



Aligning evaluation approaches with context: reductionism, system thinking, and pragmatic synthesis

Huey T. Chen, Liliana Morosanu & Victor H. Chen

To cite this article: Huey T. Chen, Liliana Morosanu & Victor H. Chen (2024) Aligning evaluation approaches with context: reductionism, system thinking, and pragmatic synthesis, Asia Pacific Journal of Education, 44:3, 540-551, DOI: [10.1080/02188791.2024.2390657](https://doi.org/10.1080/02188791.2024.2390657)

To link to this article: <https://doi.org/10.1080/02188791.2024.2390657>



Published online: 09 Oct 2024.



Submit your article to this journal [↗](#)



Article views: 129



View related articles [↗](#)



View Crossmark data [↗](#)



Aligning evaluation approaches with context: reductionism, system thinking, and pragmatic synthesis

Huey T. Chen^a, Liliana Morosanu^a and Victor H. Chen^b

^aDepartment of Public Health, Mercer University - Atlanta Campus, Atlanta, USA; ^bManagement, Tulane University, New Orleans, USA

ABSTRACT

The Campbellian validity typology has been used as a foundation for outcome evaluation and for developing evidence-based interventions for decades. As such, randomized control trials were preferred for outcome evaluation. However, some evaluators disagree with the validity typology's argument that randomized controlled trials as the best design for outcome evaluation, and with its strategy to maximize internal validity at the expense of external validity. The debates over these issues have been ongoing and intensive. This paper proposes a fresh idea for approaches to address this controversy, by bringing problem-solving philosophies into discussions. Three problem-solving philosophies have been used to develop intervention programs: reductionism, systems thinking, and pragmatic synthesis. This paper argues the validity typology is appropriate for evaluating reductionism-based programs only, but not systems thinking- or pragmatic synthesis-based programs. Mismatched applications of research designs in evaluation are sources of confusion and controversies. This insight indicates a pressing need for developing evaluation theories and approaches for systems thinking and pragmatic synthesis. Because the majority of community programs are pragmatic synthesis-based programs, this paper devotes to developing such solutions.

ARTICLE HISTORY

Received 1 August 2023
Accepted 14 November 2023

KEYWORDS

Problem-solving philosophies; reductionism; systems thinking; pragmatic synthesis; evaluation approaches

Introduction

One of the major tasks of evaluators is to evaluate the effectiveness of intervention programs. To this end, evaluators have used the Campbellian validity typology (Campbell & Stanley, 1963; T. D. Cook & Campbell, 1979; Shadish et al., 2002) to conduct outcome evaluation. The typology stresses internal validity as the top priority for outcome evaluation – to provide credible evidence for program effectiveness, and randomized controlled trials (RCTs) as the best research design for maximizing internal validity. RCTs are promoted as the gold standard for the effectiveness of new interventions or treatment research (Hariton & Locascio, 2018). When RCTs are difficult to use, the typology recommends the next best designs – quasi-experimental.

Despite its popularity and contributions, the typology is not without criticisms. There have been ongoing, intensive debates on whether RCTs are the best evaluation design for several decades (Chen et al., 2011a). Critics pointed out important assumptions and limitations of RCTs (C. E. Cook & Thigpen, 2019; Ioannidis, 2018), argued that evaluation should be focused on external validity, and promoted the use of qualitative methods instead (Greene, 2011; Howe, 2005).

This paper proposes a new way to examine these issues, by bringing problem-solving philosophies into discussions. Three problem-solving philosophies used to develop intervention programs will be discussed: reductionism, systems thinking, and programmatic synthesis (Chen, 2016). Each problem-solving philosophy has its unique postulates, strategies, and purposes. The premise of discussion in this paper is that intervention programs based on different problem-solving philosophies require different types of evaluation theories, approaches, and designs. We further argue that validity typology and RCTs are only suitable for evaluating reductionism-based programs, although they have been used for evaluating all kinds of programs, regardless of which problem-solving philosophies they use. This one-size-fits-all application of the validity typology is a source of controversy and confusion. Alternative evaluation theories, approaches, and designs, particularly for systems thinking and pragmatic synthesis-based programs are necessary. The rest of the paper will focus on discussing the validity typology merits and controversies, problem-solving philosophies and their implications for evaluation theory and approaches, and on specific evaluation theories and approaches for pragmatic synthesis-based programs, as they appear to comprise the large majority of community-based programs.

The campbellian validity typology in evaluation: contributions and controversies

The validity typology, the top-down strategy, and the evaluation

Program evaluation started as a discipline around the 1960s. During that time, evaluators were searching for research designs and methods for conducting outcome evaluation. Suchman (1967) suggested that the Campbellian validity typology (Campbell & Stanley, 1963), although originally developed for research purposes, can also serve evaluation purposes. Well-received in the evaluation community, the Campbellian validity typology became the dominant paradigm for conducting outcome evaluation. The typology introduced the concepts of internal validity – to conclude on intervention effectiveness, and external validity – to assert its generalizability to other settings or populations. Both internal and external validity are important, however, Campbell and his associates assigned top priority to internal validity, to emphasize the importance of providing credible evidence for intervention effectiveness. The best designs for obtaining credible evidence are RCTs and other experimental designs because they rule out all or most threats to internal validity. Biomedical research has extensively used RCTs for developing and testing pharmaceutical drugs, with remarkable achievements. It stood to reason that the validity typology and RCT designs might enhance the scientific reputation and achievement of program evaluation as well.

Because it provides credible evidence, the typology has been used for developing evidence-based interventions (EBIs), with the ultimate purpose of disseminating them to communities for utilization (Flay et al., 2005; Mowbray et al., 2003). RCTs are employed for EBI efficacy evaluations (in research conditions). Ideally, these should be followed by effectiveness evaluations, in real-world conditions. However, time and cost severely limit effectiveness evaluations, leading to the common practice of dissemination, immediately after efficacy evaluation.

Funding agencies encourage grantees such as local agencies and community-based organizations to use EBIs in their communities and apply the validity typology to evaluate implementation and effectiveness. Furthermore, many journals favour manuscripts that use experimental methods in evaluating intervention programs. All of these show the dominant influence of the validity typology and RCTs in program evaluation and intervention science.

Campbellian typology and RCT controversies

There have been ongoing and intensive debates among evaluators on whether RCTs are the best evaluation method for social programs (Chen et al., 2011b). Opponents presented several arguments against the Campbellian typology (Greene, 2011; Howe, 2005). Most significant arguments proposed

external validity rather than internal validity should be the focus of evaluation Furthermore, *RCTs are not the best method* for this purpose, as they exclude important contextual factors considered threats to internal validity, however of great interest to stakeholders. Finally, *persuasion, interpretation, and the subjective nature of conclusions should be emphasized, rather than validity*, to prioritize the human experience in the natural settings of a program. However intense, the debates have not reduced the influence of Campbellian validity typology, or the privilege of RCTs in outcome evaluation.

Limitations of applying the validity typology and the top-down strategy in communities

The validity typology and top-down strategy for evaluation and dissemination made significant contributions to advance program evaluation and EBIs. However, their applications for program evaluation likely create the following problems or drawbacks:

The relevancy of EBIs to community applications

The setup and requirements for producing EBIs are drastically different from community practice. Often, stakeholders do not see the relevance of EBIs to their practice. A good example to illustrate this stand is the Open Airways for Schools (OAS) intervention – an EBI providing counselling services to parents and children to control and manage asthma (Evans et al., 1999). The EBI used high-qualified counsellors with a Doctoral or Master's degree in counselling while local schools where the EBI was disseminated could not afford to hire highly qualified counsellors, therefore used volunteers – mostly parents with limited training to deliver the intervention. Unsurprisingly, the EBI was perceived as irrelevant to everyday practice. These differences maintain a void between evidence and practice.

Confusions between fidelity and adaptation

Because EBIs are created under controlled/ideal conditions, stakeholders find them difficult to implement in communities. The differences between controlled and real-world conditions create confusion on how to implement EBIs. As such, stakeholders adapt or reinvent EBIs, to match the needs and conditions of their communities. Some researchers disagree on whether adaptations should be allowed and insist on community implementation with high fidelity, to preserve intervention effectiveness (Mowbray et al., 2003). On another hand, others suggest adaptation is necessary to improve intervention fit to the local context, and to improve effectiveness (Castro et al., 2004).

One example to illustrate the widespread adaptation approach was highlighted in evaluating HIV prevention programs sponsored by the Centers for Disease Control and Prevention such as Voices/VOCES and Mpower (Venegas et al., 2009). Although fidelity was stressed in the sponsoring announcement, the community-based organizations considerably modified or redesigned key characteristics of recommended EBIs such as the number/duration of sessions, addition of extra elements, and modification of the intervention content and delivery methods. These adaptations were carried out during the pre-implementation, implementation, and maintenance phases. The controversies of fidelity vs. adaptation have yet to be resolved.

Lack of evidence of the effectiveness of EBIs in communities

As discussed previously when researchers follow the top-down approach to develop EBIs, they only conduct efficacy evaluations without a subsequent effectiveness evaluation in communities). Few EBIs completed sequential efficacy followed by real-world outcome evaluations, with mixed results. For example, Coordinated Approach to Child Health (Luepker et al., 1998) was successful in research as well as real-world settings. By contrast, the evaluation results of Reconnecting Youth (Hallfors et al., 2006) sounded real-world alarm bells. According to the initial efficacy evaluation of this program (Eggert et al., 1994), among other things, the intervention was found to decrease drug control problems and hard drug use and increase Grade Point Average (GPA). However, in a subsequent effectiveness evaluation in real-world settings, not only the program lacked desirable

effects on drug control problems and school performance, but also on peer bonding and socially desirable weekend activities (i.e., program adolescents had worse outcomes than did the control group) (Hallfors et al., 2006). The authors argued that the harmful effects might have resulted from the iatrogenic effects of grouping high-risk youth in a real-world setting.

EBIs do not address practical information that is important for stakeholders

Stakeholders have a great interest in practical issues such as how to recruit and retain clients, how to intervene to fit into existing organizational structures and schedules, how to train implementers, etc. From the researchers' standpoint point, these practical issues may be regarded as trivial, while to stakeholders they are crucial to their day-to-day activities.

However, most EBIs do not adequately address these practical issues, therefore, stakeholders do not find them useful (Green & Glasgow, 2006). These differences in focus result in significant gaps between intervention research and real-world practice (Wandersman et al., 2008).

Problem-solving philosophies and their implications for evaluation theories and approaches

The above controversies and debates surrounding the Campbellian validity typology and the top-down strategy went many years without practical solutions. Examining these issues from a new angle may help break the status quo. This paper proposes such a perspective by bringing problem-solving philosophies underlying intervention programs into discussion, as a context, to examine the issues. This section will discuss these problem-solving philosophies and their relevance for evaluation.

Problem-solving philosophies

The three problem-solving philosophies commonly used to develop intervention programs are reductionism, systems thinking, and pragmatic synthesis (Chen, 2010, 2016). Each will be discussed next.

Reductionism

Reductionism assumes that a system, complex or not, can be broken down into simpler fundamental components which are easier to analyse, understand, or explain (Kemeny & Oppenheim, 1956; Nagel, 1961). If an intervention program is based on reductionism, the program mainly focuses on a few major components. All the variables are clearly defined and measured, and the relationships among them are easy to identify. Biomedical research used reductionism successfully, to study diseases and develop treatments from initial animal studies, human studies, to medicine dissemination. The validity typology follows reductionism's principles by focusing on the rigorous examination of relationships between an intervention and a few clearly defined outcomes.

Systems thinking

Systems thinking argues a system cannot be understood by reducing it to its smaller parts, but by embracing the complexity that emerges from interconnecting parts (Meadows, 2008; Midgley, 2003). This perspective stresses that problem-solving should consider the relationships and interactions among various system components, as a whole. Intervention programs based on systems thinking consider the complex interactions and feedback loops between its different parts. Systems thinking criticizes reductionism for oversimplifying a system by focusing on a few isolated parts, ignoring interdependencies among parts, and failing to consider their context. System thinking-based

programs emphasize the entire system – components and interdependencies – to holistically address complex problems effectively.

Applying systems thinking to develop an intervention program requires identifying all system components and their interactions. Causal loops are a useful tool for this purpose (Barbrook-Johnson & Penn, 2022). For example, Butland et al. (2007) applied systems thinking to an obesity program and identified 108 variables and more than 300 connections among variables. Using the validity typology and RCTs to evaluate systems thinking-based programs would be a herculean task – almost impossible to complete. However, concrete evaluation approaches particularly for system thinking-based programs are yet to be developed.

Pragmatic synthesis

Reductionism and systems thinking appear to be at opposing ends on a scale of complexity, from simplifying or reducing programs to the cumulus of their components to adding the layers of their multiple interactions. Pragmatic synthesis represents a middle ground of the complexity scale and reflects stakeholders' views on how to design and/or implement an intervention program (Chen, 2010, 2016). Working in communities on a day-to-day basis, stakeholders realize that program outcomes are not entirely affected by interventions alone. Contextual factors such as culture, community support, accessibility, peer support, and organizational capacity can interact with interventions and affect outcomes. Stakeholders view programs as one piece of the puzzle – nevertheless essential. However, they are aware that other pieces (i.e., contextual factors) are also needed to solve the puzzle. The extent to which stakeholders address these factors in service delivery depends on funding, expertise, and other available resources in the organization. The great majority of intervention programs designed or implemented by stakeholders in communities are pragmatic synthesis-based programs. However, the evaluation literature has not paid attention to discussing evaluation theories and approaches for these programs.

Implications of problem-solving philosophies for the future direction of evaluation theories and approaches

The above discussions reveal a serious problem in program evaluation. Synthesis-based and system thinking-based programs are drastically different from reductionism-based programs. Traditional validity typology and RCTs are not suitable for evaluating these programs. Unfortunately, the literature does not suggest particular evaluation theories and approaches to evaluate system thinking or pragmatic synthesis-based programs. Consequently, funding agencies continue to require or encourage evaluators and stakeholders to apply the validity typology, fidelity evaluation, and other traditional evaluation approaches to evaluate these programs.

Developing evaluation theories and approaches suitable for systems thinking and pragmatic synthesis-based programs is urgent. The rest of the paper suggests a first step to address these problems by developing evaluation theories and approaches for evaluating pragmatic synthesis-based programs. This focus is because the great majority of real-world programs are pragmatic synthesis-based programs. Our work might guide future research and inspire others to do similar efforts for evaluating systems thinking-based programs.

Evaluation theories and approaches for pragmatic synthesis-based programs

The development of evaluation theories and approaches for evaluating pragmatic synthesis-based programs must build on stakeholders' rather than researchers' views and needs. A major difference between stakeholders' and researchers' views on intervention programs is how to address contextual factors. Researchers view contextual factors as confounding factors or sources of biases that must be controlled or eliminated in evaluation. However, stakeholders cannot eliminate or control

these contextual factors in the real-world conditions of their local communities – they must work with these contextual factors. The reality is, when stakeholders are asked to implement an intervention in communities, the first information they want is not credible evidence on its efficacy, but information on whether it could be viable in their communities, and whether it would be effective in real-world settings. This implies the existing validity typology and top-down strategy may not fit for real-world application. Alternative evaluation theories and approaches for pragmatic synthesis programs must be developed.

This paper proposes a bottom-up strategy (Chen, 2010) for evaluating real-world programs and serving stakeholders' evaluation needs. This strategy consists of a sequence of three innovative evaluation approaches: Viability, effectuality evaluation, and transferability. In actual applications, decision-makers and stakeholders can select one, two, or all three, depending on needs and interests.

Viability evaluation

When community-based organizations or local agencies are asked by funding organizations to implement an EBI, their immediate concern is the viability of the intervention, that is, whether the proposed intervention fits well with the organization and community. The purpose of viability evaluation is to assess:

- Practicality – whether an intervention can be operationalized in community settings and has the potential for solving community problems;
- Suitability – whether an intervention fits a community-based organization's mission and capacity to implement it;
- Evaluability – whether an intervention has a well-organized structure, measurable outcomes, and justifiable connections among them;
- Affordability – whether a community organization is financially capable to provide the services;
- Cultural responsiveness – whether an intervention fits clients' culture or attributes or at least is not against them;
- Helpfulness – whether implementers and ordinary clients accept the intervention and perceive it likely to work.

If the results of the viability evaluation are positive, stakeholders can proceed to implement the intervention. If the results are negative, program designers must work with stakeholders to revise the intervention.

Methodology

Viability evaluation needs to take a participatory approach. To be useful, the evaluation needs to be conducted at the planning or initial implementation stage (Chen, 2003). Evaluators need to get support from both decision-makers of the funding agency and stakeholders to conduct a viability evaluation. Before the evaluation, evaluators need to work with program managers and decision-makers to identify stakeholder representatives to participate in the evaluation. The funding agency must agree to provide a safe environment for participants to express their views without fear of consequences. The research method for data collection could be focus group meetings or face-to-face interviews.

Example of viability evaluation

The Center for Disease Control and Prevention (CDC) decided to establish a national HIV prevention monitoring and evaluation system (M&E) in 2000. This initiative required collaboration

with 65 state-level health departments and numerous community-based organizations. Because this was a challenging undertaking due to the range of organization infrastructures, capacity, and population services, a viability evaluation was conducted. State and community representatives were invited to attend meetings to provide comments and suggestions on the plan. The National Alliance of State and Territorial AIDS Directors (NASTAD) – which represents and advocates for HIV health workers – agreed to participate in the project and to collaborate in selecting representatives and designing group meetings for collecting stakeholders' suggestions and feedback for the M&E plan.

NASTAD's presence encouraged participants to express their concerns and feedback regarding the initiative. Many barriers to the implementation of the M&E were identified (i.e., the burden of reporting to the M&E as initially designed, the lack of capacity and expertise to comply with intricate report requirements, fear of ignoring stakeholders' input in developing evaluation guidelines, arbitrary use of evaluation results by the CDC to decide future funding, etc.). The CDC project team used the viability evaluation findings to revise the M&E system, develop evaluation guidance, and build the state and local agencies and CBOs' capacity. In the end, the M&E project was successfully implemented and had strong stakeholders' buy-in and support (Author).

Holistic effectuality evaluation

The second major evaluation in the bottom-up strategy is holistic effectuality evaluation. This is a very different outcome evaluation approach than the Campbellian validity typology. Holistic effectuality evaluation is to assess the real-world effects of a program by using the concept of adjuvants to better explain real-world effects. Adjuvants are strategies and activities used by stakeholders to capitalize on the contextual factors' influence to intervene and work better in communities. Examples of adjuvants include helping clients realize or admit they have a problem and need help, building relationships with clients, enhancing or supplementing the intervention, addressing cultural appropriateness issues, encouraging or incentivizing participation or retention, and taking advantage of environmental change or natural trends (Chen, 2015). Holistic effectuality evaluation is to evaluate an intervention's joint effects, that is, a combination of the intervention and adjuvants. This is opposed to Campbellian typology which regards adjuvants as threats to internal validity that need to be controlled in outcome evaluation – as it aims to document the pure, independent effects of the intervention.

Holistic effectuality evaluation uses the following steps:

Inquiring into the process of contextualizing an intervention in a real-world setting:

Evaluators need to examine how stakeholders contextualize an intervention in community settings, including the use of adjuvants for service delivery purposes, through in-depth interviews, focus group meetings, and/or field observation. For example, in evaluating an HIV counselling and testing program in China (Chen et al., 2007), evaluators found that stakeholders contextualized the intervention to fit the Chinese society and culture. Adjuvants included building trustful relationships with clients and stressing collective values during counselling.

Using an unobtrusive quantitative design that requires no manipulation of settings/conditions to assess change:

Recommended designs are pretest/post-test or quasi-experimental, but not RCTs. For example, in the HIV counselling and testing study, the evaluators used the one-group pretest/post-test design to assess the joint effects of the intervention and adjuvants.

Using an auxiliary design to triangulate evidence:

The holistic effectuality evaluation recommends auxiliary designs to triangulate the evidence provided by the primary quantitative design. The auxiliary design could be a qualitative study or pattern matching. In the HIV counselling and testing study, the evaluators used pattern matching and interviews to triangulate the evidence of the pretest/post-test. The pattern matching used smoking as a controlled construct because Chinese drug users usually are smokers. The intervention

did not intend to reduce clients' smoking, just the HIV risks, which is what the findings indicated. The qualitative data provided additional information to endorse the program's effectiveness.

Transferability evaluation

The third major type of evaluation in the bottom-up strategy is transferability evaluation. The concept of transferability is different from the concept of external validity as proposed by Campbell and Stanley (1963) and Shadish et al. (2002). They conceptualized external validity as "to what populations, settings, treatment variables, and measurement variables can this effect be generalized?" (Campbell & Stanley, 1963). According to this definition, external validity is an endless quest for generalizability. As such, this is a research rather than evaluation concept, because it is impossible for any evaluation to achieve external validity. This might explain the lacking progress in external validity evidence or evaluation in the last several decades.

The bottom-up strategy requires a practical concept and approach to address generalizability issues. Chen (2014) proposed two kinds of generalizability: Targeted generalization and exhibited generalization. Targeted generalization is for clearly defined settings in which generalization is sought (i.e., similar intervention, implementing organization, clients, and community). However, in most cases, information on the target for generalization is not precise. In this case, exhibited generalization can address this issue.

Exhibited generalization requires that evaluation provides sufficient information on the contextual factors for an intervention to be successful in a real-world application. Stakeholders can thereafter use the information to assess the intervention generalization potential to their population and settings and decide whether the intervention fits their communities. Exhibited generalization needs to provide the following information to potential users: (1) How to develop and/or set up the intervention? (2) Is the intervention viable in community settings? (3) Is the intervention effective in community settings? (4) What are the conditions that make the intervention work? Based on the information, potential users can make an informed decision on whether to adopt or adapt the intervention to their communities.

It is important to know that the exhibited generalization does not tell potential users whether the intervention effects would be generalizable to their communities. Rather, stakeholders must make an informed decision whether to adopt or adapt the intervention in their community, based on transferability evaluation findings. Evaluators may conduct a follow-up study to examine how many communities adopt or adapt the intervention to their communities and what were the results.

A good example of transferability evaluation is the evaluation of a carbon monoxide ordinance (Chen et al., 2014). A North Carolina county passed an ordinance requiring carbon monoxide (CO) alarms to be installed in residential buildings starting in 2012. Other counties and cities had frequently reached out with questions, being interested to enact similar ordinances in their communities. However, the county health department did not have sufficient information to provide adequate answers to interested parties, having never conducted a formal evaluation of the ordinance's effects. With assistance from the Centers for Disease Control and Prevention, a transferability evaluation was conducted to address dissemination issues (Chen et al., 2014). The evaluation team conducted an exhibited generalization to provide answers to the following questions:

(1) How to develop and/or set up the intervention?

New legislation requires justifying data (i.e., the county reported 124 cases of symptomatic CO poisoning cases and 4 deaths at that time) to rally support in the legislation process. The department began to build a coalition to support the ordinance including the fire department, medical service providers, hospitals, and governmental public officers. They also needed to defuse potential opposition groups such as builder's associations (concerned with the cost increase of new construction), apartment or landlord associations (concerned with the cost of alarm installation), natural gas utility companies (concerned with the burden of false alarms), etc.

(2) Is the intervention viable in the community setting?

Key partners expressed the ordinance was practical because they were capable of implementing the tasks related to the ordinance. The local health department and fire department mentioned their missions and capacity were suitable for implementing the ordinance. They also indicated the ordinance was manageable within their budgets and resources. They viewed the goals of the ordinance were clear and achievable. Overall, key stakeholders viewed the CO alarm as viable in their county.

(3) Is the ordinance effective?

The evaluation team conducted an outcome evaluation by randomly selecting a sample of 214 households and found 75.3% of these homes had working alarms after the CO legislation. The percentage was higher than the national data which showed 30% of households in the United States had working CO alarms. Furthermore, the residents of the intervention county had a 50% reduction in severe CO poisoning in comparison with nearby counties.

(4) What are the conditions that make the intervention work?

Key partners of the program made additional suggestions after the ordinance was passed. The ordinance stated that violation of the ordinance is a misdemeanour and subject to a fine of \$50. They suggested authorities not enforce the ordinance because the enforcement might diminish public support. Instead, they suggest that efforts should be placed on promoting public awareness of CO poisoning and prevention. One of the strategies is to work with the media. For example, supporters of the ordinance need to closely work with the media and encourage them to report public hearings, debates, and meetings during the legislative period. The county health department provided the evaluation information to interested parties or potential users for them to consider whether the CO alarm ordinance in this county is transferable to their communities and make their own decision.

There is a potential for evaluators to conduct a follow-up study by assessing whether those who inquired about the information made a decision to move ahead to enact legislation of CO ordinance, how did they implement the legislation, and what were the outcomes.

Conclusion and discussion

The Campbellian validity typology and the top-down strategy have been intensively used for evaluation and dissemination, not without long and intensive debates on their merits and usefulness. This paper provides a new perspective for discussion by bringing in problem-solving philosophies as a context to examine under what conditions are appropriate to apply the validity typology. The paper discusses three problem-solving philosophies traditionally used as a foundation to develop intervention programs: Reductionism, systems thinking, and pragmatic synthesis. From these perspectives, the validity typology and the top-down strategy seem appropriate only for evaluating reductionism-based programs, but not for others. However, the top-down strategy has almost unequivocally recommended the Campbellian typology for evaluating programs implemented in practice settings. Such a one-size-fits-all application of the validity typology is the source of controversy and confusion. To further complicate the issue, evaluation theories and approaches for evaluating systems-thinking or pragmatic synthesis-based programs have not been developed yet. This paper represents a step towards addressing this problem, by proposing a bottom-up strategy for evaluating and disseminating pragmatic synthesis-based programs. This strategy aims to accommodate stakeholders' needs since the majority of programs developed and/or implemented by stakeholders in communities are such programs. This paper introduced three innovative evaluation approaches under this strategy: Viability evaluation, holistic effectuality evaluation, and transferability evaluation, and provided the frameworks, methodologies, and examples for each approach.

The bottom-up strategy and its evaluation approaches for pragmatic synthesis programs have merits worth discussing.

1. Assure the recommended intervention is useful to stakeholders to avoid wasting money.

The traditional top-down approach usually begins with an expensive and time-consuming efficacy evaluation to assess an innovative intervention. After millions of dollars are spent on

efficacy evaluation, it might be found that the efficacious intervention is very difficult to implement in the real world, not of interest to stakeholders, or not effective in the real world. As a consequence, the EBI may not be useful resulting in a waste of money and other resources. By contrast, the bottom-up approach starts with viability evaluation. This first assesses the viability of an intervention as proposed by researchers or stakeholders. The use of viability evaluation assures that an intervention has a good chance to survive in the real world before undertaking an expensive effectiveness or efficacy evaluation, or both. Because interventions with low viability are screened out in the beginning, this approach could save funding agencies considerable money and resources. The bottom-up approach encourages funding agencies to fund viability evaluations in communities first, to accumulate scientific knowledge on viable interventions, and to select highly viable interventions for further rigorous studies.

2. Provide an opportunity to revise and improve an intervention in the real world before its finalization.

One top-down approach limitation is finalizing the intervention protocol or package before or during efficacy evaluation. The protocol is not supposed to change after the evaluation. When an intervention protocol is finalized at such an early stage, it prevents the intervention from gaining input and/or feedback for improvement from stakeholders. This approach may restrict the intervention's applicability to the real world.

By contrast, the bottom-up approach allows for improving an intervention during the viability evaluation. One viability evaluation purpose is to enhance the quality and usefulness of intervention from organizational and community dynamics standpoints. Intervention protocols developed from stakeholder inputs and implementation experience increase their real-world relevancy and contribution.

3. Provide a balanced view of credible evidence.

Verifiable validity subsumes credible evidence from three components: viability, efficacy/effectiveness, and generalizability. Under this perspective, evidence on intervention efficacy/effectiveness (internal validity) is not a stand-alone or context-free concept. Rather, it should be viewed or discussed regarding viability and generalizability. This perspective may aid those advocates of evidence-based intervention to move from currently narrow views of single-dimension evidence (efficacy/effectiveness) to a more balanced, multi-dimensional evidence model (viability, efficacy/effectiveness, and generalizability).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Huey Chen is a Professor of the Department of Public Health and the Director of the Center for Evaluation and Applied Research at Mercer University. He published several books and over 90 peer-reviewed articles in program evaluation.

Liliana Morosanu is a Lecturer at the Department of Public Health and Senior Evaluator at the Center for Evaluation and Applied Research at Mercer University.

Victor Chen is a Lecturer in the Management Department, A.B. Freeman School of Business at Tulane University. He is particularly interested in program evaluation and health management issues.

References

Barbrook-Johnson, P., & Penn, A.S. (2022). Causal loop diagrams. In P. Barbrook-Johnson & A.S. Penn (Eds.), *Systems mapping: How to build and use causal models of systems* (pp. 47–59). Springer International Publishing.

- Butland, B., Jebb, S., Kopelman, P., McPherson, K., Thomas, S., Mardell, J., & Parry, V. (2007). *Tackling obesity: Future choices-project report* (Vol. 10, p. 17). Department of Innovation, Universities and Skills.
- Campbell, D.T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. RandMcNally.
- Castro, F.G., Barrera, J.M., & Martinez, J.C.R. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science*, 5(1), 41–45. <https://doi.org/10.1023/B:PREV.0000013980.12412.cd>
- Chen, H.T. (2003). Theory-driven approach for facilitation of planning health promotion or other programs. *Canadian Journal of Program Evaluation*, 18(2), 91–113.
- Chen, H.T. (2010). The bottom-up approach to integrative validity: A new perspective for program evaluation. *Evaluation and Program Planning*, 33(3), 205–214. <https://doi.org/10.1016/j.evalprogplan.2009.10.002>
- Chen, H.T. (2016). Interfacing theories of program with theories of evaluation for advancing evaluation practice: Reductionism, systems thinking, and pragmatic synthesis. *Evaluation and Program Planning*, 59, 109–118. <https://doi.org/10.1016/j.evalprogplan.2016.05.012>
- Chen, H.T., Donaldson, S.I., & Mark, M.M. (2011a). *Advancing validity in outcome evaluation: Theory and practice: New directions for evaluation, number 130*. John Wiley & Sons.
- Chen, H.T., Donaldson, S.I., & Mark, M.M. (2011b). Validity frameworks for outcome evaluation. In H.T. Chen, S. I. Donaldson, & M.M. Mark (Eds.), *Advancing validity in outcome evaluation: Theory and practice* (Vol. 130, pp. 5–16). Jossey-Bass.
- Chen, H.T., Liang, S., Liao, Q., Wang, S., Schumacher, J.E., Creger, T.N., Wilson, C.M., Dong, B., & Vermund, S.H. (2007). HIV voluntary counseling and testing among injection drug users in South China: A study of a non-government organization based program. *AIDS and Behavior*, 11(5), 778–788. <https://doi.org/10.1007/s10461-007-9215-x>
- Chen, H.T., Yip, F., Lavonas, E.J., Iqbal, S., Turner, N., Cobb, B., & Garbe, P. (2014). Using the exhibited generalization approach to evaluate a carbon monoxide alarm ordinance. *Evaluation and Program Planning*, 47, 35–44. <https://doi.org/10.1016/j.evalprogplan.2014.06.003>
- Cook, C.E., & Thigpen, C.A. (2019). Five good reasons to be disappointed with randomized trials. *The Journal of Manual & Manipulative Therapy*, 27(2), 63–65. <https://doi.org/10.1080/10669817.2019.1589697>
- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Rand McNally.
- Eggert, L.L., Thompson, E.A., Herting, J.R., Nicholas, L.J., & Dicker, B.G. (1994). Preventing adolescent drug abuse and high school dropout through an intensiveschool-based social network development program. *American Journal of Health Pro-Motion*, 8(3), 202–215.
- Evans, R., Gergen, P.J., Mitchell, H., Kattan, M., Kercsmar, C., Crain, E., Anderson, J., Eggleston, P., Malveaux, F.J., & Wedner, H.J. (1999). A randomized clinical trial to reduce asthma morbidity among inner-city chil-dren: Results of the national cooperative inner-city asthma study. *Journal of Pedi-Atrics*, 135(3), 332–338.
- Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S., Mościcki, E.K., Schinke, S., Valentine, J.C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151–175. <https://doi.org/10.1007/s11121-005-5553-y>
- Green, L.W., & Glasgow, R.E. (2006). Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. *Evaluation & the Health Professions*, 29(1), 126–153.
- Greene, J.C. (2011). The construct(ion) of validity as argument. *New Directions for Evaluation*, 2011(130), 81–91. <https://doi.org/10.1002/ev.367>
- Hallfors, D., Cho, H., Sanchez, V., Khatapoush, S., Kim, H.M., & Bauer, D. (2006). Efficacy vs effectiveness trial results of an indicated “model” substance abuse program: Implications for public health. *American Journal of Public Health*, 96(12), 2254–2259.
- Hariton, E., & Locascio, J.J. (2018). Randomised controlled trials - the gold standard for effectiveness research: Study design: Randomised controlled trials. *BJOG: An International Journal of Obstetrics & Gynaecology*, 125(13), 1716. <https://doi.org/10.1111/1471-0528.15199>
- Howe, K.R. (2005). The question of education science: Experimentism versus experimentalism. *Educational Theory*, 55(3), 307–321.
- Ioannidis, J.P.A. (2018). Randomized controlled trials: Often flawed, mostly useless, clearly indispensable: A commentary on Deaton and Cartwright. *Social Science & Medicine*, 210(1982), 53–56. <https://doi.org/10.1016/j.socscimed.2018.04.029>
- Kemeny, J.G., & Oppenheim, P. (1956). On reduction. *Philosophical Studies*, 7(1–2), 6–19. <https://doi.org/10.1007/BF02333288>
- Luepker, R., Perry, C., Osganian, V., Nader, P., Parcel, G., Stone, E., & Webber, L. (1998). The child and adolescent trial for cardiovascular health (CATCH). *The Journal of Nutritional Biochemistry*, 9(9), 525–534.
- Meadows, D.H. (2008). *Thinking in systems: A primer*. Chelsea Green Publishing.
- Midgley, C. (2003). *Systems thinking*. Sage.
- Mowbray, C.T., Holter, M.C., Teague, G.B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24(3), 315–340. <https://doi.org/10.1177/109821400302400303>
- Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. Harcourt, Brace & World.

- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Suchman, E. (1967). *Evaluative research: Principles and practice in public service and social action program*. Russell Sage.
- Veniegas, R.C., Kao, U.H., & Rosales, R. (2009). Adapting HIV prevention evidence-based interventions in practice settings: An interview study. *Implementation Science*, 4, 1–9.
- Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., Blachman, M., Dunville, R., & Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology*, 41(3–4), 171–181.