tools could be used for future research on evaluation concerning CRE and logic model designs.

Dedication

I dedicate my dissertation to my family, friends, colleagues, and lifelong mentors. I thank my beautiful family for their love, support, and encouragement while I pursued my education: Ken Knight (husband), Carolyn and Michael Dupray (parents), Lisha and Lee Jones (sister and brother-in-law), Cerrissa Dumaine (sister), Charles and Carol Knight (parents by marriage), Jennifer and Matthew Sloan (siblings by marriage), and Patricia Garvey (chosen sister). Where my human family was unavailable, I always had my feline family during the latter years leading up to this dissertation. Juno, Midnight, and Moe always kept me company, even during the inhumane night and morning hours. I appreciate the new and old friends that remained close, supportive, and patient during this process. My CGU colleagues have been the most collegial group I have ever experienced in my academic tenure. I appreciated the "Portfolio and Dissertation Support Group" sessions and your willingness to provide feedback and encouragement as lived experts of Ph.D. life. I am also very thankful for my lifelong mentors who guided me throughout my academic, personal, and professional growth: Mathew Riggs, Sora Tanjasiri, and Tarek Azzam. Last but never least, I thank the late Harry Reed, who let me intern at the Lawrence Berkley Lab's Workforce Diversity Office. This opportunity introduced me to logic models and evaluation. I am ever grateful to my village for lovingly embracing me during this process while I continue to stand on the shoulders of my ancestors.

Acknowledgments

This dissertation was supported by funds from the California State University's Chancellors' Doctoral Incentive Program (CDIP) Dissertation Fellowship. The contents are solely the responsibility of the author and do not necessarily represent the official views of CDIP. The author thanks Tarek Azzam, Wanda Casillas, Stewart Donaldson, Katrina Bledsoe, and Sora Tanjasiri for their support and guidance with this dissertation.

Table of Contents

Chapter 1 Introduction

A growing interest in culturally responsive evaluation (CRE) and a renewed interest in logic models has become increasingly more salient in evaluation. Over the past two decades, the broader discussion concerning CRE has exponentially grown. This is reflected by major publications about CRE (e.g., Hood et al., 2005; Hood et al., 2015; SenGupta et al., 2004), development of frameworks and guidelines (e.g., Bledsoe & Donaldson, 2015; Bowman et al., 2015; LaFrance et al., 2012), American Evaluation Associations' (AEA) Public Statement on Cultural Competence in Evaluation (2011), and the establishment of a center that holds annual conferences solely dedicated to CRE and assessment (e.g., University of Illinois Urbana-Champaign, n.d.). CRE has influenced multiple domains in the evaluation, including survey development (i.e., Bowen & Tillman, 2015) and evaluation design (i.e., LaFrance & Nichols, 2008). However, connections to other essential evaluation domains have yet to be made more visible in common evaluation literature. One such domain is logic model design. An abundance of logic model literature exists with commonly referenced literature primarily focuses on providing guidelines that emphasize the identification of components and model creation with little empirical evidence of its effectiveness as a tool to understand a program's theory (Knowlton & Phillips, 2013; Price et al., 2009). Several well-known examples are the W.K. Kellogg Foundation's Logic Model Development Guide (WKKF, 2004), the Center for Disease Control (CDC; 2014), and resources at the University of Wisconsin Extension (UW-CE; n.d.).

Only recently have evaluators produced empirical evidence of the effectiveness of logic models as a communication tool. This is reflected in data visualization literature (Azzam et al., 2013; Jones et al., 2019). Yet, despite the attention given to CRE and logic models' important role in evaluation, minimal attention has been given to integrating both to improve the practice of

employing culturally tailored communication. This entails being intentional about the mode, expressions, and visualizations used to relay a message based on the culture of the receivers of that message. Also, it ensures the understanding, receptiveness, and credibility of the message.

Culturally Responsive Evaluation

To better understand how this integration between CRE and logic modeling aids in effective communication, both practices must be explained. Culturally responsive evaluation (CRE) is an approach supported by various social science and transdisciplinary fields that centers all evaluation processes around the culture of the secondary stakeholders (i.e., program implementation, recipients, clients, and evaluator; Hood et al., 2015; Hopson, 2009). Culture is defined more broadly in CRE as "the web of negotiated and shared meanings of a group…displayed in the process of communication, and communication as the carrier of culture reflects and shapes the structure of culture" (Herdin et al., 2020, p. 7; SenGupta et al., 2004). Accordingly, CRE encompasses a wide variety of terminology which include culturally competent evaluation and multicultural evaluation-specific terms focused on a particular social group (i.e., Indigenous evaluation or feminist evaluation). All people involved with the program and its evaluation bring their cultural backgrounds to the environment and the challenges that come with it. This includes cultural orientations spanning individualism-collectivism, politics, language, religious affiliations, and more. This entails the need for evaluators practicing CRE to be attentive to how they communicate across cultural groups.

Proponents of CRE emphasize prioritization of the voices of historically marginalized groups and cultural contexts (CDC, 2014; Hood et al., 2015). This is combined with integrating critical and evaluation theories for practice throughout all phases of the evaluation process to ensure an accurate portrayal of stakeholder experiences (Hopson, 2009). This is an expansion

upon Stakes' (2004) responsive evaluation, which emphasizes addressing the needs of secondary stakeholders rather than those who merely commissioned the evaluation. This is because they are entrenched in the program's operations and are most likely to be impacted by the evaluation. Given this, their culture often shapes the program. Subsequently, attentiveness to cultural dynamics, such as varied cultural orientations of secondary stakeholders, may minimize any misinterpretations of data using culturally tailored communication while increasing the credibility of the evaluation leading to its use (Johnson et al., 2019; Rosenstein, 2005).

While culturally tailored communication is not explicitly mentioned among the principles that guide CRE practice, foundational literature that underly them includes guidelines for effective communication with stakeholders (Hood et al., 2015). For instance, CRE experts stress that evaluators should "[Understand] and respect varying communication and relational styles" (Hood et al., 2010, p. 289; Frierson et al., 2010). This entails evaluators being aware of different cultural subtleties apparent in the language and behavior of stakeholders and evaluators because interpretation may be required to ensure the understanding, transparency, and clarity of evaluation discussions (Hood, 2009). This is aligned with recommendations from expert CRE practitioners to ensure there are linguistic and language orthography experts to translate information or ideas to ensure effective communication between stakeholders and evaluators (Frierson et al., 2010; Hood et al., 2015; LaFrance et al., 2015). Lastly, the AEA Statement on Cultural Competence in Evaluation (2011) specifically states that evaluators must "promote full participation when evaluation activities are conducted in participants' primary or preferred languages [including] considerations of culturally specific communication styles and mannerisms." Accordingly, CRE evaluators must ensure their modes and methods for communication and reporting to stakeholders are culturally tailored. Therefore, CRE is better

understood as an approach supported by various social science and transdisciplinary fields that centers all evaluation processes around the culture of the secondary stakeholders (Hood et al., 2015; Hopson, 2009). These processes include intentionally using culturally tailored communication aligned with the stakeholders' culture, which is embedded in developing a logic model, selecting evaluation designs, determining data collection methods, analyzing data, and reporting findings.

Logic Models

Similar to CRE, logic models are centered around communication. They are designed to aid in timely and effective communication. Nevertheless, failure to be attentive to their design may lead to decreased credibility and minimal use of the model (Chen, 2014; Donaldson, 2007; Evergreen & Metzner, 2013; Ghanbarpou et al., 2020; Jones et al., 2019; Tufte, 2006). Accordingly, logic models are also primarily used as a communication tool, which is an understudied area. They help evaluators visualize data from multiple sources to reflect a mutual understanding of the program's theory "…history, activities, stakeholders, and more generally the context in which the program operates" efficiently and effectively (Azzam et al., 2013, p. 13). A major benefit of logic models is that they help with program design, planning, management, and evaluation by "[developing] a common language between evaluators and stakeholders [and they also] provide a credible reporting framework" (Frechtling, 2007; Knowlton & Phillips, 2013, p. 3).

Logic models can easily be mistaken for theory of change models, but they should be differentiated. In comparison to theory of change models, "logic models are popular and relatively easy to use, and they are very useful for reducing a complicated program to a set of meaningful, manageable components [whereas theory of change models are] …more elaborate

and takes more time to learn.…" (Chen, 2015, p 58; Chouinard & Cram, 2019; Madison, 2000). The latter is useful for illustrating the underlying causal mechanisms of the program activities, outcomes, and impact. Due to the complexity of theory of change models, this study will focus on logic models, given their relative ease of design and use.

Recent research has focused on the overall effectiveness of logic models. For example, one study indicated that the logic model design alone impacted the program's credibility and the viewer's visual efficiency (i.e., the ability to understand the program quickly and accurately; Jones et al., 2019). However, minimal research still connects cultural responsiveness with logic model design.

Integration of CRE and Logic Models through Cultural Orientation

Given the strengths of CRE and logic models, a commonality between the two practices is communication. CRE stresses the role of culture in how one may relay and interpret information. At the same time, logic models are frequently used as a process and tool to understand programs. This is done through their connection of program activities to their anticipated outcomes. Yet commonly used materials to guide the logic model development and design process neglect the importance of culturally tailored communication.

Social science research highlights the importance of being attentive to culture for effective communication (Frechtling, 2007; Gudykunst, 1993; Hood et al., 2015). Culture can influence how people communicate with others that are different from their own culture. To simplify, "[it] is through communication that culture is passed down, created, and modified from one generation to another [and] communication is necessary to define cultural experiences" (Ting-Toomey & Dorjee, 2019, pp. 19-20). Moreover, when a person communicates with another, culture is often represented through mannerisms and norms. Given this, culture is

present in logic models through the language and visualizations evaluators use in their designs. Not being attentive to this can lessen the effectiveness of logic models as a communication tool.

The importance of culture in communication is further exemplified in social science literature. In particular, the constructs of individualism and collectivism (IC) have been extensively researched. It is commonly used as an operationalization for culture in psychology, communication, and various other fields (Hofstede, 1980; Oyserman, 2017; Triandis & Gelfand, 2012). Individualism and collectivism (IC) represent polar opposites of a cultural orientation continuum. This provides added support for using CRE to facilitate communication between cultures through how information is visualized (Ting-Toomey & Dorjee, 2019). Based on IC conceptualization and research, stakeholders with individualistic cultural orientation tend to navigate the world with their values placed upon self-fulfillment and success, whereas stakeholders with a collectivistic cultural orientation place value upon the group or community's success. Over the years, an enormous amount of IC research has indicated that individualistic and collectivistic people can be distinguished from each other by how they cognitively process information (Oyserman & Lee, 2008; Ueda et al., 2018). This is an important consideration for evaluators as they design logic models.

These differences along the IC continuum have been found between countries. For example, western countries display more individualistic characteristics and values than Eastern countries, which tend to be more collectivistic (Ting-Toomey & Dorjee, 2019). IC difference can also be found between genders, with women being more collectivistic than men, as represented by their common relational communication style (Kray et al., 2002; Kray et al., 2004; Kray et al., 2001; Mortenson, 2002). Moreover, while IC is not explicitly mentioned, its conceptualization is

reflected in evaluation literature on cultural competency, particularly how eastern and western cultures are distinguished (SenGupta et al., 2004).

Despite these differences, IC orientations can be nuanced because they are situational, a common challenge in studying culture (Casillas & Trochim, 2015; Oyserman, 2017). For instance, differences along the continuum are reflected within countries and among specific communities (i.e., within groups or between subgroups). Many non-white Americans tend to be more collectivistic than white Americans, who tend to be more individualistic (Vandello & Cohen,1999). This demonstrates that the IC continuum is not a generalization of cultural differences. Instead, it is a tool for guiding evaluators' understanding of how cultural differences can potentially impact the visual efficiency and credibility of program theory based on the logic model design. This is consistent with evaluation practice, especially from a CRE approach. Both programs and culture are fluid by nature. So, if one can envision stakeholders and evaluators discussing a specific program with each other, a common cultural orientation for an individual or group may emerge (Oyserman, 2017). Accordingly, these orientations have impacted attitudes, behavior, and decision-making (LeFebvre & Franke, 2013; Oyserman, 2017; Ting-Toomey & Dorjee, 2019). Therefore, understanding how the underlying values of IC individuals are represented in communication and interpretation of information provides additional research to help evaluators understand the value of a CRE approach to logic models through culturally tailoring their design.

Scope of Literature Review

This chapter reviews the evaluation and CRE-specific literature interwoven with support from IC literature to emphasize the value of integrating a CRE approach to logic model design. While many evaluation experts recommend designing logic models in close collaboration with

stakeholders, this paper focuses on design considerations and the final logic model product without discussing the collaboration process. This focus would convey to evaluators how stakeholders may understand a program's theory with high visual efficiency when using logic models designed to reflect their common cultural orientation.

Accordingly, this chapter is organized into three distinct sections. Two sections, Language and Non-textual Visualizations, represent the key aspects of logic model designs. The beginning of these sections reviews common literature about language and non-textual visualizations while making connections to logic model design. Then, each section discusses relevant CRE literature and includes examples or additional considerations for designing CRE logic models. Throughout each section, IC is discussed when appropriate. Given this organization, the Language section consists of two parts: 1) The Role of Language in CRE and 2) Language Considerations for Logic Model Designs. The Non-textual Visualization section consists of five parts: 1) Common Designs of Logic Models, 2) Complexity in Logic Model Designs, 3) The Role of Non-textual Visualizations in CRE, 4) Non-textual Visualizations in CRE Logic Model Designs, and 5) Expanded Possibilities for Non-textual Visualizations.

Ultimately, explicit connections between CRE and logic model design are supported using relevant transdisciplinary literature from evaluation, psychology (i.e., cognitive, social, & cross-cultural), communication (i.e., cross-cultural), and data visualization. This is reflected throughout this paper. However, the last section, Implications of the Literature, provides an overarching final summary of the literature presented in this paper. Moreover, it discusses the significance of the literature with the IC framework as a connector for integrating CRE into all logic model designs. Hence, this supports the full integration of CRE into common evaluation practice.

Language

Language is one of many ways to communicate (via spoken or written words and phrases) with each other. This is important for most human interactions, including interactions using visualizations such as logic models. However, it is rarely neutral in evaluation mainly because it often reflects a singular perspective (Patton, 2000). Yet, language can differ across countries or communities, age groups, and professions. This means having a shared language does not always equate to having shared meaning among those interacting with each other. For instance, certain words and phrases within a shared language can still illicit a different understanding, attitude, belief, value, and behavior across communities. Given this, many evaluators and social scientists heavily acknowledge the importance of language, particularly for its influential power when used strategically. Nevertheless, experts often recommend using simple language to help prevent confusion or misinterpretation (Few & Edge, 2007; Torres et al., 2005). While this continues to be sound advice, there is much more to consider about how evaluators use language in their logic model designs.

The power of language applies to logic models due to the use of written text to assist with establishing a shared understanding of the program theory among stakeholders. It is common for evaluators to label the various logic model components as inputs (i.e., resources used to operate the program), activities (i.e., strategies, specific tasks, or services the program provides), outputs (i.e., direct tangible products or deliverables produced from the program activities), and outcomes (i.e., short through long term goals, objectives, aims of the program anticipates achieving) and the content describing the program organized within these components (McLaughlin & Gretchen, 2015). In addition, some logic models are heavily worded while others are not. Lastly, the titles we give logic models have the potential to impact the mental effort it

takes for stakeholders to understand what the overall logic model is about (Wanzer et al., 2021). Thus, how evaluators use language in logic models is critical to shaping stakeholders' views of them.

The way language is used in logic models can impact stakeholder attitudes. In a study by Mason and Azzam (2019), participants who randomly received a positive program description tended to have significantly increased positive global attitudes after reading an evaluation report of positive program outcomes. Similarly, those randomly assigned to read the negative program description and subsequently reviewed a negative evaluation report about the program had significantly increased negative global attitudes. This demonstrates the influence of language choice (i.e., positive, negative, or neutral words or phrases) on stakeholders. Therefore, language use in logic models can influence stakeholders' attitudes about a program and its evaluation.

The language evaluators use in logic models not only invokes different meanings at times but can also solicit varied connotations. For instance, Madison (2000) emphasized this importance by reflecting on their evaluation of a statewide at-risk youth program. When grantee agencies were asked to describe "at-risk youth," many listed negative characteristics of these minors and their socio-economic environments. Some examples of the words used to describe them were "African American," "Latino," and "antisocial behavior," coupled with the following words to describe their socio-economic environments: "low income," "single parents," and "uncaring parents." These definitions were starkly different from the granter, which was "youth involved in gangs, teen parents, youth in alternative education, and youth involved in the courts" (Madison, 2000, p. 22). Therefore, the granter's use of the term "at-risk" unintentionally invoked negative and deficit-focused connotations attributed to specific racial and ethnic groups by

grantees. Again, this emphasizes the importance of selecting words in the logic model that reduces the potential for miscommunication and invocation of negative attitudes.

IC literature further exemplifies the role of words in how stakeholders may communicate with each other based on their cultural orientation. For instance, collectivistic-oriented stakeholders tend to use group-focused words, phrases, or pluralistic pronouns (i.e., us, we, our, or ours; Kashima & Kashima, 1998; Ting-Toomey & Dorjee, 2019). Even frequent words used in conversations about goals are group focused, which include words like "our goal," "our work team", or "our future goal" (Berendt & Tanita, 2011; Ting-Toomey & Dorjee, 2019, p. 207). Whereas individualistic communication often includes the words "I," "me," self-help," or "my goal," all of which are self-focused (Kashima & Kashima, 1998; Ting-Toomey & Dorjee, 2019, p. 207). This distinct trend in communication among collectivistic and individualistic-oriented people further supports the notion that evaluators should tailor the language presented in logic models to reflect the culture of the stakeholders to ensure they clearly understand it.

Given the clear demonstration of IC cultural values represented in how stakeholders communicate based on their orientation, evaluators should be even more attentive to their word selection for outcomes in logic models. The language used to describe outcomes should be written to resonate with the secondary stakeholders or audience and reflect their culture. For example, evidence from IC literature indicates that if self-efficacy is an outcome of a program, then evaluators should know that individualistic-oriented stakeholders may experience higher self-efficacy when working by themselves, while collectivistic-oriented stakeholders may experience higher self-efficacy when working in a group (Earley, 1993; Triandis & Gelfand, 2012). Moreover, collectivistic-oriented cultures tend to be more prevention-focused regarding assessing personal outcomes in a given situation, while individualistic-oriented cultures tend to

be more promotion-focused (Lalwani et al., 2009; Lee et al., 2000; Triandis & Gelfand, 2012). For instance, Lalwani et al., 2009 found that participants from Hong Kong significantly focused their shortcomings on a task (i.e., prevention-focus) compared to US participants. Oppositely, participants from the US significantly focused on their progress, accomplishments, or achievements on a given task compared to Hong Kong participants. These findings inform how evaluators should be considerate of how they write outcomes in a logic model. Specifically, outcomes should be written to reflect the shared cultural orientation across the stakeholders in the program to ensure effective communication.

Evaluators must also ensure that these IC outcomes align with the program design. For instance, programs can focus on progress at an individual level yet cater to a collectivistic community. When faced with this situation, evaluators should talk with stakeholders to clarify these outcomes and the potential measures (i.e., group or individual level) that may be used in the evaluation. Then, they should use the appropriate culturally tailored language in the logic model to ensure a shared understanding of the outcomes among stakeholders.

When the stakeholder values align with the program design, writing culturally tailored outcomes does not require much effort. For instance, individualistic framed outcomes can be rewritten to include the prefix "self." If a program aims to impact the participant's self-esteem, this is ideally framed for individualistic-oriented cultures. To change this, so it better aligns with collectivistic-oriented cultures, the prefix "self" should be replaced with "collective," "group," or "community." Additional examples can be reviewed in Table 1.

Table 1

Examples of Reframed Outcomes for Individualistic (I) and Collectivistic I Oriented Cultures

Outcome	I-Framed Outcome	C-Framed Outcome
Esteem	Self-Esteem	Collective/Group/Community Esteem
Confidence	Self Confidence	Collective/Group/Community Confidence
Efficacy	Self-Efficacy	Collective/Group efficacy
Resilience	Self-Resilience (Independence)	Community/Group Reliance (Interdependence)
Identity	Self-Identity	Collective/Group/Community Identity
Actions/knowledge gained for …	Self-Fulfillment	Collective/ Group/Community Fulfillment
	Self- Achievement	Collective/Group/ Community Achievement

These insights from evaluation and social science research are promising for enhancing logic model design. They indicate that language choice can be a powerful influence on stakeholders' attitudes, perceptions, and attentiveness. However, improper language can cause confusion and potential misinterpretation of logic model content. Evaluators tend to use the language of the most powerful and dominant stakeholder(s), such as funders or granters, as opposed to stakeholders that are more involved with the program, such as program managers, implementors, participants, or community members (Chouinard & Cram, 2019; Madison, 2000). Even commonly used labels for logic model components (i.e., inputs, outputs, and outcomes) are familiar to funders. This is problematic, considering these labels are known to be confusing for stakeholders as they often find difficulty in distinguishing between outputs and outcomes in logic models (Frechtling, 2007). This pattern of language use defeats the potential purpose and value of logic models for establishing a mutual understanding among stakeholders. Therefore, this should be addressed when designing logic models.

The Role of Language in CRE

While language is important in evaluation to influence communication and use, a CRE approach delves deeper than typical evaluation approaches. It encourages using the language of the program implementers, participants, and community that reflects their values. This goes beyond merely ensuring understanding; rather, "meaning and language are connected to culture through our mental representations, which we then use to interpret things in the world. People who share a culture often share these same meanings and constructs of the social world, which they represent through a shared language" (Chouinard & Cram, 2018, p. 138). This meaning is derived from culturally tailored language and ensures stakeholder trust, ownership, and credibility (Alkin & Vo, 2017; Bledsoe, 2005). This could influence the attitudes and behaviors of stakeholders in a meaningful way to encourage the use of evaluation information, including information presented in logic models.

Accordingly, language focuses on CRE's theoretical underpinnings, frameworks, and practices (Avant, 2020; Chouinard & Cram, 2020; Mertens, 2009). It influences the attitudes and behaviors of the stakeholders (Brown, 2000; Chouinard & Cram, 2020; Kray et al., 2002; Kray et al., 2004; Kray et al., 2001; LaFrance et al., 2012; Patton, 2000). While not explicitly labeled as CRE, the theoretical underpinnings of appreciative inquiry evaluation are centered around language that is not deficit-based, which is a key characteristic of CRE (Hood et al., 2015; Hopson, 2009; Preskill et al., 2006; Paige et al., 2015). Other theoretical underpinnings of the CRE framework related to language include the importance of language expression reflecting shared stakeholder values (Hood, 1998; 2009; Hood et al., 2015), using a cultural liaison or language translator to facilitate "translation of ideas or terms may require the assistance of linguistic or language orthography experts" (Hood et al., 2015, p. 293; LaFrance et al., 2015),

and using a culture checklist to ensure evaluators are emphasizing whose voice are being represented by means of language and expression [when working with stakeholders]" (Hood et al., 2015; Kirkhart, 2013).

The importance of language choice in CRE is reflected in frameworks and recommendations for practice in specific communities. For example, the Indigenous language is embedded in Kaupapa Maori Evaluation Framework (Cram et al., 2018). It integrates Māori words and phrases representing their cultural values and the "Taonga Tuku Iho—The Cultural Aspirations Principle that validates the Maori language in evaluation practice" (p. 66). In addition, language is emphasized in a variety of additional community-specific CRE. This includes the Culturally Responsive Indigenous Evaluation Framework (Bowman & Dodge-Francis, 2018) and recommendations for working in Latino communities (Clayson et al., 2002; Guzman, 2003).

This focus on language is evident among evaluators that integrate a CRE approach into their practice (Chouinard & Cram, 2019; 2020). For instance, Clayson et al. (2002) emphasized adjusting words and phrases to "ensure cultural relevance to the community" (p. 81) in their evaluation of a Latino immigrant serving program. Likewise, Copeland-Carson (2005) recommends adjusting concepts to "provide terms people would find more familiar and palatable" when working with communities across the African diaspora, and Carlson et al. (2017) uses the local language of Maori stakeholders in their practice as opposed to the common use of the language of the evaluator.

Similarly, LaFrance (2004, 2012) considers how they use language in the evaluation and even describes logic models. For example, they do not use the label "logic model" in evaluations with indigenous communities. Rather they use a "conceptual map," "conceptual picture," or

"conceptual model" to describe the program theory visualization to communities. LaFrance (2004) does this because logic models "…connotes an intellectualism that can come across as elitist, mysterious, and Western" (p. 45). In other words, the indigenous cultural connotation of the word "logic" implies the program does not make sense, and the evaluator needs to make sense of it. This further implies a power dynamic in which the evaluator holds most of the power in the evaluation process (Haugen & Chouinard, 2019). Therefore, this language can be viewed as offensive and counterproductive to creating a logic model. This further exemplifies the power of cultural meanings or interpretation in the written language by emphasizing the importance of careful wording and labeling used in logic models. In other words, language can play a key role in increasing or decreasing stakeholders' receptiveness to logic models.

A counterargument to LaFrance's (2004) choice to relabel "logic models" is that professional fields (including evaluation) need consistent language and approaches to identify themselves as a distinct field (Patton, 2000). This holds true primarily when professionals from the same field communicate with each other. However, this is not a valid argument for communicating with stakeholders that are not evaluators. LaFrance (2004) and practitioners that use the same practice of relabeling words or concepts are in alignment with the common ethical guidelines for evaluation practice. For instance, evaluators must focus "on understanding the unique circumstances, multiple perspectives, and changing settings of evaluations and their users/stakeholders" (AEA, 2018). This entails communicating in a manner that is appropriate for the evaluation context. In the case of LaFrance (2004), this context involved the cultural dynamics of the indigenous stakeholders involved with the evaluation. This is also aligned with the foundational literature that underly CRE principles, in which evaluators are encouraged to respect the different methods of communication by ensuring they are tailoring their

communication to adhere to the culture of the stakeholders (Hood et al., 2010; Frierson et al., 2010). This will likely involve using linguistic translators or interpreters (Clayson et al., 2002; Frierson et al., 2010; Hood, 2009; Hood et al., 2015; LaFrance et al., 2015). Hence, ensuring how evaluators describe or label logic models and their components are culturally tailored to establish a respectful interaction with stakeholders does not interfere with having a consistent language among evaluators.

Using native language or relabeling words based on the cultural orientation of stakeholders in logic models contributes to multicultural validity, which is related to credibility (Alkin & Vo, 2017; LaFrance et al., 2015). Multicultural validity concerns the "accuracy or trustworthiness of understandings and judgments, actions, and consequences, across multiple, intersecting dimensions of cultural diversity" (Kirkhart, 2010, p. 401). This entails using the language of the stakeholder when designing logic models. Yet, interacting with stakeholders whose native language differs from the evaluator(s) can be challenging. A direct translation of languages is often impossible because the word or phrase may not exist in that language. This means evaluators and stakeholders must determine what words or phrases exist to convey an equivalent meaning to ensure a shared understanding of the information provided in the logic model. An example presented by LaFrance et al. (2015) described how the word "validity" does not exist among Tewa and Ojibwe speakers. Rather the closest equivalent words were correctness (kori), truth (ta'ge), and trust (apaenimoondaugaewin). This further conveys the importance of not only using their native language, but the evaluator must select the appropriate translations to avoid confusion, miscommunication, and misinterpretation of the logic model and its portrayal of the program theory. Accordingly, evaluation practitioners and theorists agree that

"appropriate terms familiar to the participants and community members should be used in the materials for the evaluation" (WKKF, 2017, p. 8). This includes logic models.

Beyond multicultural validity, evaluators argue that being attentive to language use in evaluation practice is not just culturally responsive; but equitable. Using a CRE lens to the language used in logic models puts culture to the forefront and helps remove barriers to obtaining a shared understanding of the program (Avant, 2020). Through a language justice framework, language has historically been used as a deliberate or unintended tool of power that can "either include or exclude specific communities from fully integrating into society and gives guidelines for creating multilingual spaces where no language dominates, and people can engage with each other as equals in the language in which they feel most comfortable" (Ghanbarpou et al., 2020; p. 41). Accordingly, being attentive to language choice in evaluation is integral for enhancing clear and transparent communication across stakeholders and is critical to social justice. It can empower meaningful action and create a multilingual environment that welcomes diverse worldviews. This argument supports the notion that language choice in developing a logic model using a CRE lens has the potential to contribute to equitable communication among stakeholders and evaluators.

Language Considerations for CRE LM Designs

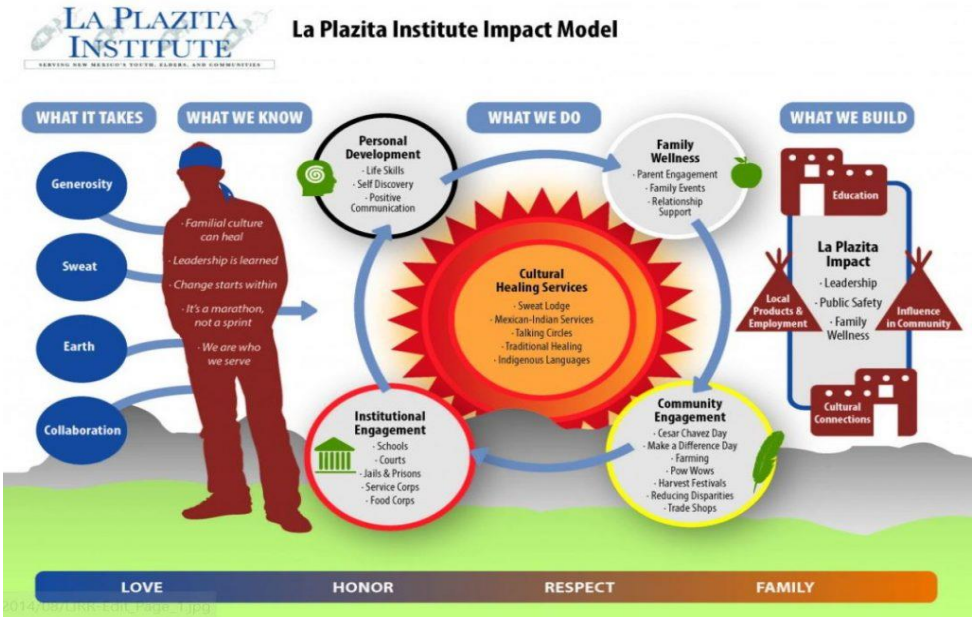
While most guidebooks reflect a preference for individualistic-oriented language in logic model designs, there are examples of designs that include culturally relevant language reflecting the values of the community the program serves. For instance, culturally relevant language is used in the La Plazita Institute (n.d.) Impact Model (e.g., see Figure 1). This institute serves Mexican and American Indian populations in New Mexico that are known to be collectivistic through their values centering around the family, community, and interpersonal relationships as

opposed to the self-fulfillment nature of individualistic cultural orientations (SenGupta et al., 2004). The language used in their logic model design does not reflect the common language used in logic models. Rather the language they use reflects the collectivistic culture of the communities that this institute serves, emphasizing the program's anticipated impact on the community instead of the individual (Triandis & Gelfand, 2012). Consistent with a CRE lens, the designers labeled this as an impact model instead of a logic model (LaFrance, 2012). This highlights their careful selection of labels to reflect cultural values. Secondly, the language used in their model reflects a language justice approach, which focuses on establishing equitable communication (Ghanbarpou et al., 2020). The common use of inputs, outputs, outcomes, and impact to describe logic model components is not used. Instead, the model uses "what it takes," "what we know," "what we do," and "what we build."

Additionally, at the bottom of the model, there are the words "love," "honor," "respect," "and "family." This selection of words communicates the cultural beliefs and understanding of the beneficiaries of the institute and "honors the centrality of culture and adds the importance of communicating across multiple boundaries" (p. 44) and stakeholders that further contributes to language justice (Ghanbarpou et al., 2020; Hall & Hood, 2005). Lastly, the language used in this model reflects the collectivistic values of Mexican and Indigenous communities that place the community's needs above one's individual needs. This is consistent with a CRE lens regarding using the use of language and the values of the community recommended in guidelines for working with Latino and Indigenous stakeholders (Carlson et al., 2017; Chouinard & Cram, 2019; 2020; Clayson et al., 2002).

Figure 1

La Plazita Institute Impact Model



Note. La Plazita Institute (n.d.). *La Plazita institute impact model*. Retrieved October 1, 2020, from https://laplazitainstitute.org/.

The La Plazita Institute Impact Model exemplifies the application of a CRE lens through its use of language that reflects the shared culture of the secondary stakeholders. This intentional use of language may increase the accurate interpretation and credibility of the information presented in the logic model (Frechtling, 2007; LaFrance et al., 2015). However, being intentional about language, even with a CRE approach to evaluation, is not enough. Rather misunderstandings can still occur, especially when there are language and literacy barriers among evaluators or stakeholders. Due to this, including non-textual visualization with written text is often recommended by data visualization and evaluation practitioners (Azzam et al., 2013).

Non-textual Visualizations

Despite the rise in the literature on language and non-textual data visualization in evaluation, there is not nearly as much literature on the sole use of non-textual visualizations beyond graph or chart selection for reporting quantitative data. This is likely because visualizations without written descriptions can contribute to misinterpretations (Goagoses et al., 2020). For instance, Goagoses et al. (2020) gave participants a non-textual sketchbook illustrating two key concepts in ethical research practice (i.e., consent, benefits, representation, authorship/ownership, and partnership), and they provided their written interpretation of them. The participants had low accuracy in their interpretation of the sketchbook, which suggests that textual explanations are necessary for visualizations to enhance accurate interpretation of them. Accordingly, Goagoses et al. (2020) and evaluation experts Torres et al. (2005) advise evaluators to avoid relying solely on written communication when working with stakeholders with literacy or language challenges to minimize the chance of misinterpretation. This is beneficial with data collection using visual response options for participant responses to an oral survey (Connor, 2004). This recommendation is relevant to the design of logic models because without words, logic models are purely a set of interwoven non-textual visualizations that form a larger image. The stakeholders shared culture should play a leading role in what visualizations are chosen for the logic model design, yet common designs of logic models often do not reflect this.

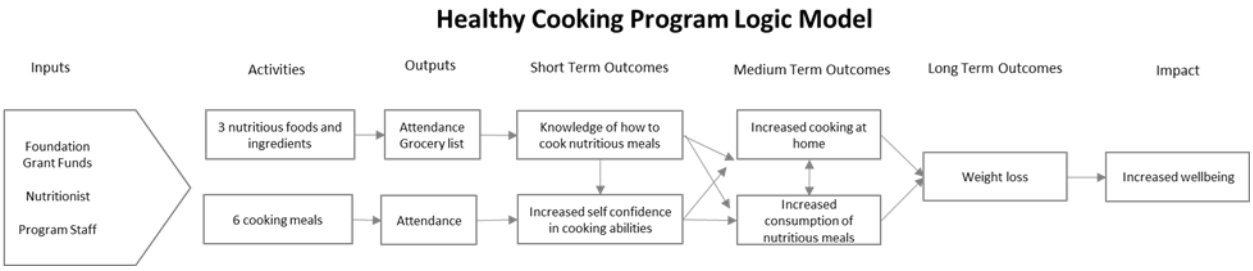
Common Designs of Logic Models

While there is no official gold standard for logic model creation, the common visualizations present in logic models primarily consist of geometric visualizations, which include shapes, lines, or arrows. This is also represented in how logic models are presented in guidebooks for logic model design, such as the W.K. Kellogg Foundation's Logic Model

Development Guide (2004), guidelines from the University of Michigan Extension, and the CDC (2014) materials. However, these well-known materials and approaches are often unidirectional regarding achieving a larger goal, impact, or outcome (e.g., see Figures 2 and 3). Accordingly, these common designs are linear and reflect an individualistic-oriented culture, which is beneficial if this is the culture of the stakeholders. For instance, individualistic-oriented communities interpret and communicate the notion of time linearly (CDC, 2014; Sengupta et al., 2004; Ting-Toomey & Dorjee, 2020). Oppositely, these same designs may not resonate with various collectivistic-oriented communities that may perceive time as cyclical and dynamic, which could impact their interpretation and receptiveness to linear logic model designs.

Figure 2

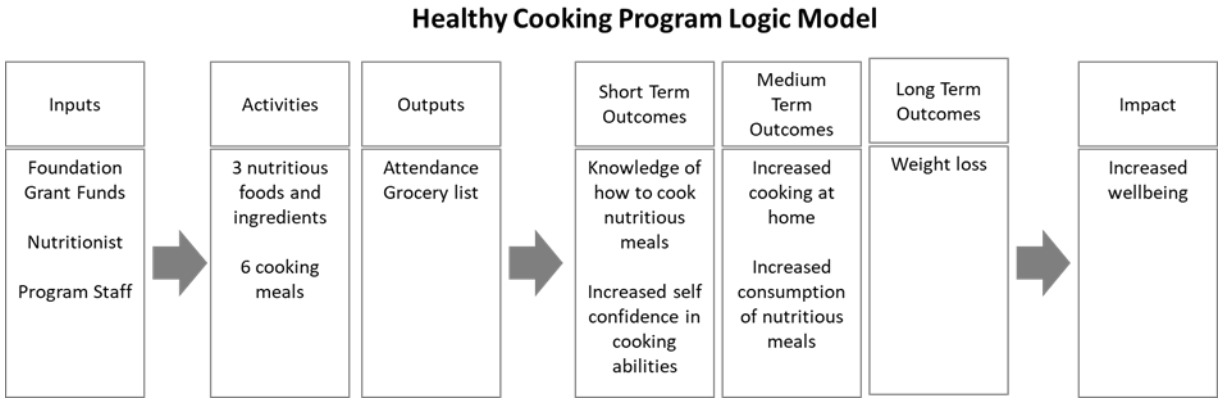
Hypothetical Example of a Common Logic Model Design Utilizing Arrows and Boxes



Note. This design is in the likeliness of logic models presented in W.K. Kellogg Foundation's Logic Model Development Guide (2004).

Figure 3

Hypothetical Example of a Common Logic Model Design Utilizing Columns and Arrows



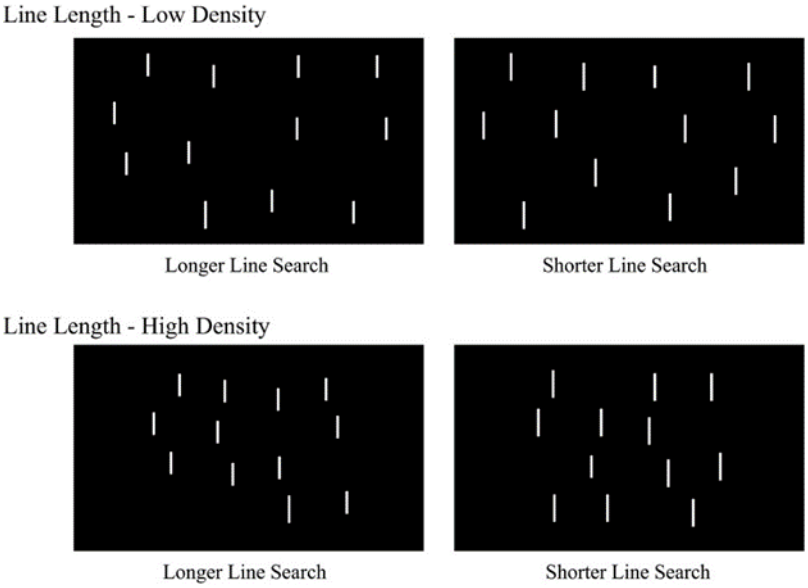
Note. This design is based upon material presented in W.K. Kellogg Foundation's Logic Model Development Guide (2004).

Furthermore, cognition research supports the use of linear designs for individualistic cultures. For example, Ueda et al. (2018) found that North Americans (a predominately individualistic-oriented culture) had a more reliable asymmetric visual processing (i.e., the difference in timely visual search and accurate interpretation of either short or long line length) than Japanese (a predominately collectivist-oriented culture). Specifically, when participants were presented with a basic geometric visualization of lines, they were tasked with identifying either a short line among longer lines or a long line among shorter lines among either high-density or low-density visualization of lines (e.g., see Figure 4). North Americans' performance on the tasks indicated they could search for a long line among short lines faster than vice versa and faster than Japanese participants. In contrast, no search asymmetry was present in any condition for Japanese people. Therefore, not only may common linear logic model designs resonate more with individualistic communities, but the line lengths (when in logic models) also should be strategically arranged for these communities. Given this and how IC cultural

orientations communicate time, it is critical for evaluators to consider how smaller visualizations are used and the recommendations for using them to convey a larger message. This is in addition to the added value of using a CRE lens to guide evaluators in what non-textual visualizations should enhance communication, particularly with complex logic model designs.

Figure 4

Examples of the search displays used in Experiment 1.



Note. Used with permission. From "Cultural differences in visual search for geometric figures," by Y. Ueda, L. Chen, J. Kopecky, E. S. Cramer, R. A. Rensink, D. E. Meyer, S. Kitayama & P. Saiki, 2018, Cognitive Science, 42(1), p. 290 (https://doi.org/10.1111/cogs.12490). Copyright 2017 by John Wiley and Sons.

Complexity in Logic Model Designs

The difference in how IC cultures process line lengths reflects the logic model design limitations presented in the evaluation literature. This limitation concerns reflecting program complexity in a manner that does not overwhelm stakeholders. Accordingly, this has been noted

by evaluators that offered solutions. For instance, Castillo and Grim (2020) argue that the two-dimensionality design of logic models limits one's understanding of the complexity of a program, which often concerns contextual dynamics such as the culture and politics of a program. They used this argument to support the use of 3D logic models.

Similarly, Azzam et al. (2013) presented an example of a multidimensional interactive logic model aligned with the two-dimension limitation of logic models. Their multidimensional model enabled them to reflect the following perspectives from the macro to micro level: community organizational, family, and child-related program activities and outcomes. Moreover, their logic model used a lot of circles and ovals, or ellipses, as opposed to boxes, lines, and arrows. This further shed light on the limitation of only using geometric visualizations in logic model guidelines and may contribute to the lack of contextual complexity reflected in many logic model designs.

Determining how much contextual complexity should be conveyed in logic models is relevant to IC cultural orientations. For instance, stakeholders with individualistic cultural orientations tend to focus on content during communication which entails explicit or direct messaging (Gendron et al., 2017; Gudykunst, 1996; Oyserman, 2017; Reynolds & Valentine, 2011). On the opposite end of the continuum, collectivistic, culturally oriented stakeholders tend to be sensitive to contextual information and communicate in a highly contextually complex manner in which one would need to consider all of the contextual pieces of the message to understand its meaning clearly. Moreover, “visual cues appear to have more importance for members of collectivistic, high-context, relationship-oriented cultures than for members of individualistic, low-context, task-oriented cultures” (Fussell, Zhang, & Setlock, 2009, p. 1801). The nature of this low and high-context communication is demonstrated in Figure 5. These

images both convey adult and youth fishing. Low-context visualization would have an image of the two people fishing. In contrast, a high-context visualization would have the two people fishing and a complete visualization of their surroundings. Accordingly, this provides additional support for the expansion of logic model designs to reflect the complexity of a program's environment that they operate within and with consideration given the cultural orientation of the stakeholders.

Figure 5

Example of Low and High Context Visualization

Low Context High Context

The Role of Non-textual Visualization in CRE

To address the limitations of logic model designs, using a range of non-textual visualizations beyond shapes may be a more straightforward solution. It can influence stakeholders' receptiveness, understanding, and use of the logic model. The symbolic non-textual visualizations exemplify support for this in reporting and descriptions of community-specific

evaluation frameworks (Frazier-Anderson et al., 2012). For instance, Bowman (2020) highlighted how Indigenous nations have historically and presently visualized information. They referenced a visualization used by the Urban Indian Health Institute (2018) of an Indigenous woman wearing a tribal skirt that included statistics of missing and murdered indigenous women and girls (MMIWG; see Figure 6). At the same time, a common yet individualistic-orientated visualization would mostly use various graphs or charts (Douville et al., in press).

Visualizations used in describing community-specific frameworks are not linear. Instead, they reflect a collectivistic cultural orientation using non-linear shapes such as triangles and circles. For instance, Indigenous evaluation frameworks are accompanied by non-textual visualizations that include circles and triangles (Bowman, 2018). Circles reflect the holistic and systems-thinking nature of many Indigenous and collectivistic-oriented communities, in which the past, present, and future, as well as the contextual environment, play a role in communication and understanding (Lee, 1997; Sengupta et al., 2004; Ting-Toomey & Dorjee, 2019). Moreover, evidence from cognitive researchers indicates that individualistic-oriented cultures have a more reliable visual perception of line lengths than collectivistic-oriented cultures (Ueda et al., 2018). Therefore, relying on common linear designs of logic models may not be appropriate for all communities that work with evaluators.

Figure 6

Urban Indian Health Institute Visualization of MMIWG Statistics from a Survey of 71 Cities

Note. From "Missing and Murdered Indigenous Women & Girls" by the Urban Indian Health Institute, 2018. Retrieved January 20, 2020, from https://www.uihi.org/resources/missing-and-murdered-indigenous-women-girls/.

Across the U.S.

Non-Textual Visualizations Used in CRE Logic Model Designs

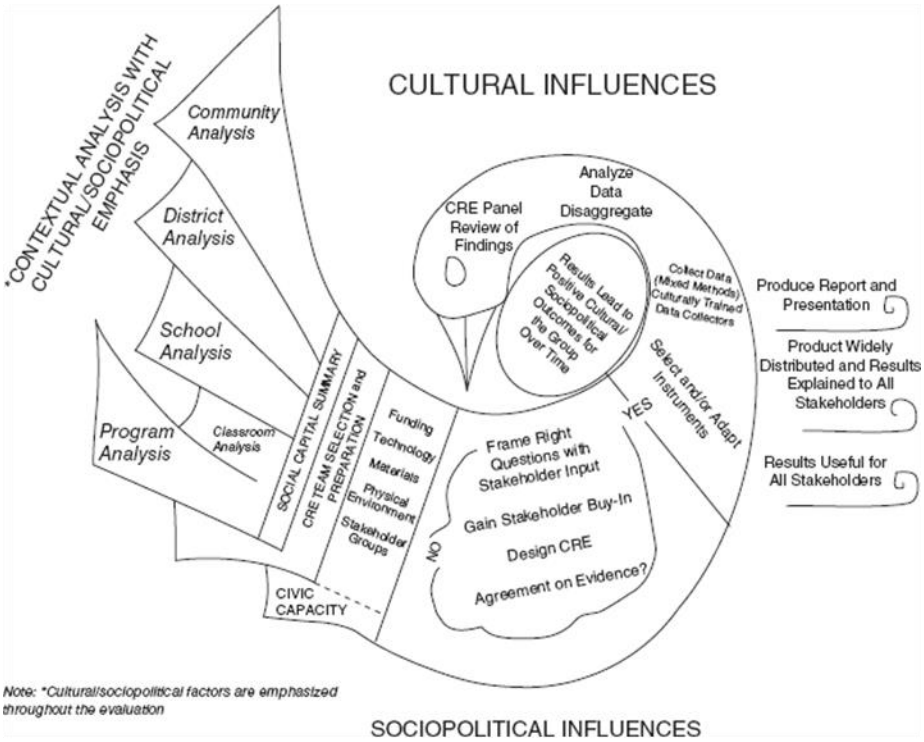
Beyond the use of specific shapes in community-specific evaluation frameworks, imagery can reflect shared cultural ancestry. For example, the African American Culturally Responsive Evaluation System for Academic Settings (ACESAS) is a community-specific CRE framework representing an Afrocentric logic model design in the shape of a symbolic bird called Sankofa (Frazier-Anderson et al., 2012; see Figure 7). Sankofa is an Akan word from Ghana

"represented by Adrinka symbols of a backward-facing bird plucking an egg from its back, or a heart-shaped connection of past and present" (Jones & Leitner, 2015, p. 201). This represents the value of looking towards the past (one's heritage) to guide the direction of one's future. In other words, this logic model represents CRE principles for conducting evaluations within African American communities and utilizes a non-textual visualization. This represents shared ancestry and values and the non-linear manner of collectivistic-oriented communication of time (Ting-Toomey & Dorjee, 2019).

As previously noted, as an example of a logic model whose language reflects the cultural values and orientation of the community the program serves, the non-textual visualizations in the La Plazita Impact Model also reflect them. The overall design depicts Indigenous and Mexican cultural values of relationships with communities, nature, etc. (Clayton et al., 2002; Waapalaneexkweew, 2018). Like Waapalaneexkweew's (2018) visualization for the Culturally Responsive Indigenous Evaluation Model (Stockbridge-Munsee/Lunaape Framing), La Plazita use triangles in their model, which similarly reflects the collectivist nature of the community this program serves as well as symbolizes historical values and customs (i.e., representation of home structures). The model's overall layout is non-linear, reflecting the stakeholder's relational perspective of time as an interaction between the past, present, and future as opposed to the individualistic orientation of time as focusing on the present and future (Ting-Toomey & Dorjee, 2019).

Figure 7

Sankofa Bird Model of the ACESAS



Note. Used with permission. From "Preliminary consideration of an African American culturally responsive evaluation system," by P. Frazier-Anderson, S. Hood, and R. K. Hopson, 2012, Qualitative research: An introduction to methods and designs (pp. 347–372). Copyright 2012 by John Wiley and Sons.

This image of a Sankofa bird and the non-textual visualizations used in Indigenous evaluation frameworks exemplify culturally centered non-textual visualizations used in evaluation reporting for specific cultural communities designed to facilitate communication and CRE practice. Hence, CRE-aligned frameworks incorporate a variety of non-textual visualizations, beyond boxes and lines, as appropriate for the culture of the stakeholders. This emphasizes the importance of evaluators increasing their range of visualization used to suit the stakeholders' appropriate cultural orientation.

Expanded Possibilities for Non-textual Visualization for Logic Model Designs

With the understanding of how stakeholders communicate and interpret information, evaluators must stay abreast of the diverse range of non-textual visualizations they can use to design culturally responsive logic models. Other options have potential, especially when working with collectivistic-oriented communities responsive to high-context communication (Gudykunst et al., 1996). These additional design options include cartoons, photos, or icons representing the shared cultural representations among the stakeholders. These have more potential for adding contextual complexity to logic models in a manner reflective of either the low context nature of individualistic or the high context nature of collectivistic-oriented cultures despite the dimensionality of the design.

Cartoons. While not commonly seen, cartoons (or comics) have been used in evaluation not only to enhance communication among stakeholders but also to increase their attention and interest in the evaluation (Chin, 2005)., There is evidence of cartoons used for evaluation training and reporting (Chin, 2003; Hutchinson, 2017). Cartoons are also ideal for conveying meaning and contextual dynamics of the program theory and triggering emotions that may influence action among stakeholders in a visual storytelling format (Hutchinson, 2017). Additional evidence derived from data visualization experts also indicates that this visualization method can help stakeholders quickly understand complex yet temporal information using a very minimal text to accompany it (Bach et al., 2016). This is ideal for collectivistic-oriented cultures that may engage in high-context communication (Gendron et al., 2017; Gudykunst, 1996; Oyserman, 2017; Reynolds & Valentine, 2011). Moreover, this is particularly beneficial when there are linguistic challenges that evaluators must navigate to be culturally responsive to stakeholders (Hood et al., 2015). Lastly, cartoons can easily be tailored in logic models to either

reflect low-context visualizations aligned with individualistic cultures or high-context visualizations aligned with collectivistic cultures (See Figure 5).

Photos. While it is rare that evaluators utilize photos in their practice, they are a promising practice (Fang, 1985; Patton, 1990). For example, evaluators have used photos as a data source to facilitate communication and to influence or build credibility among stakeholders (McAlindon et al., 2019; Torres et al., 2005). For instance, Bessell, Deese, and Medina (2007) highlighted the potential advantages of using photos in evaluation practice to help elicit desired responses or actions from stakeholders through their emotions. To do this, the evaluators used Protolanguage, a relatively new strategy therapists use to present black-and-white photos to clients to elicit an emotional response or assist them with self-expression. A key outcome of using this strategy included increased engagement of low-literacy stakeholders. These stakeholders spoke more and provided more detailed responses than stakeholders that did not have the photolanguage option. Also, photolanguage elicited more feedback and enthusiasm among youth stakeholders. Lastly, the evaluators noted that they successfully used photolanguage with diverse stakeholders across cultural, racial, socioeconomic, and age groups. Therefore, this supports using photos in evaluation practice, especially in designing logic models from a CRE lens.

Icons. While photos and cartoons are ideal for conveying context in logic models design, icons are great with text for conveying cultural representation and meaning. Icons are visual and often symbolic labels used in place of words or numbers. They are commonly used in flow charts, concept maps, mind maps, infographics (a combination of icons and words), pictographs, or pictograms to communicate the meaning derived from data (Azzam et al., 2013; Evergreen, 2019; Hutchinson, 2017). Moreover, "…all cultures have symbols that convey meaning. For

example, many North Americans understand that a light bulb means an idea, crossed swords mean conflict, and linked hands mean harmony. However, these symbols are cultural and may have no meaning or different meaning in another context" (Knowlton & Phillips, 2013, p. 88). Relatedly, La Plazita Institute Impact Model is an infographic version of a typical logic model (See Figure 1). They use icons of a teepee, adobe buildings, a sun, and a feather to tell the story of their program theory with culturally based visualizations. Accordingly, icons are visualizations valued for their potential to present the program theory in a storytelling manner to help stakeholders understand complex processes or relationships presented in logic models (Hutchinson, 2017). Icons are also designed to aid in quick comprehension of information, which aligns with a key benefit of using logic models (Azzam et al., 2013; Jones et al., 2019).

Recommendations for Using Cartoons, Photos, and Icons.

Although the differences in the visual design and uses of cartoons, photos, and icons, evaluators have provided guidelines for general reporting and communication that apply to all of them. These recommendations include identifying the larger story, selecting representative visualizations, using text for clarification, and monitoring the use of humor. These recommendations further enhance culturally tailored communication when coupled with CRE principles and expert insights.

For icons, photos, and cartoons to contribute to quick and accurate interpretations of the program theory, evaluators must first identify the story that needs to be represented in the logic model (Knaflic, 2015). For instance, icons, particularly in infographics, have a bad reputation for being poorly done. As a result, these visualizations tend to be presented overwhelmingly, which may contribute to cognitive overload. This partly explains why experts advise designers to prioritize visualizations pertinent to communicating the identified story (Hutchinson, 2017;

Knaflic, 2015). Hence, this story should align with the cultural values, perspectives, historical antecedents, and relational styles of the lived experience of the stakeholders (Hood, 1998; 2009; Frierson et al., 2002; Frierson et al., 2010; Frazier-Anderson et al., 2012; Kirkhart, 2013). Such an alignment means that the visualizations are biased towards the culture of the stakeholders, which typically calls for a CRE evaluator to engage in the reflexivity element for assurance of multicultural validity (Hood et al., 2015; Kirkhart, 2013). This permits the evaluator to tailor communication beyond their cultural values to the stakeholder's cultural orientation to ensure an accurate understanding and interpretation of the program theory.

This alignment of the story with the culture of the stakeholders also applies to how evaluators should select icons, cartoons, or photos for use in logic models. Torres et al. (2005) recommend using photos representative of stakeholders' shared values and perspectives. This recommendation applies to icons and cartoons. For instance, humor is cautioned when working with specific cultural communities, particularly in cartoons. Humor is only sometimes translatable and appropriate in certain cultures and can be detrimental to the credibility of the evaluation (Alkin & Vo, 2018; Torres et al., 2005). Therefore, humor should not be used with these visualizations if the evaluator cannot align or determine if it is appropriate for the cultural values and norms of the stakeholders.

The cultural representation of selected visualizations is aligned with the voice element of multicultural validity that underlies CRE. This involves the engagement and collaboration of marginalized community members, "[welcoming] new participants or audiences…[employing] culturally appropriate communication strategies…. [and] using accessible strategies for clear oral, visual, and written communication of findings to primary perspectives authentically" (Kirkhart, 2013, p. 150). Nevertheless, selecting such visualizations for a culturally responsive

logic model entails an in-depth collaboration between the evaluator and stakeholders. This collaboration is necessary to adhere to the ethical recommendation and practice of obtaining permission to take and use photos as part of communication and reporting (AEA, 2018; Torres et al., 2005). Furthermore, consulting stakeholders is necessary to ensure appropriate cultural interpretation and that icons or cartoons are respectful, accepted, and deemed credible (Hood et al., 2015; Clayson et al., 2002).

A final recommendation for using photos, cartoons, and icons concerns using text to accompany non-textual visualizations. Torres et al. (2005) suggest that call-outs, captioning, or text used within the cartoon illustrations should be written using straightforward language at an elementary reading level to ensure clear communication. This includes using narrative text to explain the cartoon's content to ensure the message is conveyed to stakeholders. This recommendation is consistent with the widespread practice of including a written narrative to provide an additional explanation of the logic model (Jones et al., 2020; UW-CE, n.d.). CRE principles and evaluators further clarify this recommendation by advocating for the use of significant words or phrases that are preferably in the native language of the stakeholders when applicable (Clayson et al., 2002; Frierson et al., 2010; Hood, 2009; Hood et al., 2015; LaFrance et al., 2015). This likely further increased the attention, interest, and accuracy of stakeholder interpretation of the program theory and enhanced the clear and quick communication of complexity, especially with cartoons, across cultural orientations. It also aids in designing a culturally responsive logic model that resonates with the stakeholders.

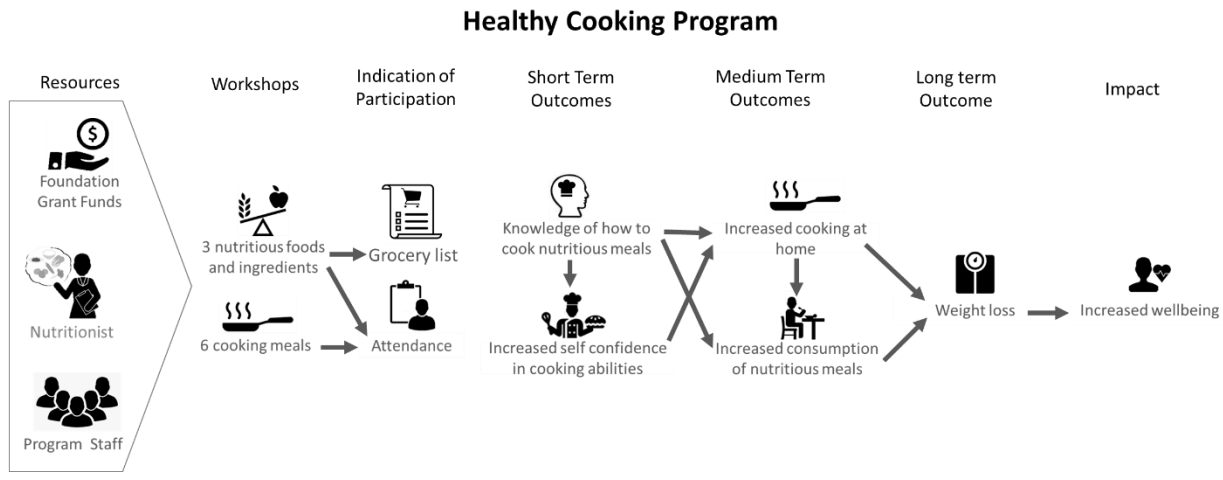
Extended Non-Textual Visualizations Used in CRE Logic Model Designs.

Using photos, icons, and cartoons in logic model designs also stresses the importance of evaluators being aware of the vast possibilities of non-textual visualizations that can be used to

enhance communication through a CRE lens. Whether these additional possibilities are considered when designing logic models, it is crucial to understand what a CRE-driven logic model design may look like based on IC cultural orientations. Given this, along with stakeholder language and literacy considerations, a CRE logic model based upon an individualistic cultural orientation of stakeholders would resemble many logic models present in common guidebooks without utilizing any extended non-textual visualizations in its design. However, all individualistic-oriented logic model designs would have direct or low-context messaging (Gendron et al., 2017; Gudykunst, 1996; Oyserman, 2017; Reynolds & Valentine, 2011). In addition, they would be linear in their overall design (i.e., they may include many lines and arrows; Ueda et al., 2018). Lastly, the outcomes would be self-promotive or individual-focused in the selected language or phrasing (i.e., increased self-confidence in Figure 8) and visualizations (Lalwani et al., 2009; Lee et al., 2000; Triandis & Gelfand, 2012).

Figure 8

Hypothetical Example of an Individualistic-Oriented Logic Model Using Icons



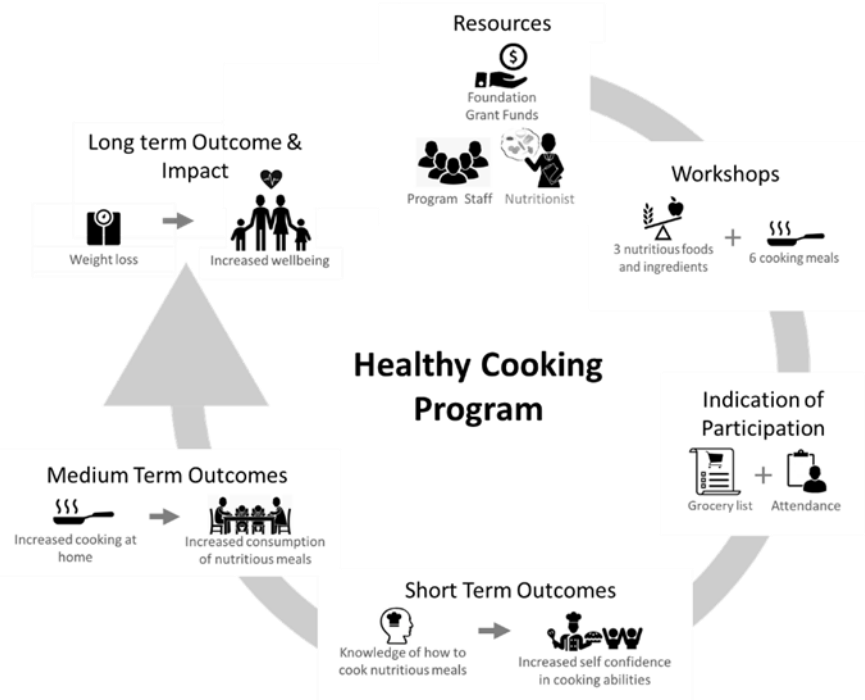
This understanding of individualistic-oriented cultures, reflected in Figure 8, represents a simple hypothetical example of a CRE logic model using icons in its design. The linear logic model reads from left to right and includes multiple arrows to show linear-based connections between components. Icons visualizing a single individual were strategically placed throughout the logic model. Specifically, they are used to represent increased self-confidence in cooking abilities (i.e., short-term outcome), increased consumption of nutritious meals (i.e., medium-term outcome), and increased well-being (i.e., impact). These icons were selected based on the values and communication of individualistic-oriented cultures, which are person-centered (e.g., Figure 8).

On the other end of the spectrum, CRE logic models based upon a collectivistic cultural orientation of stakeholders would have indirect or high-context messaging (Gendron et al., 2017; Gudykunst, 1996; Oyserman, 2017; Reynolds & Valentine, 2011). Their overall design would be primarily cyclical or non-linear (e.g., see Figure 1). They may include minimal lines or arrows, and they may be curvilinear. Lastly, the outcomes would be group, family, or community focused on the selected language, phrasing (i.e., increased confidence of community), or visualizations (Lalwani et al., 2009; Lee et al., 2000; Triandis & Gelfand, 2012). While the La Plazita Institute Impact Model exemplifies a CRE logic model using icons designed for collectivistic, culturally oriented communities. Figure 9 also represents a hypothetical CRE logic model using icons.

Unlike Figure 8, the design of Figure 9's logic model is non-linear, with the components reading along a circular pattern. Also, the icons used in the short- and medium-term outcomes and the program impact reflect a family. Since collectivistic-oriented cultures heavily value family and the community, these icons will likely draw their attention and resonate with them.

Figure 9

Hypothetical Example of a Collectivistic-Oriented Logic Model Using Icons



Therefore, Figure 9 would likely draw the attention of collectivistic-oriented stakeholders and assist with a timely and accurate interpretation of the program's theory using icons to enhance this communication by emphasizing collectivistic values.

Despite the IC orientation of stakeholders, evaluators should use simple language that lay communities would understand (Few & Edge, 2007; Torres et al., 2005). Accordingly, both hypothetical examples presented in Figures 8 and 9 have similarities based on common guidelines for effective communication using written text. The title of both models is the Healthy Cooking Program, which excludes the "logic model" from it. Logic models are typically well-known among evaluators, funders, and stakeholders representing upper management (Chouinard

& Cram, 2019; Madison, 2000). Often evaluators must explain what logic models are, define their purpose, and describe their components to stakeholders.

Given that CRE centers its focus on secondary stakeholders, including the "logic model" in the title is inappropriate. This was further indicated in Indigenous evaluation literature, where the word logic may illicit negative connotations (LaFrance, 2004; 2012). Nevertheless, visualizations should have a clear, simple, and informative title to reduce the mental effort needed to understand the logic model's overall purpose (Wanzer et al., 2021). Given this, the Healthy Cooking Program is an appropriate title that addresses these recommendations.

To further increase the understanding of both hypothetical logic models, the descriptions of the components were slightly altered in Figures 8 and 9. Both used resources, workshops, and an indication of participation as labels for inputs, activities, and outputs of logic model components. As indicated in the rationale for the title, these labels are also clear and direct so stakeholders may understand them (Few & Edge, 2007; Torres et al., 2005). However, it is essential to note that using inputs, outputs, outcomes, and impact is suitable when logic models are created for funders as they are familiar with this language (Chouinard & Cram, 2019; Madison, 2000). Hence, evaluators should thoroughly know their audience of the logic models they create rather than constantly using the same language in their designs.

Lastly, the language could further differ between the two hypothetical logic models given the community-specific values of the program as well as its outcomes and goals (See Figures 8 and 9). For instance, if the stakeholders collectively speak a different language, the evaluator should, at minimum, strategically place words in the logic model in the community's language. If the evaluator worked within a Spanish-speaking community with collectivistic values, the word "familia" (i.e., family) could be placed throughout the logic model. Many Spanish-

speaking communities have a collectivistic cultural orientation, and their values are firmly centered around family (Clayson et al., 2002; Hofstede, 1991; Ting-Toomey, & Dorjee, 2019). At the same time, evaluators must be attentive to the overall program goals and outcomes when strategically determining whether to place these words and phrases in the logic model. Outcomes and impact help evaluators determine what should be measured and how to do it (Knowlton & Phillips, 2013). Accordingly, a program with community-level ambitions should also have community-level (i.e., group-level) measures (Zumbo & Forer, 2011). For instance, a program with community-level ambitions and a collectivistic cultural orientation should have language in the program outcomes and impact that reflect this. In Figures 8 and 9, the short-term outcome of "Increased self-confidence in cooking ability" is worded to reflect individualistic values. For stakeholders that represent collectivistic values and have community-level outcomes, this short-term outcome should be changed to "Increased confidence in family's cooking abilities." However, when a program has individual-level outcomes for collectivistic-oriented stakeholders, various visualizations such as icons, photos, and cartoons can also communicate collectivistic values. In contrast, the outcomes remain at the individualistic level of measurement, as depicted in Figure 9.

In all, this entails evaluators should be knowledgeable about the visualization tools available to them beyond geometric images. Unfortunately, this competency is lacking among the AEA's (2018) core competencies for evaluation, which are commonly used in the U.S. to determine evaluators' training needs and qualifications. They include elements relevant to designing CRE and communication tools, such as logic models. For instance, one professional competency concerns exhibiting respect for stakeholders of diverse backgrounds. This is important for selecting visualizations for culturally responsive logic model designs.

Additionally, two competencies directly relate to logic model design and its use as a communication tool. One is a methodological competency that involves "[Using] program logic and program theory as appropriate." At the same time, the other is a context competency that concerns "[communicating] evaluation processes and results in timely, appropriate, and effective ways" (AEA, 2018). Yet a core competency concerning evaluators' utilization of visualizations for creating culturally appropriate logic models as an effective communication tool is missing. Accordingly, this domain is rare among the common training and development of evaluators.

Implications of the Literature

The body of literature covered in this review exemplifies how culture is unavoidable in evaluation. Culture was represented among both CRE-specific and non-specific evaluation literature alike. Insights from general evaluation and CRE-specific evaluation literature, along with support from IC literature, emphasized the feasibility and need for logic models that reflect the culture of the stakeholders. In addition, there are more critical aspects of language, and non-textual visualization selection evaluators should address them when designing logic models. To further support this, the evaluation literature was aligned with literature on IC cultural orientations. This added more depth to understanding how underlying cultural values are represented in how people communicate and interpret information, which is important for designing logic models.

Consequently, a lot of common evaluation literature predominantly reflects individualistic cultural orientations. Specifically, common evaluation literature highlighted the importance of language use in evaluation in enhancing understanding primarily with simple words or phrases and using non-textual visualizations to ensure stakeholders with either language or literacy barriers can understand the program theory (Torres et al., 2005). However, CRE

literature builds upon these recommendations for language and non-textual visualization use in logic model designs by adding insights about collectivistic-oriented cultures, equitable communication, and facilitation of a meaningful program theory by reflecting the stakeholders' cultural values. This can be done by using the native language, selecting words or phrases in the native language, or conveying cultural values in the non-native language of the stakeholders. Accompanying this careful language selection with culturally relevant non-textual visualizations not only addresses language and literacy barriers but also draws interest, encourages action, and invokes appropriate emotions from the stakeholders. Such careful selection of both the words and non-textual visualizations in logic model design ultimately ensures evaluators present the program theory in a manner that reflects the stakeholder's epistemological perspective, contributing to multicultural validity and increased credibility of the logic model.

Given the above, the IC literature provides specific information about how to design logic models based on the cultural orientation of the stakeholders. In addition, this information includes specific values, words, and visualizations that people representing each orientation use when communicating with others. Lastly, this literature base aligned with common and CRE-specific literature, making this critical for minimizing its disconnect with logic model design guidelines. Accordingly, additional research is necessary to explore using IC as a mechanism for bridging CRE with logic model design.

Chapter II Research Questions for the Study

This study explored measuring learning from logic models, determining valid measures for IC interpretation, and investigating whether utilizing culturally tailored logic models impacts stakeholders' interpretation. A brief recount of the literature is discussed. Lastly, specific research questions for this study are detailed in this chapter.

Measuring Learning from Logic Models

A premise for using logic models in evaluations is that they help stakeholders understand the program theory and could clarify knowledge about the relationship between activities and outcomes or goals and how program information is organized into logic model components (Alkin & Vo, 2018; Knowlton & Phillips, 2013). This can influence perceived program credibility, likely leading to use (Jones et al., 2019; Kaplan & Garret, 2005). Much of the evaluation literature includes narrative recounts to support the logic model's influence on stakeholders' perceived credibility of a program's theory with minimal quantitative measures. However, no agreed-upon measure is used to determine whether stakeholders understand the program theory presented in a logic model. This is the first step towards accessing credibility and using a logic model. Due to this, the following exploratory research question (RQ) is:

RQ 1: How can learning from a logic model be assessed?

Individualism and Collectivism and Logic Model Interpretation

Unlike logic models, individualism and collectivism (IC) are key constructs used to operationalize culture that support standardizing culturally responsive evaluation (CRE) to guide effective communication with stakeholders. Measures of IC have been used to prime people to adhere to either individualistic or collectivistic cultural values, which undergirds how people communicate (Chouinard & Cram, 2018; McBride, 1998; Oyserman & Wing-sing Lee, 2008;

Triandis et al., 2012). Many scaled measures have also been used to distinguish between people with individualistic and collectivistic cultural values in addition to assessing demographic differences (Chen et al., 2015; Gudykunst et al., 1996; Oyserman, 1993; Triandis et al., 1988; Wagner & Moch, 1986; Wagner, 1995). However, since evaluation involves working with stakeholders' pre-established cultural values, priming is not a realistic practice. Given the availability of multiple measures for IC, the following exploratory research question is:

RQ 2: Which measures of IC would be more likely to help inform how people interpret logic models differently?

Language in Logic Models

Evaluation literature also supports the notion of being attentive to language to influence the stakeholders' credibility of the program (Hood et al., 2015; Mason & Azzam, 2019). CRE further supports this by providing explicit principles for communicating with stakeholders based on their cultural values (Chouinard & Cram, 2018). When using IC as an operationalization for culture, studies have found that individuals tend to differ in their communication through the written language they use, with individualistic-oriented people tending to emphasize the self in their communication and collectivistic-oriented people emphasizing the group or their group membership (Kashima & Kashima, 1998; Ting-Toomey & Dorjee, 2019). This has potential applications to how language may be culturally tailored in logic model designs which lend to the exploratory research questions:

RQ 3: What impact does culturally tailored language in logic model designs have on the effective communication of a program theory to stakeholders?

RQ 4: What impact does culturally tailored language in logic model designs have on the credibility of the program theory?

Visualizations in Logic Models

People in IC cultures differ in how they visually communicate (Bowman, 2018; Gudykunst et al., 1996; Ueda et al., 2018). This is made much more feasible due to the increased availability of accessible data visualization technology and strategies (i.e., use of cartoons, icons, and photos) that can quickly minimize evaluators' reliance on linear and geometric logic model designs, which may not always resonate stakeholders (Bledsoe & Donaldson, 2015; Azzam et al., 2013). These various visualization strategies have also impacted credibility, receptiveness, and attentiveness. Given this, the following exploratory research questions are:

RQ 5: What impact do culturally tailored visualizations in logic model design have on the effective communication of a program theory to stakeholders?

RQ 6: What impact do culturally tailored visualizations in logic model design have on the credibility of the program theory?

Combining Language and Visualizations in Logic Models

Altogether, logic models often consist of written language and visualizations, a supported practice among data visualization specialists and researchers (Goagoses et al., 2020). This combination of communication strategies should also aid in effective communication and thee credibility of the program. Therefore, the following research questions are also for this study:

RQ 7: What impact does the combination of culturally tailored language and visualizations have on the effective communication of a program theory to stakeholders?

RQ 8: What impact does the combination of culturally tailored language and visualizations in a logic model design have on the credibility of the program?

Chapter III Exploratory Multiphase Mixed Methods and Analyses

An exploratory multiphase mixed methods design was conducted to address the eight research questions. This included three phases, each with a different design: 1) exploratory sequential design (Phase 1), 2) quantitative two-group comparison design (Phase 2), and 3) explanatory experimental design (Phase 3). Each phase was conducted consecutively because the previous study informed the subsequent study (See Figure 10). The description of methods, participants, measures, and procedures for each phase are described in this chapter.

Figure 10

Exploratory Multiphase Mixed Methods Design

Phase 1 Exploratory Sequential Design: LM Knowledge Test Development and Content Validation

	Procedures	*Product*
Data Collection	Quant: 2 structured measures & demographic survey QUAL: Semi-structured focus group	Quant: Numerical item scores QUAL: Notes
Data Analysis	QUAL: Content Analysis QUANT: Fleiss Kappa	QUAL: Codes, categories, and themes QUANT: Logic Model Knowledge Test

Phase 2 Quantitative Two-Group Design: IC Measure Selection

	Procedures	*Product*
Data Collection	QUANT: 6 structured measures & demographic survey	QUANT: Numerical item scores
Data Analysis	QUANT: T-Test	QUANT: IC measure

Phase 3 Explanatory Experimental Design: CRE Logic Model

	Procedures	*Product*
Data Collection	QUANT: 7 structured measures & demographic survey Qual: 3 open-ended questions	QUANT: Numerical item scores Qual: Open-ended responses
Data Analysis	QUANT: ANOVA Qual: Content Analysis	QUANT: Numerical item scores Qual: Codes, categories, and themes

Instruments and Measures

A total of 11 measures were used for this multiphase mixed methods study using various data collection formats (i.e., focus group and online surveys). In addition, Phase 3 included three logic model conditions plus a control condition aligned with each of two cultural orientations (i.e., seven different logic model designs), individualistic-collectivistic (IC) orientation as an independent variable, and visual efficiency (VE) and perceived credibility of the program (CP) was dependent variables. Accordingly, each measure, instrument, and condition for Phase 3 are described.

Procedures

Each phase was approved by Claremont Graduate University's (CGU) Internal Review Board (IRB) to ensure ethical research practice as starting procedural point. Participant recruitment and data collection did not begin before each approval was granted. Additional details about the procedures involved in each phase are described.

Analyses

Quantitative and qualitative analyses were conducted with various data sources across the three phases. Phase 1 consisted of pre and post-test agreement analyses of experts, qualitative insight from experts, and test attributes to address the research question (RQ) 1. Accordingly, qualitative and quantitative strands were equally prioritized for the study. Phase 2 prioritized the quantitative strand, which consisted of analyses of three individualism and collectivism (IC) measures using MTurk data to address RQ 2. Finally, phase 3 consisted of an experiment and prioritized the quantitative strand using Mturk data to address RQ 3-8. The qualitative strands for Phases 2 and 3 were analyzed via a content analysis of qualitative data derived from the program description, stated contributions to their mental effort (ME) rating, manipulation check, and

opinion about the study prompts (Creswell & Poth, 2017). In addition, emerging codes and themes that provided additional context to quantitative findings were discussed. Also, critical cases were scrutinized based on responses from the manipulation check to determine if they should be removed from the dataset. The specific methods, results, and discussion are detailed in the subsequent sections organized by each phase.

Chapter IV Methods, Analysis, and Discussion of Each Phase

Phase 1 Exploratory Sequential Design: Logic Model Knowledge Test Development and Content Validation

Quantitative and qualitative data were collected concurrently from an online survey and concurrently from a subsequent focus group data collection. Then, the last data collection only used a quantitative strand to assess the internal reliability of the measure. Given that the purpose of this Phase 1 was to develop and validate a logic model knowledge test (i.e., RQ 1), this method was appropriate since it is commonly used for this purpose (Creswell & Clark, 2017). Sequentially, the qualitative feedback from experts was quantitively tested.

For this phase, qualitative and quantitative strands were equally prioritized, along with a concurrent timing of both strands from expert data. In addition, the interpretation of each strand was done interactively for the survey and focus group of experts to inform better the changes made to the Logic Model Knowledge Test (LMKT). Lastly, the two strands were merged for the expert data during the analysis. This did not apply to the final quantitative data collection, which was for assessing the reliability of the measure.

Participants

Data from two kinds of participants were sampled for this multiphase mixed methods study. Phase 1 included subject matter experts to provide insight into the LMKT development and validation. The second aspect of Phase 1 and all of Phase 2 used the same sample of Amazon Mechanical Turk (MTurk) crowdworkers. Lastly, Phase 3 used a separate sample of MTurk participants. Details about each sample are described below.

Expert Participants. For the pretest, participants were recruited utilizing a purposive reputational sample (in which the researcher knows the individuals invited to participate in the

study; Bamberger & Mabry, 2020). They included eight non-unique advanced CGU evaluation students and an alumnus who completed their course requirements. Most of the participants were women (87.5%), non-Latino White (62.5%), and had a graduate education (75%; See Appendix A). Furthermore, most of them identified as evaluators (62.5%), and all indicated either proficient or expert level of competence in using logic models, communicating logic model benefits, survey editing, survey development, measurement editing, and measurement development (e.g., see Appendix B). Hence, they are identified as experts in this study. Lastly, of those that participated in the survey, five non-unique online experts gave further insight in the focus group session.

Amazon Mechanical Turk (MTurk) Participants. Data from 87 MTurk participants were collected for Phases 1 and 2. Eight percent (n = 7) of responses were removed from the dataset because they were bots (n = 5). For instance, responses merely copied words and phrases from the logic model that do not fully represent the description of the program theory (e.g., "nutritionist," "foundation grant funds 5 personal nutrition and foods"). Other responses suggested the participants either did not understand the question or they were impersonators (n = 2; Sharpe Wessling et al., 2017). For instance, responses appeared to be taken from existing text elsewhere (e.g., "'Make in India' had three stated objectives: to increase the manufacturing sector's growth rate to 12-14% per annum; to create 100 million additional"). The remaining 80 responses were used for data analysis. Half of the MTurk participants were from India and the other half were from the United States (US). Most MTurk participants were men (68%), with 75% of men in India and 60% of men in the US. This group was highly educated, with most having college degrees (89%). The overall median age was 30.5 years old (M = 33.73, SD = 9.56), with the age of MTurk participants in India (M = 29.95, SD = 5.17) significantly younger

than those in the US (M = 37.5, SD = 11.36), t(54.48) -3.83, p < .001, Cohen's d = .86. Lastly, many MTurk participants perceived they were above the median point range of perceived socio-economic class (SSCM; M = 4.77, SD = 1.91). However, MTurk participants in India had a mean score of 5.46 (SD = 1.14), which was significantly higher than the US participants MTurk participants who had a mean score of 4.09 (SD = 2.26), t(57.57) = 3.42, p < .001, Cohen's d = 77. Those that identified themselves as Latinx also had significantly higher perceived social class scores (M = 5.79, SD = 1.19) in comparison to Non-Latinx participants (M = 4.41, SD = 2.00), t(59.60) = 3.73, p < .001. Additional demographic characteristics can be viewed in Appendix C.

Measures and Instruments

Five instruments were used for Phase 1. The first version of the newly developed LMKT was incorporated into the first online survey for the subject matter experts and discussed in a focus group. A questionnaire was used to guide the focus group session. This was followed by a second version of the LMKT based on expert feedback used in an online survey for non-expert participants. Lastly, two demographic surveys were used to collect data from subject matter experts and non-subject matter experts, respectfully. Each instrument is described in detail.

LMKT. This 16-question test was developed based on the three isolated benefits of logic models and two purposes. The purposes of determining whether respondents understand the logic model components were captured using nine questions. Their understanding of the content of the program theory was captured using seven questions adapted from Jones et al. (2020) to reflect the current logic model. The lowest possible score was 0, meaning the participant did not answer questions correctly. The highest possible total score on this measure was 16, meaning the participant correctly answered all the questions and demonstrated an understanding of the program theory and logic model components.

Lastly, an attention check question from Jones et al. (2020) was included in this test. The question asks, “What is the title of this image?” (e.g., Appendix D). This question determines whether participants read or understood the questions instead of selecting random responses (Crano et al., 2015). This was a vetted technique used in many measure designs.

Online Pretest Survey for Subject Matter Experts. The online survey via Qualtrics was used for the pretest. It had 34 questions, with four inquiring about the benefits of logic models and 16 about whether specific LMKT questions should be removed, kept, or edited. When presented with an LMKT question, the experts were asked, “Based upon your understanding of the benefits of logic models and the purpose of this test, what should be done with Question #1?” They were given the following response options: 1) It should be kept in the test, 2) It should be removed from the test, 3) It should be edited (then a space for editing suggestions was provided), and 4) I do not know. The survey concluded with 13 demographic questions about the experts’ background and evaluation experience (e.g., see Appendix E and F).

Focus Group Posttest Questionnaire. The semi-structured focus group l was used for the post-test. The focus group protocol included six questions designed to elicit feedback regarding the alignment of LMKT questions with the stated benefits of using logic models. The guidelines for asking the questions in the focus group depended on the initial analysis of the online survey data. See Table 2 for sample open-ended focus group questions based on the online survey Fleiss Kappa (κ) levels.

Table 2

Focus Group Process and Questions Based Upon Online Survey Fleiss Kappa (κ) Levels

Agreement	κ	Action	Sample Questions
1. **Benefits of logic models**	.80 or less	Discuss with the focus group	What are the additional benefits of logic models that should be reflected in this test? Why?
2. **Question removal**	.61-1.00	Discuss with the focus group, then remove the test question	Why should this question be removed?
3. **Question edits**	.41-1.00	Discuss with the focus group	Why should this question be edited?
4. **Final LMKT**	.61-1.00	Finalize the test with edits made by the focus group	N/A

Procedures

Upon approval from CGU's IRB, evaluation experts were recruited via email and LinkedIn to participate in an online survey followed by a virtual focus group via Zoom. They were given informed consent and notified of the purpose, procedures, benefits, risks, and voluntary nature of their participation in the survey and focus group (e.g., see Appendix F). For the online survey, the experts were asked to review the benefits of logic models as criteria for the LMKT. Then, they were asked to review a logic model used in Phase 3 (e.g., see Appendix G, Condition 2 Individualistic Language). The subsequent questions asked whether each LMKT should be removed, edited, or kept. Lastly, they were asked about their personal and professional background using demographic questions. Afterward, they were invited to a follow-up focus group session, where they were reminded of the informed consent. With verbal consent from the experts, they reviewed the survey findings for each benefit and LMKT question at a time. Upon reviewing each finding, they were guided into a discussion about whether the benefits or questions should be added, removed, or edited. The next set of online survey findings was

presented when a consensus was reached. This process repeated itself until the focus group time was completed. The evaluators were thanked for their participation, and the session concluded.

The LMKT was updated to reflect the expert feedback. Then, with CGU's IRB approval, MTurk participants were recruited to complete the test to complete Phase 1. First, they were presented with the title Survey about Understanding a Data Visualization followed by the following description:

> Review an image describing a program and answer questions concerning understandability and your personal perceptions. ***Disclaimer*** This HIT must be taken on a computer or laptop, not on a phone or tablet.

Then, they are redirected to the LMKT posted on Qualtrics.com. Participants were first presented with an informed consent which described the research and IRB contact information, the purpose of the study, a summary of the tasks, state the voluntary nature of the study, and notified of the payment of $4.00 they will receive upon completion of the study (e.g., see Appendix H). If they did not consent to participate, they were redirected to the end of the survey and thanked for their interest in the study. Participants that consented to the study were asked to review an image of a logic model and answer questions about that image. As part of Phase 2, they were also asked to complete each of the three IC measures (e.g., see Appendices I-K) followed by a demographic survey (e.g., see Appendix L). Lastly, they were thanked for their participation in the study and electronically paid for their participation.

Results

This study used quantitative and qualitative techniques to develop and analyze the content validity of the LMKT to address RQ1: How can learning from a logic model be assessed? Quantitative responses from the online survey and semi-structured focus group were

analyzed using the multi-rater Fleiss Kappa (κ) to measure interrater reliability. Unlike Cohen's κ, Fleiss κ assesses the interrater reliability of more than two raters (Crano et al., 2015; Nichols et al., 2010). This is appropriate since the survey respondents and focus group subject matter experts involved more than two people. The variability in responses to the survey questions guided the focus group questioning and determined the questions used in the final measure. The goal was to increase the online survey Fleiss κ based upon the 100% agreement of survey questions during the follow-up focus group. An overall total Fleiss κ with moderate (.41 - .60), substantial (.61 - .80), or almost perfect agreement (.81 – 1.00) was preferred for the final test (Nichols et al., 2010).

A content analysis of the qualitative data from the online survey and focus groups was conducted. The salient trends in the responses concerning how and why logic model benefits should be edited, removed, or added were identified (Creswell, 2013). The same was done regarding responses to test questions. Lastly, descriptive statistics and reliability analysis were conducted with MTurk participant data from the updated LMKT. This was in conjunction with a content analysis of qualitative data from the open-ended questions.

Pretest Expert Agreement Results. Fleiss' κ was conducted to determine if there was an agreement between the expert judgment on whether to include, remove, or edit two of three stated benefits of logic models as a basis for the LMKT. The findings indicated an overall low agreement between the judgments, κ = -.091 (95% *CI*, -.305 to .092), $p = .123$ (e.g., see Appendix M). However, one benefit had 100% agreement for inclusion: "B2. Logic models help clarify evaluation collaborators' knowledge about the program's anticipated outcomes or goals" (κ = 1.00, $p < .001$). Despite the low agreement between the two benefits (i.e., B1 and B3), the experts did not provide qualitative responses to explain why they should be edited.

There was 100% agreement between the expert judgment on whether to add to the list of benefits of logic models ($\kappa = 1.00$, $p < .001$). Suggestions for additional benefits from qualitative feedback included the use of logic models as a communication tool to facilitate "…a sense of shared understanding of program operations and goals," "…discussion amongst program collaborators," and "…communication to outside [organizations] (e.g., funders) [and] communication of program goals so everyone is on the same page." In addition, responses continued to reflect the beneficial aspects of the logic model as a tool for various functions. They include it being a tool for "…refining the program theory," "gaining consensus on the components," "outlining the assumptions of a program …and unintended consequences," "living [documentation] or at least items for records-keeping," "[helping]internal staff with strategy," "[aligning] internal resources and activities with intended outcomes," and "prioritizing evaluation questions or eval [planning]" (e.g., see Appendix M).

Based upon the three benefits of logic models, Fleiss' κ was conducted to determine if there was an agreement between the expert judgment on whether to include, remove, or edit the 17 LMKT questions. The Fleiss' κ showed a low agreement between the evaluators' judgments, $\kappa = .031$ (95% *CI*, -.030 to .092), $p = .319$. Typical recommendations for question edits mainly concerned listing exact logic model wording in test questions and answers for clarity (e.g., see Appendix N). For instance, one original test question was "What short-term outcome is a result of the physical activity log?" yet "individual physical activity" was used in the logic model. Hence the question was reworded as "What short-term outcome is directly expected from the 'individual physical activity log' output?"

Additionally, several responses recommended the inclusion of quotes when mentioning exact logic model wording in test questions. This is also demonstrated in the previous example

concerning the "individual physical activity log." Hence, the experts emphasized the precision and coherence of the wording of the test questions and response options. An additional salient recommendation concerned being specific with directional language. A true/false question included in the LMKT was, "Each medium-term outcome leads to one long-term outcome." An evaluator recommended adding the word "directly" in front of "leads" because "…the answer is TRUE, but I think you are searching for FALSE because some have other pathways to get to a long-term outcome." Therefore, this recommendation also relates to ensuring that questions are written concisely.

Post-test Expert Agreement Results. Based upon the extremely low agreement among experts in the survey, the focus group focused on reviewing, discussing, and editing all three original benefits of logic models. They came to a 100% agreement on the final three core benefits of logic models ($\kappa = 1.00$, $p < .001$). Next, the experts discussed ways to state the benefits clearly and concisely. For B1, a participant asked if the words "' knowledge about' add any additional value here as opposed to just helps clarify the relationship between program goals and components [or] is there anything that you can do to strike words?" Another inquiry mirrored this, "Can we get these down to five words a piece?" Hence, this focus on minimizing unnecessary words is reflected in the final agreed-upon phrasing of benefits listed in Appendix P. Similarly, the participants discussed whether to include verbs such as anticipated, expected, or intended about the outcomes and goals in a logic model to specify the benefits further. This led to a unanimous vote on using the verb "expected" when an evaluator inquired, "Looking with anticipated it. Is that also in itself more of an academic term…Can we use the word expected instead?" They determined this verb was more appropriate given how evaluators will likely present logic models to lay communities.

Beyond the wording, the experts discussed the order of the benefits based on their relationship. For example, one person said the adjusted B1 was the "…precursor stuff like program components…I think it's the relationship between both…And actually, just flip it all around. Three is one. Two is two. One is three." Thus, this emphasized the hierarchical nature of the agreed-upon benefits.

This discussion of the hierarchy naturally extended to determining whether additional benefits should be considered in the design of the LMKT. Despite the online survey indicating a 100% agreement for adding to the list of logic model benefits, the experts did not believe they should be listed among the "core" benefits of logic models. One evaluator noted how the additional benefits generated from the online survey data "feels like creating the logic model of logic models [and]a lot of [the additional benefits are] how logic models can be used as opposed to the benefits of them as a tool for strategic planning as a tool for building internal capacity, as a tool…." Likewise, another expert mentioned, "We're really looking at the logic model to be that instigator of conversation. So that is a key use of logic model because on its own it doesn't do anything." Therefore, the experts decided not to add to the list of core benefits of logic models, with one saying the list of additional benefits are actually "…ways we should be using [logic models] to enhance the outcomes of [the core] benefits."

During the focus group, only 10 LMKT questions were discussed, edited, and 100% agreed upon for inclusion in the study ($\kappa = 1.00$, $p < .001$). These questions are reflected in Appendix N. The focus of the agreed-upon edits to these questions heavily overlapped with the online survey recommendations and the focus group discussion about the core benefits. The experts highlighted language and word changes to emphasize the clarity and conciseness of questions and response options on the LMKT. This further informed the need to tailor the LMKT

for the different logic model conditions used for Phase 3. Therefore, the Fleiss Kappa was conducted again with the adjusted agreement on the 10 questions, resulting in a moderate agreement between the judgments [$\kappa = .646$ (95% *CI*, .576 to .717), $p = .001$], indicating an overall increase in agreement.

LMKT Internal Consistency Results Based on MTurk Responses. An MTurk sample was used to assess the internal consistency of responses for the LMKT. Before any analyses were conducted, the LMKT and time variables were checked for linearity, normality, outliers, and homogeneity. No abnormalities were found among either variable. However, 4 participants were removed from the dataset because they did not correctly answer the LMKT Q7, an attention check question (e.g., see Appendix E, Logic Model Knowledge Test Question 6). These participants were all located in India, had the lowest average test score on the exam ($M = 2.75$, $SD = .96$), and had the lowest amount of time spent on it ($M = 1.11$ minutes, $SD = .32$). Therefore, data from 76 MTurk participants were used for the analyses. This adjusted sample took an average of 7.08 minutes ($SD = 5.57$) to complete the LMKT and had an average score of 7.79 ($SD = 3.71$). The questions with the most incorrect answers were Q10 *(*$n = 67$, 88.2%), while Q14 had the most correct answers ($n = 74$, 97.4%).

A Cronbach alpha (α) reliability analysis was conducted on the 16 quantitative LMKT. This yielded a Cronbach α coefficient of .819. However, when assessed by country, India yielded a Cronbach α of .697 while the US had .841. Yet, removing Q12 would increase the α to .849, with .756 for India and .854 for the US (e.g., Table 3). Given this, Q12 was removed from the final LMKT.

Table 3

Cronbach α Coefficients for LMKT

	Total	India	US
α With all Questions	0.819	0.697	0.841
α Without Q12	0.849	0.756	0.854

Discussion

Overall, this phase provided insight into what should be emphasized for assessing one's understanding of a logic model. This was based upon insight from subject matter experts, which was used to develop and finalize a viable LMKT with strong content validity based upon core benefits logic model, test purpose, and test questions. This basis also included additional insight from the findings of MTurk participants that completed the LMKT.

Previous work about the benefits of logic models was partially validated by survey feedback. The three core benefits of logic models agreed upon by the experts that participated in the focus group were also reflected in logic model training materials (Alkin & Vo, 2018; Chen, 2018; Kellogg, 2004; Knowlton & Phillips, 2013; McLaughlin & Gretchen, 2015; Rus-eft & Preskill, 2009; Preskill & Rus-eft, 2015). Moreover, the subject matter experts mentioned the additional beneficial nature of logic models as a tool for communication, documentation, and evaluation planning (among the many listed). However, they were not listed among the core benefits of logic models. This warrants additional research into the potential of how logic models are regularly used.

Phase 2 Quantitative Two-Group Design: Individualism and Collectivism Measure Selection

Phases 1 and 2 use the same quantitative sample to address their questions meaning data collection occurred simultaneously for both phases. However, Phase 2 only uses a quantitative

strand to address RQ 2, which focused on which individualism and collectivism (IC) measure would be ideal for the final study in Phase 3. Each IC measure's internal reliability was assessed across two countries associated with individualistic and collectivistic cultures: the United States (US) and India. Given this, a two-group comparison design was used, in which all participants received the same measures. The IC scores from participants located in India were compared to those located in the US to determine if the nature of their scores were consistent with existing IC literature.

Participants

The same data from the 87 Mturk participants used for Phase 1 was used for Phase 2. In addition, the same screening conducted in Phase 1 was used. Accordingly, the remaining 80 responses were used for Phase 2 data analysis. See Appendix C for detailed participant demographics.

Measures and Instruments

Three IC measures were reviewed using a two-group comparison design to address RQ 2 (Creswell & Clark, 2017). These measures included a demographic survey and three existing scale measures of IC from Wagner (1995), Chen (2015; CISC), and Oyserman (1993). Based on the findings of this phase, an IC measure was selected for the final experiment in Phase 3. Additional details about each measure are described.

IC Orientation. The existing scale measures of IC included statements that MTurk participants respond to using Likert-type scales ranging from 1 (strongly disagree) to 5 (strongly agree). The first is a 20-item IC measure by Wagner (1995) for their study exploring the effects of IC on cooperation in student groups. This measure was derived from items adopted and adapted from previous IC research (i.e., Erez & Earley, 1987; Hui, 1988; Triandis et al., 1988;

Wagner & Moch, 1986) and yielded Cronbach coefficient alphas ranging between .72 - .83 based upon each of the five factors in the measure. This measure's lowest possible mean score was 1, which indicated a collectivistic orientation, while the highest score of 5 indicated an individualistic orientation.

The second measure used in this study is the Concise Scale of IC (CSIC) developed by Chen et al. (2015). CSIC contains 18 items, with several of their items derived from previously published scales (i.e., Chen et al., 2009, 2013; Oyserman et al., 2002; Singelis et al., 1995; Singelis, 1994; Triandis et al., 1990; Wang et al., 2014) and yielded a total Cronbach alpha of .91 with .86 for the individualism subscale and .83 for the collectivism subscale (Chen et al., 2015). The last measure used for this study was Oyserman's (1993) 18-item measure of IC, which had a Cronbach alpha coefficient of .49 for the individualism subscale and .74 for the collectivism subscale.

Demographic Questionnaire. Demographic questions were also asked to account for age, education, gender, race/ethnicity, country of residence, social and economic status (SES), and immigration status (e.g., see Appendix L). The Subjective Social Class Measure (SSCM; Belmi & Neale, 2014) was used to account for SES. It included six items in which MTurk participants rated their level of agreement on each using a rating scale ranging from 1 (strongly disagree) to 7 (strongly agree). The highest possible score of 7 indicated a high perceived SES, while a 1 was the lowest possible score which indicated a low perceived SES. They were also used to determine the consistency of IC measures with demographic measures common to the literature (Gudykunst et al., 1996).

Procedures

Since this phase utilized the same data as Phase 1, the procedures were the same. See Appendix H for the Informed Consent.

Results

While Phase 1 focused on the development and validation of the LMKT, Phase 2 addressed RQ 2: Which measures of IC would be more likely to help inform how people interpret logic models differently? Data collection for Phase 2 occurred at the same time as Phase 1. Hence, to address this question, data collected on three existing IC measures (i.e., CISC, 2015; Oyserman, 1993; Wagner, 1995) was used to assess the reliability and validity of each measure's characteristics to relevant constructs and demographic variables. Accordingly, only quantitative data was collected, and the details of the findings are described.

Internal Consistency Results. A Cronbach α reliability analysis was conducted on the three IC measures (i.e., CISC, 2015; Oyserman, 1993; and Wagner, 1995). India was .601, and the US was .795, with an overall Cronbach α coefficient of .801 for Wagner (1995). However, the coefficients increased when question 13 (i.e., "Outcomes can be defined as") was removed from the analysis. It yielded a total acceptable Cronbach α coefficient of .83, so question 13 was removed from measure for any subsequent analysis. Oyserman (1993) and CISC (2015) had much smaller Cronbach α coefficients of .167 and .480, respectively. Only two measures were significantly negatively correlated with each other, Wagner (1995) and CSIC (2015), $r = -.444$, $p < .01$. See Table 4 for additional details.

Table 4

Cronbach α Coefficients and Pearson r Correlations for IC Measures

Predictor	*α (India)*	*α (US)*	*1*	*2*	*3*
1. CSIC (2015)	.120	.584	.480		
2. Oyserman (1993)	.159	.413	.173	.312	
3. Wagner (1995)	.681	.855	-.444*	-.062	.830

* $p > .01$

Assumption Checks. Before conducting the analyses, the variables were checked for missing data, linearity, outliers, and normality. Equal variances cannot be assumed for Wagner's (1995) data; $F(1,78) = 19.81$, $p < .01$. Otherwise, all other assumptions were met across the IC variables.

Independent Samples T-Test. Three independent sample t-tests were conducted to determine if Indian and US scores significantly differed on each IC measure. A Bonferroni adjusted α level of .017 per test (.05/3) was used to account for a potentially inflated Type I error due to the multiple independent sample t-test analyses (Howell, 2012). There was a statistical mean difference in Wagner's (1995) scores; $t(60.90) = 3.308$, $p < .01$, Cohen's d = .514). Crowdworkers from India had larger Wagner (1995) scores ($M = 3.41$; $SD = .35$) than the US ($M = 3.03$; $SD = .64$). However, there were no mean differences between the countries across CSIC (2015) scores [t (72) = -1.88, $p = .064$, Cohen's $d = -.437$] and Oyserman (1993) scores [t (69) = -.598, $p = .552$, Cohen's $d = -.142$].

Supplemental Quantitative Results: Simple Linear Regression. A simple linear regression was conducted to better understand the relationship between perceived SES via SSCM (2015) and IC using the Wagner (1995) measure. The fitted regression model was $y = 2.055 + .244x$. The overall regression was statistically significant ($r^2 = .731$, F (1, 78) = 212.42, $p <$

.001). Accordingly, it was found that SSCM significantly predicted IC ($\beta = .855$, $p < .001$). SSCM accounted for 73% of the variance in IC scores.

Discussion

This phase attempted to identify the most appropriate scaled measure of IC by exploring the reliability and validity characteristics of three existing measures (i.e., CISC, 2015; Oyserman, 1993; and Wagner, 1995) using an MTurk sample. The reliability coefficient of Wagner (1995) was above the commonly accepted level of .70 for measures of social science constructs (Field. 2013). Based on the findings, there were more internally consistent responses on Wagner (1995), thus, making it a more reliable IC measure than Oyserman (1993) and CSIC (2015).

Another interesting finding was that responses from Wagner (1995) significantly differed based on country of residence. However, India's scores were more individualistic than the US. While this appears to be the opposite of common findings of IC research, socio-economic status can moderate that relationship (Chadda & Deb, 2013). Gouveia and Ros (2000) found that countries with high domestic growth products and low illiteracy (i.e., macro-level measures of SES) were correlated with individualism. The strong positive correlation between perceived social class and the IC measures, in addition to the predictiveness of PSC of IC from the supplemental analyses, is consistent with their findings. Accordingly, Phase 2 contributes to the measures' concurrent and predictive validity, which further adds to the body of empirical evidence in support of the psychometric soundness of this measure. Also, this suggests Wagner (1995) was the better measure of IC using crowdsourced data (Burr & Bacharach, 2008). Moreover, it is a better measure for assessing cultural differences among stakeholders in evaluation research and practice. Given this, this measure is deemed appropriate to distinguish between participants' cultures in Phase 3.

Phase 3 Explanatory Experimental Design: Culturally Responsive Evaluation Logic Model Experiment

The findings of Phases 1 and 2 were used to determine the measures used in Phase 3, an explanatory experimental design. This design consisted of 7 conditions in which participants were randomly assigned to address RQ 3-8. In each condition, they responded to quantitative and qualitative measures concurrently. Nevertheless, qualitative and quantitative analyses were conducted separately because the quantitative strand was prioritized, with the qualitative strand used only to support quantitative findings or provide contextual details.

Participants

Data was collected from 361 MTurk participants for Phase 3 (Stopher, 2012). Half of the participants resided in the US (50%) and India (50%). Over half (53%) of the MTurk participants were individualistic (as identified by the IC survey). Most of the participants were white (44%), men (76%), and highly educated, with most having college degrees (87%). The overall median age was 31 years old (M = 33.68, SD = 7.90), with the age of individualistic MTurk participants (M = 32.42, SD = 5.99) significantly younger than collectivistic participants (M = 34.79, SD = 9.15), t(330.91) 2.93, $p < .01$, Cohen's d = .30. There was a small but significant negative correlation between age (M = 32.98, SD = 6.48) and IC (M = 3.42, SD = .42), r (351) = -.14, $p <$ .01. As age increases, IC scores decrease. Lastly, many MTurk participants perceived they were above middle socio-economic class (SSIC; Mdn = 5.83, M = 5.30, SD = 1.46). However, individualistic MTurk participants had a mean score of 6.15 (SD = .65), which was significantly higher than the collectivistic participants, who had a mean score of 4.55 (SD = 1.56), t(260.79) = -13.01, $p < .001$, Cohen's d = -1.32. Lastly, over half of the participants indicated they had a

visual impairment (55%), yet 65% indicated they found reading the English language easy or very easy. Additional demographic characteristics can be viewed in Appendix O.

Conditions, Measures, and Instruments

This design included seven logic model conditions, including a control. Each condition was compared to see if there was a difference in visual efficiency (VE) and credibility of the program theory (CP) based on the IC orientation of the participants. The details regarding these conditions, measures, and instruments are described.

Conditions. The first of seven logic model conditions was the control condition. (e.g., see Appendix G). For this condition, participants viewed a commonly designed logic model. Next, the control logic model was adapted for manipulating conditions based on IC values reflected in language and visualizations. These manipulation conditions included logic models with culturally tailored: 1) language for individualistic orientations (IL), 2) language for collectivistic orientations (CL), 3) visualizations for individualistic orientations (IV), 4) visualizations for collectivistic orientations (CV), 5) language and visualizations for individualistic orientations (ILV), and 6) language and visualizations for collectivistic orientations (CLV). Accordingly, these manipulated conditions represented three independent variables: 1) culturally-tailored language, 2) culturally-tailored visualizations, and 3) culturally-tailored language and visualizations.

Measures. Qualitative and quantitative measures were used to address RQ 3-8. The qualitative measures provided descriptive insight into the MTurk participants' interpretation of logic models, explanation for any mental effort they may have experienced while interpreting the logic model, description of their experience in the study, and experience while engaging in the study. It also helped determine if any participants were aware of the study manipulation. The

quantitative measures used to assess differences between individualistic and collectivistic-oriented participants were mental effort (ME), LMKT accuracy, the response time (RT) on the knowledge test, and VE between individualistic and collectivistic individuals within and between the randomly assigned condition (i.e., IL, CL, IV, CV, ILV, and CLV). This section further details each measure (e.g., see Figure 11).

Program Description. To ensure that participants have processed the content of the logic model, they were prompted to respond to the following open-ended question: What is the story behind this program? This strategy was successfully utilized by a past logic model study (Jones et al., 2018).

Credibility of the Program (CP). A single item was used as a measure for the dependent variable to gauge the MTurk participants' level of CP. Participants were asked to rate their level of agreement with the following statement: This image is helpful in developing the credibility or believability of a program. Their ratings were based on a Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Visual Efficiency (VE). As indicated from previous literature, logic models are beneficial for facilitating effective communication of the program theory. One construct that addresses this is VE, which concerns generating a quick and accurate understanding of a visualization with minimal mental effort (Huang et al., 2009). Furthermore, VE has been used in previous research to assess the effective communication of a logic model (Jones et al., 2018). Given this, the following formula developed by Huang et al. (2009, p. 142) was used as a dependent variable to calculate the VE of each participant:

$$\text{Visual efficiency} = \frac{Z_{accuracy} - Z_{ME} - Z_{RT}}{\sqrt{3}}$$

In the formula, ME represents mental effort. It is "the level of attention required to interpret the logic model and answer questions about the program" (Jones et al., 2018, p. 7). A single item was used to measure ME (Jones et al., 2018; Paas, 1992). Participants were asked to use a Likert-type scale ranging from 1 (low ME) to 9 (high ME) to respond to this item. Participants were also asked what contributed to their ME rating to better understand the ratings.

Accuracy was measured using the LMKT developed and validated in Phase 1 in addition to two equivalent versions, which include individualistic or collectivistic language to match the accompanying logic model. The wording of the 15 test items matched the language within each manipulated logic model design (e.g., see Appendices G and P). The Cronbach alpha (α) reliability coefficient for version A ranged from .71 to .79, version B ranged from .78 to .81, and version C from .68 to .75. See Table 5 for additional Cronbach α coefficient details. The minimum score one could receive on this test was 0, which indicated no knowledge of logic models, versus the maximum of 15 points which indicated high knowledge of the logic model.

Table 5

Summary of IC Cronbach α Coefficients on LMKT by Condition

Version	Conditions	Total α	*Individualistic α (n)*	*Collectivistic α (n)*
A	Control	0.71	0.65 (26)	0.76 (28)
A	IV	0.79	0.67 (31)	0.79 (22)
A	CV	0.71	0.19 (23)	0.73 (30)
B	IL	0.78	0.55(25)	0.88 (22)
B	ILV	0.81	0.22 (20)	0.87 (32)
C	CL	0.68	0.61 (18)	0.72 (31)
C	CLV	0.75	0.49 (27)	0.81(26)

Lastly, RT was referred to as response time to measure how long the participant took to describe the story of the logic model and complete the LMKT. A function in Qualtrics was used to measure this in minutes.

IC Orientation. Based upon the findings of Phase 2, Wagner's (1995) IC measure was used as an independent variable to assess participants' cultural orientation. Participants were categorized as either individualistically oriented or collectivistically oriented using the 50th percentile score marker of 3.53 for distinguishing between the two. Those below 3.53 were denoted as collectivistically oriented, and those above it were denoted as individualistically oriented.

Demographic Questions. The demographic questions used in this study were SES via SSCM (2014), age, education, gender, race/ethnicity, country of residency, immigration status, visual ability, and English language comprehension (e.g., see Appendix L).

Manipulation Check. A manipulation check was employed to ensure an accurate interpretation of the data, given the experimental manipulation of conditions and the purpose of the study (Crano et al., 2015). The following question was asked of participants to determine the influence of the manipulation: "What do you think the purpose of this study was?" Should a participant indicate knowledge of the purpose of the study, their quantitative responses were compared to those that do not know the purpose of the study to determine that data should be used in the final analysis.

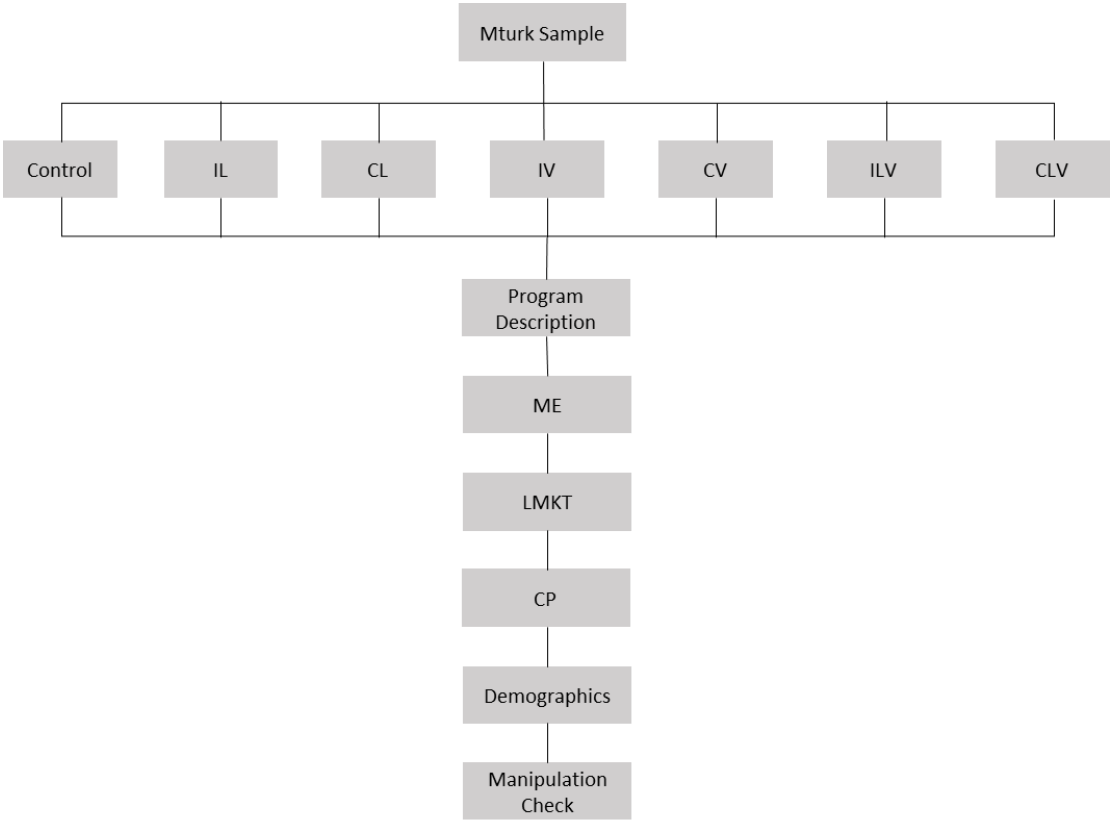
Procedures

Upon CGU IRB approval, participants were recruited to participate in the study via MTurk. They were given an informed consent which described the research and IRB contact, the purpose of the study, a summary of the tasks that they were asked to perform, stated that the

study is voluntary, and notified them of their $4.00 payment upon completion of the study (e.g., see Appendix Q). Finally, they were directed to the end of the survey and thanked for their interest in the study if they did not consent. Participants that consented to the study were randomly assigned to one of seven logic model conditions (e.g., see Appendix G). They were first asked to review the logic model, briefly describe the program in their own words, assess their ME, and then CP. Following this, the participants took the LMKT (e.g., see Appendix L) and were timed to gauge how long it took to complete it (i.e., RT). Then, they were presented with the Wagner (1995) IC measure and demographic survey (e.g., see Appendix G), followed by the manipulation check. Lastly, they were thanked for their participation in the study and electronically paid. See Figure 11 for an outline of the experimental procedures.

Figure 11

Procedures for Experimental Design



Phase 1 and Phase 2 findings were used to determine the measures for knowledge of logic models and IC orientation to address the RQ 3-7, which concerns assessing whether there are cultural differences in logic model knowledge based upon one control and six culturally responsive logic model designs (See Appendix G). The six culturally responsive designs differed based on language, visualizations, and a combination of the two. Quantitative and qualitative analyses were performed to address these questions. Specifically, a 2x3 plus a control analysis of variance (ANOVA) was performed on two dependent variables: visual efficiency and perceived image usefulness. The independent variables were LM conditions (i.e., control, culturally tailored language, culturally tailored visualization, and culturally tailored language and

visualizations) and IC orientation. In addition, a content analysis of the qualitative data was conducted to identify salient trends in the responses to provide supportive contextual information related to the manipulation check and each RQ (Creswell & Poth, 2017). Details of the finding are explained below.

Pilot Test

Before official data collection began, this study was piloted to ensure participants were randomized into appropriate study conditions and general ease of survey flow (Stopher, 2012). An MTurk sample of 129 (56% from India and 44% from the US) and 3 CGU students majoring in program evaluation completed the pilot test. Initially, participants were not randomized to specific conditions. Instead, they received all conditions. Once this was evident, it was fixed, and additional pilot data were collected to ensure randomization was occurring. Feedback specifically from the CGU participants indicated that some demographic questions needed edits to ensure clarity. These changes were made. Then, 28 pilot test MTurk participants were used in the final analysis because they were correctly randomized into conditions after all edits to the survey were made.

Results

The official data were screened to determine whether participants should be removed from the dataset and whether statistical assumptions were met for quantitative analysis. The manipulation check revealed a few qualitative responses that needed to be reviewed to ensure those participants were not aware of the purpose of the study (n = 3). For instance, one participant wrote, "To test out logic models." Similarly, several responses related to describing the program represented by the logic model (n = 5) and mental effort in interpreting it (n = 5) indicated that participants were familiar with logic models or may know the exact purpose of the

study. For example, a person wrote "present a logic model" in response to the inquiry concerning the program's story represented by the logic model. However, another mentioned the logic model when describing their mental effort, "I found this logic model to be very easy to understand." Additional quotes are listed in Appendix R. When comparing their quantitative responses to those that did not mention logic models, there was no difference in VE [$t(349) = -.52$, $p = .60$)] and CP [$t(359) = -1.60$, $p = .11$)]. Given this, their data was used in the final analysis.

A total sample of 361 MTurk participants was available for analysis. This sample was reduced to 351 due to the deletion of univariate outliers on VE. The assumptions of normality, linearity, multicollinearity, and homogeneity of variance were satisfactory.

Quantitative Results. A 2x3 ANOVA with a control condition was conducted to determine if there were any differences in VE between IC and logic model condition (i.e., RQ 3, 5, and 7) and CP by IC orientation and logic model condition (i.e., RQ, 4, 6, and 8). See Appendix S for IC descriptive statistics by condition. There was no main effect of condition on VE, $F(1, 6) = .91$, $p = .49$. Similarly, there was no main effect of condition on CP, $F(1, 6) = 1.08$, $p = .37$. Also, there was no main effect of IC orientation on CP, $F(1, 2) = 3.12$, $p = .08$. However, there was a significant main effect of IC orientation on VE, $F(1,2) = 4.59$, $p < .05$. Participants with a collectivistic cultural orientation had higher VE scores ($M = .05$, $SD = .64$) than those with an individualistic cultural orientation ($M = -.08$, $SD = .61$). Only 1% of the variance was explained by IC orientation after excluding variance explained by the seven conditions. However, there was no interaction between condition and IC orientation on VE, $F(1,6) = .96$, $p = .45$. Likewise, there was no interaction between condition and IC orientation on CP, $F(1,6) = .68$, $p = .67$ (e.g., see Table 6).

Table 6

Summary of 2x3 Plus a Control ANOVA Results

IV	DV	Type III sum of squares	*df*	M^2	*F*	*p*	η_p^2
Condition	VE	2.14	6	0.36	0.91	0.49	0.02
	CP	5.13	6	0.89	1.08	0.37	0.02
IC Orientation	VE	1.8	1	1.8	4.59	0.03	0.01
	CP	2.55	1	2.55	3.12	0.08	0.01
Condition x IC Orientation	VE	2.25	6	0.38	0.96	0.45	0.02
	CP	3.34	6	0.56	0.68	0.67	0.01

Qualitative Findings. Qualitative data included responses to the inquiry about the program story and a description of their mental effort to understand the logic model. Many participants did not write lengthy responses to qualitative questions. For instance, some listed a few words to describe the program's story (e.g., "life circle diet") or the mental effort involved in interpreting the logic model (e.g., "Very, very, very much"). Others wrote very short sentences to describe the program's story (e.g., "It teaches people to eat better and exercise to ultimately improve their well-being") and their mental effort (e.g., "I feel the chart little bit difficult to understand the activities"). Yet, many responses to the qualitative questions were either incoherent (e.g., "TENSION AND PREESAR.SEAK"), copied from the study description (e.g., "The purpose of this study is to determine people's understandability and perceptions of an image about a program using a crowdsourcing platform"), or taken from an online definition (e.g., "A program is a set of instructions that a computer follows in order to perform a particular task"). Given this, many of the coherent responses did not yield any meaningful trends nor added contextual insights to address any of the RQs, so the qualitative findings were not prioritized in this mixed methods study.

Overall, there were mixed findings concerning whether participants were able to adequately describe the story of the program (e.g., "It appears to be some kind of a multi-faceted health and wellness program, involving educating clients about living healthier by restructuring their eating habits as well as exercise regimen") or not (e.g., "Substance Abuse and Mental Health Services Administration"). However, many of the responses related to health in some manner.

This pattern of responses was similar for the descriptions about mental effort where participants either found interpreting the logic model challenging (e.g., "…when you see the image it looks like big and very difficult to understand so we have to take time to read and understand the image") or easy (e.g., "This program was very very very easy to understand"). While there were minimally detailed explanations for their mental effort scores, some participants provided details. For instance, one person described how understanding the logic model was difficult for them:

"…we have to see many pictures to understand the meaning and there are many arrow marks that goes up and down so we have to see the arrow marks by one by one to understand the meaning of the image that is very difficult because when you see the image it looks like big and very difficult to understand so we have to take time to read and understand the image."

Yet another participant initially had similar sentiments:

> "At first, the flowchart itself looked a bit intimidating, since there were a lot of different boxes and connections to multiple which made me confused. However, once I actually read through the content of the boxes and looked at the program, along with the pictures included, it made the purpose and steps of it a lot easier to understand."

Both participants were from the CLV conditions, but their feedback mirrored those that provided detailed descriptions of the other six conditions. See Appendices T and U for additional descriptions of the program story and mental effort used to interpret the logic models.

Supplemental Quantitative English Language and LMKT Results. Due to 26% of the sample indicating that they found reading the English language either difficult or very difficult, a Pearson r analysis was conducted to determine the relationship between English language proficiency, VE, and perceived program credibility. There was a small yet significant correlation between English language proficiency and VE, $r(359) = .13$, $p < .05$. There was no significant relationship between English language proficiency and CP, $r(359) = -.05$, $p = .37$.

Discussion

Phase 3 assessed whether culturally tailored language and visualizations in logic models contributed to effective communication (i.e., RQ 3, 5, and 7). The quantitative findings indicated an overall effect of IC orientation on visual efficiency. Despite logic model conditions, collectivistic-oriented people scored higher than those with individualistic orientation. While there were no meaningful IC differences in VE among the conditions, the findings were consistent with existing literature. For instance, logic models provide more contextual information than words alone. The connections and order of the information indicate a relationship among the words. This contextual nature of visualizing information is consistent with the communication patterns among collectivistic-oriented cultures (Gendron et al., 2017; Gudykunst, 1996; Oyserman, 2017; Reynolds & Valentine, 2011). Likewise, logic models can convey complex information through the context or amount of content within them, which is also aligned with collectivistic-oriented communication.

Despite these findings, there were no differences between each logic model condition on VE and CP. This may be due to the higher internal consistency on the LMKT for collectivistic orientations than individualistic orientations. In fact, the LMKT was internally inconsistent across the conditions for individualistic participants. This warrants a re-examination and possible redesign of the LMKT, so it has comparable Cronbach alpha coefficients for individualistic and collectivistic participants.

Furthermore, CP across cultural orientations and logic model conditions was the same. Rather both cultural orientations across the conditions equivalently indicated that they believed the logic models in their assigned conditions were credible. While this does not support the notion that there are differences among cultural perceptions of credibility based on whether the logic model is culturally tailored, it does support previous research findings. For instance, Jones et al. (2019) found logic models more beneficial than a written program description alone. Accordingly, a logic model, despite its design, is perceived as credible.

Chapter V Discussion

Altogether, this multiphase mixed methods study aimed to explore the integration of culturally responsive evaluation (CRE) in logic model designs to inform effective communication in evaluation practice using culturally tailored language and non-textual visualizations. This final chapter briefly reviews the research questions (RQ) used to guide the three study phases to address this purpose. Moreover, this chapter's major sections further summarize each phase's key results. Then, the implications for evaluation practice and research on evaluation are discussed. This is followed by study limitations. These latter two sections are organized by each phase. Lastly, the conclusions of the research are reviewed.

Research Questions

Eight research questions (RQ) were created to explore culturally tailored communication through logic model designs based upon the underlying principles of CRE. These questions concerned how learning can be assessed from a logic model (RQ1) because no existing tools were readily available to assess this. This was addressed in Phase 1. The subsequent RQ 2 concerned identifying which individualism-collectivism (IC) measure would best inform interpretations of logic models, which was addressed in Phase 2. The final set of RQs was addressed in Phase 3. They inquired about the impact of culturally tailored language in logic models on effective communication (RQ 3) and the credibility of a program theory (RQ 4). Similarly, two additional questions concerning the impact of culturally tailored non-textual visualizations in logic models on effective communication (RQ 5) and the credibility of a program theory (RQ 6). The last two questions inquired about using culturally tailored language and non-textual visualizations in logic models on effective communication (RQ 7) and the credibility of a program theory (RQ 8).

Findings

The first two study phases yielded findings that were integrated into Phase 3. Findings from Phase 1, which addressed RQ 1, resulted in a reliable and valid measure of learning from logic models. Edits at the item level, as well as content validity, were determined by evaluation experts with their qualitative feedback. The final Logic Model Knowledge Test (LMKT) was created and used for Phase 3. Findings from Phase 2 identified Wagner (1995) as the existing IC measure with the highest and most acceptable internal consistency compared to two additional measures.

Moreover, Wagner (1995) better distinguished IC between individuals across two nations (India and US) on IC scores, which made it the most suitable measure to use in Phase 3's experiment. Phase 3 integrated the LMKT and Wagner (1995) into its design to address RQ 3-8. This set of questions directly informed the purpose of the study. In particular, RQ 3, 5, and 7 assessed the difference in VE by IC and logic model condition. The findings from this experiment indicated overall higher VE scores among collectivistic-oriented people across the seven logic model conditions. This means collectivistic participants had a quicker and more accurate interpretation of the program theory with lesser mental effort than individualistic-oriented participants. Thus, logic models are a more effective communication tool for evaluators when working with collectivistic-oriented stakeholders despite their culturally tailored language and visualizations. This partially supports the stance for culturally tailoring logic models to ensure effective communication among stakeholders and evaluators.

Contrary to this, the findings addressing RQ 4, 6, and 8 yielded no significant differences. While participants' average perceived credibility of the program scores was high across logic model conditions, there were no differences between those conditions or participant IC

orientation. The are several potential reasons for this. One reason concerns the participants' potential familiarity with logic models or similar visualizations (i.e., flow charts). Very few qualitative responses indicated a familiarity with logic models, yet this does not mean they were the only ones. Since the publication of W.K. Kellogg Foundation's Logic Model Development Guide (2004), the rise in literature, training, and funder requests for logic models indicates its visibility and participant exposure to them.

In addition to the potential familiarity with logic models, the source of the logic models may have impacted participant ratings. For example, when a logic model is presented by a researcher or funder (or someone with authority), one may automatically find the program theory trustworthy (Mayr et al., 2019). Hence, a researcher at an academic institution presenting the logic models as part of a study may have contributed to the perceived program credibility.

Another reason for the findings regarding the program's credibility may be due to the program theory or subject matter itself. The activities listed in the logic model were associated with physical and mental health outcomes. In the design of the hypothetical logic model, the content in each component was based upon common recommendations from physicians and health research. This may have been common knowledge among the study participants. Therefore, the credibility may have been due to their knowledge of the logic model as a visualization of health and well-being resources, activities, and outcomes rather than the culturally tailored logic model designs.

Phase 1 Implications and Limitations

The findings of Phase 1 alone have several key implications for evaluation practice and add to the literature on evaluation and IC alike. The creation of the LMKT adds to the existing logic model literature and toolkits. Tests and quizzes to check one's understanding of logic

models are absent in many training materials (e.g., Alkin & Vo, 2018; Donaldson, 2022; McLaughlin & Gretchen, 2015). Rather many provide discussion questions, reflective questions, and exercises (e.g., Chen, 2018; Kellogg, 2004; Knowlton & Phillips, 2013; Rus-eft & Preskill, 2009; Preskill & Rus-eft, 2015). While these assessments are valuable, they focus on assessing the ability to apply, analyze, evaluate, or even create logic models (Darwazeh & Branch, 2015). They skip over the first levels of learning, ensuring factual, conceptual, and procedural knowledge of logic models, which are critical when working with stakeholders. Hence, the creation of the LMKT fills this gap and adds to the existing logic model literature and toolkits.

Given that the LMKT fills this gap, it can help determine the effectiveness of logic model training and education. A trend in how stakeholders respond can indicate their understanding of 1) the program theory, 2) the program components, 3) the logic model components, or 4) any combinations of these outcomes. This will provide direction to the next steps in the process, whether it includes additional explanations of the logic model or moving forward with the next evaluation process.

The LMKT can also be used as a tool to communicate "evaluation processes and results in timely, appropriate, and effective ways" (AEA, 2018). For example, MTurk participants took less than ten minutes to complete the LMKT. This is an ideal amount of time for evaluators with limited time blocks (especially during hour-long staff meetings) to interact with stakeholders. In addition, their results can be revealed during the meeting when the test is presented in an online survey format or other software that can calculate scores within seconds of test submission. Accordingly, the LMKT can contribute to the enhancement of evaluator competencies.

The LMKT can be used as a reference for evaluators to develop measures to inform them whether their logic models are effective. As previously mentioned, several questions used in the

LMKT were adapted from the measure created by Jones et al. (2020), which only addressed the core benefits of a logic model. At the same time, the current measure expanded this by providing questions that also reflected the conceptual knowledge of logic models. Similarly, the LMKT can be adapted to appropriately reflect a program's logic model's content.

Lastly, the LMKT can be used in research on evaluation. This may primarily cover research on data visualizations or logic models. Nonetheless, that research can inform logic model training and practice. More importantly, this can contribute to a significant advancement of the entire field.

While Phase 1 yielded many benefits, several limitations should be noted. For the expert feedback, neither participants for the online survey nor the focus group were randomly selected, which is required for Fleiss κ analyses (Nichols et al., 2010). Moreover, this sample is limited because it does not reflect the vast array of cultures represented among evaluators, nor does it reflect evaluators not part of the national association. This indicates the risk of continuing the cycle of creating visualization guidelines based on white cultural norms, which can limit an accurate interpretation of a logic model (AEA, 2011; Bowman et al., 2020; Douville et al., in review.). However, highly specialized experts were necessary to get quality feedback for validation of the LMKT. Given this, it is important to emphasize that the experts in this study were highly representative of AEA (2018) membership which includes predominantly women. This provides some representation of evaluation experts within a renowned professional association contributing to a high internal validity.

The experts received the incorrect logic model during the pre and post-test data collection. Instead of receiving the control logic model with neutral language that lacks individualistic and collectivist words, the experts received a logic model with individualistic

language. Due to this, the individualistic logic model was used for piloting for methodological consistency. Yet this mistake enabled the experts to note the need for more clarity in LMKT questions and responses. This informed the need for tailoring the LMKT for the different logic model conditions that manipulated language. Therefore, for subsequent studies using the LMKT with different logic model language conditions, the language in the test should be changed to reflect the language used in the logic model presented. The LMKT questions affected by these language modifications are questions 2, 3, and 5.

Another limitation of this study is that only 10 of the 16 LMKT questions were reviewed during the focus group session. Due to the enriching discussion among experts, despite staying for an extra 30 minutes for the focus group, they needed more time. This prevented the potential for further increasing the final Fleiss κ and getting additional expert feedback. Hence the measure was piloted to account for anything not addressed during the focus group. Future studies should ensure diverse expert and participant insight is solicited.

Similar to the pre and post-test, the sample characteristics of the MTurk participants may limit the findings' reach. For example, the participants in India took less time to take the LMKT and yielded lower test scores than US participants. A possible reason for this may be a language barrier. The participants in India may have a different English reading comprehension than those in the US. Future replication studies should control for this potential issue.

Another interesting characteristic was that most of the participants were men, which does not adequately represent the 75% of women that work in the areas that receive the most amount of money for evaluation in the US (i.e., education, health, and social assistance; American Association of University Women, 2018; Lemire et al., 2018). Past MTurk demographic research has shown more women participants than men (Burnham et al., 2018). Since this study used a

convenience sample, men might have been mainly available when recruitment opened for this phase. This limitation of the sample should be addressed in future data visualization studies since gender may influence how one interprets an image (AEA, 2011; Bowman et al., 2020; Kray et al., 2002; Kray et al., 2004; Kray et al., 2001; Mortenson, 2002). Despite this limitation, the sample consisted of many highly educated participants, which is representative of many evaluation stakeholders (Lemire et al., 2018).

Phase 2 Implications and Limitations

Similar to the LMKT, Wagner's (1998) IC measure was identified as a measure with good validity and reliability that distinguishes between cultural nuances among evaluation stakeholders. In addition, the highly educated nature of MTurk samples makes them very similar to program staff often engaged in evaluations (Lemire et al., 2018). Given this, there are implications for using Wagner's (1995) IC measure in evaluation practice in light of the limitations of Phase 2. Both are discussed in this section.

Practical Implications

Based upon Phase 2 findings, the Wagner (1995) IC measure can apply CRE's underlying principle of connecting with the stakeholders to build trust and interpersonal relationships (Hood et al., 2015). This study did not focus on the process of developing a logic model. Yet, this scale can facilitate a conversation about cultural values and communication to better connect program staff, community members, and evaluators with each other before beginning the process of describing the program theory. For instance, an evaluator can state their CRE approach and explain the role of culture in one's understanding and experience with a program. Then, introduce IC and Wagner's (1995) measure as one way to understand the cultural orientation among stakeholders and evaluators concerning their work. Findings from Wagner

(1995) could be used to "clarify diverse perspectives, interests…and cultural assumptions," better understand those that may use the logic model, and "facilitate a shared understanding of the program and its evaluation…" in a culturally competent manner (AEA, 2011, 2018; Azzam et al., in press). Subsequently, this could lead to a conversation about the values and expectations of the program that the evaluator can use to understand the program theory and design a logic model that reflects the cultural orientation of the stakeholders. This process could ultimately help ensure that stakeholders accurately interpret the logic model in a timely manner (AEA, 2018).

Additionally, the findings from Phase 2 add to the evidence of the reliability and multicultural validity of the Wagner (1995) measure. Namely, Wagner (1995) was the most reliable using a crowdsourcing sample, which is gaining popularity among researchers that do not want to rely on student participants (Buhrmester et al., 2011). Yet, this was also a binational sample of people residing in the US and India. Hence, this study further adds to the multicultural validity of this measure concerning its use with a binational and nonstudent sample (Kirkhart, 2010).

Limitations

While the findings of this phase provided good validity and reliability support for the Wagner (1995) IC measure, all three IC measures had significantly lower reliability for participants in India than in the US. These differences may be due to cross-cultural differences in interpreting scaled measures (Cozma, 2011). For instance, some cultures' use of the extreme sides of the scale by participants is important in conveying a truthful response (Cozma, 2011; Triandis, 1995). The same could be occurring with the India sample and thus impacting the psychometrics of the Concise Scale of Individualism and Collectivism (CSIC; 2015) and Oyserman (1993) IC measures.

Another possible reason for lower reliability coefficients among the India sample and the overall lower reliability of Oyserman (1993) and CSIC (2015) may be due to the changes made to the measures. Specifically, several items on each original measure were double-barreled in which they were addressing two topics with one response scale. Items of this nature cannot be accurately interpreted because one would need to know which aspect of the item the participant was addressing (Stopher, 2012). In addition to this, items were adapted from Oyserman (1993) to make them more gender inclusive by using "his/her" or "he/she." Before the change, the items only used the pronouns "he" or "his." These changes may have affected how participants interpreted and responded to the items. Additional research should be conducted to determine if the original Oyserman (1993) and CSIC (2015) measures replicate these findings to decide if they are ideal for cross-cultural participants or stakeholders.

Phase 3 Implications and Limitations

Finally, the findings of Phase 3 provided partial empirical support using a CRE lens for logic model designs to facilitate effective communication of a program's theory among evaluators and stakeholders alike. Collectivistic participants outperformed individualistic participants across the logic model conditions. This finding provides partial support for integrating CRE into logic model designs. While tailoring the language and non-textual visualization in the logic model conditions did not affect collectivistic participants, the overall contextual nature of logic models may have contributed to this finding. Logic models, as opposed to a written description of a program theory, provide an additional contextual visualization of the program theory. Considering the study's limitations, all these findings have implications for evaluation practice and theory. Each of these aspects is discussed.

Implication for Evaluation Practice

This finding aligns with collectivistic trends in communication as they communicate in a more context-rich manner in comparison to the direct nature of individualistically oriented stakeholders (Gendron et al., 2017; Gudykunst, 1996; Oyserman, 2017; Reynolds & Valentine, 2011). Similar to the recommendations for evaluation practice for Phase 2, Wagner (1995) would be a good tool to assess the IC orientation of stakeholders. Given the findings of Phase 3, if the majority of stakeholders are collectivistic, evaluators should use more contextually rich visualizations and language in logic model designs to endure the effective communication of the program theory. On the contrary, evaluators should design a more simplistic logic model if the stakeholders are more individualistically oriented.

While the findings of Phase 3 suggest logic models may be a communication tool that yields more VE benefits for collectivistic-oriented stakeholders, this does not erase the benefits of using logic models. Outside of cultural orientation, there could have been other participant characteristics that would make logic models and the overall use of visualizations a better option for effective communication. For example, a logic model with extra iconic or imagery-based visualizations may be more accessible to those with learning (i.e., Douville et al., in press), literacy, and language challenges (Torres et al., 2005).

Despite cultural orientation, written descriptions of a program theory alone may not be as beneficial as accompanying it with visualizations. For instance, Jones et al. (2018) found that logic models were more beneficial than written program theory descriptions alone. Given this, logic models remain a better communication tool for both cultural orientations as opposed to not using them altogether. Hence, Phase 3 findings have practical implications and add to the evaluation's literature and the value of logic models.

Theoretical Implications

Phase 3 findings also add support to evaluation theory. Stake's (2004) responsive evaluation is foundational to CRE. This theory, which lies among the evaluation theories focused on valuing, emphasizes that the contextual rich nature of evaluation and the beliefs and values of program staff and participants are critical to evaluations. They should be represented in the reporting of evaluation results (Alkin, 2012). Accordingly, Phase 3 focused on the values that underlie IC orientations in presenting a program theory via logic model. The findings provided partial support for the basis of responsive evaluation, by extension, CRE.

Limitations

Several limitations to this study were important to note as they may have impacted the findings. They concern the quality of the qualitative data, sample size, the complexity of logic models, and sample characteristics. Each limitation provides insight for further research on culturally responsive logic model designs.

Throughout the study, participants focused on quantitative responses. This initially was not a major concern because the study prioritized the quantitative strand. However, having saturated qualitative feedback of quality may have yielded a contextual and explanatory depth of their quantitative responses (Creswell & Plano, 2017). This could help evaluators understand what makes a logic model credible and what parts of its design contribute mainly to effective communication. Given this, future research should focus on the how and why questions concerning culturally tailored logic model designs by prioritizing qualitative methods.

The sample size in this study may have also contributed to the findings. There were unequal participants in each condition (Howell, 2010). For instance, the sample size for each condition ranged between 45 to 55 participants (e.g., See Appendix S). Moreover, there were

more collectivistic-oriented participants than individualistic participants in the study. There were 19 individualistic participants in one condition compared to 31 collectivistic participants. Given this, the findings may be driven by the number of collectivistic participants. Therefore, future research should replicate the Phase 3 findings using a larger sample with equal amounts of individualistic and collectivistic members in each group.

To complicate the potential limitations of sample size, this study did not address the levels of complexity of logic models. A hypothetical logic model based on existing programs and their logic models was used rather than an actual logic model. The level of IC cultural saturation in the logic models is closely related to complexity. This may have been limited or overwhelming. Each logic model included 30 pieces of information about the program theory. A total of 30 icons were in each manipulated logic model to distinguish between IC cultural orientations. Moreover, there were 14 language adjustments made in each manipulated logic model to distinguish between IC cultural orientations (e.g., see Appendix G).

Given the language and visual manipulations of the logic models, the LMKT also had to be adjusted. Each condition did not use the same LMKT. Instead, there were a total of three equivalent versions. Control, IV, and CV used the same test, whereas the other four conditions used equivalent tests in which the words in the test matched the words in the logic model condition. For instance, IV and ILV had the same version of the LMKT that utilized words (i.e., "individual physical activity log") to mirror the individualistic language used in the logic model condition (e.g., see Appendix G). On the other hand, CV and CLV had the same version of the test that used collectivistically oriented words (i.e., "household physical activity log") to match their logic models. Moreover, the interplay of complexity and sample sizes could have also impacted this discrepancy in the reliability of the three equivalent LMKT versions.

Lastly, sample characteristics may have influenced the findings. The supplemental analysis indicated a significant relationship between English language reading proficiency and VE. Future studies should ensure that participants are fluent in the language used for the study. Also, insight from the qualitative responses lends to future study recommendations. Several participant responses to questions about the program description and project purpose indicated that a surprising amount of Mturk participants had been exposed to logic models. This may have been due to participants conducting online image searches. However, their quantitative responses did not differ from those unfamiliar with logic models. This has potential benefits for guiding future studies about logic models; participant familiarity with logic models may not be necessary. Future research should explore this further to add validity to the Phase 3 study and the LMKT.

Limitations Across Phases

Across the three study phases, there are two major limitations to note. The first concerns the sample used across the three phases, and the second concerns specific characteristics that the study did not address. Each phase used a MTurk sample. This sample was limited to only two nations (i.e., India and the US), each of which represented either individualistic or collectivist cultural orientations according to IC research as opposed to utilizing multiple nations to yield more representative results of international stakeholders that may be involved in evaluations (Gudykunst et al., 1996). Also, the ethnic diversity of the US population posed a possible limitation (Burnham, 2018). While the US is predominantly individualistically oriented, collectivistic communities reside within them (Vandello & Cohen, 1999). This may explain why the US samples were more collectivistic than the India sample.

Furthermore, while this study attempted to recruit participants fluent in English, this may not have occurred. There was no method of assessing this in Phases 1 and 2, but a question was asked about participants' ability to read English fluently in Phase 3. Accordingly, many participants indicated they could not fluently read the English language. This means some likely participants had a limited understanding of the study instructions and measures. Unfortunately, this could have impacted the reliability of their responses (Stopher, 2012). Given this, future research should ensure that participants are fluent in the language of the study.

Finally, this study did not address several design characteristics of logic models. Firstly, this study did not account for the different layouts of logic models. The logic models used in this study were still linear concerning both time and direction of interpretation, which aligns with individualistic communication. Conversely, collectivistic individuals have more of a highly contextual and cyclical way of thinking (Ting-Toomey & Dorjee, 2020). Secondly, this study did not use visualization design principles for the logic models (Few & Edge, 2008; Tufte, 2001). Previous research has found that simple changes to a logic model based on design principles, such as minimizing non-data ink, improved VE (Jones et al., 2019). Using these principles may have impacted the interpretation of the program theory. Thirdly, the IC construct and Wagner (1995) measure only capture one cultural aspect.

Moreover, this study could have used other dimensions of culture to assess differences among stakeholders to guide effective communication of a program's theory, such as power distance (e.g., Hofstede & Bond, 1984), long versus short-term orientation (e.g., Hofstede, 1991), approaches to conflict (e.g., Ting-Toomey & Oetzel, 2001), uncertainty avoidance values (e.g., Hofstede, 1980). Overall, these elements that were not addressed in this study could have impacted its findings. Given this, future research should focus on replicating this study using

different constructs to assess different aspects of culture, varied layouts of logic models, and manipulation of various cultural aspects of logic model designs (i.e., color, complexity, etc.).

Recommendations for Evaluation Practice

Based on the study findings, several concrete recommendations exist for how evaluators could use every phase in their practice. Firstly, when engaging with the stakeholders, the IC measure should be used as a tool for facilitating the CRE principle of connection or establishing interpersonal relationships among evaluation collaborators (Hood et al., 2015). The evaluator should explain their CRE approach to the evaluation, the role of culture in evaluation in terms of communication, values, and how evaluation collaborators relate to one another. This should be followed by a description of how IC is one of many methods to identify cultural differences. Next, evaluators and stakeholders should complete the IC measure. If the evaluator has previously completed the measure, then they should be prepared to share those responses with the stakeholders.

The sharing of IC findings should further facilitate the application of the CRE principle of plasticity or "the ability to…receive new information, reorganize and change in response to new experiences, and evolve new ideas and forms" (Hood et al., 2015, p. 308). If the total mean score of all the stakeholders is a 4 or 5, they are predominantly individualistically orientated. If the total mean score is 1 or 2, they are predominantly collectivistically oriented. If the mean score is 3, they are a mixture of both IC orientations. Discuss what this means in terms of communication and values with the stakeholders. This discussion should generate new knowledge among the evaluator and stakeholders to guide the logic model design based on their shared cultural values and communication.

Secondly, to explain what logic models are and describe their benefits to ensure transparency and evaluation capacity building in accordance with evaluation principles and competencies (AEA, 2018). Following this explanation, give stakeholders the LMKT with language and visualizations matching their collective IC orientation. Utilize the LMKT to determine where the most misunderstanding lies based on which questions they got incorrect. If most of the incorrect questions were the content questions concerning the comprehension of the program theory (i.e., Questions 1-8; see Appendix P), then the evaluator should provide further details about that material. Then, have stakeholders determine the correct answers. This same procedure should be conducted if most logic model component comprehension questions (i.e., Questions 9-15; see Appendix P) are incorrect. If there is no trend in what questions stakeholders got incorrect, then there should be a review of logic models overall followed by the stakeholders discussing which answers are correct.

Thirdly, given the Wagner (1995) IC scores, evaluators should use culturally tailored language and non-textual visualization in the logic model design that aligns with the cultural orientation of the stakeholders. While this study did not focus on the process of working with stakeholders to develop a logic model, evaluators should ensure the design is a participatory process with them (Chen, 2015). Ask the stakeholders if the language or visualizations used in the logic model resonates with them (Torres et al., 2005; American Psychological Association, 2020). Also, discuss and reconcile any misalignment between program goals and outcomes with stakeholder cultural orientation. For instance, if program goals and outcomes reflect the individualistic language and the stakeholders are collectivistic, the evaluator should discuss this with the stakeholders to reconcile this misalignment. The design process ends when there is consensus among the stakeholders indicating the logic model reflects their program.

Conclusions

Overall, this exploratory study partially supported the notion that culturally responsive evaluation (CRE) should be integrated into logic models. The findings contribute to facilitating the evaluator competencies and the overall advancement of the evaluation field in several ways. One way was to develop a measure to assess learning from logic models in Phase 1. This test can be used to teach evaluation and practice alike to determine whether stakeholders understand a logic model when presented to them. Thus, this would aid evaluation capacity building as part of the planning and management domain of evaluation practice (AEA, 2018).

Also, this study replicated findings from individualism-collectivism (IC) research in Phases 2 and 3, adding to the validity of the Wagner (1995) IC measure and the overall construct. Moreover, its use in this study highlights its potential as a tool for evaluators to better understand and engage in the different perspectives among stakeholders as part of the interpersonal domain of competent evaluation practice (AEA, 2018). Consistent with IC literature, individualistic and collectivistic-oriented participants differed in visual efficiency (VE) scores. This is encouraging as it provides further support for attentiveness to culture regarding rapid and accurate interpretations of logic models.

Accordingly, this facilitation of improved VE further informs the importance of using logic models to "communicate evaluation processes and results in a timely, appropriate, and effective way" by incorporating CRE into its design (AEA, 2018). With this, the findings of this study encourage the prioritization of integrating CRE materials to facilitate competent evaluation training and practice concerning the design of logic models. This also extends to designing data visualizations overall.

Appendix A

Phase 1 Demographic Characteristics of Subject Matter Experts

	Frequency	Percent
Gender		
Woman	7	87.5
Man	1	12.5
Race		
Non-Latinx	8	100
Ethnicity		
White	5	62.5
Asian/Asian American	2	25
Decline to state	1	12.5
Education		
Associates	1	12.5
MA/MS	3	37
Doctorate Candidate	1	12.5
Doctorate	3	37.5

Appendix B

Phase 1 Professional Demographics Characteristics of Subject Matter Experts

	Frequency	Percent
Professional Identity		
Evaluator (in any capacity)	5	62.5
A student involved in evaluation (paid or unpaid)	1	12.5
College or university faculty member or instructor	1	12.5
Dual researcher-evaluator identity	1	12.5
Competency with using Logic Models		
I am proficient	4	50
I am an expert	4	50
Competency with communicating logic model benefits		
I am proficient	5	62.5
I am an expert	3	37.5
Competency with Survey Editing		
I am proficient	4	50
I am an expert	4	50
Competency with Survey Development		
I am proficient	4	50
I am an expert	4	50
Competency with Measurement Editing		
I am proficient	7	88
I am an expert	1	12.5
Competency with Measurement Development		
I am proficient	7	88
I am an expert	1	12.5
Evaluation Development		
Graduate level evaluation courses at a university	8	100
Internship	3	37.5
Work Experience	1	12.5
Professional Workshops	6	75

Appendix C

Phase 1 and 2 MTurk Participant Demographic Characteristics

	All (N = 80)				India (n = 40)				US (n = 40)			
	f	%	*M (SD)*	*Mdn*	*f*	%	*M (SD)*	*Mdn*	*f*	%	*M (SD)*	*Mdn*
Age			33.73 (9.56)	30.51			30 (5.17)	28.5			37.5 (11.36)	34.5
Gender												
Women	26	32.5			10	25			16	40		
Men	54	67.5			30	75			24	60		
Race												
Latinx/Hispanic	21	26.3			12	42.5			4	10		
Non-Latinx/Non-Hispanic	59	73.8			23	57.5			36	90		
Ethnicity												
American Indian/Alaskan Native	6	7.5			6	15			-	-		
Asian/Asian American	30	37.5			26	65			4	10		
Black or African American	6	7.5			-	-			6	15		
White	34	42.5			6	15			28	70		
Prefer to describe:	1	1.3			1	2.5						
Decline to state	3	3.8			1	2.5			2	5		
Citizenship Status												
U. S.	41	51.2			8	20			33	82.5		
Permanent US resident	9	11.3			3	7.5			6	15		
Foreign national	11	13.8			11	27.5			-	-		
Undocumented in the US	7	8.8			7	17.5			-	-		
Decline to state	12	15			11	27.5			1	2.5		
US Birth Country												
Yes	41	51.2			8	20			33	82.5		
No	39	48.8										
No response					32	80			7	17.5		

	All (N = 80)				India (n = 40)				US (n = 40)			
	f	%	*M (SD)*	*Mdn*	*f*	%	*M (SD)*	*Mdn*	*f*	%	*M (SD)*	*Mdn*
Education												
High school or equivalent	3	3.8							3	7.5		
Some college, no degree	5	6.3							5	12.5		
Associate degree	2	2.5							2	5		
Bachelor's Degree	50	62.5			30	75			20	50		
Masters degree	17	21.3			8	20			9	22.5		
Professional degree	2	2.5			2	5			1	2.5		
Decline to state	1	1.3							1	2.5		
Highest Education of Head of Family												
Less than a high school diploma	1	1.3							1	2.5		
High school or equivalent	8	10			2	5			6	15		
Some college, no degree	7	8.8			2	5				5	12.5	
Associate degree	2	2.5			1	2.5			1	2.5		
Bachelor's Degree	41	51.2			22	55				19	47.5	
Masters degree	19	23.8			11	27.5				8	20	
Professional degree	2	2.5			2	5						
Domestic servants in family			1.37 (1.26)	1			1.89 (1.26)	2			0.85 (1.04)	1
SSCM			4.77 (1.91)	5.5			5.46 (1.14)	5.83			4.09 (2.26)	5.33
Visual impairment(s)												
Yes	24	30			20	50			4	10		
No	55	68.8			20	50			35	87.5		
Decline to state	1	1.3							1	2.5		

Appendix D

Phase 1 Online Survey

Describe this is what the experts say in their online survey as instructions.

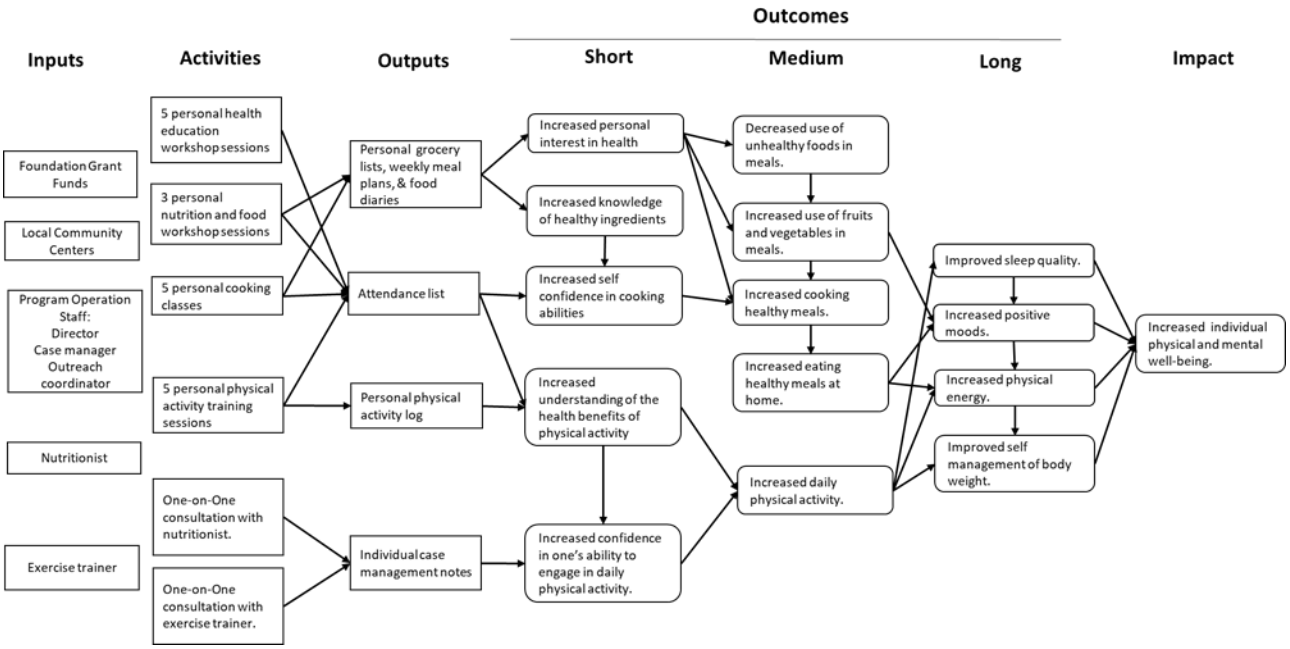
Please reflect on your knowledge of the benefits of logic models. Review the below list of benefits and determine whether they should be kept, reworded, or removed by selecting the appropriate response.

1) Logic models help clarify evaluation collaborators' knowledge about the relationship between program components
    a. Keep this benefit
    b. Reword this benefit
    c. Remove this benefit
2) Logic models help clarify evaluation collaborators' knowledge about the program's anticipated outcomes or goals.
    a. Keep this benefit
    b. Reword this benefit
    c. Remove this benefit
3) Logic models help provide an understanding of the program theory.
    a. Keep this benefit
    b. Reword this benefit
    c. Remove this benefit
4) Should this list of benefits be expanded?
    a. Yes
    b. No

    Skip logic: If "yes" is selected then show the following question:

    What additional benefits of logic models should be included in the definition?

Please review the present logic model.



This logic model will be used as a basis for a knowledge test. Using this logic model (which will be visible throughout this section) with your understanding of the benefits logic models, please provide your recommendation for what should be done with each test question by answering the following question:

Based upon your understanding of the benefits of logic models, what should be done with ***Logic Model Knowledge Test Question*** *#?*

a. It should be removed from the logic model knowledge test.
b. It should be kept in the logic model knowledge test.
c. It should be reworded. Please explain: _____
d. I am not sure what should be done.

Test Questions:

Logic Model Knowledge Test Question 1.
What short-term outcome is a result of physical activity log?

a. Physical activity training session (1)
b. Increased understanding of the health benefits of physical activity. (2)
c. Increased daily physical activity. (3)
d. Increased confidence in ability to engage in daily physical activity. (4)
e. All the above (5)
f. None of the above (6)

Logic Model Knowledge Test Question 2
What is the overall impact of the program?

a. Decrease use unhealthy foods in meals. (1)
b. Increased positive moods. (2)
c. Increased physical and mental well-being. (3)

d. Improved management of body weight. (4)
e. All the above (5)
f. None of the above (6)

Logic Model Knowledge Test Question 3.
Each medium-term outcome leads to one long-term outcome.
a. True (1)
b. False (2)

Logic Model Knowledge Test Question 4.
Which activity is connected to a grocery list, weekly meal plans, and a food diary?
a. One-on-one consultation with nutritionist. (1)
b. Physical activity training sessions. (2)
c. One-on-one consultation with personal trainer (3)
d. Nutrition and food workshops and cooking classes (4)
e. All the above (5)
f. None of the above (6)

Logic Model Knowledge Test Question 5.
What inputs are necessary to achieve the program outcomes?
a. Foundation grant funds (1)
b. Program staff, director, case manager, and outreach coordinator (2)
c. Nutritionist (3)
d. Personal trainer (4)
e. Local community centers
f. All the above (5)
g. None of the above (6)

Logic Model Knowledge Test Question 6.
What is the title of the image?
a. Saving the Baby Seals (1)
b. Reducing and Preventing Youth Tobacco Use (2)
c. Improve Reading Skills Through Dancing (3)
d. Increasing Happiness Through Puppies (4)
e. All the above (5)
f. None of the above (6)

Logic Model Knowledge Test Question 7.
Increased cooking healthy meals leads to increased eating healthy meals at home.
a. True (1)
b. False (2)

Logic Model Knowledge Test Question 8.
Inputs can be best defined as…
a. The supportive resources that make the program possible. (1)
b. People that are employed by the program. (2)

c. Results from program activities. (3)
d. People leading the program activities. (4)
e. All the above (5)
f. None of the above (6)

Logic Model Knowledge Test Question 9.
After viewing this image, activities can be described as
a. all the actions of the program that creates the program outputs. (1)
b. program tasks supported by the outcomes. (2)
c. tasks carried out by program staff. (3)
d. program tasks supported by the outputs. (4)
e. All the above (5)
f. None of the above (6)

Logic Model Knowledge Test Question 10.
Outputs can be described as the
a. documented evidence of program activities. (1)
b. items created by program activities. (2)
c. items given by program activities. (3)
d. size and range of items created by program activities. (4)
e. size and range items given by program activities. (5)
f. All the above (6)
g. None of the above (7)

Logic Model Knowledge Test Question 11.
Short-term outcomes can be described as the potential resources of the program.
a. True (1)
b. False (2)

Logic Model Knowledge Test Question 12.[1]
Which aspect of this image includes beneficial behaviors and actions people take due to the program's activities and outputs.
a. Outputs
b. Short-term outcomes
c. Medium-term outcomes
d. Long-term outcomes
e. Impact
f. All the above
g. None of the above

Logic Model Knowledge Test Question 13.
Outcomes can be defined as
a. beneficial changes resulting from activities and outputs. (1)

[1] Removed from measure in Phase 3

b. documented evidence of program activities. (2)
c. potential resources of the program. (3)
d. tasks completed by program staff. (4)
e. All the above (5)
f. None of the above (6)

Logic Model Knowledge Test Question 14.
Long-term outcomes can be described as beneficial changes to a condition or situation due to the program activities, short-, and medium-term outcomes.

a. True (1)
b. False (2)

Logic Model Knowledge Test Question 15.
Which aspect of this image refers to the program's purpose, mission, or vision?

a. Outputs (1)
b. Short-term outcomes (2)
c. Medium-term outcomes (3)
d. Long-term outcomes (4)
e. Impact (5)
f. All the above (6)
g. None of the above (7)

Logic Model Knowledge Test Question 16.
Overall, this image is designed to provide a picture of

a. how a program is intended to work (1)
b. the current success of a program. (2)
c. the current results of a program. (3)
d. All the above (4)
e. None of the above (5)

Logic Model Knowledge Test Question 17.
What short-term outcome is a result of physical activity log?

a. Physical activity training session (1)
b. Increased understanding of the health benefits of physical activity. (2)
c. Increased daily physical activity. (3)
d. Increased confidence in ability to engage in daily physical activity. (4)
e. All the above (5)
f. None of the above (6)

Appendix E

Phase 1 Subject Matter Experts Demographic Survey

Professional Demographic Survey
To ensure transparency of about the various backgrounds and perspectives represented among those that have participated in this survey, I'd like to learn more about you. Please provide me with more information about your personal background and professional experience, by answering the following questions.

What is the highest degree or level of school you have completed?

a. Less than a high school diploma
b. High school degree or equivalent (e.g., GED)
c. Some college, no degree
d. Associate degree (e.g., AA, AS)
e. Bachelor's degree (e.g., BA, BS)
f. Master's degree (e.g., MA, MS, Med)
g. Professional degree (e.g., MD, DDS, DVM)
h. Doctorate (e.g., PhD, EdD)
i. Prefer to describe: _____
j. Prefer not to describe.

What evaluation development experience do you have? (Select all that apply)

a. Professional workshops
b. Graduate level evaluation courses at a university
c. Undergraduate level courses at a college or university
d. Internship
e. Prefer to describe: _____
f. Prefer not to describe.

What is currently your primary professional identity in the evaluation field?

a. Evaluator (in any capacity)
b. Student involved in evaluation (paid or unpaid)
c. College or university faculty member or instructor
d. Researcher
e. Retired, but no longer active in the evaluation field
f. No longer active in the evaluation field.
g. Retired but still active in the evaluation field in some way(s)
h. Trainer
i. Unemployed or currently seeking employment
j. Prefer to describe: _____
k. Prefer not to describe.

Please tell more about your evaluation and research competencies by answering the following questions.

How competent are you with using a program logic model?

a. I do not have any knowledge of how to do this.
b. I am a novice.
c. I am proficient.
d. I am an expert.

How competent are you with communicating the benefits of using a program logic model?

a. I do not have any knowledge of how to do this.
b. I am a novice.
c. I am proficient.
d. I am an expert.

How competent are you with creating a set of survey items?

a. I do not have any knowledge of how to do this.
b. I am a novice.
c. I am proficient.
d. I am an expert.

How competent are you with editing an existing survey?

a. I do not have any knowledge of how to do this.
b. I am a novice.
c. I am proficient.
d. I am an expert.

How competent are you with developing a test to measure knowledge on a subject?

a. I do not have any knowledge of how to do this.
b. I am a novice.
c. I am proficient.
d. I am an expert.

How competent are you with editing test questions designed to measure knowledge on a subject?

a. I do not have any knowledge of how to do this.
b. I am a novice.
c. I am proficient.
d. I am an expert.

Lastly, I'm interested in knowing more about your demographic background to determine who is represented among the survey responses. Please answer the following questions.

What is your gender?

a. Male
b. Female
c. Prefer to describe: _____
d. Prefer not to answer.

Are you of Hispanic, Latino, or of Spanish origin?

a. Yes
b. No

How would you describe yourself?

a. American Indian or Alaska Native
b. Asian
c. Black or African American
d. Native Hawaiian or Pacific Islander
e. White
f. Prefer to describe: _____
g. Prefer not to describe.

Thank you for providing details about your personal and professional background. I look forward to speaking with you during the focus group and sharing the survey findings.

Appendix F

Phase 1 Informed Consent for Online Survey and Focus Group

You are being asked to participate in a research project conducted by Ciara Knight, a graduate student studying Evaluation and Applied Research Methods in the School of Behavioral and Organizational Sciences at Claremont Graduate University, under the guidance of Tarek Azzam, Ph.D., Assistant Professor of Evaluation at Claremont Graduate University.

Purpose of the Study
The aim of the study is to develop and validate a newly developed logic model knowledge test. You were selected due to your knowledge of evaluation, program theory, logic models, survey research, or test and measure development. This is addition to your affiliation with Claremont Graduate University's evaluation community

Procedures
If you volunteer to participate in this study, you will be asked to indicate your consent at the end of this page by selecting " I have read the informed consent statement and give my consent to participate in this study" and then click the ">>." Please reflect on your knowledge of the benefits of logic models. Participants will be asked to review the benefits of logic models as criteria for the knowledge test and review the base logic model for the knowledge test. Then, you will be asked to use the base logic model and your understanding of the benefits of logic models to determine whether each test question should be kept, reworded, or removed. Lastly, you will be asked to provide additional information about your personal background and professional evaluation experience. In all, the survey should take approximately 20 minutes to complete. After completing the online questionnaire, please press the ">>" button on the last page to submit your responses. Upon completion of the survey, you will be able to participate in a one-hour focus group to discuss the survey findings and help finalize the logic model knowledge test.

Potential Risks
The potential risks of taking part in this study are extremely low. You may feel slightly uncomfortable answering any demographic questions, which is typical in many studies.

Potential Benefits
Your participation in the study may provide the field with information on how to assess stakeholders understanding of a logic model. This information may also help you to reflect on your own practice.

Compensation
There is no compensation for your participation in this study.

Participation and Withdrawal
Please understand that participation is completely voluntary. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind.

Confidentiality
Any information that is obtained about this your survey response and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. Confidentiality will be maintained by means of using aggregated data, password protection of files, and names will not be attached in any way to survey responses.

Rights of Research Subjects
You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal rights because of your participation in this research study. The Claremont Graduate University Institutional Review Board, which is administered through the Office of Research and Sponsored Programs (ORSP), has reviewed this project. You may also contact ORSP at (909) 607-9406 with any questions.

Identification of Investigator
If you have any questions or concerns about the research, please feel free to contact:

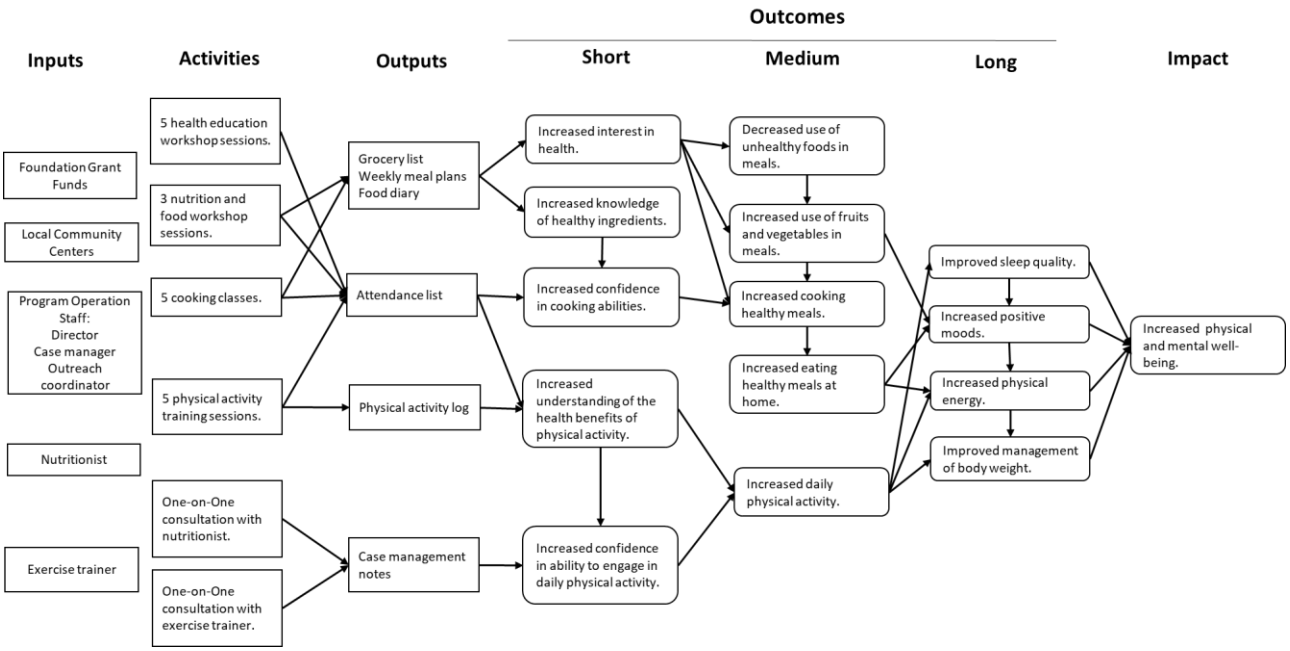
Tarek Azzam, Ph.D.
Assistant Professor
School of Behavioral and Organizational Sciences
Claremont Graduate University
123 East 8th Street
Claremont, CA 91711
Tel: (909) 374-5355
Fax: (909) 607-9009

- o I have read the informed consent statement and give my consent to participate in this study.
- o I do not want to participate in this study.

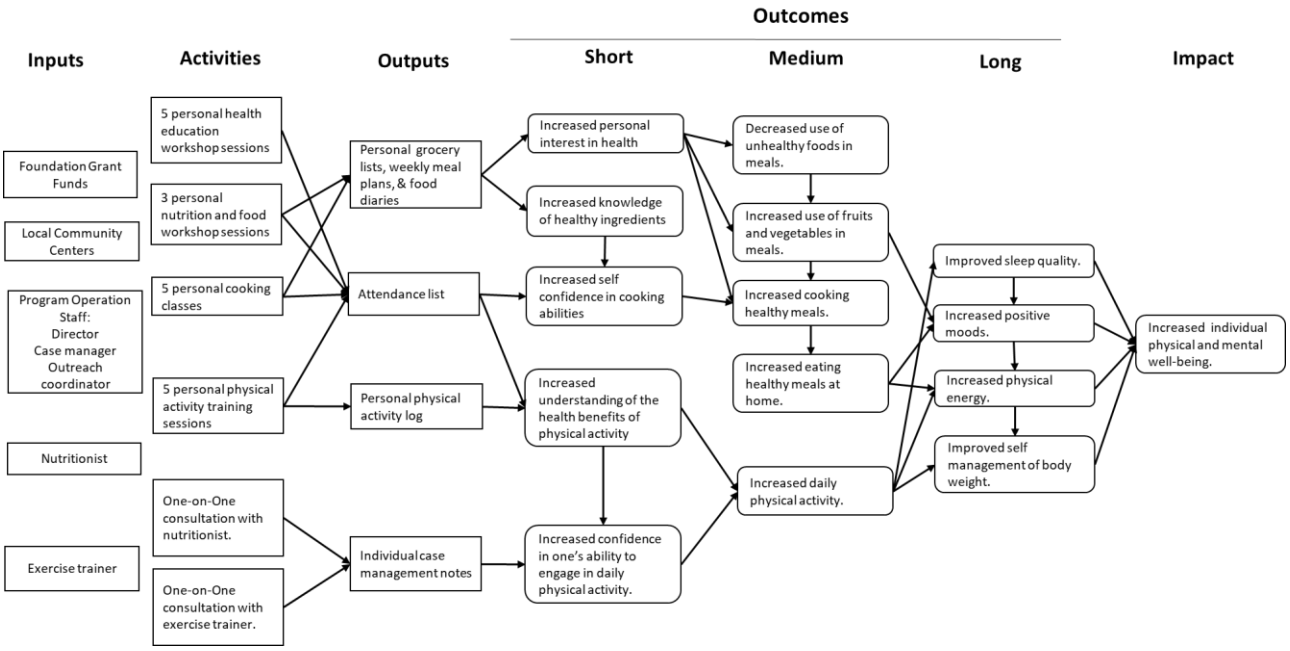
Appendix G

Logic Model Conditions

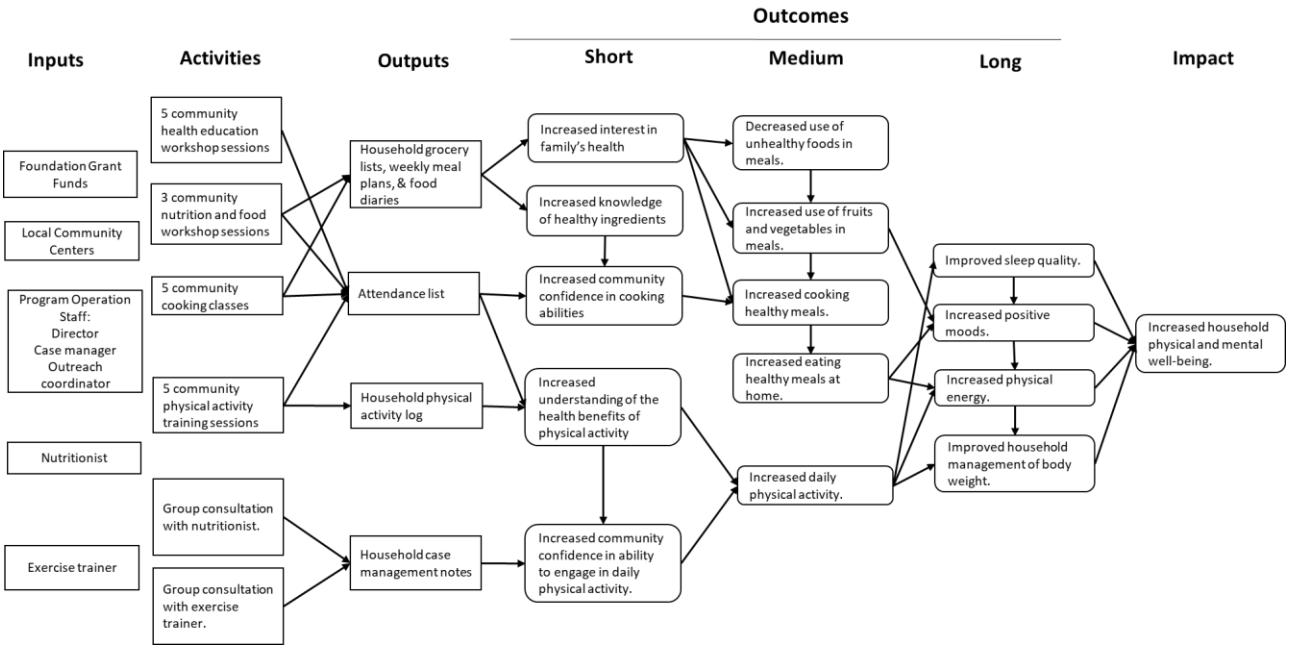
Condition 1 Control



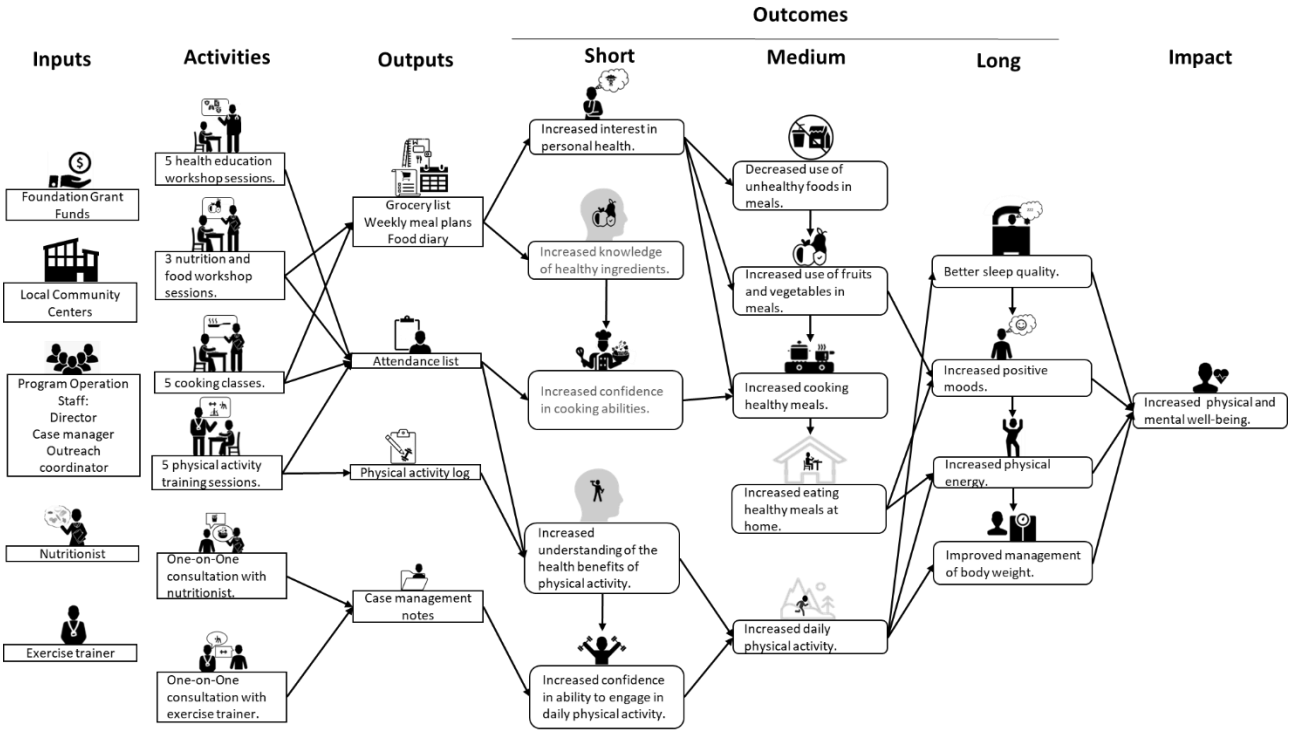
Condition 2 Individualistic Language (IL)



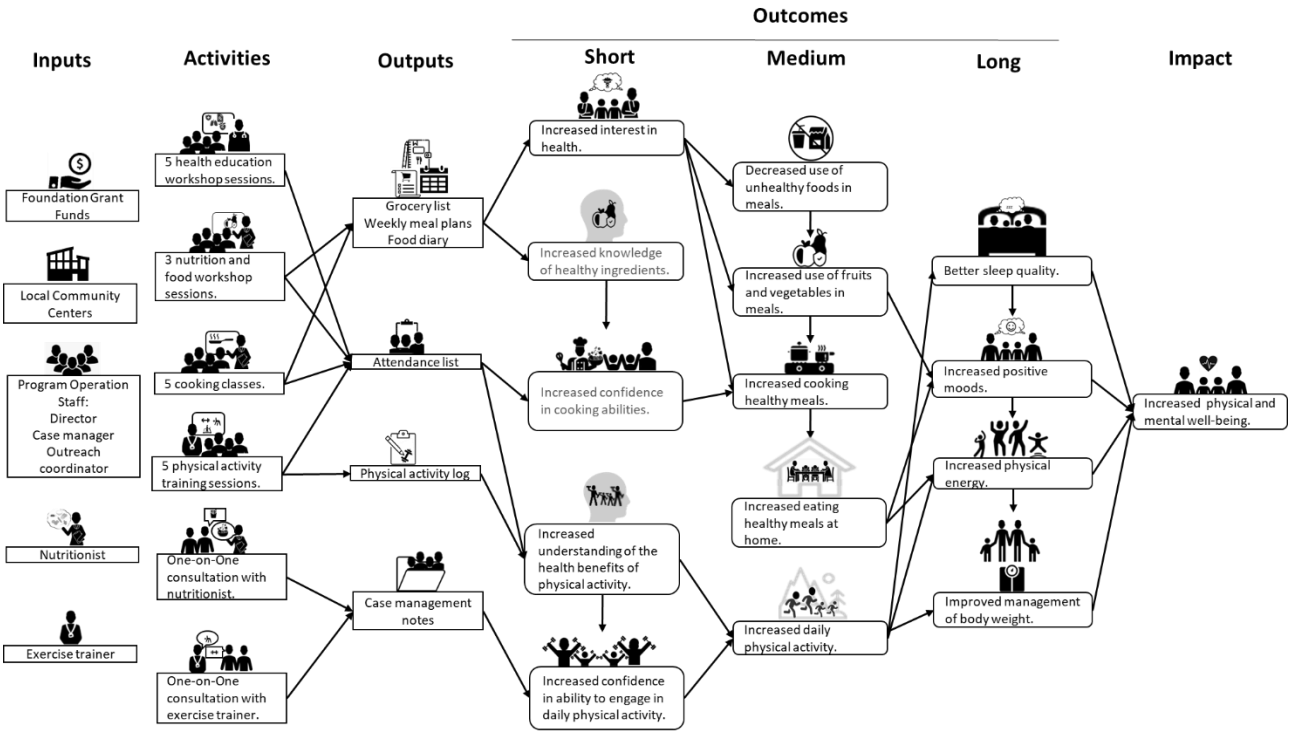
Condition 3 Collectivistic Language (CL)



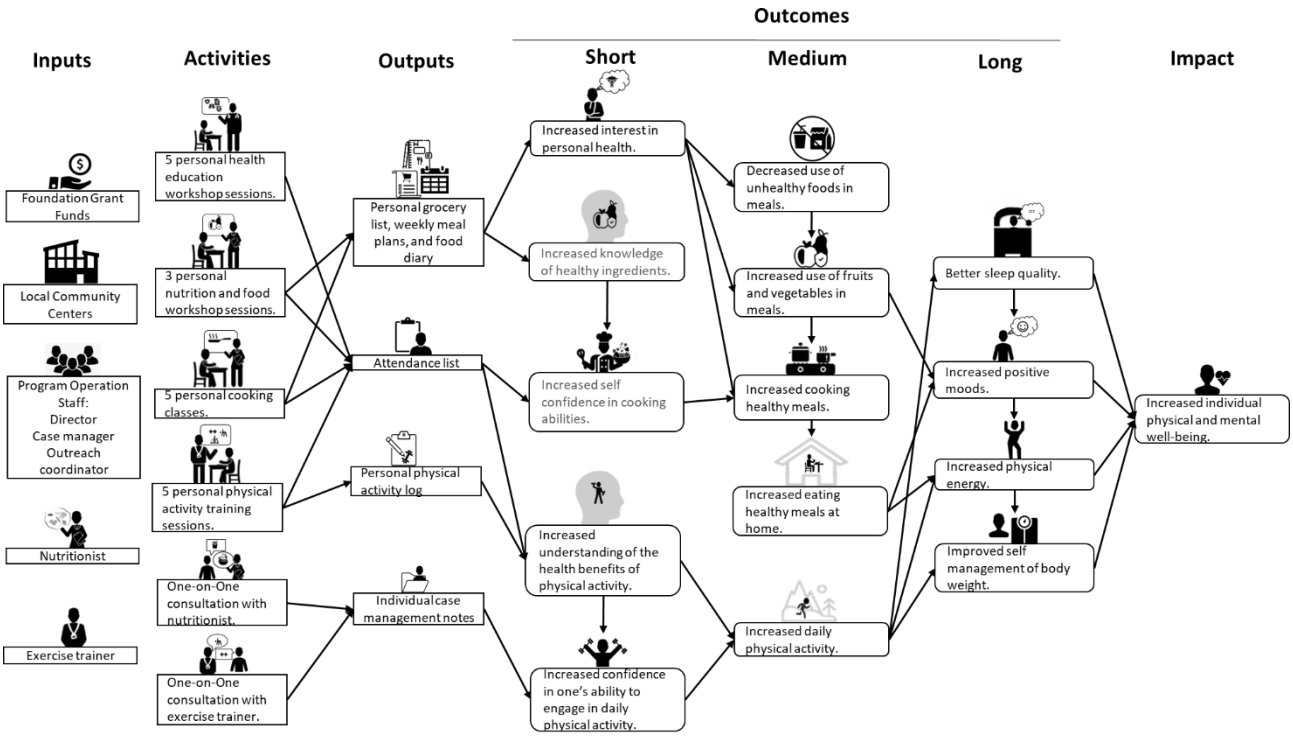
Condition 4 Individualistic Visualization (IV)



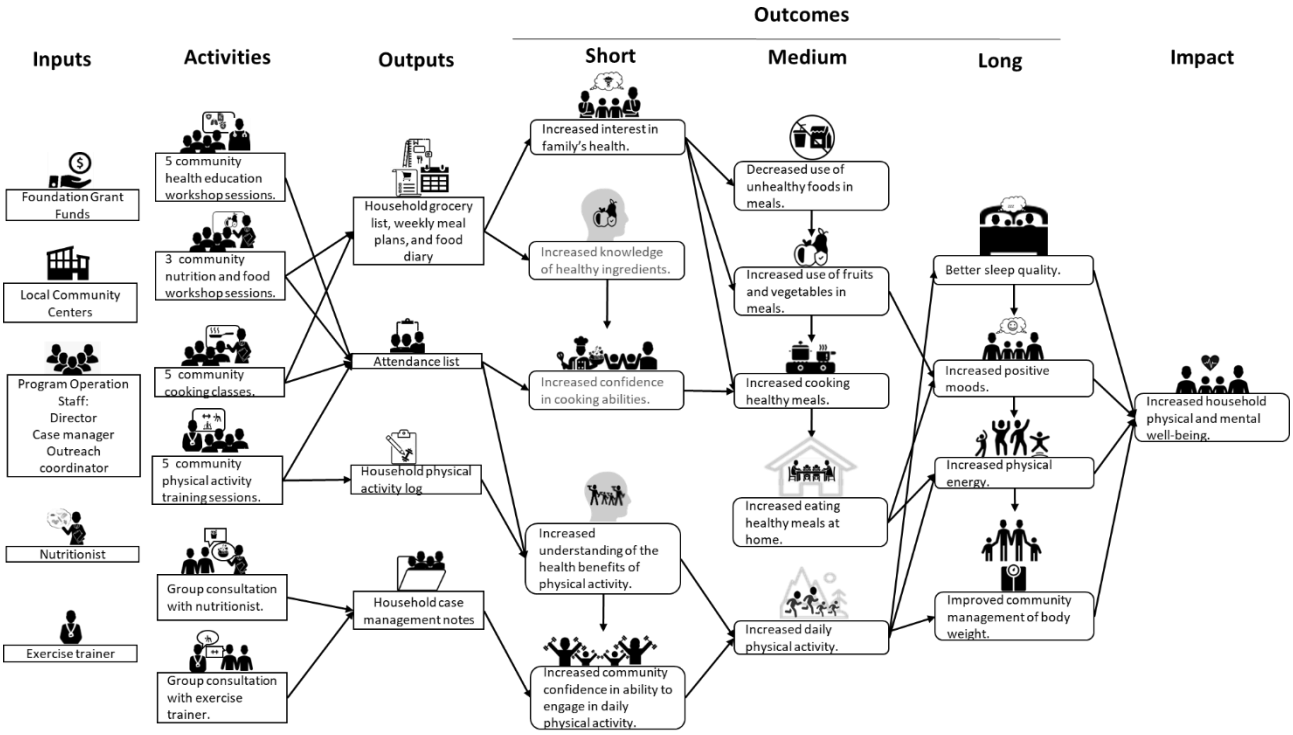
Condition 5 Collectivistic Visualization (CV)



Condition 6 Individualistic Language and Visualization (ILV)



Condition 7 Collectivistic Language and Visualization (CLV)



Appendix H

Phase 1 and 2 MTurk Online Survey Informed Consent and Focus Group Protocol

Introduction

"Thank you so much for sharing your time with me. I appreciate it! I would like to give you a quick orientation and overview of the informed consent for our session."

Informed Consent

[The researcher will read this aloud and a copy will be emailed before the session]

You are being asked to participate in a one-hour follow-up focus group conducted by me, Ciara Knight, a graduate student studying Evaluation and Applied Research Methods in the School of Behavioral and Organizational Sciences at Claremont Graduate University, under the guidance of Tarek Azzam, Ph.D., Assistant Professor of Evaluation at Claremont Graduate University.

Purpose of the Study. The aim of the study is to develop and validate a newly developed logic model knowledge test. You were selected due to your availability and recent completion of the survey about test questions for a logic model knowledge test.

Procedures. During this focus group session, you will be presented with the survey findings. Based upon these findings, the group will be guided into a discussion about potential mixed survey findings about the benefits of logic models and test questions. This will be done to gain a consensus about what should be included in the logic model knowledge test.

Potential Risks. The potential risks of taking part in this study are extremely low. You may feel slightly uncomfortable if you do not like group discussions.

Potential Benefits. Your participation in the study may provide the field with information on how to assess stakeholders understanding of a logic model. This information may also help you to reflect on your own practice.

Compensation. There is no compensation for your participation in this study.

Participation and Withdrawal. Please understand that participation is completely voluntary. If you volunteer to be in this study, you may still withdraw at any time without consequences of any kind.

Confidentiality. Due to the group setting, I am unable to guarantee confidentiality. Therefore, it is recommended that you so not mention any information that you do not feel comfortable with others knowing. This session will be recorded, and a third-party transcriber will have access to it. Once transcripts are completed, the recording will be deleted.

Rights of Research Subjects. You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal rights because of your participation in this research study. The Claremont Graduate University Institutional Review Board, which is

administered through the Office of Research and Sponsored Programs (ORSP), has reviewed this project. You may also contact ORSP at (909) 607-9406 with any questions.
Identification of Investigator. If you have any questions or concerns about the research, please feel free to contact:

Tarek Azzam, Ph.D.
Assistant Professor
School of Behavioral and Organizational Sciences
Claremont Graduate University
123 East 8th Street
Claremont, CA 91711
Tel: (909) 374-5355
Fax: (909) 607-9009

Do you understand and consent to participating in this recorded focus group session? (If anyone says no, they will be thanked for their time).

Administrative

"A copy of this informed consent was emailed to you when you were invited to participate in this focus group. If you have not received it or would like me to resend it, please let me know. I am happy to do this. Before we begin, do you have any questions for me?"

Survey Findings

"Thank you for completing the online survey. It really helped me to focus this session on the mixed or inconsistent findings. Here are the survey response trends."

[Show aggregated survey response descriptive statistics.]

Based upon these trends, the following questions had mixed responses:

[List the questions that had mixed responses.]

Focus Group Questions

"Our session will focus on each of these so we can gain a consensus about what should be included in the logic model knowledge test. We will discuss one question at a time."

[Present first question with mixed responses. Based upon the question, each of the below focus group questions will be asked.]

1. What are additional benefits of logic models that should be reflected in this test? Why? Now, that we have discussed this benefit. What do you think should be done with it?
    a. Keep this benefit
    b. Reword this benefit
    c. Remove this benefit
2. Why should this question be removed?
3. Why should this question be edited?

4. How should this question be phrased differently for clarity?
5. Now, that we have discussed this question. What do you think should be done with it?
    a. It should be removed from the logic model knowledge test.
    b. It should be kept in the logic model knowledge test.
    c. It should be reworded. Please explain: _____
    d. I am not sure what should be done.
6. What questions should be added to this test and why?
7. Now, that we have discussed this question. What do you think should be done with it?
    a. It should be removed from the logic model knowledge test.
    b. It should be kept in the logic model knowledge test.
    c. It should be reworded. Please explain: _____
    d. I am not sure what should be done.
8. How does each question reflect the stated benefits of logic models?
9. Do you believe this test is ready for use?

Thank you so much for lending me your time and expert insights about how this logic model knowledge test should be developed. It is much appreciated.

Appendix I

IC Measure 1 – Wagner (1995)

Please tell me what characteristics contributes to a successful life. Using the rating scale of 1 (strongly disagree) to 5 (strongly agree), please your level of agreement with each statement.

1. Only those who depend on themselves get ahead in life.
2. To be superior a person must stand alone.
3. If you want something done right, you've got to do it yourself.
4. What happens to me is my own doing.
5. In the long run the only person you can count on is yourself.
6. Winning is everything.
7. I feel that winning is important in both work and games.
8. Success is the most important thing in life.
9. It annoys me when other people perform better than 1 do.
10. Doing your best isn't enough; it is important to win.
11. I prefer to work with others in a group rather than working alone.
12. Given the choice, I would rather do a job where I can work alone rather than doing a job where I have to work with others in a group.
13. Working with a group is better than working alone.
14. People should be made aware that if they are going to be part of a group then they are sometimes going to have to do things they don't want to do.
15. People who belong to a group should realize that they're not always going to get what they personally want.
16. People in a group should realize that they sometimes are going to have to make sacrifices for the sake of the group as a whole.
17. People in a group should be willing to make sacrifices for the sake of the group's well-being.
18. A group is more productive when its members do what they want to do rather than what the group wants them to do.
19. A group is most efficient when its members do what they think is best rather than doing what the group wants them to do.
20. A group is more productive when its members follow their own interests and concerns.

Appendix J

IC Measure 2 -Adapted Concise Scale of IC (CSIC; Chen et al., 2015)

Please describe people's role in society. Using the rating scale of 1 (strongly disagree) to 5 (strongly agree), please your level of agreement with each statement.

1. All individuals in the society are absolutely independent from each other.
2. Individuals are much more important than the group.
3. Everyone must put his/her own interests first.
4. One can do better by working alone than in a group.
5. Groups, including the government, should not interfere with personal behavior.
6. The value of a person is solely determined by his/her personal achievements.
7. A person must follow only his/her own ideas of how to act
8. A person must follow only his/her own ideas of how to behave.
9. It is more than enough to focus only on one's own business.
10. It is essential to maintain one's personal characteristics in work
11. It is essential to maintain one's personal characteristics in daily life.
12. All individuals in a society are closely related to each other.
13. Individuals may not be able to survive if there is no group
14. Individuals may not be able to survive if there is no country.
15. To ensure group interests are met, self-interests must be sacrificed.
16. An individual's talents can be realized only through teamwork.
17. An individual's talents can be realized only through group collaboration.
18. Individuals should be unconditionally submissive to the group.
19. Individuals should be unconditionally submissive to the nation.
20. The value of a person is determined primarily by assessments of oneself that are made by others.
21. The value of a person is determined primarily by assessments of oneself that are made by the society.
22. Every one of us must consult others about how to act.
23. Every one of us must consult others about how to behave.
24. It is much more important to help others than it is to mind your own business.
25. One must conform to the opinion of the majority in work.
26. One must conform to the opinion of the majority in daily life.

Appendix K

IC Measure 3 – Adapted Oyserman's (1993) Measure

Please describe how character contributes to fulfillment in various aspects of life. Using the rating scale of 1 (strongly disagree) to 5 (strongly agree), please your level of agreement with each statement.

1. In the end, a person feels closer to members of his/her own group than to others
2. A mature person understands that he/she must act in accordance with the honor of the group
3. A person of character helps his/her group before all else
4. A mature person understands the needs of the group so he/her she may act to fulfill them
5. In order to really understand who I am, you must see me with members of my group
6. If you know what groups I belong to, you know who 1am
7. What is good for my group is good for me
8. Without group loyalty there is no self-actualization
9. My personal goals match those of my group
10. Advancement in life are dependent on self-initiative
11. In the end, achievements define the person
12. A mature person knows his/her abilities to obtain maximum utility from them
13. A person of character attempts to act on his/her values to attain goals without depending on others
14. A person of weak character forms opinions in consultation with his/her friends
15. To advance, a person must be willing to sacrifice social relations
16. The decisions I make on my own are better
17. Investing a lot of time in social relationships makes achieving one's potential harder
18. I feel uncomfortable if I find I am very similar to the others in my group

Appendix L

MTurk Sample Demographic Survey

Subjective Social Class Measure (SSCM; Belmi & Neale,2014)

Now, we are interested in how you view your role in society. Please use the rating scale of 1 (strongly disagree) to 7 (strongly agree), please your level of agreement with each statement.

1. I feel as though I know what it's like to belong to a high social class.
2. I feel as though I am part of the elite group in society.
3. I feel as though I can identify with the lives of the rich.
4. I feel I can identify with those who have a lot of money.
5. I feel as though I am part of with those who live a life of wealth and privilege.
6. I feel as though I am part of the elite.

To ensure transparency of about the various backgrounds and perspectives represented among those that have participated in this survey, I would like to learn more about you. Please provide me with more information about your personal background, by answering the following questions.

7. What is the highest degree or level of school you have completed?
    k. Less than a high school diploma
    l. High school degree or equivalent (e.g., GED, certificate)
    m. Some college, no degree (e.g., post high school diploma)
    n. Associate degree (e.g., AA, AS)
    o. Bachelor's degree (e.g., BA, BS)
    p. Master's degree (e.g., MA, MS, Med)
    q. Professional degree (e.g., MD, DDS, DVM)
    r. Doctorate (e.g., PhD, EdD)
    s. Prefer to describe: _____
    t. Prefer not to describe.

8. What is the highest level of education of your head of family?
    a. Less than a high school diploma
    b. High school degree or equivalent (e.g., GED, certificate)
    c. Some college, no degree (e.g., post high school diploma)
    d. Associate degree (e.g., AA, AS)
    e. Bachelor's degree (e.g., BA, BS)
    f. Master's degree (e.g., MA, MS, Med)
    g. Professional degree (e.g., MD, DDS, DVM)
    h. Doctorate (e.g., PhD, EdD)
    i. Prefer to describe: _____
    j. Prefer not to describe.

9. How many domestic servants (e.g., in-home childcare providers, cooks, cleaners, etc.) does your family have?
    a. _____
    b. Prefer not to answer.

10. What is your gender?
    e. Male
    f. Female
    g. Prefer to describe: _____
    h. Prefer not to answer.

11. How old are you?
    a. _____
    b. b. Prefer not to answer.

12. What is your citizenship status?
    a. I am a United States citizen
        - **Display logic.** If selected, the following questions will appear:
            1. Were you born in the United States?
                a. Yes
                b. No
                c. I prefer not to answer
    b. I am a permanent resident in the United States.
    c. I am a foreign national
        - **Display logic.** If selected, the following question will appear:
            1. What is your country of citizenship? _____.
    d. I am undocumented in the United States.
    e. I prefer not to answer

22. Are you of Hispanic, Latino, or of Spanish origin?
    c. Yes
    d. No
    e. I prefer not to answer

23. How would you describe yourself? Please select all that apply.

    h. American Indian or Alaska Native
    i. Asian
    j. Black or African American
    k. Native Hawaiian or Pacific Islander
    l. White
    m. Prefer to describe (e.g., Indian Punjabi, Chinese, Samoan, Afro Cuban, Mestizo): _____
    n. Prefer not to describe.

24. I understand that people identify themselves differently than common questions on surveys which are often very limited. Please feel welcome to tell me how you personally describe yourself. _____

25. Do you have any visual impairments that may have impacted your ability to accurately complete this survey?
    a. Yes
    b. No

26. How would you describe your ability to read in the English language? (Included in Phase 3)
    a. Reading the English language is very difficult for me.
    b. Reading the English language is difficult for me.
    c. Reading the English language is neither easy nor difficult for me.
    d. Reading the English language is easy for me.
    e. Reading the English language is very easy for me.

Lastly, what do you think is the purpose of this Mturk project? (Included in Phase 3)

Appendix M

Phase 1 Logic Model Revised Benefits Table

	Original Benefits	Revised Benefits
	B1. Logic models help clarify evaluation collaborators' knowledge about the relationship between program components	B3. Logic models clarify the relationship(s) between program components and goals.
	B2. Logic models help clarify evaluation collaborators' knowledge about the program's anticipated outcomes or goals.	B2. Logic models clarify the program's expected outcomes or goals.
	B3. Logic models help provide an understanding of the program theory.	B1. Logic models clarify the program's components.
Fleiss Kappa	κ = -.091 (95% *CI*, -.305 to .092), p =.123	κ = 1.00, p < .001

Appendix N

Phase 1 LMKT Question Revisions Table

	Original Questions κ = .031 (95% *CI*, -.030 to .092), *p* = .319	Revised Questions κ = .646 (95% *CI*, .576 to .717), *p* = .001
1	What short-term outcome is a result of physical activity log?	What short-term outcome is directly expected from the "[individual/household] physical activity log" output?
2	What is the overall impact of the program?	What is the expected impact of the program?
3	Each medium-term outcome leads to one long-term outcome.	Each medium-term outcome directly leads to one long-term outcome.
4	Which activity is connected to a grocery list, weekly meal plans, and a food diary?	Which activities are directly connected to "a grocery list, weekly meal plans, and a food diary?"
6	What is the title of the image?	What is the topic of this program?
7	Increased cooking healthy meals leads to increased eating healthy meals at home.	"Increased cooking of healthy meals" leads to "increased eating healthy meals at home."
10	Outputs can be described as the a. documented evidence of program activities. b. items created by program activities. c. items given by program activities. d. size and range of items created by program activities. e. size and range items given by program activities. f. All the above g. None of the above	Outputs can be described as the a. documented evidence of program activities. b. behavior changes as a result of the program. c. resources that the program needs. d. ultimate goal of the program. e. All the above. f. None of the above
11	Short-term outcomes can be described as the potential resources of the program.	Short-term outcomes can be described as the inputs of the program.
12	Which aspect of this image includes beneficial behaviors and actions people take due to the program's activities and outputs.	Which aspect of this image includes beneficial behaviors and actions people take due to the program.
14	Long-term outcomes can be described as beneficial changes to a condition or situation due to the program activities, short-, and medium-term outcomes.	Long-term outcomes can be described as beneficial changes to a condition or situation due to the program's collective activities, outputs, short-, and medium-term outcomes.

Appendix O

Phase 3 MTurk Participant Demographic Characteristics

	All (N=361)		Individualistic (n = 170)		Collectivistic (n = 191)	
	f	%	*f*	%	*f*	%
Gender						
Women	86	23.8	53	31.2	30	16.1
Men	275	76.2	117	68.8	156	83.9
Race						
Latinx/Hispanic	124	34.3	88	51.8	36	19.4
Non-Latinx/Non-Hispanic	236	65.4	82	48.2	149	80.1
Decline to state	-	-	-	-	1	0.5
Ethnicity						
American Indian/Alaskan Native	27	7.5	20	11.8	7	3.8
Asian/Asian American	141	39.1	81	47.6	60	32.3
Black or African American	12	3.3	6	3.5	5	2.7
White	157	43.5	52	30.6	102	54.8
Prefer to describe	23	6.4	11	6.5	12	6.5
Location						
U. S.	182	50.42	71	41.76	111	58.12
India	179	49.58	99	58.24	80	41.88
US Citizenship						
US citizen	195	54	79	46.5	111	59.7
Permanent US resident	26	7.2	11	6.5	15	8.1
Foreign national	55	15.2	20	11.8	35	18.8
Undocumented in the US	34	9.4	33	19.4	1	0.5
Decline to state	50	13.9	26	15.3	24	12.9
US Birth Country						
Yes	187	51.8	76	44.7	107	57.5
No	7	1.9	3	1.8	3	1.6
Decline to state	1	0.3	-	-	1	0.5
Education						
Less than a high school diploma	4	1.1	2	1.2	1	0.5
High school or equivalent	20	5.5	10	5.9	10	5.4
Some college, no degree	11	3	3	1.8	8	4.3
Associate degree	13	3.6	3	1.8	10	5.4
Bachelor's Degree	221	61.2	111	65.3	109	58.6
Masters degree	90	24.9	41	24.1	47	25.3
Professional degree	1	0.3	-	-	1	0.5
Decline to state	1	0.3	-	-	-	-

	All (N=361)		Individualistic (n = 170)		Collectivistic (n = 191)	
	f	%	*f*	%	*f*	%
Highest Education of Head of Family						
Less than a high school diploma	7	1.9	2	1.2	5	2.7
High school or equivalent	32	8.9	13	7.6	18	9.7
Some college, no degree	20	5.5	9	5.3	11	5.9
Associate degree	19	5.3	10	5.9	8	4.3
Bachelor's Degree	181	50.1	86	50.6	93	50
Masters degree	94	26	46	27.1	48	25.8
Professional degree	5	1.4	4	2.4	1	0.5
Doctorate	2	0.6	-	-	1	0.5
Decline to state	1	0.3	-	-	1	0.5
Visual impairment(s)						
Yes	198	54.8	101	59.4	97	52.2
No	160	44.3	69	40.6	86	46.2
Decline to state	3	0.8	-	-	3	1.6
Ability to Read English						
Very difficult	51	14.1	24	14.1	26	14
Difficult	41	11.4	15	8.8	26	14
Neither Easy nor Difficult	35	9.7	20	11.8	15	8.1
Easy	78	21.6	53	31.2	25	13.4
Very easy	156	43.2	58	34.1	94	50.5

	All (N=361)		Individualistic (n = 170)		Collectivistic (n = 191)	
	M (SD)	*Mdn*	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
Age	33.68 (7.90)	31	32.42 (5.99)	30.5	34.8 (9.18)	31
Domestic servants in family	6.70 (55.75)	2	8.97 (76.52)	2	4.78 (26.25)	2
SSCM	5.30 (1.46)	5.83	6.15 (.65)	6.33	5 (1.54)	6

Appendix P

Phase 3 Final Logic Model Knowledge Test

1. What short-term outcome is directly expected from the "[individual/household] physical activity log" output?
   a. "[Personal/Household] Physical activity training session"
   b. "Increased understanding of the health benefits of physical activity"
   c. "Increased daily physical activity"
   d. "Increased [community] confidence in [one's] ability to engage in daily physical activity"
   e. All the above
   f. None of the above

3. What is the expected impact of the program?
   a. "Decrease use unhealthy foods in meals"
   b. "Increased positive moods"
   c. "Increased physical and mental well-being"
   d. "Improved [self/community] management of body weight"
   e. All the above
   f. None of the above

4. Each medium-term outcome directly leads to one long-term outcome.
   a. True
   b. False

5. Which activities are directly connected to a "[Household/Individual] grocery list, weekly meal plans, and a food diary?"
   a. "[One-on-one/Group consultation with nutritionist"
   b. "[Personal/Household] Physical activity training sessions"
   c. "[One-on-one/Group] consultation with personal trainer"
   d. "[Personal/Community] Nutrition and food workshops and cooking classes"
   e. All the above
   f. None of the above

6. What inputs are necessary to achieve the program outcomes?
   a. "Foundation grant funds"
   b. "Program staff, director, case manager, and outreach coordinator"
   c. "Nutritionist"
   d. Exercise trainer"
   e. "Local community centers"
   f. All the above
   g. None of the above

7. What is the topic of this program?
   a. Saving the Baby Seals
   b. Reducing and Preventing Youth Tobacco Use

c. Improve Reading Skills Through Dancing
d. Increasing Happiness Through Puppies
e. All the above
f. None of the above

8. "Increased cooking healthy meals" leads to "increased eating healthy meals at home."
   a. True
   b. False

9. Inputs can be best defined as…
   a. The supportive resources that make the program possible.
   b. People that are employed by the program.
   c. Results from program activities.
   d. People leading the program activities.
   e. All the above
   f. None of the above

10. After viewing this image, activities can be described as
   a. all the actions of the program that creates the program outputs.
   b. program tasks supported by the outcomes.
   c. tasks carried out by program staff.
   d. program tasks supported by the outputs.
   e. All the above
   f. None of the above

11. Outputs can be described as the
   a. documented evidence of program activities.
   b. b. behavior changes as a result of the program.
   c. c. resources that the program needs.
   d. d. ultimate goal of the program.
   e. All the above
   f. None of the above

12. Short-term outcomes can be described as the inputs of the program.
   a. True
   b. False

13. Long-term outcomes can be described as beneficial changes to a condition or situation due to the program's collective activities, short-, and medium-term outcomes.
   a. True
   b. False

14. Which aspect of this image refers to the program's purpose, mission, or vision?
   a. Outputs
   b. Short-term outcomes
   c. Medium-term outcomes

d. Long-term outcomes
e. Impact
f. All the above
g. None of the above

15. Overall, this image is designed to provide a picture of
    a) how a program is intended to work
    b) the current success of a program.
    c) the current results of a program.
    d) All the above
    e) None of the above

Appendix Q

Phase 3 Informed Consent

You are being invited to participate in a research project conducted by Ciara Knight, a graduate student studying Evaluation and Applied Research Methods in the School of Behavioral and Organizational Sciences at Claremont Graduate University, under the guidance of Tarek Azzam, Ph.D., Senior Visiting Fellow at Claremont Graduate University. Volunteering will probably not benefit you directly, but you will be helping meet my graduate academic requirements and contribute to the knowledge-base concerning interpretation of images. If you volunteer, your identity will be anonymous. You will be given an image describing a program to look at and asked to answer questions concerning that image. This will take about 20 to 30 minutes of your time. Volunteering for this study involves no more risk than what a typical person experiences on a regular day. Your involvement is entirely up to you. You may withdraw at any time for any reason. Please continue reading for more information about the study.

PURPOSE:
The purpose of this study is to determine people's understandability and perceptions of an image about a program using a crowdsourcing platform.

ELIGIBILITY:
Participants must be located either in the US or India. They also must be over the age of 18 and complete the survey using a computer, not a mobile device.

PARTICIPATION:
During the study, you will be given an image describing a program to look at and asked to answer questions concerning that image. I expect your participation to take approximately 20-30 minutes to complete.

RISKS OF PARTICIPATION:
The risks of taking part in this study are minimal and are not greater than those ordinarily encountered in daily life or during the performance of routine work.

BENEFITS OF PARTICIPATION:
I do not expect the study to benefit you personally. This study will benefit the researcher because it will meet doctorate degree requirements and may result in publication of findings.

COMPENSATION:
You will receive $4.00 for your participation as compensation.

VOLUNTARY PARTICIPATION:
Please understand that participation is completely voluntary. Your decision whether or not to participate will in no way affect your current or future relationship with CGU or its faculty, students, or staff. You have the right to withdraw from the research at any time without penalty. You also have the right to refuse to answer any question(s) for any reason, without penalty.

CONFIDENTIALITY:
Your individual privacy will be anonymous for all aspects of this study and use of data for all publications, presentations, or future research resulting from this study. If this data is shared with other researchers, your identity will be kept anonymous, meaning any identifiers will be removed before other researchers gain access to the dataset. All responses will be aggregated and stored in a password protected computer, accessible only to the principal investigator.

FURTHER INFORMATION:
If you have any questions or concerns about the research, please feel free to contact me or my faculty supervisor:

Ciara Knight, MA, MS
ciara.knight@cgu.edu
(909) 556-0719

Tarek Azzam, Ph.D.
tarek.azzam@cgu.edu
(909) 374-5355

The Claremont Graduate University Institutional Review Board (IRB), which is administered through the Office of Research and Sponsored Programs (ORSP), has certified this project as exempt. You may also contact the IRB at (909) 607-9406 with any questions if you have questions about your rights as a research subject.

CONSENT:
Do you want to participate in this study?
o I have read the informed consent statement and give my consent to participate in this study.
o I do not want to participate in this study

Appendix R

Phase 3 Manipulation Check Sample Quotes

Story of the Program	Mental Effort	Study Purpose
present a logic model	This type of chart is learned from university as part of graduation and this type of chart is used while designing programs at the workplace. I am a professional social worker involved in such projects.	To test out logic models
This program may serve well to help an evaluator identify and account for people's outcomes for a short, medium, and long period according to the various sessions they maintain. The primary goal of this program is to maintain person's physical and mental well-being.	I could follow the logic pretty easily using the boxes and arrows listed and the headers in the logic model. It seemed like a logistical flow of constructs. Plus my background is in public health so the story is a familiar one.	To analyze the model and the understanding level about it.
logic model.	According to me, this logical model is quite simple. When you initially look at this it seems jumbled, but when you just look left side and follow sequence it become super easy, that's the reason I have selected this.	To understand if logic models are culturally responsive?
PERSONAL ACTIVITY LOGIC MODEL	I found this logic model to be very easy to understand. The activities are very clearly defined, and their output is also clearly defined. Few activities have one or more outputs, while some activities have just one. The benefit of these activities is segregated based on their duration. For example, a change in behavior can be measured, whether it is short-term, medium-term, or for long duration.	
health education and knowledge exercise trainer in logic model	lengthy model, The flow is hard to understand	

Appendix S

Phase 3 Variable Descriptive Statistics by Logic Model Condition and IC Orientation

DV	Condition	IC Orientation	*N*	*M*	*SD*
VE	Control	Col	27	0.29	0.61
		Ind	26	-0.16	0.66
	IL	Col	21	0.01	0.58
		Ind	24	-0.04	0.63
	CL	Col	31	-0.08	0.50
		Ind	17	-0.28	0.48
	IV	Col	21	-0.07	0.64
		Ind	31	0.03	0.55
	CV	Col	29	0.09	0.67
		Ind	22	-0.03	0.52
	ILV	Col	31	-0.03	0.62
		Ind	19	-0.13	0.79
	CLV	Col	26	0.17	0.79
		Ind	26	-0.04	0.66
	Total	Col	186	0.05	0.64
		Ind	165	-0.08	0.61
CP	Control	Col	27	4.37	0.88
		Ind	26	4.23	0.91
	IL	Col	21	4.14	0.65
		Ind	24	4.29	0.69
	CL	Col	31	4.23	0.84
		Ind	17	3.88	1.05
	IV	Col	21	4.48	0.75
		Ind	31	4.10	1.22
	CV	Col	29	4.31	0.76
		Ind	22	4.32	0.84
	ILV	Col	31	4.03	1.11
		Ind	19	3.95	0.78
	CLV	Col	26	4.54	0.76
		Ind	26	4.12	1.03
	Total	Col	186	4.29	0.85
		Ind	165	4.14	0.96

Appendix T

Phase 3 Example Descriptions of the Program Story by Condition

Condition	Descriptions of the Program Story
Control	Understanding - It is like a computer block diagram. In short, computer logical data flow diagram...By using the inputs, we have to process our life style to achieve the good healthy, wealthy and mental effort one, that is the output a person can get. Understanding - The flow chart is all about fitness program and step by step process and its results. It starts from food program with proper dieting and some methods of physical activity management, eating habits and good mental well-being. Several routes shown in the chart for a healthy physical and mental well-being.
IL	Understanding - There are seven types of algorithms in the programs in inputs, activities, outputs, short, medium ,long and impact. This program shows what is the need of main factors and individual physical mental health well being. Mental and physical health are equally important components of overall health for examples, depression increases the risk for many types of physical health problems. this programs clearly show how to handle it thought a diagram. Understanding - Objective of this program to improve individual's healthy life style to improve overall wellbeing. Which is not just physical but mental as well. This program also break down and shows different step involved for improving life style. So various activities, which person needs to follows. There are activities which leads to some outcome further these outcome classified in short, long and medium outcomes and finally what would the impact on individual.
CL	Understanding - By strengthening the social environment in communities and by helping to address employment, education, safety, physical exercise, and nutrition, many community. Understanding - It appears to be the program of perhaps a community center whose goal is to increase people's physical and mental well-being. The model outlines how the different components ("inputs") of the program work together to achieve these goals.
IV	Understanding - The program teaches and recommends a healthy lifestyle to improve the users overall health. Understand - The main objective of this project is to improve a normal person's mental and physical health.

Condition	Descriptions of the Program Story
CV	Understanding - This picturize the importance and awareness of physical and mental fitness. This is a effective miniature that anyone could easily understand and follow the protocols for a healthy life. Mental and physical health are interrelated. The steps to start up with has been clearly mentioned and the outcome is obviously rational. Appropriate measures will yield better results always. increased physical activity, of any kind, can improve depression symptoms experienced by people across the lifespan. Engaging in regular physical activity has also been shown to reduce the risk of developing depression. Following health and nutrients diet will keep you health and active . Plan , act and impact is the major components of this flowchart. Understand - When you prepare your own meals, you have more control over the ingredients. By cooking for yourself, you can ensure that you and your family eat fresh, wholesome meals. This can help you to look and feel healthier, boost your energy, stabilize your weight and mood, and improve your sleep and resilience to stress.
ILV	Understanding: It takes into account a person's lifestyle, career, exercise habits, health habits, nutrition, and helps them find a system that can help them balance their mental and physical well-being. Understand - This program shows what is the need of main factors for individual physical and mental well being. Mental and physical health are equally important components of overall health. For example, depression increases the risk for many types of physical health problems. This program clearly shows how to handle it though a diagram.
CLV	Understanding - The program tries to enhance the overall well being of communities with help from foundations and local community centres. Workshops that focus on physical activity training, nutrition, and health education are organized to improve the physical and mental health of communities. Understanding - This program shows what is the need of main factors for Household physical and mental wellbeing. Mental and physical health are equally important components of overall health. For example, depression increases the risk for many types of physical health problems. This program clearly shows how to handle it though a diagram.

Appendix U

Phase 3 Example Descriptions of Mental Effort by Condition

Condition	Description of Mental Effort
Control	"Yeah the outcome is good and easy to understand." "No I did not took any mental effort because the diagram is so understandable."
IL	Good "This is a simple chart diagram and it gives very simple explanation of big and critical plan and planning." "Initially the task seemed slightly overwhelming in that I have no formal training (or real personal experience) with either healthy eating OR exercising. Once I reviewed it for a few minutes, it eventually became more evident what kind of program it is, so I rated it a 3." "A lot of information to read"
CL	"Lengthy model. The flow is hard to understand" “All of the arrows and boxes made it complicated for me to really figure out, though eventually I think I realized it’s simpler than it appears.(explain)”
IV	“The title of each column helped to understand what all the rows beneath them are about and by going from left to right direction, it was to easy to understand the flow of the program functionality.” "It was a detailed chart, so it took some time to read it all. And not all the subtleties are immediately apparent, but it isn't highly technical. It took some effort to digest everything, but it was pretty self-explanatory once I got the jist of it."
CV	"I mean, there are like 30 different boxes with many wrds to read" “Through the diagram we clearly understand that the diagram tells us a basic human life good activities. Physical activities may help us to lead a good mental effort and cooking activities creates a patient quality among us. This diagram tells us what are thing we eat and lead a peaceful life. Through the picture we understand that the human life’s are very precious.”
ILV	“This type of chart is learned from university as part of graduation and this type of chart is used while designing programs at the workplace. I am a professional social worker involved in such projects.” "All the pictures were distracting - would have liked only a few to help me understand - not everything as a picture. Also, would want a one sentence description" "Various interconnections which appears as a web make it somewhat difficult to get the path in the first view, had to look on a few times to understand the whole picture."

Condition	Description of Mental Effort
CLV	"While I believe it was somewhat straight forward, it is still challenging because each step increases the amount of information as well as expectation of understanding how prior inputs change the results going forward. The arrows and titles help very much to have a strong reference point, but each individual step can change outcomes especially with short- and long-term goals. Ultimately I found it to be more difficult to analyze rather than easier which is why I went with this choice." "It is just based on basic understanding & common sense and finally time, that influenced the mental effort score, that is i did rate this way just based on my knowledge and understanding of the above given data." "all the pictures were distracting - would have liked only a few to help me understand - not everything as a picture. Also, would want a one sentence description" "Various interconnections which appears as a web make it somewhat difficult to get the path in the first view, had to look on a few times to understand the whole picture."

References

Alkin, M. C. (Ed.). (2012). *Evaluation roots: A wider perspective of theorists' views and influences*. Sage.

Alkin, M. C., & Vo, A. T. (2017). *Evaluation essentials: From A to Z.* Guilford Publications.

American Association of University Women AAUW (2018). *Broken ladders: barriers to women's representation in nonprofit leadership.* Retrieved August 25, 2022, from https://www.aauw.org/app/uploads/2020/03/women-in-leadership.pdf

American Evaluation Association (2018). *Guiding principles for evaluators.* Retrieved March 25, 2023, from https://www.eval.org/About/Guiding-Principles

American Evaluation Association (2018). *The 2018 AEA evaluator competencies*. Retrieved August 8, 2021, from https://www.eval.org/About/Competencies-Standards/AEA-Evaluator-Competencies

American Evaluation Association 2018 member survey report. Retrieved August 25, 2022, from https://www.eval.org/Portals/0/Docs/2018%20AEA%20Survey%20Report%20Final%20(002).pdf

American Evaluation Association. (2011). *American Evaluation Association: Statement on cultural competence in evaluation.* Retrieved January 25, 2021, from https://www.eval.org/ccstatement

American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: The official guide to APA style* (7th ed.). American Psychological Association.

Avant, C. (2020, August 18). *Language and equity in stem evaluation practice.* [Lecture notes]. American Evaluation Association STEM Topical Interest Group, Virtual.

Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. *New Directions for Evaluation, 139*, 7–32. https://doi.org/10.1002/ev.20065

Bach, B., Kerracher, N., Hall, K. W., Carpendale, S., Kennedy, J., & Henry Riche, N. (2016, May). Telling stories about dynamic networks with graph comics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3670–3682). https://doi.org/10.1145/2858036.2858387

Bamberger, M., & Mabry, L. (2019). *RealWorld evaluation: Working under budget, time, data, and political constraints*. Sage

Bamberger, M., & Mabry, L. (2020). *Realworld evaluation: Working under budget, time, data, and political constraints.* Sage.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*, 800. https://doi-org.ccl.idm.oclc.org/10.3758/s13428-011-0081-0

Berendt, E. A. & Tanita, K. (2011). The "Heart" of things: A conceptual metaphoric analysis of heart and related body parts in Thai, Japanese, and English. *Intercultural Communication Studies, 20,* 65-78.

Bessell, A. G., Deese, W. B., & Medina, A. L. (2007). Photolanguage: How a picture can inspire a thousand words. *American Journal of Evaluation, 28*(4), 558–569. https://doi.org/10.1177/1098214007306372

Bledsoe, K. (2005). Using theory-driven evaluation with underserved communities: promoting program development and program sustainability. In S. Hood, H. Frierson, and R. Hopson (Eds.). *The role of culture and cultural context in evaluation: A mandate for inclusion, the discovery of truth and understanding* (pp. 179–199). IAP.

Bledsoe, K., & Donaldson, S. (2015). Culturally responsive theory-driven evaluation. In S. Hood, R. Hopson, and H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 3–28). IAP.

Bowen, M. L., & Tillman, A. S. (2015). Developing culturally responsive surveys: Lessons in development, implementation, and analysis from Brazil's African descent communities. *American Journal of Evaluation, 36*(1), 25–41. https://doi.org/10.1177/1098214014539831

Bowman, N. R., Evergreen, S., O'Connor, J., & Parman, M. (2020). *Indigenous data visualization* [Conference session]. Annual Conference of the American Evaluation Association, Virtual.

Bowman, N. R., Francis, C. D., & Tyndall, M. (2015). Culturally responsive indigenous evaluation. In S. Hood, R. Hopson, and H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 335–360). IAP.

Brown, C. L. (2000). Sociolinguistic dynamics of gender in focus groups. *New directions for Evaluation, 86,* 55–67. https://doi.org/10.1002/ev.1172

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. https://doi.org/10.1037/e527772014-223

Burnham, M. J., Le, Y. K., & Piedmont, R. L. (2018). Who is MTurk? Personal characteristics and sample consistency of these online workers. *Mental Health, Religion & Culture*, *21*(9-10), 934-944.

Carlson, T., Moewaka Barnes, H., & McCreanor, T. (2017). Kaupapa Maori evaluation: A collaborative journey. *Evaluation Matters-He Take to Te Aronmatawai, 3,* 67–99. https://doi.org/10.18296/em.0023

Casillas, W. & Trochim, W. (2015). A systems approach to culturally responsive evaluation practice: Culturally responsive uses of the systems evaluation protocol (SEP). In S. Hood, R. Hopson, and H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 29-48). IAP.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*, 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Castillo, I. & Grim, E. (2020). *3D logic models.* [Conference session]. Annual Conference of the American Evaluation Association, Virtual.

Centers for Disease Control and Prevention (2014). *Practical strategies for culturally competent evaluation*. US Department of Health and Human Services. https://www.cdc.gov/dhdsp/docs/Cultural_Competence_Guide.pdf

Chadda, R. K., & Deb, K. S. (2013). Indian family systems, collectivistic society and psychotherapy. *Indian journal of psychiatry*, *55*(Suppl 2), S299.

Chen, H. T. (2014). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective*. Sage.

Chen, X., Gong, J., Yu, B., Li, S., Striley, C., Yang, N., & Li, F. (2015). Constructs, concept mapping, and psychometric assessment of the Concise Scale of Individualism–Collectivism. *Social Behavior and Personality: An International Journal, 43*(4), 667–683. https://doi.org/10.2224/sbp.2015.43.4.667

Chen, X., Stanton, B., Gong, J., Fang, X., & Li, X. (2009). Personal Social Capital Scale: An instrument for health and behavioral research. *Health Education Research, 24*, 306–317. http://doi.org/c9zkrm

Chin, M. C. (2005). Cartooning. In S. Mathison (Ed.), *Encyclopedia of Evaluation.* https://doi.org/10.4135/9781412950558.n65

Chouinard, J. A., & Cram, F. (2019). *Culturally responsive approaches to evaluation: Empirical implications for theory and practice* (Vol. 4). Sage. https://doi.org/10.4135/9781506368559

Clayson, Z. C., Castañeda, X., Sanchez, E., & Brindis, C. (2002). Unequal power—changing landscapes: Negotiations between evaluation stakeholders in Latino communities. *American Journal of Evaluation*, 23(1), 33-44. https://doi.org/10.1016/s1098-2140(01)00169-2

Coghlan AT, Preskill H, & Catsambas T. T. (2003). An overview of appreciative inquiry in evaluation. *New Directions for Program Evaluation,* 100, 5–22. https://doi.org/10.1002/ev.96

Cooper, E. A., & Farid, H. (2016). Does the Sun revolve around the Earth? A comparison between the general public and online survey respondents in basic scientific knowledge. *Public Understanding of Science, 25*, 146–153. https://doi.org/10.1177/0963662514554354

Copeland-Carson, J. (2005). Promoting diversity in contemporary black philanthropy: Toward a new conceptual model. *New Directions for Philanthropic Fundraising, 2005*(48), 77–87. https://doi.org/10.1002/pf.107

Cozma, I. (2011). How are individualism and collectivism measured*? Romanian Journal of Applied Psychology*, *13*(1), 11–17.

Cram, F., Pipi, K., & Paipa, K. (2018). Kaupapa Māori evaluation in Aotearoa, New Zealand. *New Directions for Evaluation*, 159, 63-77. https://doi.org/10.1002/ev.20331

Crano, W. D., Brewer, M. B. & Lac, A. (2015). Principles and Methods of Social Science Research (3rd ed). Routledge.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches*. Sage.

Darwazeh, A. N., & Branch, R. M. (2015). A revision to the revised Bloom's taxonomy. *2015 Annual Proceedings–Indianapolis*, *2*, 220-225.

Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications.* Routledge.

Donaldson, S. I. (2022). *Introduction to theory-driven program evaluation: Culturally responsive and strengths-focused applications*. Routledge.

Douville, S., Grandjean Targos, P., Jones, N, Knight, C., Azzam, T. (in press). Data visualization process and challenges: Key thoughts from data visualization experts.

Earley, C. (1993). East meets West meets Mideast: Further explorations of collectivistic and individualistic workgroups. *Academy of Management Journal, 36*, 319–348. http://dx.doi.org/10.2307/256525

Erez, M., & Earley, P. C. (1987). Comparative analysis of goal-setting strategies across cultures. *Journal of applied psychology*, *72*(4), 658.

Evergreen, S. D. (2019). *Effective data visualization: The right chart for the right data.* SAGE.

Evergreen, S., & Metzner, C. (2013). Design Principles for Data Visualization in Evaluation: Design Principles for Data Visualization. *New Directions for Evaluation*, 140, 5–20. https://doi.org/10.1002/ev.20071

Fang, W. (1985). Using photographs in an evaluation report. *Evaluation News, 6*(2), 24–27. https://doi.org/10.1177/109821408500600207

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

Few, S., & Edge, P. (2007). Data visualization: past, present, and future. *IBM Cognos Innovation Center.*

Few, S., & Edge, P. (2008). *What ordinary people need most from information visualization today. Perceptual Edge: Visual Business Intelligence Newsletter, 1–7.*

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Frazier-Anderson, P., Hood, S., & Hopson, R. K. (2012). Preliminary consideration of an African American culturally responsive evaluation system. In S. Lapan, M. Quartaroli, and F. Riemer (eds.), *Qualitative research: An introduction to methods and designs* (pp. 347–372). Jossey-Bass.

Frechtling, J. A. (2007). *Logic modeling methods in program evaluation* (Vol. 5). John Wiley & Sons.

Fuller, T., Edwards, J., Vorakithphokatorn, S., Sermsri, S. (1993). Using focus groups to adapt survey instruments to new populations. In

Furr, R. M., & Bacharach, V. (2008). Psychometrics and the importance of psychological measurement. *Psychometrics.* Sage

Fussell, S. R., Zhang, Q., & Setlock, L. D. (2009). Global culture and computer mediated communication. In *Human Computer Interaction: Concepts, Methodologies, Tools, and Applications* (pp. 1801-1816). IGI Global. DOI: 10.4018/978-1-87828-991-9.ch118

Gendron, M., Hutwelker, C., & Kisz, K. (2017). Best practices for culturally sensitive data visualizations. In S. Lapan, M. Quartaroli, and F. Riemer (Eds.), *Qualitative research: An introduction to methods and designs* (pp.1777-1798). Jossey-Bass.

Ghanbarpour, S., Noguez Mercado, A. P., & Palotai, A. (2020). A language justice framework for culturally responsive and equitable evaluation. *New Directions for Evaluation, 166,* 37-47. https://doi.org/10.1002/ev.20412

Goagoses, N., Winschiers-Theophilus, H., & Zaman, T. (2019). Community protocols for researchers: Using sketches to communicate interaction guidelines. *AI & SOCIETY,* 1-13. https://doi.org/10.1007/s00146-019-00914-x

Gouveia, V. V., & Ros, M. (2000). Hofstede and Schwartz s models for classifying individualism at the cultural level: their relation to macro-social and macro-economic variables. *Psicothema*, *12*(Su1), 25-33.

Gudykunst, W. B. (1993). Toward a theory of effective interpersonal and intergroup communication: An anxiety/uncertainty management (AUM) perspective. In R.L. Wiseman & J. Koester (Eds.), *Intercultural communication theory* (pp. 33−71). Sage.

Gudykunst, W. B., Matsumoto, Y., Ting-Toomey, S., Nishida, T., Kim, K., & Heyman, S. (1996). The influence of cultural individualism-collectivism, self construal's, and individual values on communication styles across cultures. *Human Communication Research, 22*(4), 510–543. https://doi.org/10.1111/j.1468-2958.1996.tb00377.x

Herdin, T., Faust, M., & Chen, G. M. (2020). *De-westernizing visual communication and culture: Perspectives from the global south.* Nomos Verlag. https://doi.org/10.5771/9783748906933-7

Hofstede, G. & Bond, M. (1984). Hofstede's culture dimensions. Journal of Cross-Cultural Psychology, 15, 417-433.

Hofstede, G. (1980). Culture's consequences: International differences in work-related values. Beverly Hills, CA: SAGE. Hofstede, G. (1991). *Cultures and organizations: Software of the mind*. McGraw-Hill.

Hood, S., Frierson, H., & Hopson, R. (2005). *The role of culture and cultural context in evaluation: A mandate for inclusion, the discovery of truth and understanding.* IAP.

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. In K. Newcomer, H. Hatry, and J. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 281–317). John Wiley & Sons.

Hood, S., Hopson, R., & Frierson, H. (2015). Introduction: This is where we continue to stand. In S. Hood, R. Hopson, and H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. ix-xvii). IAP.

Hopson, R. (2009). Reclaiming knowledge at the margins: Culturally responsive evaluation in the current evaluation moment. In K. Ryan and J. Cousins (Eds.), *The Sage international handbook of educational evaluation* (pp. 429–446). Sage. https://www-doi-org.ccl.idm.oclc.org/10.4135/9781452226606

Huang, W., Eades, P., & Hong, S. H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization, 8*, 139–152.

Hui, C. H. (1988). Measurement of individualism-collectivism. *Journal of Research in Personality, 22*, 17–36.

Hutchinson, K. (2017). A short primer on innovative evaluation reporting. *Kylie Hutchinson.*

Jones J.M., Leitner J.B. (2015). The Sankofa effect: Divergent effects of thinking about the past for blacks and whites. In M. Stolarski, N. Fieulaine, and W. van Beek (Eds.), *Time perspective theory; review, research and application* (pp. 197-211). Springer. https://doi.org/10.1007/978-3-319-07368-2_13

Jones, N. D., Azzam, T., Wanzer, D. L., Skousen, D., Knight, C., & Sabarre, N. (2020). Enhancing the effectiveness of logic models. *American Journal of Evaluation, 41*(3), 452-470. https://doi.org/10.1177/1098214018824417

Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning*, *28*(2), 167-172. https://doi.org/10.1016/j.evalprogplan.2004.09.002

Kashima, E. S., & Kashima, Y. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-cultural Psychology*, *29*(3), 461–486.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. A*merican Journal of Evaluation, 31*(3), 400-413. https://doi.org/10.1177/1098214010373645

Knaflic, C. (2015). *Storytelling with data: A data visualization guide for business professionals.* Wiley.

Knowlton, L. W., & Phillips, C. C. (2013). *The logic model guidebook: Better strategies for great results* (2nd ed). Sage. https://doi.org/10.4087/foundationreview-d-12-00050.1

Kray, L. J., Galinsky, A. D., & Thompson, L. (2002). Reversing the gender gap in negotiations: An exploration of stereotype regeneration. *Organizational Behavior and Human Decision Processes, 87,* 386–409. Doi:10.1006/obhd.2001.2979

Kray, L. J., Reb, J., Galinsky, A. D., & Thompson, L. (2004). Stereotype reactance at the bargaining table: The effect of stereotype activation and power on claiming and creating value. *Personality and Social Psychology Bulletin, 30,* 399–411. Doi: 10.1111/1467-9280.00272

Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology, 80,* 942–958. https://doi.org/10.1037/0022-3514.80.6.942

La Plazita Institute (n.d.). *La Plazita Institute Impact Model*. Retrieved October 1, 2020, from https://laplazitainstitute.org/

LaFrance, J. (2004). Culturally competent evaluation in Indian country. *New Directions for Evaluation, 102,* 39–50. https://doi.org/10.1002/ev.114

LaFrance, J., & Nichols, R. (2008). Reframing evaluation: Defining an Indigenous evaluation framework. *The Canadian Journal of Program Evaluation*, 23(2), 13-31.

LaFrance, J., Nichols, R., & Kirkhart, K. E. (2012). Culture writes the script: On the centrality of context in indigenous evaluation. *New Directions for Evaluation*, 135, 59–74. https://doi.org/10.1002/ev.20027

Lalwani, A. K., Shrum, L. J. & Chiu, C-Y (2009). Motivated response style: The role of cultural values, regulatory focus, and self- consciousness in socially desirable responding. *Journal of Personality and Social Psychology, 96,* 870-882. http://dx.doi.org/10.1037/a0014622

Lee, A. Y., Aaker, J. L. & Gardner, W. L. (2000). The pleasure and pains of distinct self-construals: The role of interdependence in regulatory focus. *Journal of Personality and Social Psychology, 78,* 1122–1134. http://dx.doi.org/10.1037/0022-3514.78.6.1122

Lee, E. (1997). Overview: The assessment and treatment of Asian American families. In E. Lee (Ed.), *Working with Asian Americans: A guide for clinicians* (pp. 3–36). Guilford Press. https://doi.org/10.4135/9781452232713.n5

LeFebvre, R., & Franke, V. (2013). Culture matters: Individualism vs. collectivism in conflict decision-making. *Societies, 3*(1), 128-146. https://doi.org/10.3390/soc3010128

Lemire, S., Fierro, L. A., Kinarsky, A. R., Fujita-Conrads, E., & Christie, C. A. (2018). The US federal evaluation market. *New Directions for Evaluation*, *2018*(160), 63-80.

Madison, A. M. (2000). Language in defining social problems and in evaluating social programs. In R. K. Hopson (Ed.), *New Directions for Evaluation, 86,* 17–28. https://doi.org/10.1002/ev.1169

Mason, S., & Azzam, T. (2019). In need of an attitude adjustment? The role of data visualization in attitude change and evaluation influence. *American Journal of Evaluation, 40*(2), 249-267. https://doi.org/10.1177/1098214018778808

Mayr, E., Hynek, N., Salisu, S., & Windhager, F. (2019, June). Trust in Information Visualization. In *Proceedings of the 21st Eurographics Association Conference on Visualization* (pp. 25-29).

McAlindon, K., Neal, J. W., Neal, Z. P., Mills, K. J., & Lawlor, J. (2019). The BOND framework: A practical application of visual communication design and marketing to advance evaluation reporting. *American Journal of Evaluation, 40*(2), 291–305. https://doi.org/10.1177/1098214018771219

McBride, A. (1998). Television, individualism, and social capital. *PS: Political Science & Politics*, *31*(3), 542–552.

McLaughlin, J. A., & Gretchen, B. J. (2015). Using logic models. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 62–87). Jossey-Bass.

Mertens, D. (2009). *Transformative research and evaluation.* Guilford Press.

Meyer, M. L., Louder, C. N., & Nicolas, G. (2021). Creating with, not for people: theory of change and logic models for culturally responsive community-based intervention. *American Journal of Evaluation.* https://doi.org/10.1177/10982140211016059

Morgan, D. (ed.) Successful Focus Groups: Advancing the State of the Art, pp. 89–104. Sage.

Mortenson, S. (2002). Sex, communication values, and cultural values: Individualism-collectivism as a mediator of sex differences in communication values in two cultures. *Communication Reports, 15,* 57-70. Doi:10.1080/08934210209367752

Nichols, T. R., Wisner, P. M., Cripe, G., & Gulabchand, L. (2010). Putting the kappa statistic to use. *The Quality Assurance Journal*, *13*(3-4), 57-61.

Oyserman, D. & Lee, S. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological Bulletin, 134,* 311–342. Doi: 10.1037/0033-2909.134.2.311

Oyserman, D. (1993). The lens of personhood: Viewing the self and others in a multicultural society. *Journal of Personality and Social Psychology*, *65*(5), 993–1009.

Oyserman, D. (2017). Culture three ways: Culture and subcultures within countries. *The Annual Review of Psychology, 68,* pp. 435–463. Doi: 10.1146/annurev-psych-122414-033617

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology, 84,* 429–434.

Paige, C., Peters, R., Parkhurst, M., Beck, L. L., Hui, B., May, V. T. & Tanjasiri, S. P. (2015). Enhancing community-based participatory research partnerships through appreciative

inquiry. *Progress in Community Health Partnerships: Research, Education, and Action, 9(3),* 457–463.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*, 184–188.

Patton, M. Q. (1990). Qualitative evaluation and research methods. Sage.

Patton, M. Q. (2000). Overview: Language matters. *New Directions for Evaluation, 86,* 5–16.

Price, R. M., Alkema, G. E., & Frank, J. C. (2009). California Geriatric Education Center logic model: an evaluation and communication tool. *Gerontology & geriatrics education, 30*(4), 317–331. https://doi.org/10.1080/02701960903347691

Reynolds, S., & Valentine, D. (2011). *Guide to cross-cultural communication* (2nd ed.). Prentice Hall.

Rosenstein, B. (2005). Culturally responsive evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 97-101). Sage. https://www-doi-org.ccl.idm.oclc.org/10.4135/9781412950558.n131

SenGupta, S., Hopson, R., & Thompson-Robinson, M. (2004). Cultural competence in evaluation: An overview. *New Directions for Evaluation,* 102, 5–19. https://doi.org/10.1002/ev.112

Sharpe Wessling, K., Huber, J., & Netzer, O. (2017). MTurk character misrepresentation: Assessment and solutions. *Journal of Consumer Research*, *44*(1), 211-230.

Shultz, K., & Whitney, D. (2005). Measurement theory in action: Case studies and exercises. Sage.

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and social psychology bulletin*, *20*(5), 580-591.

Singelis, T. M., Triandis, H. C., Bhawuk, D. P., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-cultural research*, *29*(3), 240-275.

Stake, R. E. (2004). *Standards-based and responsive evaluation*. SAGE. https://doi.org/10.4135/9781412985932

Stickl Haugen, J., & Chouinard, J. A. (2019). Transparent, translucent, opaque: Exploring the dimensions of power in culturally responsive evaluation contexts. *American Journal of Evaluation, 40*(3), 376-394. https://doi-org.ccl.idm.oclc.org/10.1177/1098214018796342

Stopher, P. (2012). *Collecting, managing, and assessing data using sample surveys*. Cambridge.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Pearson.

The University of Illinois Urbana-Champaign (n.d.). Center for Culturally Responsive Evaluation and Assessment. Retrieved January 20, 2020, from https://crea.education.illinois.edu/

Ting-Toomey, S., & Dorjee, T. (2019). *Communicating across cultures* (2nd ed). Guilford Press.

Ting-Toomey, S., & Oetzel, J. (2001). Managing intercultural conflict effectively. Thousand Oaks, CA: SAGE.

Torres, R. T., Preskill, H., & Piontek, M. E. (2005). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations.* Sage.

Triandis, H. (2001). Individualism and collectivism: Past, present, and future. In D. R. Matsumoto (Ed), The Handbook of Culture and Psychology (pp. 35–50). Oxford University Press.

Triandis, H. C. (1995). A theoretical framework for the study of diversity. In M. M. Chemers, S. Oskamp, & M. A. Costanzo (Eds.), *Diversity in organizations: New perspectives for a changing workplace* (pp. 11–36). Sage

Triandis, H., & Gelfand, M. (2012). A theory of individualism and collectivism. In P. A. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 2, pp. 498-520). Sage. https://www-doi-org.ccl.idm.oclc.org/10.4135/9781446249222.n51

Tufte, E. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Tufte, E. R. (2006). *Beautiful evidence* (Vol. 1). Graphics Press.

Ueda, Y., Chen, L., Kopecky, J., Cramer, E. S., Rensink, R. A., Meyer, D. E., ... & Saiki, J. (2018). Cultural differences in visual search for geometric figures. *Cognitive Science, 42*(1), 286-310. https://doi.org/10.1111/cogs.12490

University of Wisconsin Extension (n.d.). Program Development and Evaluation: Division of Extension. Retrieved January 20, 2020, from https://fyi.extension.wisc.edu/programdevelopment/logic-models/

Urban Indian Health Institute (2018, November). *Missing and Murdered Indigenous Women & Girls.* Retrieved January 20, 2020, from https://www.uihi.org/resources/missing-and-murdered-indigenous-women-girls/.

Vandello, J. A., & Cohen, D. (1999). Patterns of individualism and collectivism across the United States. *Journal of Personality and Social Psychology, 77*(2), 279-292. https://doi.org/10.1037/0022-3514.77.2.279

W.K. Kellogg Foundation (2004). W.K. Kellogg Foundation's Logic Model Development Guide. W.K. Kellogg Foundation. https://www.wkkf.org/resource-directory/resources/2004/01/logic-model-development-guide

Waapalaneexkweew (Nicole Bowman, Mohican/Lunaape), & Dodge-Francis, C. (2018). Culturally responsive Indigenous evaluation and tribal governments: Understanding the relationship. *New Directions for Evaluation, 159,* 17-31. https://doi.org/10.1002/ev.20329

Wagner III, J. A. (1995). Studies of individualism-collectivism: Effects on cooperation in groups. *Academy of Management Journal*, *38*(1), 152–173.

Wagner, J. A., & Moch, M. K. (1986). Individualism-collectivism: Concept and measure. *Group & Organization Studies*, *11*(3), 280-304.

Wang, K. T., Castro, A. J., & Cunningham, Y. L. (2014). Are perfectionism, individualism, and racial color-blindness associated with less cultural sensitivity? Exploring diversity awareness in White prospective teachers. *Journal of Diversity in Higher Education*, *7*(3), 211.

Wanzer, D. L., Azzam, T., Jones, N. D., & Skousen, D. (2021). The role of titles in enhancing data visualization. *Evaluation and Program Planning, 84,* 1-10. https://doi.org/10.1016/j.evalprogplan.2020.101896

Zumbo, B. D., & Forer, B. (2011). Testing and measurement from a multilevel view: Psychometrics and validation. In J. A. Bovaird, K. F. Geisinger, & C. W. Buckendahl (Eds.), *High stakes testing in education, science, and practice in K-12 settings* (pp. 177-190). American Psychological Association.