 DOI: 10.1111/risa.12545

Improving the Impact and Implementation of Disaster Education: Programs for Children Through Theory-Based Evaluation

Victoria A. Johnson,[1,*] Kevin R. Ronan,[2] David M. Johnston,[1] and Robin Peace[3]

A main weakness in the evaluation of disaster education programs for children is evaluators' propensity to judge program effectiveness based on changes in children's knowledge. Few studies have articulated an explicit program theory of how children's education would achieve desired outcomes and impacts related to disaster risk reduction in households and communities. This article describes the advantages of constructing program theory models for the purpose of evaluating disaster education programs for children. Following a review of some potential frameworks for program theory development, including the logic model, the program theory matrix, and the stage step model, the article provides working examples of these frameworks. The first example is the development of a program theory matrix used in an evaluation of ShakeOut, an earthquake drill practiced in two Washington State school districts. The model illustrates a theory of action; specifically, the effectiveness of school earthquake drills in preventing injuries and deaths during disasters. The second example is the development of a stage step model used for a process evaluation of *What's the Plan Stan?*, a voluntary teaching resource distributed to all New Zealand primary schools for curricular integration of disaster education. The model illustrates a theory of use; specifically, expanding the reach of disaster education for children through increased promotion of the resource. The process of developing the program theory models for the purpose of evaluation planning is discussed, as well as the advantages and shortcomings of the theory-based approaches.

KEY WORDS: Children; drills; earthquake; evaluation; tsunami

1. INTRODUCTION

In response to growing costs and consequences of disasters, and predictions of communities' increased vulnerability to hazards due to the effects of climate change and population settlement patterns, there is an increasing need for communities to prepare for and proactively mitigate disaster risks to prevent catastrophic damages, injuries, and deaths. Education continues to be a cornerstone of disaster risk reduction efforts, as many policymakers and practitioners view education as a vehicle to instigate individual and community-initiated actions that reduce their own vulnerability. In 2005, 168 Member States of the United Nations endorsed the *2005–2015 Hyogo Framework For Action* (HFA), agreeing to five priority actions to reduce disaster risks globally, including Priority for Action #3: *Use knowledge, innovation and education to build a culture of safety and resilience at all levels* (Ref. 1, p. 18). The HFA states the intended outcomes and impact of this

[1]Massey University–School of Psychology, Joint Centre for Disaster Research, Wellington, New Zealand.
[2]CQ University Australia–School of Human, Health and Social Sciences, Rockhampton, Queensland, Australia.
[3]Massey University–School of People, Environment and Planning, Wellington, New Zealand.
[*]Address correspondence to Victoria A. Johnson, Massey University–School of Psychology, Joint Centre for Disaster Research, P.O. Box 756, Wellington 6140, New Zealand; johnson.va@gmail.com.

 0272-4332/15/0100-0001$22.00/1 © 2015 Society for Risk Analysis

priority action: "Disasters can be substantially reduced if people are well informed and motivated towards a culture of disaster prevention and resilience, which in turn requires the collection, compilation and dissemination of relevant knowledge and information on hazards, vulnerabilities and capacities" (Ref. 1, p. 9). This same ethos is contained in the recently signed Sendai Framework for Disaster Risk Reduction, 2015–2030, including the idea of children and youth as "agents of change" (Ref. 2, p. 20).

The dissemination of knowledge and information that the HFA and Sendai Framework refers to includes a wide variety of educational activities internationally that are focused on teaching people about disaster risks and actions to reduce their vulnerability. These include practicing self-protective actions to prevent injuries, creating family communication plans, securing disaster insurance, stockpiling food, water, and supplies at home, and preventing hazards through efforts such as updating building codes to improve the safety of buildings. Although adults have always been a primary audience for these initiatives, the HFA stimulated more recent efforts to focus on children as an audience for disaster education. At the heart of the U.N, promotion of children's disaster education, particularly school-based education, is the idea that children represent the future generation of adults who will embody the collective values and culture of disaster prevention.[3] UNICEF and UNISDR (Ref. 4, p. 19) state that disaster education programs for children aim to "contribute to a drastic shift in mentalities and perceptions as well as behavioral change towards a more proactive preventative approach to disasters." Children are viewed as vehicles of disaster preparedness and prevention in the future as well as in the present; this optimism is reflected in many U.N. reports that discuss the active role of children in "child-centered" disaster preparedness activities and their role in influencing adults to take action.[3,5,6]

Internationally, a wide range of disaster education programs for children has been documented, including formal and informal community, school-based, and extracurricular programs, and school emergency drills.[5,6] One of the oldest forms of disaster education for children is school emergency drills for fires and sudden-onset disasters, although scholars and practitioners emphasize that even today, most school drills serve only as basic, perfunctory practice of school safety procedures.[7,8] To meet HFA goals, some countries, including France, Georgia, Russia, and Turkey, have taken steps to integrate curriculum-based disaster education into schools through efforts that include content reorganization, curriculum requirements, and large-scale teacher training schemes.[5] In other countries such as United States and New Zealand, where curriculum content choices are the province of individual school districts or schools, children receive disaster education primarily through voluntary school teaching and ad hoc activities. These include after-school programs, summer camps, and classes developed by nonformal educators, such as emergency management agencies and child advocacy organizations.[9,10]

The significant investment in disaster education for children is based on a consensus that these efforts produce some gain in individual and community resilience to disasters.[6] However, a review of evaluations of disaster education programs for children identified major gaps in the evidence base on the effectiveness of these programs.[11] Johnson and colleagues[11] concluded that most of what is known about the effectiveness of disaster education programs for children is based on the results of a body of quasi-experimental and correlational studies that primarily measured children's correct answers to knowledge-based questions about disaster risks and protective actions. Some studies went a bit further and measured the achievement of goals such as improved attitudes towards preparedness and children's household preparedness activities. Overall, most studies concluded that a program was successful because it produced positive, short-term outcomes, such as knowledge gain, which were assumed to cause other intended long-term impacts, such as disaster mitigation. Similarly, only a few studies evaluated the implementation of programs that were intended to be scaled nationally. Consequently, it is difficult to know if those program theories were correct. Therefore, our research aimed to address a gap in the current evidence base by building new theories of evaluative outcome indicators that could be used to test the underlying assumptions and theoretical constructs of programs.

Johnson and colleagues' review of evaluations concluded that little attention has been paid to the theoretical models of these education interventions and the mechanisms that facilitate changes in attitudes and behaviors.[11] The authors found that the gaps in the literature are not due to a lack of research, but a lack of conceptually framed program theories and meaningful outcome indicators that explicitly seek to validate if and how programs result in the intended outcomes and desired long-term impacts

(Ref. 11, p. 121). Donaldson (Ref. 12, p. 8) suggests that "program theory-driven evaluation science" comprises a three-step model that includes: "developing program impact theory; formulating and prioritizing evaluation questions; [and] answering evaluation questions." It is these three elements, systematically applied, that are missing from the extant explorations into causality in disaster education studies. In this article, we focus on practical ways to execute these steps.

The use of theory-based evaluation has the potential to improve the quality of evaluations of disaster education programs for children by providing a framework to help define program theories and identify and refine more meaningful outcome indicators and success criteria. If a culture of more systematic evaluation of programs is encouraged, long term, the application of theory-based evaluation tools to children's disaster education may help generate a cumulative body of knowledge that demonstrates how disaster risk reduction can be achieved through curricular integration and children's programming.[13]

The following sections explore the key challenges to evaluation of disaster education programs for children and ways theory-based evaluation could enhance evaluation practice in this particular field. After a review of some potential frameworks for program theory development, including the logic model,[14] the program theory matrix,[15] and the stage step model,[16] we provide real-world examples of these frameworks in practice. The first example is an evaluation focused on a theory of action—specifically, the effectiveness of school emergency drills in preventing injuries and deaths during disasters. The second example is an evaluation focused on a theory of use—specifically, expanding the reach of disaster education for children through the national distribution of a voluntary teaching resource, an approach used by several countries to integrate disaster education in school curricula.[5] The examples illustrate ways to construct program theories from central assumptions underlying disaster education programs for children for the purpose of evaluation planning.

2. THE ROLE OF THEORY-BASED EVALUATION

Theory-based evaluation goes under a number of different names and descriptions ranging from "theory-oriented evaluation" to "logic modeling."[12] It is most commonly referred to as program theory, theory-based, or theory-driven evaluation. A broad definition of theory-based program evaluation suggested by Fitz-Gibbon and Morris (Ref. 17, p. 177) is "one in which the selection of program features to evaluate is determined by an explicit conceptualization of the program in terms of a theory, a theory which attempts to explain how the program produces the desired effects." Rogers and colleagues define it as "an explicit theory or model of how the program causes the intended or observed outcomes and an evaluation that is at least partly guided by this model" (Ref. 18, p. 6).

Roots of program theory began as early as the 1930s in the work of Ralph Tyler (Ref. 12, p. 9), and it is associated with scholars such as Edward Suchman,[19] who articulated two distinct reasons for a program's failure: *theory failure*, when the intended outcomes and effects of a program do not occur, and *implementation failure*, when the operation of the program does not work as intended (as cited in Ref. 18, p. 6). In the 1980s, Huey-Tsyh Chen and Peter Rossi[20–22] discussed the advantages of theory-based evaluation by arguing that the explicit theorizing of a program's central cause-and-effect mechanisms provides useful guidance for an evaluation's planning, execution, and interpretation.

As theory-based program development and evaluation became more common in the 1990s, particularly in the fields of health promotion and risk prevention, Carol Weiss (Ref. 23, p. 57) further defined program theory, stating that it "refers to the mechanisms that mediate the delivery (and receipt) of the program and the emergence of the outcomes of interest." Scholars clarified that the mechanism of change, also known as change theory, is the process of change that leads to the attainment of the program's goals, which are intended to facilitate significant social impacts.[15,24–26] In plain language, theory-based evaluation is different from other models of evaluation in that it is mainly concerned with discerning the explicit theory or model of *how* the program causes the observed outcomes (Ref. 18, p. 5). Previous approaches to evaluation of disaster education programs for children did not look critically at the programs' underlying assumptions about how the programs were intended to work or be implemented. Consequently, while immediate positive effects of these programs, such as knowledge acquisition, are easy to document, these studies do not provide evidence that the programs have achieved their ultimate, critical goals, such as lifesaving and disaster mitigation.

Some evaluation scholars argue that the articulation and testing of a program's theory of action or use is not necessary for an effective evaluation, particularly if evaluation resources are in short supply and the goal is simply to judge a program's worth, merit, or significance (Ref. 27, p. 59). Even so, there is consensus that theory-based evaluation methods are beneficial for uncovering faulty assumptions about a program, improving the collaboration of program developers and external evaluators, identifying appropriate data-collection and analysis methods, and developing better quality outcome indicators to measure program impacts and processes.[16,21,26,28–30]

Theory-based evaluation was developed to address the lack of emphasis on testing a program's underlying assumptions. Such assumption testing is necessary as a first step in evaluation if program stakeholders aim to understand *how* a social intervention works or fails.[31] Often, program managers do not articulate a program theory at the outset of program development. Therefore, when theory-based evaluation is applied, central assumptions need to be unpacked retrospectively.[32]

There are a number of frameworks for developing or reconstructing the underlying theories of how a program is intended to work and what it is intended to achieve.[33] The most common framework used by many government and nongovernmental organizations is the logic model, which is a visual chart that depicts the sequential process of a program's inputs, activities, outputs, and outcomes.[14] The last step in the logic model may be program impacts, which are the longer-term outcomes expected to be achieved through the immediate and intermediate program outcomes. The simplest form of a logic model depicts a single, linear chain that illustrates a sequential series of variables from inputs to impacts. In more complex logic models, variables like program activities may be differentiated and depicted in several different boxes, or the models may depict linkages across and between variables to illustrate the ways in which variables influence each other (see models in Ref. 14). Fig. 1 depicts a basic, linear logic model of the HFA Priority for Action #3 (Ref. 1, p. 9) stated in the article's opening paragraph.

The primary criticism of logic models is that they do not illustrate causal links among components; the simplistic, linear trajectory for social change is incapable of showing where, how, and at what scale outcomes and impacts are achieved.[34] In other words, stakeholders and evaluators often assume a program output, like an educated populace (Fig. 1, line 3), directly results in an outcome, like a culture of disaster prevention and resilience (Fig. 1, line 4). Thus, the achievement of the immediate output is evaluated without a critical analysis of the validity of the assumed link between the outputs and intended outcomes. Oftentimes, the achievement of the immediate outcome results in a conclusion that the program is successful and worthwhile.

To look critically at program theories, evaluators have developed and refined more comprehensive program theory models. Two of the more common models discussed in the academic literature include the program theory matrix[15] and the stage step model.[16] These models were chosen as planning tools for two original pieces of evaluation research conducted by the authors. Later in this article, the functionality and application of these two models are described in more detail.

Theory-based evaluation and related frameworks offer a pathway toward more systematic approaches to identifying if and how disaster education programs for children are facilitating the goal of positively changing the culture of disaster preparedness and prevention. The most pressing need is a more critical examination of underlying assumptions and alternative casual explanations for program outcomes. Johnson *et al.*[11] found that one of the most common goals of disaster education programs for children was to increase children's household preparedness activities. Sixteen evaluation studies, almost half the evaluations identified for the review, measured household preparedness as a program outcome and indicator of program success (Ref. 11, p. 118). In contrast, a theory-based evaluation may seek to validate or disprove the assumption that short-term household preparedness actions stimulated by the program result in long-term preparedness for disasters.

Similarly, a theory-based evaluation approach may help evaluators identify the mechanisms of change facilitated by education programs, if they exist at all. Jacobs *et al.* (Ref. 36, p. 356) note that many, if not most, information-based public education programs change social norms of the way people *speak* of program goals, such as the need to prepare for disasters, but they do not necessarily cause changes in social behaviors, particularly when there are no social consequences for failure to act (Ref. 36, p. 362). Similarly, information-based disaster education programs for children may be ineffective in instigating behavior changes like household preparedness. Most of the evaluations reviewed by Johnson *et al.*[11]

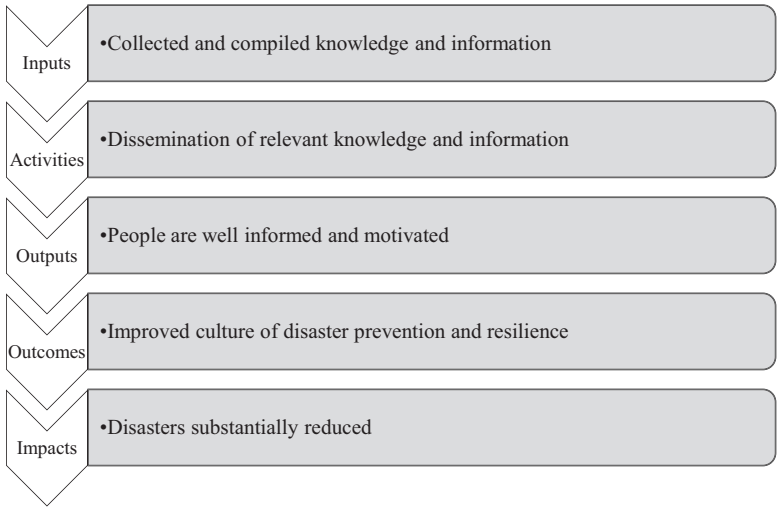

Fig. 1. Basic logic model of the HFA Priority for Action #3.
Note. Text adapted from "Hyogo Framework for Action 2005–2015: Building the Resilience of Nations and Communities to Disasters: Priority for Action #3" (Ref. 1, p. 9). Model adapted from "The Program Logic Model as an Integrative Framework for a Multimethod Evaluation" (Ref. 14, p. 120).

identified positive changes in children's knowledge and attitudes towards preparedness immediately following a program, but few examined instrumental actions that would change disaster outcomes or prevent disaster risks.

A theory-based evaluation approach also offers evaluators the opportunity to refine and iterate outcome indicators that could better validate if and how programs result in the intended outcomes and longer-term impacts. With the exception of the small number of countries executing national curriculum integration of disaster risk reduction education (e.g., Turkey and Russia), disaster education programs for children are being disseminated in an inconsistent, ad hoc manner to relatively small pockets of people.[5] The geographically inconsistent spread of disaster education programs globally reinforces a tendency toward individualistic program evaluations that do little to produce knowledge that is generalizable and meaningful for theory iteration. Because evaluation of disaster education programs has remained limited to few empirical studies, a substantial gap has emerged between program theory and program development. A concerted effort to systematically test program theory across programs may help generate more meaningful outcome indicators of program effectiveness.[13]

One way to expand the scope of evaluation research to include more critical appraisal of program theories and causal factors is to look at some worked examples of program theory construction and modeling. Astbury and Leeuw (Ref. 32, p. 365) identified three purposes for program theory modeling. First, if used as part of an initial evaluability assessment, the theoretical framing can help determine the feasibility of a study. Second, it can be used to facilitate collaborative program planning with stakeholders and, third, help clarify the design of a program. Finally, a theoretical approach can be used for evaluation planning to identify appropriate research questions, data-collection tools, and analysis techniques.

In order to explore the applicability and relevance of theory-based evaluation approaches for the purpose of evaluation planning, we investigated Suchman's[19] idea of evaluating "theory failure" and "implementation failure" using two real-life examples of disaster education programs for children. The first example is in relation to the role of school-based emergency drills in teaching children self-protective actions for disasters. In this instance, we applied a *program theory matrix* to model the generally unexamined assumptions in the drill activities that related to the drill's theory of action. The program theory matrix was used in the planning of an evaluation of ShakeOut, an earthquake and tsunami drill in two Washington State school districts.[36,37] The second example was used for planning a process evaluation of *What's the Plan, Stan?*, a national teaching resource first disseminated to New Zealand primary schools in 2006.[39] In this instance, we applied a *stage step model* to examine the factors that influence awareness, use, and nonuse of the resource that were relevant to the program's implementation theory. Both of these cases are discussed in more detail below, including the value of the theory-based approaches.

3. A PROGRAM THEORY MATRIX FOR SCHOOL EARTHQUAKE DRILLS

The program theory matrix, which originated in 1985 in the state of New South Wales, Australia,[40] is a visual representation of a hierarchy of intended program outcomes, each of which includes a series of questions and answers embodied in a complementary matrix that help identify potential data sources, evaluative criteria, and external factors that may influence the outcomes (Ref. 15, pp. 92–93). Funnell argues that the program theory matrix helps illustrate that the immediate and intermediate outcomes do not always explicitly link with the desired long-term impacts. By answering questions in the matrix such as *What would success look like?* and *What are the factors that influence the achievement of each outcome?*, evaluators can produce better-quality measurements of success at all levels of the hierarchy of outcomes (Ref. 15, p. 92).

To illustrate an example of program theory matrix, we used one of the most common forms of disaster education for children, the school emergency drill. School emergency drills normally entail students and staff practicing the school's emergency response procedures such as evacuation for fires, "drop, cover, and hold" for earthquakes, "shelter-in-place" drills for tornados, and high ground evacuation for tsunamis, among other scenarios.[7] A central assumption of emergency drills is that children can be effectively taught safety procedures and self-protective actions for disasters, and when children practice these procedures often, disaster-related injuries and deaths will be prevented.[7,41,42] Over the past 50 years there has been a dramatic decrease in fire-related injuries and deaths in school buildings in the United States, which has been primarily credited to schools' execution of routine fire drills.[43] However, school drills tend to be brief and perfunctory in nature, and some scholars and practitioners argue that this may inhibit learning outcomes.[7,8] To minimize time and disruption to school routines, schools often conduct the same drill at expected times and locations, typically during class when students are at their desks.[7,44–46] Often, drills do not incorporate lessons on the reasons why the procedures are practiced and how they protect against injury.[7,8] There has been very little study of school drills beyond visual observations. Therefore, it is an assumption that rote practice of protective actions provides children with the knowledge and skills needed to successfully protect themselves in an emergency that happens whether they are inside or outside the classroom.

In 2012, we undertook an evaluation of ShakeOut, an earthquake drill that took place in two Washington State school districts.[37,38] Working with local stakeholders, we established that the goal of the evaluation was to identify children's learning outcomes from the drill, particularly learning outcomes that help children successfully protect themselves from preventable injuries in an earthquake. To begin, we created a basic, linear logic model that illustrates the central underlying theory of the ShakeOut drill (Fig. 2). The figure illustrates the difference between the input (information) and the activity, namely, when children are prompted in the classroom, they practice "drop, cover, and hold" under a desk. "Drop, cover, and hold" is a protective action that is used to protect against injuries from falling and flying objects during ground shaking. As mentioned previously, basic logic models typically do not expand on the presumed links between activities, outputs, outcomes, and impacts that describe how the output (here, the correct demonstration of "drop, cover, and hold") enacts the impact (fewer earthquake-related injuries and deaths). It seems logical that children's knowledge of "drop, cover, and hold" would achieve the goal of children successfully protecting themselves from preventable injuries in an earthquake. The challenge with evaluating the achievement of this goal is the fact that real earthquakes are extremely rare. Therefore, in drill evaluations, an intermediate outcome, such as a visual observation of children correctly demonstrating the protective actions, typically serves as the evidence that injuries and deaths will be prevented during a real emergency.[43–46]

Unlike the basic logic model, a program theory matrix provides a framework for testing an underlying program assumption through the introduction of questions that probe a more detailed hierarchy of intended outcomes. Funnel (Ref. 15, pp. 92–95) describes the components of a program theory matrix, which includes a sequenced hierarchy of intended outcomes (immediate, intermediate, and ultimate), a series of questions for each outcome, and a table where variables are listed for each outcome. The matrix includes variable categories such as success criteria, program factors affecting success, nonprogram factors affecting success, activities and resources of the program, performance information, and sources of data. The variables in the matrix help evaluators

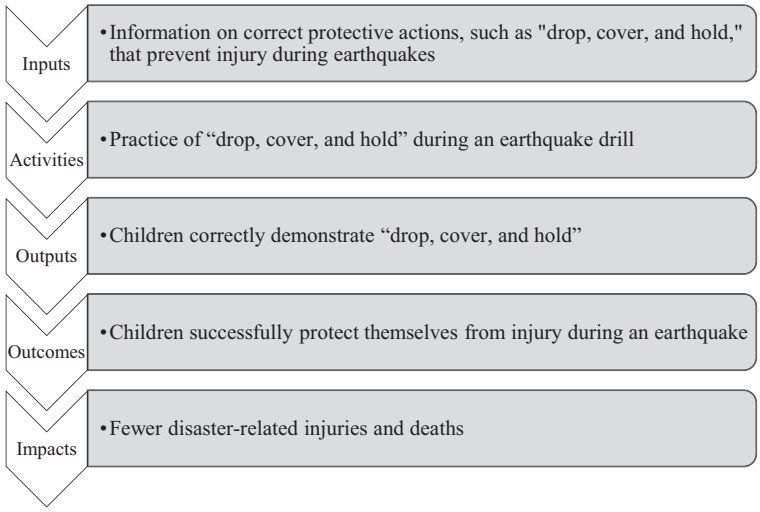

Fig. 2. Basic logic model of Shakeout, a school-based earthquake drill.
Note. Model adapted from "The Program Logic Model as an Integrative Framework for a Multimethod Evaluation" (Ref. 14, p. 120).

consider appropriate outcome indicators for evaluations assessing program impacts, processes, or both.

The biggest missing piece in the basic logic model in Fig. 2 is the expression of the casual link between the output (children correctly demonstrate "drop, cover, and hold") and the outcome (children successfully protect themselves from injury during an earthquake). If children's demonstration of "drop, cover, and hold" is the success criteria of a drill, one must consider: Does correct demonstration of "drop, cover, and hold" mean that children understand what this protective action is for? Do children understand what types of injuries "drop, cover, and hold" protects against? Will children be able to perform an action they have practiced in a drill during a real earthquake? Will children perform "drop, cover, and hold" in settings where it has not been practiced, such as outside? Can children apply their knowledge of earthquake risks and protective actions to protect themselves in unfamiliar settings? Do they have adequate knowledge to make good response decisions?

With these questions in mind, we proposed a hierarchy of intended outcomes for the ShakeOut drill displayed in Fig. 3. If the ultimate outcome is fewer injuries and deaths during earthquakes, even for the specific circumstance of earthquakes that occur during school hours, the intermediate outcomes must be the criteria for children's ability to successfully protect themselves from preventable injuries in a variety of scenarios, including indoors when they are not near a desk or other cover, and outdoors. For our evaluation framework, we proposed intermediate outcomes in Fig. 3, which are based on principles of learning theory described in Vosniadou's *How Children Learn,*[47] namely: active involvement, social participation, meaningful activities, engaging in self-regulation and self-reflection, and knowledge transfer to real-life situations. A strong theme across all the principles of learning theory described by Vosniadou is the need for children to participate in active problem solving: "People learn by employing effective and flexible strategies that help them to understand, reason, memorize, and solve problems" (Ref. 47, p. 14).

While children's participation in "drop, cover, and hold" with their peers and teachers provides elements of active involvement and social participation, it is questionable whether drills are an effective learning technique because they usually lack opportunities for problem solving, self-reflection, and knowledge transfer. Based on the hierarchy of intended outcomes displayed in Fig. 3, we theorized other intermediate outcomes of the evaluation could include: (1) children's comprehension of the most common causes of injury during earthquakes, a prerequisite for children's ability to strategically choose an appropriate protective action in an unfamiliar scenario; (2) children's comprehension of the purpose of practicing "drop, cover, and hold"; and (3) children's ability to identify correct and incorrect protective actions in different earthquake scenarios, an indicator of their comprehension of how and why protective actions are used. It was also theorized that children's levels of anxiety when thinking or talking about earthquakes and tsunamis

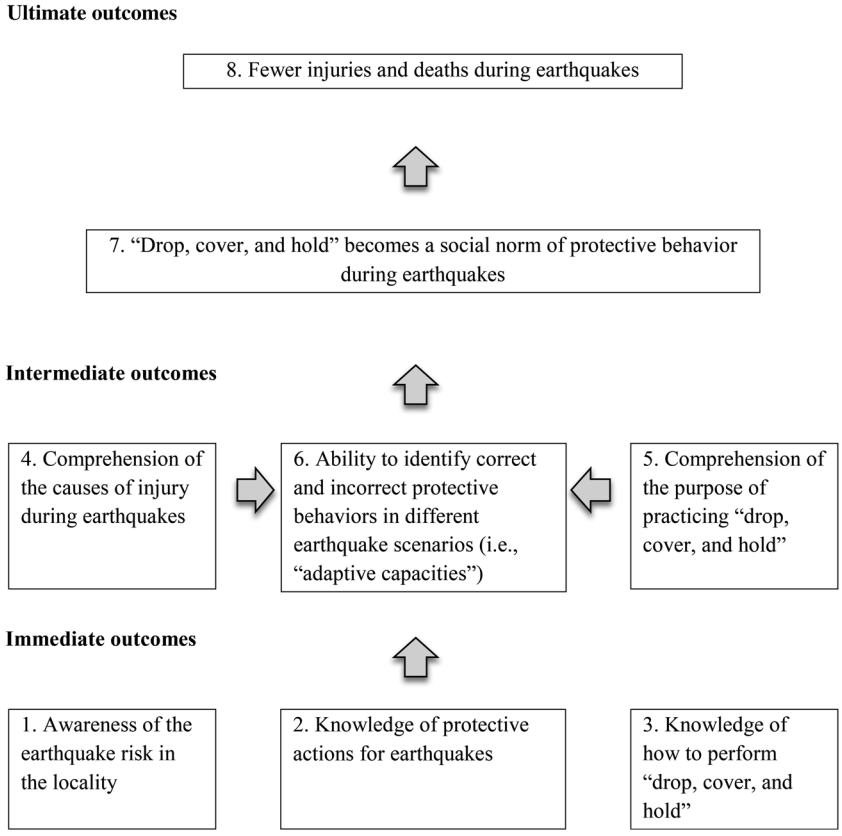

Fig. 3. Hierarchy of intended outcomes for ShakeOut, a school-based earthquake drill.
Note. Adapted from "Developing and Using a Program Theory Matrix for Program Evaluation and Performance Monitoring" (Ref. 15, p. 93).

could indicate whether the drills impacted children's self-confidence in their ability to self-protect during an earthquake.

In the top row of program theory matrix provided in Table I are the questions proposed by Funnell (Ref. 15, p. 92) to develop the variables that fill the matrix. The central question for developing the outcome indicators is: *What would success look like?* (Table I, column 2). Our theories, stated previously, are listed here. Column 3 includes the program factors affecting success, including teacher leadership and comprehension of the response actions, peer and teacher participation in the drill, and annual repetition. Some nonprogram factors that may affect the learning outcomes of the drills (column 4) include anxiety produced by the topic or activities, children's self-confidence in their ability to self-protect, past experiences with earthquakes, and a lack of earthquake conditions that cannot be simulated (e.g., ground shaking, anxiety, dangers), among other factors. Column 5 lists the activities involved in the drill, which typically include a predrill notice to teachers and children, the classroom-based drill (led by teachers), and repetition of the drill annually. Drill activities that are not always included but are often recommended include teacher explanations of "drop, cover, and hold," explanation of the causes of injuries during earthquakes, and review of alternative response actions for different scenarios, among others. Column 6 includes the proposed pieces of performance information, or outcome indicators. Typically, classroom-based drills are evaluated using a visual observation of children demonstrating "drop, cover, and hold." For the ShakeOut evaluation, we proposed measuring before and after the drill: (1) the percentage of children who know the causes of earthquake injuries; (2) the percentage of children who recognize correct and incorrect protective actions in different scenarios, including indoors with a desk, indoors without a desk, and outdoors; and (3) children's levels of disaster-related anxiety before and after the drill.[37] Other potential outcome indicators that were outside the scope of the ShakeOut evaluation included the percentage of children with high self-confidence in their ability to protect themselves during an earthquake, and the percentage of children who have high trust in authorities. Because a visual observation would not be able to capture children's

Table I. Application of a Program Theory Matrix to School Earthquake Drills

Questions:	What would success look like?	What are the factors that influence the achievement of each outcome?	Which factors are outside the direct influence of the program?	How does the program address these factors in order to bring about the outcome?	What performance information should we collect (quantitative and qualitative indicators and comparisons?)	How can we gather this information?
1. Intended Outcome	2. Success Criteria	3. Program Factors Affecting Success	4. Nonprogram Factors Affecting Success	5. Activities and Resources of the Program	6. Performance Information Examples for Columns 2 to 5	7. Sources of Data
Comprehension of correct and incorrect protective actions in different earthquake scenarios	In a disaster, no preventable injuries among children In absence of a disaster, children can identify how and when they should respond to prevent injury in different scenarios In absence of a disaster, children can identify how	Teacher leadership during the drill Teacher comprehension of the rationale for school safety procedures and protective actions Participation of peers and teachers Annual repetition of the drill	Anxiety produced by the topic or activities Self-confidence in ability to protect oneself Trust in authorities Beliefs and past experiences with earthquakes Lack of earthquake conditions that cannot be simulated: e.g., ground shaking, anxiety, dangers Parent support for drills	Predrill notice to teachers and children Drill Repeat drill annually *Not always included in drills:* "Drop, cover, and hold" clearly explained by teachers in age-appropriate manner to children Risks and causes of injuries clearly explained	*Typically:* All children can demonstrate "drop, cover, and hold" during a classroom-based simulation *Proposed for the evaluation:* Percentage of children who know the causes of earthquake injuries Percentage of children who recognize correct and incorrect actions in different	*Typically:* Observation of children's individual and group behaviors during drills *Other potential methods:* Surveys with children Interviews with children Focus groups with children Surveys of children's parents protective

Table I. (Continued)

Questions:	What would success look like?	What are the factors that influence the achievement of each outcome?	Which factors are outside the direct influence of the program?	How does the program address these factors in order to bring about the outcome?	What performance information should we collect (quantitative and qualitative indicators and comparisons?)	How can we gather this information?
1. Intended Outcome	2. Success Criteria	3. Program Factors Affecting Success	4. Nonprogram Factors Affecting Success	5. Activities and Resources of the Program	6. Performance Information Examples for Columns 2 to 5	7. Sources of Data
	not to respond in different scenarios			Simulation of different scenarios: inside with cover, inside with no cover, outside, etc. Review of alternative response actions—protective and not protective	scenarios, including indoors with a desk, indoors without a desk, and outdoors Levels of disaster-related anxiety *Other possible indicators:* Percentage of children with high self-confidence in ability to protect themselves during an earthquake Percentage of children who indicate high trust in authorities	

knowledge of appropriate response actions in different scenarios, we chose to conduct the ShakeOut evaluation using a pretest-posttest questionnaire, specially designed for children age 10 and older.[35]

Ultimately, the evaluation identified some outcomes that are indicative of children's adaptive capacities for protecting themselves in a real earthquake. Most of the children who participated in the study identified correct protective actions for both familiar and less familiar earthquake scenarios (Ref. 37, p. 21). However, approximately a third of children chose an incorrect action or indicated uncertainty in scenarios not practiced by the participating schools during ShakeOut or in previous earthquake drills, namely, scenarios where the children would be outside, or inside but not near a table or desk (Ref. 37, p. 21). These findings and others prompted us to conclude that it is problematic to assume that ShakeOut and similar school drills result in learning outcomes that will effectively mitigate injuries and deaths among children during an earthquake (Ref. 37, p. 21). Children may be able to successfully repeat actions they are directed to do during a drill, but we cannot assume they understand how to effectively apply what they have learned in other scenarios.

The main challenge that the program theory matrix helps to overcome is the tendency to overlook the measurement of inputs, processes, and outputs needed to achieve the ultimate outcomes (Ref. 15, p. 96). The value of the program theory matrix in this example was the provision of a practical framework in which to deeply examine and articulate the program theory of how the Shakeout drill was expected to achieve reduced injuries. Through this process, we discerned the need to test children's ability to apply knowledge in different scenarios. It became clear that a visual observation of the ShakeOut drill would not be adequate for collecting data on this outcome, and that another data-collection method, such as a pretest-posttest questionnaire, would be needed.

The theory-based approach was also useful in identifying ways to answer key evaluation questions about the effectiveness of drills, and in producing preliminary data for more in-depth examinations of learning outcomes in future evaluations. For example, we theorized if evaluation of a school earthquake drill incorporated a measurement of children's ability to choose or demonstrate both correct and incorrect actions in different earthquake scenarios, evaluators would have stronger evidence that children are successfully learning and applying knowledge that can prevent injuries, as opposed to rehearsing a memorized action when prompted. With this information, evaluators could determine whether drills are effective in enhancing children's ability to protect themselves during earthquakes, and delve deeper into questions of why or why not. The mechanism of learning (or lack of learning) could then be theorized and tested. For example, if evaluators find that drills are effective in enhancing children self-protective skills, the potential mechanisms of the learning that could be investigated include the active experience of practicing "drop, cover, and hold," the promotion of "drop, cover, and hold" by children's trusted authorities, the repetition of the drills that provides for ongoing reflection and practice, or other factors. As Astbury and Leeuw argue (Ref. 32, p. 375), the mechanisms may not be discerned or easily measured, and they are sensitive to variations in context; however, "theorizing with mechanisms strengthens our understanding of how and why programs work, with whom and under what circumstances," which allows opportunities to develop universal knowledge about social programs.

4. A STAGE STEP MODEL OF THE IMPLEMENTATION OF A VOLUNTARY DISASTER TEACHING RESOURCE

Our second exploration of a theory-based evaluation approach aimed to test theories of use and implementation; specifically, theories that explain how a program is intended to occur in terms of process components such as uptake of the teaching resources, adherence to the program, barriers to use, and participants' experience with the program. Although there is a tendency among program evaluators to focus predominately on measuring program outcomes, it is also critical to assess the validity of a program's implementation theory, the success or failure of which will have a direct impact on a program's reach and long-term impact.[16] Lipsey and Pollard (Ref. 16, p. 321) describe a stage step model of program theory introduced by Runyan,[48] which is a visual depiction of the major stages and statuses through which people progress in the context of interest, such as participation in a program.

We used a stage step model in a process evaluation of *What's the Plan, Stan?*, a national, voluntary disaster teaching resource developed by New Zealand's Ministry of Civil Defence and Emergency Management (MCDEM).[39] *What's the Plan, Stan?* is a multimodal resource for teaching disaster science

and preparedness to students in years 1 through 8 in New Zealand primary schools.[49] In 2006 and 2009, respectively, MCDEM distributed one hard copy of the original and updated version of the teaching package, which includes unit plans, fact sheets, and classroom activities, to all primary schools in New Zealand. The purpose of the resource is to help primary school teachers voluntarily incorporate disaster-related topics into the English, social studies, science, and health and physical Education curricula. The development and distribution of the resource is a key component of New Zealand's goal to integrate disaster risk reduction in school curricula, a core indicator of achievement for the *Hyogo Framework for Action* Priority for Action #3.[50] Use of the resource remains voluntary since disaster risk reduction education is not a required school subject in New Zealand.

The basic logic model depicting the program's theory of use of *What's the Plan, Stan?* (Fig. 4) was reconstructed from MCDEM's *What's the Plan, Stan? Communications Strategy for 2009 Launch.*[51] When the *Communications Strategy* was published, uptake and use of the resource had remained low since the resource was first released in 2006. The policy document focused heavily on the notion that promotion of the resource through advertising would maximize awareness of the resource, which would in turn increase uptake and use of the resource. Therefore, the program's theory of use was the assumption that the awareness of the resource would lead to its use. Long term, the intended outcome of the program's implementation was the integration of disaster education in the New Zealand curriculum and the intended impact was all New Zealand school children would learn about disaster preparedness.

In 2012, six years after the initial distribution of *What's the Plan, Stan?* to New Zealand schools, the Ministry of Education conducted a national survey of primary schools that included some questions on teachers' awareness and use of the resource.[52] This was a conventional approach to evaluating the program's implementation. The survey found that approximately 31% of schools used the resource at some time since 2006. It also found that another third of survey respondents were aware of the resource but had not used it. The remaining third were not aware of the resource. These results were limited by the fact that a single school administrator responded on behalf of each school and, therefore, the survey did not gather detail about the percentage of individual teachers engaged with the resource or the frequency of their engagement. In any case, the survey revealed that there were challenges to awareness and use of the resource in New Zealand schools, but this survey did not provide enough information to identify and address those challenges.

Our process evaluation of *What's the Plan, Stan?* aimed to understand how to increase curriculum integration of disaster education in New Zealand through the distribution of a voluntary teaching resource. At the outset of our evaluation, we questioned the assumption that primary school teachers who were aware of the resource would use the resource, particularly since use of the resource is voluntary. There are other unknown intervening factors within this implementation theory. These unknown factors exist between resource promotion and teachers who continue to lack awareness of the resource. There are also unknown factors that both facilitate and deter teachers' uptake. The stage step model presented in Fig. 5 was used as a starting point for planning the evaluation,

The stage step model highlights the unknown intervening factors, which may not be clear when starting from the assumption that an increase in resource promotion would increase resource use. From an evaluation planning standpoint, the visualization prompts the questions: *Why are teachers not aware of the resource when it is being promoted?*, *Why do teachers who are aware of the resource use it?*, and *Why do teachers who are aware of the resource not use it?* The resulting evaluation of *What's the Plan, Stan?* aimed to determine the key intervening factors and their relative strength, in order to postulate what facilitating factors could be influenced and what deterring factors could be removed, if possible, to increase the use of the resource.[39]

The stage step model helps guide the choice of the evaluator's research methods, data-collection tools, and analysis techniques. For example, because the influencing factors at different stages of the implementation process are unknown, an evaluator could develop theories of what those factors may be and test the existence and strength of those theorized factors through research designs such as surveys or interviews with teachers and other stakeholders. A major disadvantage to this approach is evaluators may not anticipate key factors that should be tested. The intervening factors may be unknown even to teachers, the key informants. Therefore, we felt a more promising approach to evaluating the implementation of *What's the Plan, Stan?* would be to interpret factors through a thematic analysis of qualitative and quantitative data available from other studies. For example, one source of data for

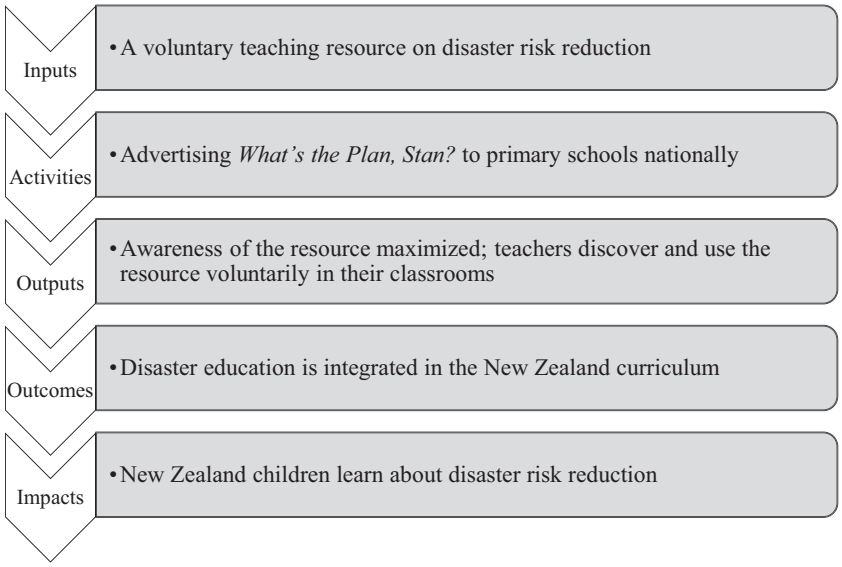

Fig. 4. Basic logic model of *What's the Plan, Stan?* theory of use.
Note. Model adapted from "The Program Logic Model as an Integrative Framework for a Multimethod Evaluation" (Ref. 14, p. 120).

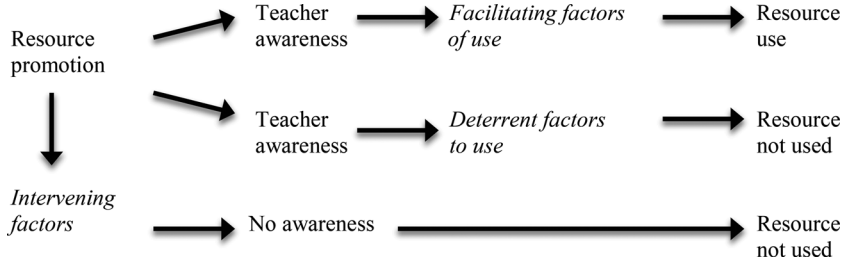

Fig. 5. A stage step model of the implementation theory of *What's the Plan, Stan?*, a voluntary disaster teaching resource distributed to NZ primary schools
Note. Adapted from "Implementing Disaster Preparedness Education in New Zealand Primary Schools."[39] Model originally adapted from "Driving Toward Theory in Program Evaluation: More Models to Choose From" (Ref. 16, p. 322).

the *What's the Plan, Stan?* evaluation was focus groups with teachers who discussed their use or nonuse of the resource,[10] and another source was the quantitative data from a national survey of schools mentioned previously.[52] A qualitative, thematic analysis of the two sets of available data uncovered some of the more common factors and the strengths of those factors.[39] Some of the facilitating factors included school-wide use of the resource, promotion by other teachers, a personal interest in the subject matter, and direct engagement with civil defense staff. Some of the deterrent factors included the voluntary nature of the resource and competing priorities, lack of awareness of the resource, and the perception that training was needed to use the resource, which stemmed from teachers' discomfort with the subject matter.

The value of the stage step model in this example was the framework it provided for organizing a theory-driven qualitative analysis of the data to answer key evaluation questions. With this framework, the analysis identified factors beyond the promotion of the resource that affected its use.[41] Based on these findings, we exposed a potentially faulty assumption about the program's theory of use and concluded that increased promotion of the resource would not be adequate enough to increase its uptake. The framework also allowed us to identify some of the mechanisms of implementation; specifically, the factors that facilitated teachers' uptake beyond their awareness of the resource, such as school-wide use of the resource, student interest in the topic, and recent disasters. Ultimately, we used the total findings to develop recommendations on how to improve the implementation of disaster education in New Zealand schools. In sum, the stage step model served as a practical tool for both evaluation planning and data analysis.

5. DISCUSSION

In these two cases, we found that the main advantage to using visual theory-based evaluation models was the framework they provided to critically analyze underlying theories and assumptions about how the programs work. Our approaches uncovered what Suchman would describe as a potential "theory failure" in the case of ShakeOut, and a potential "implementation failure" in the case of *What's the Plan, Stan?* In the evaluation of ShakeOut, the hierarchy of intended outcomes helped to identify what the school drill would actually need to accomplish to achieve the goal of preventing injuries and deaths during an earthquake or tsunami—namely, children's capacity to identify correct and incorrect protective behaviors in different earthquake scenarios, among other immediate and intermediate outcomes. Similarly, we used a program theory matrix to determine how our evaluation could retrieve that information. The matrix helped identify questions and the best data-collection tool—an age-appropriate pretest–posttest questionnaire. Consequently, the resulting evaluation uncovered that a significant portion of children did not know how to protect themselves in scenarios they had not practiced, indicating drills may not adequately achieve the goal of reducing some children's vulnerability in an earthquake. A conventional approach to evaluating the drill, such as a drill observation, might have led to the conclusion that the drill was a success and not in need of improvement. In contrast, our approach to the evaluation provided more nuanced information about children's learning outcomes that point to ways in which the drill can be improved.

The potentially faulty assumption we uncovered in our evaluation of *What's the Plan, Stan?* was that greater awareness of the resource would lead to greater use of the resource, an assumption that supported the strategy to increase advertising. Although we lack data on exactly how many teachers were exposed to advertising of *What's the Plan, Stan?,* our evaluation identified several deterrent factors to teacher use of the resource that would likely remain unaffected by an increase in advertising, such as teachers' discomfort with the subject matter. Similarly, we identified a number of facilitating factors that could increase teachers' use of the resource, such as engagement with civil defense staff, which may require different resources than advertising. In this case, the stage step model was most useful in helping to identify evaluation questions that would produce valuable information that the program administrators could act upon.

There are shortcomings to the theory-based approaches we described in these two examples. These approaches mainly supported the activity of critically analyzing the programs' underlying assumptions. In our two respective evaluations we did not achieve an understanding of what mechanisms in these programs cause positive long-term outcomes of interest; instead, our evaluations focused on assessing the validity of the programs' theories and how flawed assumptions may impact the programs' presumed value. Also, while we were able to identify potentially faulty assumptions about the respective programs' theories of action and use, we did not collect data that proved with absolute certainty that those assumptions are wrong. Yet, evaluation research is typically highly interpretative. Evaluation research includes both an empirical aspect of inquiry and a normative aspect of judging the value of something, and the value feature is what distinguishes evaluation from other forms of research.[53] Thus, it remains the responsibility of evaluators to be transparent to stakeholders about their approach to data collection and analysis, particularly when this process is not fully exposed in the resulting literature. On this note, we suggest one additional advantage to using visual theory-based evaluation models is that they provide useful information about evaluators' methods, choices, and biases, which help keep evaluators accountable to their stakeholders and can be used to critique evaluation in practice.

6. CONCLUSION

The *2005–2015 Hyogo Framework for Action* (HFA) urged member countries of the United Nations to use disaster education, including disaster education programs for children, to build a culture of safety and resilience at all levels.[1] This same ethos is contained in the recently signed Sendai Framework for Disaster Risk Reduction, 2015–2030, including the idea of children and youth as "agents of change" (Ref. 2, p. 20). Currently, there is little empirical research to inform how children's education improves individual and community resilience to disasters, if at all. This article argues for the application of theory-based evaluation approaches to test underlying assumptions of educational initiatives and improve the theoretical and conceptual constructs of disaster education programs for children. The use of visual program theory models can help identify, test,

and refine more meaningful outcome indicators during program and evaluation planning and iteration. Two examples of program theory models using existing programs were provided here to illustrate the practical application of theory-based evaluation and its benefits for executing program evaluations of disaster education programs for children. Long term, a more comprehensive effort to test program theories could help generate a cumulative body of knowledge that demonstrates how disaster risk reduction can be achieved through children's education.

ACKNOWLEDGMENTS

This research was supported with funding from GNS Science (New Zealand) and the New Zealand Earthquake Commission. The opinions expressed in this article are the personal views of the authors and do not represent in whole or part the opinions of GNS Science or any New Zealand government agency.

REFERENCES

1. United Nations International Strategy for Disaster Reduction [UNISDR]. Hyogo Framework for Action 2005–2015: Building the Resilience of Nations and Communities to Disasters. Geneva: United Nations, 2005.
2. United Nations International Strategy for Disaster Reduction [UNISDR]. Sendai Framework for Disaster Risk Reduction 2015–2030. Geneva: United Nations, 2015.
3. United Nations International Strategy for Disaster Reduction [UNISDR]. Towards a Culture of Prevention: Disaster Risk Reduction Begins at School. Geneva: United Nations, 2007.
4. United Nations Children's Fund and United Nations International Strategy for Disaster Reduction. Children and Disasters: Building Resilience Through Education. Geneva: United Nations, 2011.
5. Selby D, Kagawa F. Disaster Risk Reduction in School Curricula: Case Studies from Thirty Countries. Geneva: United Nations Children's Fund, 2012.
6. Wisner B. Let Our Children Teach Us! A Review of the Role of Education and Knowledge in Disaster Risk Reduction. Bangalore, India: Books for Change, 2006.
7. Ramirez M, Kubicek K, Peek-Asa C, Wong M. Accountability and assessment of emergency drill performance at schools. Family & Community Health, 2009; 32(2):105–114.
8. Lund JK. School Fire Drills: Incorporating Fire Prevention Education to Ensure They Are a Learning Experience. Emmitsburg, MD: National Fire Academy, 2013.
9. Federal Emergency Management Agency. Catalogue of Youth Disaster Preparedness Education Resources (updated). Washington, DC: FEMA, January 25, 2013.
10. Johnson VA. Disaster Preparedness Education in Schools: Recommendations for New Zealand and the United States. Wellington, New Zealand: Fulbright New Zealand, 2011.
11. Johnson VA, Ronan KR, Johnston DM, Peace R. Evaluations of disaster education programs for children: A methodological review. International Journal of Disaster Risk Reduction, September 2014; 9:107–123.
12. Donaldson SI. Program Theory-Driven Evaluation Science: Strategies and Applications. New York: Routledge, 2012.
13. Turnbull B. Program theory building: A strategy for deriving cumulative evaluation knowledge. American Journal of Evaluation, 2002; 23(3):275–290.
14. Cooksy LJ, Gill P, Kelly PA. The program logic model as an integrative framework for a multimethod evaluation. Evaluation and Program Planning, 2001; 24(2):119–128.
15. Funnell SC. Developing and using a program theory matrix for program evaluation and performance monitoring. New Directions for Evaluation, 2000; Autumn 2000(87):91–101.
16. Lipsey MW, Pollard JA. Driving toward theory in program evaluation: More models to choose from. Evaluation and Program Planning, 1989; 12(4):317–328.
17. Fitz-Gibbon CT, Morris LL. Theory-based evaluation. Evaluation Practice, 1996; 17(2):177–184.
18. Rogers PJ, Petrosino A, Huebner TA, Hacsi TA. Program theory evaluation: Practice, promise, and problems. New Directions for Evaluation, 2000; Autumn 2000 (87):5–13.
19. Suchman EA. Evaluative Research: Principles and Practice in Public Service & Social Action Programs. New York: Russell Sage Foundation, 1967.
20. Chen HT, Rossi PH. The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. Social Forces, 1980; 59(1):106–122.
21. Chen HT, Rossi PH. Evaluating with sense: The theory-driven approach. Evaluation Review, 1983; 7(3):283–302.
22. Chen HT, Rossi PH. The theory-driven approach to validity. Evaluation and Program Planning, 1987; 10(1):95–103.
23. Weiss CH. Evaluation. Upper Saddle River, NJ: Prentice Hall, 1998.
24. Donaldson SI, Christie CA, Mark MM (eds). What Counts as Credible Evidence in Applied Research and Evaluation Practice? Thousand Oaks, CA: Sage, 2009.
25. Chen HT (ed). Practical Program Evaluation: Assessing and Improving Planning, Implementation, and Effectiveness. Thousand Oaks, CA: Sage, 2005.
26. Weiss CH. Theory-based evaluation: Past, present, and future. New Directions for Evaluation, 1997; Winter 1997 (76):41–55.
27. Scriven M. Minimalist theory: The least theory that practice requires. American Journal of Evaluation, 1998; 19(1):57–70.
28. Riemer M, Bickman L. Using program theory to link social psychology and program evaluation. Pp. 104–138 in Mark MM, Donaldson SI, Campbell B (eds). Social Psychology and Evaluation. New York: Guildford Press, 2011.
29. Birckmayer JD, Weiss CH. Theory-based evaluation practice: What do we learn? Evaluation Review, 2000; 24(4): 407–431.
30. Bickman L. The functions of program theory. New Directions for Program Evaluation, 1987; Spring 1987 (33):5–18.
31. Lipsey MW. Theory as method: Small theories of treatments. New Directions for Program Evaluation, 1993; Spring 1993 (57):5–38.
32. Astbury B, Leeuw FL. Unpacking black boxes: Mechanisms and theory building in evaluation. American Journal of Evaluation, 2010; 31(3):363–381.
33. Leeuw FL. Reconstructing program theories: Methods available and problems to be solved. American Journal of Evaluation, 2003; 24(1):5–20.
34. Rogers PJ, Weiss CH. Theory-based evaluation: Reflections ten years on. New Directions for Evaluation, 2007; Summer 2007 (114):63–81.
35. Bell A. Designing and testing questionnaires for children. Journal of Research in Nursing, 2007; 12(5):461–469.
36. Jacobs WJ, Sisco M, Hill D, Malter F, Figueredo AJ. Evaluating theory-based evaluation: Information, norms, and adherence. Evaluation and Program Planning, 2012; 35(3):354–369.
37. Johnson VA, Johnston DM, Ronan KR, Peace, R. Evaluating children's learning of adaptive response capaci-

ties from ShakeOut, an earthquake and tsunami drill in two Washington State school districts. Journal of Homeland Security and Emergency Management, 2014; 11(3):347–373.
38. Johnson VA. An Impact Evaluation of ShakeOut, an Earthquake and Tsunami Drill in Two Coastal Washington State School Districts. GNS Science Report, 2013.
39. Johnson VA, Ronan KR, Johnston DM, Peace R. Implementing disaster preparedness education in New Zealand primary schools. Disaster Prevention and Management, 2014; 23(4):370–380.
40. Lenne B, Cleland H. Describing program logic. Program Evaluation Bulletin, 1987; 2:87.
41. Ronan KR, Johnston DM. Promoting Community Resilience in Disasters: The Role for Schools, Youth, and Families. New York: Springer, 2005.
42. Jones RT, Kazdin AE, Haney JI. Social validation and training of emergency fire safety skills for potential injury prevention and life saving. Journal of Applied Behavior Analysis, 1981; 14(3):249–260.
43. Hull B. Changing realities in school safety and preparedness. Journal of Business Continuity & Emergency Planning, 2011; 5(1):440–451.
44. Johnston DM, Tarrant RA, Tipler K, Coomer MA, Pedersen S, Garside R. Preparing schools for future earthquakes in New Zealand: Lessons from an evaluation of a Wellington school exercise. Australian Journal of Emergency Management, 2011; 26(1):24–30.
45. Green R, Petal M. Lesson Learned from School Participation in the 2008 ShakeOut. Bellingham, WA: Western Washington University, 2010.
46. Central U.S. Earthquake Consortium. Great Central U.S. ShakeOut: Overview and Final Report, 2011.
47. Vosniadou S. How children learn. Pp. 16–33 in Rao DB (ed). Successful Schooling. New Delhi: Discovery Publishing House, 2003.
48. Runyan WM. A stage–state analysis of the life course. Journal of Personality and Social Psychology, 1980; 38(6):951–962.
49. Public Education: What's the Plan, Stan? Wellington, New Zealand: Minstry of Civil Defence and Emergency Management, c2006. Available at: http://www.civildefence.govt.nz/memwebsite.nsf/wpg_url/for-the-cdem-sector-public-education-whats-the-plan-stan?
50. Hamilton J. New Zealand: National Progress Report on the Implementation of the Hyogo Framework for Action (2011–2013). Wellington: Government of New Zealand, 2013.
51. What's the Plan, Stan? Communications Strategy for 2009 Launch. Wellington, New Zealand: Ministry of Civil Defence and Emergency Management, 2009.
52. Renwick J. Report of the 2012 "What's the Plan, Stan?" Survey of New Zealand Primary Schools. Wellington, New Zealand: Research and Evaluation Services, Strategy and Governance Branch of the Department of Internal Affairs, 2012.
53. Fournier DM. Evaluation. Pp. 139–140 in Mathison S (ed). Encyclopedia of Evaluation. Thousand Oakes, CA: Sage, 2005.