

The bottom-up approach to integrative validity: A new perspective for program evaluation[☆]

Huey T. Chen^{*}

Centers for Disease Control and Prevention, National Center for Environmental Health, Division of Environmental Hazards and Health Effects, Air Pollution and Respiratory Health Branch, Atlanta, GA, United States

ARTICLE INFO

Article history:

Received 20 March 2009

Received in revised form 25 September 2009

Accepted 11 October 2009

Keywords:

Viable validity

Viability evaluation

Internal validity

External validity

Bottom-up approach

Top-down approach

Integrative validity model

Credible evidence

ABSTRACT

The Campbellian validity model and the traditional top-down approach to validity have had a profound influence on research and evaluation. That model includes the concepts of internal and external validity and within that model, the preeminence of internal validity as demonstrated in the top-down approach. Evaluators and researchers have, however, increasingly recognized that in an evaluation, the over-emphasis on internal validity reduces that evaluation's usefulness and contributes to the gulf between academic and practical communities regarding interventions. This article examines the limitations of the Campbellian validity model and the top-down approach and provides a comprehensive, alternative model, known as the integrative validity model for program evaluation. The integrative validity model includes the concept of viable validity, which is predicated on a bottom-up approach to validity. This approach better reflects stakeholders' evaluation views and concerns, makes external validity workable, and becomes therefore a preferable alternative for evaluation of health promotion/social betterment programs. The integrative validity model and the bottom-up approach enable evaluators to meet scientific and practical requirements, facilitate in advancing external validity, and gain a new perspective on methods. The new perspective also furnishes a balanced view of credible evidence, and offers an alternative perspective for funding.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

For several decades, the Campbellian validity model (Campbell & Stanley, 1963) and the traditional top-down approach to validity (FDA, 1992; Flay, 1986; Flay et al., 2005; Greenwald & Cullen, 1984; Mrazek & Haggerty, 1994) have been widely discussed and used. But the literature has yet to examine systematically and critically their conceptual and methodological soundness in the context of program evaluation or to provide an alternative to any of its perceived limitations. This article examines the weaknesses of the Campbellian model and the top-down approach and provides an alternative—not only for program evaluation, but for improved validity as well.

2. The Campbellian model and the top-down approach to validity

2.1. The Campbellian validity model

That validity is an essential element of any research or evaluation is axiomatic (Campbell & Stanley, 1963; Cook &

Campbell, 1979; Shadish, Cook, & Campbell, 2002). Over 40 years ago Campbell and Stanley (1963) proposed a validity model that continues to exert a profound influence on research and evaluation. They identified two principal validity types: internal and external.¹ Internal validity asks whether, in this specific experimental instance, the intervention made a difference. External validity, on the other hand, asks whether an experimental effect is capable of generalization, and if so, to what populations, settings, or treatment and measurement variables. Although both internal and external validity are crucial for research, Campbell and Stanley found an inverse relationship between the two: an increase of internal validity tends to decrease external validity, and vice versa. Thus in any study, this trade-off complicates any attempt to maximize both internal and external validity. Campbell and Stanley accepted the tradeoff as a given, then forcefully argued for the primacy of internal validity. They viewed internal validity as the *sine qua non* of research. That is, the legitimacy of any

[☆] The views expressed in this article do not necessarily represent the views of the Centers for Disease Control and Prevention.

^{*} Correspondence address: Centers for Disease Control and Prevention, APRHB/EHHE/NCEH/CDC, 4770 Buford Highway, MS F-58, Chamblee, GA 30341, United States. Tel.: +1 770 488 3732; fax: +1 770 488 1540.

E-mail address: hbc2@cdc.gov.

¹ Cook and Campbell (1979) and Shadish et al. (2002) have revised the internal and external classification. One important feature in their revision is the subdivision of the values of internal and external validity into two additional categories: statistical conclusion validity and construct validity. The former refers to the appropriateness of drawing conclusions from the statistical evidence. The latter refers to generalizing about higher order constructs from research operations. Popular use of the validity types is, however, still Campbell and Stanley's original version.

research finding rests first on internal validity; careful elimination of all internal validity spuriousness so there is a true result that can be generalized (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002). Campbell and Stanley asserted that without rigorously verifiable internal validity, any attempt at external validity would be misleading or irrelevant. Their arguments provide a justification for research models or evaluations that concentrate on internal validity.

Without question, the Campbellian validity model has made an important contribution to evaluation theories and methodologies. Its concepts of internal and external validity have become fundamental terms in program evaluation (Alkin, 2004; Shadish, Cook, & Leviton, 1991). Major evaluation textbooks have extensively discussed experimental and quasi-experimental methods recommended by the model for strengthening internal validity claims (Posavac & Carey, 2007; Rossi, Lipsey, & Freeman, 2004; Weiss, 1998).

2.2. Top-down approach for advancing validity

Due to internal validity's primacy, most validity issues are addressed by a top-down approach (FDA, 1992; Flay, 1986; Flay et al., 2005; Greenwald & Cullen, 1984; Mrazek & Haggerty, 1994). The top-down approach is a series of evaluations, beginning with maximizing internal validity by efficacy evaluations, then moving on to effectiveness evaluations aimed at strengthening external validity (Flay, 1986; Flay et al., 2005; Kellam & Langevin, 2003). This strategy has been intensively and successfully used in biomedical research (FDA, 1992), where a potential medicine is often identified through laboratory animal studies. The experimental intervention moves through four phases of research as outlined by the U.S. Food and Drug Administration (FDA, 1992). Phases I and II are research on dose and safety issues. After a drug is deemed safe for human trials, it is formally evaluated through a sequence of evaluations, namely efficacy evaluations (Phase III) and effectiveness evaluations (Phase IV).

2.2.1. Efficacy evaluations (efficacy studies)

Efficacy evaluations (or efficacy studies) assess treatment effects in an ideal, highly controlled, clinical research setting. Randomized Controlled Trials (RCT) are typically used to maximize internal validity (Flay, 1986; Flay et al., 2005; Kellam & Langevin, 2003). This type of evaluation usually involves

- Recruiting homogeneous and cooperative patients,
- A random assignment of patients to an experimental or control group,
- Achievement of objectivity through the use of blinding techniques (e.g., patients and therapists do not know which treatment condition they are assigned to),
- Assurances of the exact amount of dosage delivery, and
- Highly trained therapists.

Efficacy studies can provide highly credible evidence of causality between intervention and outcome. RCTs are in fact recognized as the backbone of such studies.

2.2.2. Effectiveness evaluations (effectiveness studies)

If the efficacy evaluations find the treatment has the desired effect on a small homogenous sample, effectiveness evaluations then estimate treatment effects in ordinary patients in real-world, clinical practice environments (Flay, 1986; Flay et al., 2005; Kellam & Langevin, 2003). Unlike efficacy evaluations, however, effectiveness evaluations require addressing both internal and external validity. To reflect the real-world, recruitment and eligibility criteria are loosely defined to create a heterogeneous and

representative sample of the targeted populations. Treatment delivery and patient adherence are less tightly monitored and controlled than in efficacy evaluations. The central idea is that to enhance external validity, effectiveness studies must resemble real-world environments.

Because the efficacy evaluation has already established a causal relationship between the drug therapy and the health outcome, the effectiveness evaluation can focus on issues related to generalizing the evaluation's findings to the real-world. RCTs that require an intensive manipulation of setting are not suitable for effectiveness evaluation—evaluators often need to resort to non-RCT methods. If, however, a medicine is ultimately proven efficacious and effective and receives FDA approval, that medicine is ready for clinicians to use. Many scientists traditionally regard such a top-down approach as the true scientific path to validity (FDA, 1992; Flay, 1986; Flay et al., 2005; Greenwald & Cullen, 1984; Mrazek & Haggerty, 1994).

3. Limitations on the Campbellian validity model and the top-down approach in the context of program evaluation

Because of the influence of Campbellian validity model and the long acceptance of the top-down approach, these methods have had great appeal to evaluators. The problem is whether that model and that approach are universally suitable for program evaluation.

3.1. Limitations of the Campbellian validity model

Campbell and Stanley initially developed their validity model for academic research (Campbell & Stanley, 1963). Through the introduction provided by Suchman (1967), evaluators became familiar with the Campbellian validity model and methods (Shadish & Luellen, 2004). In terms of nature, logic, and purposes, however, program evaluation might differ greatly from academic research (Alkin, 2004; Chen, 1990, 2005; Patton, 1997; Rossi et al., 2004; Shadish et al., 1991; Weiss, 1998). Because the Campbellian validity model was developed for academic research, ideas and principles proposed by the model might not be wholly applicable or relevant to program evaluation. Similarly, issues crucial to program evaluation but not to academic research are, in the Campbellian validity model, most likely ignored.

Cronbach (1982) was probably the first to bring to evaluators' attention the idea that the Campbellian validity model's emphasis on internal validity does not suit program evaluation. Cronbach argued that external validity – rather than internal validity – was the most important factor in evaluation. He viewed a profitable evaluation as one that could draw stakeholders' attention to relevant facts and maximally influence their decisions. Thus in his view, evaluations had to allow extrapolation of information from domains such as populations, treatments, measurements, and settings (Cronbach, 1982). Emerging evaluation practice and experience have resulted in a growing recognition that the over-focus on internal validity, or the over-reliance on the traditional scientific approach for advancing internal validity, have reduced the value of some evaluations or research (Chinman et al., 2005; Cunningham-Sabo et al., 2007; Glasgow, Lichtenstein, & Marcus, 2003; Glasgow et al., 2006; Green & Glasgow, 2006; Miller & Shinn, 2005; Sandler et al., 2005; Spoth & Greenberg, 2005; Wandersman, 2003; Wandersman et al., 2008; Weiner, Helfrich, Savitz, & Swiger, 2007).

Another Campbellian validity model limitation is external validity itself. As discussed earlier, external validity is defined in the model (Campbell & Stanley, 1963) to answer the question "To what populations, settings, treatment variables, and measurement variables can this effect be generalized?" But this definition may be more appropriate for academic research than for program

evaluation. Academic research has great interest in testing a proposition's or a theory's generalization potential. Academic researchers often search for a law-like proposition capable of generalization across different populations, settings, treatment variables, and measurements. Campbell and Stanley's definition of external validity fits such academic interests very well. In real-world evaluation, however, such an open-ended quest for law-like propositions is extremely difficult or even impossible to achieve—such evaluation is usually constrained by the resources, timelines, and purposes imposed by stakeholders. Thus, when external validity is conceptualized as an endless quest, the concept itself becomes a barrier that hinders evaluators from achieving effective external validity.

An even more serious Campbellian validity model limitation is that it does not adequately address stakeholders' view and need. From a stakeholders' viewpoint, a health promotion/social betterment program is created not merely for addressing scientific concerns, but for meeting political, organizational, and community requirements (Weiss, 1973/1993, 1998). The importance of practical concerns to program evaluation is clearly reflected in the four standards of program evaluation: utility, feasibility, propriety, and accuracy (Sanders & Evaluation, 1994). The Campbellian validity model does not adequately address the first three standards, which primarily concern practical issues. In other words, the Campbellian validity model's scope is not sufficiently comprehensive to serve adequately in specified evaluation situations.

3.1.1. Limitations of the top-down approach

The top-down approach prescribes a sequence of steps. The progression begins with efficacy evaluation, then proceeds to effectiveness evaluation, dissemination, and finally to wide, real-world application. Such a linear progression, however, is not the path most health promotion/social betterment programs travel (Sussman, Valente, Rohrbach, Skara, & Pentz, 2006). Efficacy studies, for example, are usually not followed by effectiveness studies (Glasgow et al., 2003, 2006; Green & Glasgow, 2006; Wandersman, 2003; Wandersman et al., 2008). This results in a huge gap between intervention research and practice (Chinman et al., 2005; Glasgow et al., 2003; Green & Glasgow, 2006; Rotheram-Borus & Duan, 2003; Wandersman, 2003; Wandersman et al., 2008). Today, practitioners, decision makers, and consumers find that traditional scientific evaluation results tend not to be relevant or useful to the everyday issues or situations about which those groups are concerned (Wandersman, 2003; Wandersman et al., 2008). Strategies such as increased effectiveness evaluations, dissemination studies, or translational research (Sussman et al., 2006) may be helpful in narrowing the gulf between research and practical communities, but they alone are a technical fix and do not adequately address the root cause of the problem. To narrow that gulf effectively, health promotion/social betterment programs require another perspective.

4. The integrative validity model and the bottom-up approach as an alternative perspective

4.1. The integrative validity model

The integrative validity model for program evaluation is proposed here as an alternative to the Campbellian model. The integrative validity model builds on, and by addressing certain limitations, enhances the Campbellian model's strengths. Simply stated, integrative validity requires that evaluator-supplied intervention information must not only be scientifically credible, but must be relevant to and useful in stakeholder practice. The major features of the integrative validity model and its contribution in

developing the bottom-up approach for advancing validity are discussed below.

4.1.1. Revision and expansion of the Campbellian validity model

The integrative validity model expands the Campbellian model's internal and external validity into three types: internal, external, and viable. Internal validity is the strength of the Campbellian validity model (Campbell & Stanley, 1963; Shadish et al., 2002). The integrative validity model slightly modifies Campbell and Stanley's definition of internal validity by stressing objectivity. Internal validity becomes the extent to which an evaluation provides objective evidence that an intervention causally affects specified outcomes. External validity remains an essential element of the model, but a discussion of its conceptual revisions is deferred to a later point in this section.

Viable validity – the third element of the integrative validity model – expands the Campbellian model's scope by adding stakeholder views and interests. Viable validity is the extent to which an evaluation provides evidence that an intervention is successful in the real world. Here, viable viability refers to stakeholders' views and experience regarding whether an intervention program is practical, affordable, suitable, evaluable, and helpful in the real-world. More specifically, viable validity means whether ordinary practitioners – rather than research staff – can implement an intervention program adequately, and whether the intervention program is suitable for coordination or management by a service delivery organization such as a community clinic or a community-based organization. An additional inquiry is whether decision makers view the intervention program as affordable. And, can the intervention program (1) recruit ordinary clients without paying them to participate, (2) does it have a clear rationale for its structure and linkages connecting an intervention to expected outcomes, and (3) do ordinary clients and other stakeholders regard the intervention as helpful in alleviating clients' problems or in enhancing their well-being as defined by the program's real-world situations. In this context, *helpful* is whether stakeholders can notice or experience progress in alleviating or resolving a problem.²

In the real-world, stakeholders organize and implement an intervention program. Thus, they have real viability concerns (Chen, 2005; Cronbach, 1982; Suchman, 1967; Weiss, 1973/1993). Viability alone might not guarantee an intervention's efficacy or effectiveness, but in real-world settings, viability is essential for an intervention's overall success. That is, regardless of the intervention's efficacy or effectiveness, unless that intervention is practical, suitable to community organizations' capacity for implementation, and acceptable to clients and implementers, it has little chance of survival in a community.

Furthermore, the integrative validity model revises and expands the Campbellian model's external validity. As discussed previously, under the Campbellian model, external validity is an endless quest for confirmation of an intervention's universal worth—impossible for any evaluation to address. The integrative validity model proposes reconceptualizing external validity from a stakeholders' perspective. Because stakeholders implement real-world programs,

² As in other types of validity, viable validity takes into account threats to its integrity. For example, a threat of partiality means viewpoints of a major stakeholder group such as implementers or clients are neglected in evaluation. Partiality may cause evaluators to reach an improper conclusion on the viability of an intervention program. Fear of evaluation consequences means that stakeholders might be unwilling to provide candid information on viability because a negative result might damage the future of the program. Examples of strategies for dealing with such threats are an inclusion of representatives from major stakeholder groups in planning the evaluation, a triangulation of qualitative and quantitative data, an effective communication between evaluators and stakeholders on the purpose of the evaluation and use of evaluation data, or a combination thereof.

their concerns for generalization are mainly on those issues transferable to real-world situations. External validity for program evaluation should be defined according to such concerns. Thus, the integrative validity model defines external validity as the extent to which evaluation findings of effectiveness can be generalized from a research setting to a real-world setting or from one real-world setting to another targeted setting. This definition stresses that external validity for program evaluation has a boundary—the real-world. That boundary provides evaluators with a useful tool by which to address external validity issues or enhance external validity.

Both exhibited and targeted generalization can advance external validity. Exhibited generalization of an evaluation itself provides sufficient contextual factors for an intervention to be effective in real-world applications. Potential users can adapt the information on the effectiveness of the intervention together with the contextual factors. Users can thereby assess its generalization potential with regard to their own populations and settings and decide whether to apply the intervention in their communities. Exhibited generalization can be enhanced through the “action model-change model” framework as described in the theory-driven approach (Chen, 1990, 2005; Chen et al., 2009). Stakeholders sometimes have a particular real-world target population or setting to which they want to generalize the evaluation results. This is targeted generalization, that is, the extent to which evaluation results can be generalized to a specific population and real-world setting. Targeted generalization is strengthened through methods such as sampling (Shadish et al., 2002), surface similarity (Shadish et al., 2002), Cronbach’s UTOS approach (Cronbach, 1982), or the dimension resemblance approach (Chen, 1990). Thus through exhibited or targeted generalization, external validity adds a workable evaluation concept to traditional external validity.

Furthermore, with viable validity, external validity can be expanded into two additional categories: effectiveness and viability. External validity for effectiveness is the traditional meaning of external validity. It refers to the ability of intervention effectiveness to generalize from research to real-world settings or from one real-world setting to another. External validity for viability, however, is an emerging concept that asks the question “To what extent can evaluation findings of an intervention’s viability be generalized from a research setting to a real-world setting or from one real-world setting to another targeted setting”? The distinction is important; that an intervention *effect* might generalize to another setting does not guarantee that an intervention’s *viability* will similarly generalize.

4.1.2. Viability evaluation and the evaluation types

The integrative validity model’s additional evaluative dimension focuses on viable validity. This type of evaluation – identified here as viability evaluation – assesses the extent to which an intervention program is viable in the real-world. More specifically, it evaluates whether the intervention

- Can recruit and/or retain ordinary clients,
- Can be adequately implemented by ordinary implementers
- Is suitable for ordinary implementing organizations to coordinate intervention-related activities,
- Is affordable,
- Is evaluable, and
- Enable ordinary clients and other stakeholders to view and experience how well it solves the problem.

Evaluation literature has discussed those evaluations that stress a practical dimension. Such evaluations include needs assessments (Altschuld, 1995), feasibility studies (Chen, 2005), capacity assessments (Gibbon, Labonte, & Laverack, 2002), and evaluability

assessments (Wholey, 1987). Traditionally, these evaluations are individually discussed. But viability evaluation may provide an effective conceptual framework for unity and synthesis. For example, a program’s viability is enhanced if the program targets community needs, if the program’s feasibility is demonstrated, if implementing organizations have the capacity to implement the program, and if the program is capable of evaluation. Subsuming these separate evaluations under the viability evaluation rubric could create a comprehensive approach to viability issues and could synergize their contributions to evaluation.

All this points up the fact that viability evaluation requires mixed methods (Chen, 2006; Greene & Caracelli, 1997; Tashakkori & Teddlie, 2003). On the one hand, evaluation relies on quantitative methods to collect data by which it monitors progress on recruitment, retention, and outcome. On the other hand, evaluation requires an in-depth understanding of stakeholders’ views on, and their experience with, an intervention program. For example, qualitative methods are necessary to assess viability’s helpfulness component and to understand clients and implementers’ views and experience on intervention performance. For helpfulness, data are triangulated by quantitative measures of progress. Similarly, affordability is assessed by cost data on money spent on the project and on a subjective assessment of affordability provided by program managers and other decision makers.

Recall that the integrative validity model expands the two traditional evaluation types (efficacy and effectiveness) into three (efficacy, effectiveness, and viability). And remember that in an efficacy evaluation, the fact that an intervention might prove efficacious does not in itself imply success in an effectiveness evaluation; implementation failure (Rossi et al., 2004; Suchman, 1967) or contextual differences between controlled and real-world settings might intervene. Furthermore, an intervention that successfully passes an efficacy evaluation or effectiveness evaluation does not necessarily mean that intervention will pass a viability evaluation. Stakeholders may reject the intervention for practical reasons such as capacity limitation, cultural insensitivity, complexity, and expense.

4.2. The bottom-up approach to advance validity

Viable validity is of prime importance to stakeholders. To enhance that validity, a bottom-up approach for program evaluation is preferable to a top-down approach. The bottom-up approach provides a sequential route for strengthening validity in the reverse order of the top-down approach. To maximize viable validity (i.e., whether the intervention is practical, affordable, suitable, evaluable, and helpful), the evaluation sequence begins with a viability evaluation. If this real-world intervention is in fact viable, a subsequent effectiveness evaluation provides sufficient objective evidence of the intervention’s effectiveness in the stakeholder’s real world. If necessary, the effectiveness evaluation could also address issues of whether such effectiveness is generalizable to other real-world settings. After the intervention is deemed viable, effective, and generalizable in real-world evaluations, an efficacy evaluation using methods such as RCTs will rigorously assess a causal relationship between intervention and outcome.

Fig. 1 highlights the differences between the bottom-up and top-down approaches:

As indicated in the left-hand side of Fig. 1, the top-down approach ideally proceeds from efficacy evaluation, to effectiveness evaluation, then to dissemination. The arrow from efficacy evaluation to dissemination indicates that in reality, dissemination is often carried out after efficacy evaluation. The right-hand side of Fig. 1 illustrates that the bottom-up approach starts with viability evaluation, goes through effectiveness evaluation and efficacy

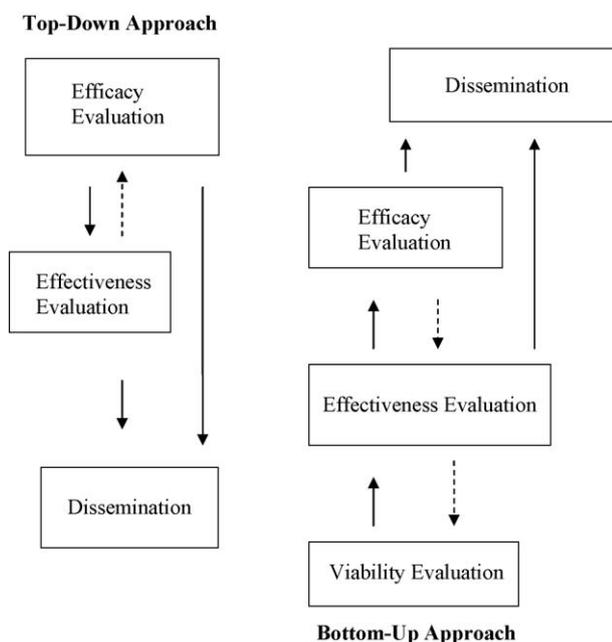


Fig. 1. Top-down approach versus bottom-up approach.

evaluation, and ends at dissemination. In the event of time and resource constraints, the approach can start with effectiveness evaluation and end with dissemination. The point to remember here is that whereas top-down interventions usually represent researchers' ideas, bottom-up interventions come from different sources. Bottom-up interventions could be stakeholders' own idea, stakeholders' idea combined with ideas from literature, stakeholders' adaptation of an evidence-based intervention, or a joint venture of stakeholders and researchers. The bottom-up approach is particularly relevant to health promotion/social betterment programs, but has not yet been formally and systematically discussed in the literature.

4.2.1. The bottom-up approach and health promotion/social betterment programs

A health promotion/social betterment program usually starts as a real-world response to a pressing problem (Bamberger, Rugh, & Mabry, 2006; Chen, 2005; Patton, 1997; Rossi et al., 2004; Weiss, 1998). The literature might provide some guidance, but stakeholders themselves are usually responsible for developing or putting together the program and usually under time and monetary constraints. Given such conditions, stakeholders ask that an evaluation provide field evidence regarding whether the intervention could successfully reach and recruit participants, whether it has broad support from the community, whether an organization can smoothly run it, and whether clients or other key stakeholders feel the intervention is helpful in the field. Such information is useful to stakeholders for strengthening the same program in the future or for designing and implementing similar programs in other regions.

In such instances, an initial viability evaluation is an appropriate methodology. Conducting a viability evaluation first to address viable validity issues makes sense, especially to demonstrate whether the program is practical and helpful in the real-world. This does not mean, however, that stakeholders are uninterested in whether a causal relationship connects the intervention and the outcomes. Rather, they may feel that a lengthy and resource-intensive efficacy evaluation does not fit their immediate evaluation needs. A program is deemed "viable" then is evaluated by using effectiveness evaluation to assess its effec-

tiveness in a real-world setting and/or whether its effectiveness and viability are generalizable to another real-world setting. If an intervention is found to be viable, effective, and generalizable in the real-world, efficacy evaluation would be conducted to rigorously assess the causal relationship between the intervention and outcomes in order to advance scientific knowledge.

Needle exchange programs (NEPs) provide one example of a bottom-up approach. These programs prevent HIV by providing injection drug users (IDUs) with new, sterile syringes in exchange for their used syringes—an exchange that reduces HIV transmission by discouraging needle sharing. In 1984 a *Junkiebond* (Junkie League), a not-for-profit organization formed by and for illicit drug users, introduced the first NEP in Amsterdam, The Netherlands (Van den Hoek, van Haastrecht, & Coutinho, 1989). The intervention program quickly became widespread. Many community-based organizations were attracted to NEPs because the intervention shared their harm-reduction philosophy and was relatively easy to apply in communities. In addition to sterile syringes, many NEPs now provide condoms and clean sterile equipment or paraphernalia (e.g., cottons, cookers, bleach) that promote safe injection. The drug users themselves support the program's viability. They view this program as nonthreatening and accessible. Thus since the mid-1980s, a number of developed countries and a growing number of developing countries have introduced NEPs as a core component of HIV prevention targeting IDUs (Stimson, 1996).

And NEPs have been subjected to effectiveness evaluations. Several studies have found NEPs effective in reducing injection-related risk behaviors (Bluthenthal, Kral, Erringer, & Edlin, 1998; Des Jarlais et al., 1996; Lurie, 1997; Vlahov et al., 1997) as well as reducing the incidence of HIV (Des Jarlais & Friedman, 1998; Hagan, Friedman, Purchase, & Alter, 1995). Meta-analyses of a large number of effectiveness evaluations (Cross, Saunders, & Bartelli, 1998; Ksobiech, 2003) concluded that NEPs contribute to a reduction in needle sharing. Recently, Masson et al. (2007) used RCTs to evaluate NEPs and to obtain the strongest evidence yet of NEP effectiveness.

For NEPs, the initial intervention idea came from stakeholders. In many situations, however, stakeholders may select components of other evidence-based interventions and combine them with their own ideas to form a new intervention. Similarly, they may alter some components of evidence-based interventions to form a new intervention to fit their local needs. As previously discussed, the bottom-up approach can accommodate such adapted or modified programs. Similarly, when researchers evaluate their own innovative interventions, they can, to assure their interventions have both practical and scientific value, follow the bottom-up rather than the traditional top-down sequence.

4.2.2. Issues on do-no-harm

Any evaluation approach should do-no-harm to participants/clients, or it should at least do more good than harm. Discussing how the top-down approach and bottom-up approach address this issue is important. Before its real-world application, the top-down approach relies on an efficacy evaluation to serve as a safeguard for assuring an intervention does no harm. This strategy works well for evaluating biomedical interventions (FDA, 1992). The assumption usually is that this strategy also works well in health promotion/social betterment programs. The evidence, however, does not support that assumption.

A recent sequential evaluation of a substance abuse program following the top-down approach (Hallfors et al., 2006) illuminates this issue. According to the initial efficacy evaluation of this program (Eggert, Thompson, Herting, Nicholas, & Dicker, 1994), the intervention was found, among other things, to effectively decrease drug control problems, decrease hard drug use, and increase GPA. But a subsequent effectiveness evaluation of the

real-world intervention showed that not only did the program have no desirable effects on drug control problems and school performance—the program actually had worse outcomes than did the control group on conventional peer bonding, high risk peer bonding, and socially desirable weekend activities (Hallfors et al., 2006). The authors argued that the harmful effects may have resulted from the iatrogenic effects of grouping high-risk youth in a real-world setting. Still, this sequential process indicates that contrary to common belief, an efficacy evaluation is not always a reliable safeguard by which to address the do-no-harm issue via the top-down approach. Health promotion/social betterment programs are likely to interact with contextual factors. The bottom line is that a top-down intervention found to do-no-harm in an efficacy evaluation does not mean the intervention will do-no-harm in the real-world.

By contrast, the bottom-up approach uses viability evaluation to address the do-no-harm issue. Viability evaluation is appropriate for this task because the issue is within its scope and the evaluation is conducted in a real-world setting. Clients/participants in real-world settings do not have the artificial benefits provided in efficacy evaluations such as monetary or other incentives, research staff's attention, and intensive supervision. Thus if an intervention has a potentially harmful effect, it is more likely to appear in real-world and rather than controlled settings as demonstrated in the above substance abuse program. Viability evaluation is right on the target.

Furthermore, viability evaluation uses mixed methods. On the one hand, outcome monitoring are used to monitor whether clients are getting better or worse on intended outcomes. If an intervention has harmful effects on intended outcomes, outcome monitoring could detect them. On the other hand, qualitative methods are used to address the do-no-harm issue related to unintended outcomes. Qualitative methods such as in-depth interviews, participant observation, and focus groups are sensitive to harmful effects of unintended outcomes. If a viability evaluation finds an intervention program has any harmful effect, that information is promptly sent to stakeholders for remedial or replacement action. And unless the issue is satisfactorily resolved, the program should stop and subsequent effectiveness and efficacy evaluations are not carried out.

4.2.3. Viability evaluation and dissemination model issues

A viability evaluation assesses a set of components that include practicality, affordability, suitability, evaluability, and helpfulness. Dissemination literature has also discussed how innovation characteristics may affect adoption. For example, Rogers (2003) argued that adoption of an innovation depends on its relative advantage, compatibility, complexity, trial potential, and ease of observation. Similarly, Zaltman and Duncan (1977) identified factors that influence the adoption process, such as reversibility, modifiability, communicability, commitment, the time required for implementation, and the adoption's effect on social relations. Viability evaluation and dissemination models are related, but they have different purposes with different objectives. The point is, however, that many current dissemination models are developed as part of a top-down approach. After an intervention is assessed by an efficacy or an effectiveness evaluation – or both – the intervention dissemination is either carried out or becomes a topic for discussion. As illustrated in Fig. 1, however, viable validity reflects the bottom-up approach: a viability evaluation is conducted first—before an effectiveness or efficacy evaluation, or before dissemination.

The dynamics of the top-down versus bottom-up approach illustrate the relationship between dissemination models and viability evaluation. Under the top-down approach, dissemination of health promotion/social betterment programs is often challenging. Because few effectiveness evaluations are carried out after

efficacy evaluations, dissemination frequently means dissemination of an efficacious intervention only. Stakeholders are skeptical of this kind of intervention because it lacks practical relevancy and value. Dissemination of an intervention after an effectiveness evaluation can overcome some of the limitations. Nevertheless, an effectiveness evaluation still may not adequately address viability issues raised by stakeholders; for example, dissemination of an effectiveness intervention might still be problematic. But under the bottom-up approach, dissemination would be easier and smoother, given that all interventions would have to pass a viability test first—before any other evaluations. Because interventions with effectiveness or efficacy evaluations have already proved viable, stakeholders might be more easily persuaded to use them; stakeholders are attracted to viable interventions. In fact, stakeholders are so attracted to interventions with high viability potential that they can, adopt or adapt a viable intervention without rigorous and systematic evidence indicating its viability and effectiveness. One example is the Resolving Conflict Creatively Program (Aber, Brown, Chaudry, Jones, & Samples, 1996). The bottom-up approach emphasizes that before dissemination, an intervention needs to be at least assessed by viability and effectiveness evaluations and the evaluations must demonstrate the intervention has desirable outcomes.

4.3. Concurrent validity approaches

Under the conceptual framework of the integrative validity model, concurrent validity approaches contemplate dealing with multiple validity issues in a single evaluation. A concurrent approach has important implications for program evaluation. Outcome evaluation is often time-consuming. For example, the turnaround time for an efficacy or effectiveness evaluation of a program could easily be a few years. A long turnaround time plus the related expense are major reasons why stakeholders ask for only one outcome evaluation as opposed to multiple outcome evaluations for a new or an existing program.

In conducting a concurrent evaluation, evaluators face a challenging question: what type of evaluation is preferable for addressing validity issues? General guidance for concurrent approaches follows.

4.4. Maximizing internal validity

When stakeholders need strong, objective proof of a causal relationship between an intervention and its outcomes, when they are willing to provide abundant financial resources to support the evaluation, and when they are willing to accept a relatively long timeline for conducting the evaluation, internal validity is a priority. Evaluators will use efficacy evaluation methodology (FDA, 1992; Greenwald & Cullen, 1984; Mrazek & Haggerty, 1994), and, as discussed previously, in the efficacy evaluation tradition, RCT is the gold standard.

4.5. Maximizing viable validity

If stakeholders have a program with multiple components that are difficult to implement in a community, and if they need evaluative information to assure the survival of the program, viable validity should be a priority. If stakeholders need information about whether a program is practical or helpful in the real-world or whether real-world organizations, implementers, and clients favor the program, an appropriate choice is to maximize viable validity. Evaluators could apply a viability evaluation for this purpose. As discussed earlier, mixed (qualitative and quantitative) methods (Chen, 2006; Greene & Caracelli, 1997; Tashakkori & Teddlie, 2003) are particularly appropriate for viability evaluation.

4.6. Optimizing

If stakeholders prefer that an evaluation provide evidence of two or three types of validity (e.g., viable, internal, and external), researchers should adopt the optimizing approach. This approach focuses on finding an optimal solution for multiple validities in an evaluation (Chen, 1988, 1990; Chen & Rossi, 1983). The effectiveness evaluation is an example of applying the optimizing approach for addressing issues raised by multiple validity types.

5. Advantages of the integrative validity model and the bottom-up approach

The integrative validity model and the bottom-up approach provide an alternative validity perspective for program evaluation. The merits of the new perspective in meeting scientific and practical requirements were illustrated in the previous sections. The other advantages include.

5.1. Providing an useful framework for advocating or working toward external validity

Currently, as seen in editorials and journal articles, many advocate the importance of external validity in evaluation (Klesges, Dziewaltowski, & Glasgow, 2008; Patrick, Scutchfield, & Woolf, 2008; Persaud & Mamdani, 2006; Steckler & McLeroy, 2008). These editorials and articles have made contributions in identifying the problems that result from maximizing internal validity and increasing awareness of the importance of external validity in evaluation. To further strengthen the arguments, however, external validity advocates need to further address issues on the role of internal validity in the context of emphasizing external validity or how to deal with a trade-off relationship between internal and external validity: For example, do they favor a switch from the primacy of internal validity to the primacy of external validity, or do they argue for a balance between internal and external validity? Do they mean a sequential or concurrent approach for strengthening validity? The integrative validity model and the bottom-up approach are useful for external validity advocates to clarify these issues and enhance the coherence of arguments.

Similarly, current approaches for enhancing external validity have mainly focused on generalizability of intervention effectiveness (Shadish et al., 2002). However, as discussed previously, stakeholders' concern on external validity goes beyond effectiveness generalizability. They are greatly interested in knowing whether the original study provides information on viability and whether the viability of the intervention is generalizable to their communities. Both effectiveness generalizability and viability generalizability have to address in order to effectively narrow the gap between the academic and practical communities on interventions. The integrative validity model is useful for program evaluation to expand the existing scope of external validity by covering viability issues. For example, the RE-AIM Framework (Glasgow et al., 2003, 2006; Green & Glasgow, 2006) stressed that generalizability of effectiveness in an evaluation can be strengthened by collecting and reporting information on components such as reach, adoption, implementation, and maintenance. Yet, these components are conceptually more relevant to viability than effectiveness. The integrative validity model may be useful for RE-AIM to clarify its relationship with effectiveness and viability for making a greater contribution to external validity.

5.2. Providing a contingency perspective on methods

Ongoing and sometimes heated debate continues between qualitative and quantitative camps on issues such as RCTs or on

which qualitative methods are the best evaluation tools (Donaldson, Christie, & Mark, 2008). The bottom-up approach provides a contingency perspective (i.e., a context-based outlook) that offers fresh insight into the role of RCTs and other methods that might be useful for qualitative and quantitative evaluators to narrow differences or identify common ground. The contingency perspective argues that for evaluation, a universally best method is simply unavailable. Different methods are useful for different evaluation circumstances. For example, the bottom-up approach recognizes the power of RCTs in maximizing internal validity, but is against a wide application of RCTs in evaluation as proposed by the top-down approach. Instead, RCTs must be applied discreetly and only for those popular interventions already assessed by viability and by effectiveness evaluation. An application of expensive and time-consuming RCTs to evaluate innovative health promotion/social betterment programs without knowing the program's viability wastes money and other valuable resources.

Similarly, in addressing viable and external validity issues and advocating for their greater application in evaluation, the bottom-up approach recognizes the essential value of qualitative methods. Unless, however, validity is conceptualized differently, or unless quantitative methods are combined with qualitative methods, qualitative methods alone are regarded in general as less powerful than are RCTs in ruling out threats to internal validity. Because the contingency perspective emphasizes strengths and limitations of different methods under different evaluation contexts, this view might be more receptive to quantitative and qualitative camps and might be useful for reconciling differences between these two camps.

5.3. Providing a balanced view on credible evidence

The evidence-based intervention movement originally started in medicine in the 1990s (Atkins, Fink, & Slutsky, 2005; Sackett, Rosenber, Gray, Hayes, & Richardson, 1996). The movement has spread to public health (Kohatsu, Robinson, & Torner, 2004; McGinnis & Foege, 2000) and to many social and behavioral disciplines under a general label of evidence-based interventions (Donaldson et al., 2008). In the evidence-based intervention movement, evidence usually means evidence of internal validity, that is, evidence produced by RCTs (Nutbeam, 1999; Speller, Learmonth, & Harrison, 1997; Stephenson & Imrie, 1998; Tilford, 2000). The integrative validity model raises an important issue regarding whether only evidence of internal validity counts as credible evidence. This question is crucial: if internal validity is just one portion of the totality of credible evidence, current evidence-based intervention movements may not have built their arguments on a solid foundation. Even researchers/scientists in medicine have argued that credible evidence should be a multi-dimensional concept. For example, Sackett et al. (1996) pointed out that evidence-based medicine means an integration of individual, clinical expertise with the best available external evidence from systematic research. They argue evidence-based medicine does not mean to apply the best external evidence in slavish, cookbook fashion. Similarly, Atkins et al. (2005) and Haynes (1999) have noted that when policymakers assess scientific evidence of a medical intervention, they ask not only "Can it work?," but also "Will it work?" and "Is it worth it?"

On issues related to credible evidence, the integrative validity model provides a balanced view. It posits that credible evidence of health promotion/social betterment programs is a set of three related types of evidence: viable validity (viability), internal validity (effectiveness), and external validity (generalizability). Under this model, evidence on intervention effectiveness (internal validity) is not a stand-alone or context-free concept. Rather, it

should be viewed or discussed with reference to generalizability and to viability.

Viewing the evidence of intervention effectiveness as a totality of evidence without reference to generalizability, however, can be problematic. This is illustrated in the effectiveness of a health promotion/social betterment intervention being contingent on contextual factors such as the types of implementers, the implementing organizations, and clients (Chen, 1990, 2005). If the context of an intervention is changed, the effectiveness of the intervention is also likely to change. For example, an innovative intervention is evaluated in a controlled setting. Clients are paid to assure their participation and retention. Furthermore, intensively trained, highly paid, and highly motivated research staff implement the intervention to assure its fidelity. In the controlled setting, the evaluation provides strong evidence of effectiveness. Because, however, the current evidence-based approach counts as evidence of internal validity only, it will classify the intervention as an evidence-based intervention.

But factor in external validity, and the credible evidence picture is altered. In the real-world, clients are not paid for participation, and the intervention is typically implemented by a staff of community-based organizations rather than by research staff. Because of the drastic difference between the controlled setting and the real-world, controlled setting effectiveness is not likely replicated. From the standpoint of stakeholders, controlled setting effectiveness may be an artificial effect, relevant or generalizable only to that artificial situation. Accordingly, if the evidence-based intervention movements count only internal validity as an evidentiary criterion, for real-world use they might promote an intervention with strong but so-called artificial evidence of effectiveness. A real-world situation is very different from an artificial situation in terms of factors such as participants, incentives, and implementers. Thus, the intervention will in all likelihood not work in the real-world. An even worse scenario is that a funding agency may require community-based organizations to adopt the artificial, evidence-based intervention as a condition for receiving funds. In such a case, if the intervention does not work in the real-world, community-based organizations may be forced to implement an ineffective intervention in their own communities.

Similarly, viability evidence should be factored in as part of credible evidence. In the real-world, an effective intervention in a controlled setting is not necessarily a viable intervention. And when stakeholders are not able or not willing to implement a nonviable intervention – no matter how strong the evidence of effectiveness produced in controlled settings – the intervention is useless to them. Again, if the evidence-based intervention movement uses only internal validity as credible evidence, it may mistakenly promote interventions that have little chance of real-world survival. Pushing an effective intervention without evidence of real-world viability is not only a waste of valuable resources—it is also unscientific. The integrative validity model argues that credible evidence must include viability, effectiveness, and generalization. The model may aid those advocates of the evidence-based intervention movements to move from current narrow and rigid views of evidence to a well-balanced, credible-evidence model.

The above discussion also indicates that the scope of current efforts for improving report quality is too narrow. For example, the Consolidated Standards of Reporting Trials (CONSORT) provides a 22-item checklist for transparent reporting of RCTs (Moher, Schulz, & Altman, 2001). Only one of these 22 items relates to external validity; none address viable validity. Similarly, the transparent reporting of evaluation with non-randomized designs (TREND) provides a compatible 22-item guideline for improving quality of reporting non-randomized evaluations. Like CONSORT, TREND only has one item relating to external validity and no items that

incorporate viable validity. With its focus on internal validity, TREND cannot address the merits of non-randomized quantitative methods and qualitative methods when dealing with the ability to generalize and with viability issues. In their current forms, CONSORT and TREND may enhance the quality of report in research, but not in service. To improve their usefulness, CONSORT and TREND should systematically address issues of external and viable validity.

5.4. Providing an alternative perspective for funding

In theory, funding agencies are interested in both scientific and viability issues. They want to see their funded projects to be successful in communities or to have capability in solving real-world problems. Perhaps due to the reasons such as an influence from the top-down approach or unavailability of a better alternative for guidance, many agencies tend to heavily emphasize scientific factors and pay insufficient attention on or ignore viability issues in funding consideration and practice. For example, some funding agencies are increasingly interested in using RCTs or other randomized experiments as a qualification criterion for grant application (Donaldson et al., 2008; Huffman & Lawrenz, 2006). As discussed previously, if a funding policy excessively stresses internal validity issues, it could contribute to the gulf between academic and practical communities or waste money on projects that might be rigorous and innovative but that have little practical value. The bottom-up approach provides an alternative perspective for funding agencies to address scientific and viability issues in funding process. This perspective suggests three levels of funding:

1. Funding for viability evaluation.

This funding level provides funds for assessing the viability of existing or innovative interventions. It will formally recognize a stakeholder's contribution in developing real-world programs. Researchers can also submit their innovative interventions for viability testing. In doing so, however, they will have to collaborate with stakeholders in addressing practical issues.

2. Funding for effectiveness evaluation.

The second level of funding is an effectiveness evaluation for viable and popular interventions. Ideally, these evaluations should address both internal and external validity issues.

3. Funding for efficacy evaluation.

The third level of funding is efficacy evaluation for those interventions proven viable, effective, and generalizable in the real-world. Efficacy evaluation provides the strongest evidence of an intervention's actual effect, with practical value as an added benefit.

These three levels of funding will promote collaborations between stakeholders and researchers and ensure that evaluation results meet both scientific and practical demands.

6. Conclusion

Evaluation literature has a long history of stressing the principle that evaluation must respond to stakeholders' views and needs (CDC, 1999; Guba & Lincoln, 1989; Patton, 1997; Rossi et al., 2004; Sanders & Evaluation, 1994). These concerns, however, are yet to be systematically integrated into a validity model with a demonstrated usefulness for guiding evaluation design and practice. Proposed here are the integrative validity model and the bottom-up approach, which provide a new perspective for evaluators systematically to advance validity in their evaluations and meet scientific and practical requirements.

The integrative validity model and the bottom-up approach help to advance external validity and provide a contingency

perspective on methods, a balanced view on credible evidence, and a new perspective for funding. Researchers should not be the only source of scientific knowledge. Rather, stakeholders' program efforts, knowledge, and experience in helping clients, as well as their evaluation priorities, should be recognized and included as an integral part of scientific knowledge. This view ties in with the argument for disseminating successful indigenous programs that fit community capacity and values (Chinman et al., 2005; Glasgow et al., 2006; Miller & Shinn, 2005; Wandersman, 2003; Wandersman et al., 2008) and with the interest in examining stakeholders' program theories in evaluations (Chen, 1990, 2005; Donaldson, 2007; Rossi et al., 2004; Weiss, 1998).

These approaches also suggest a potential horizon for developing a real-world health promotion/social betterment science in the future. Real-world science aims at accumulating a body of knowledge about how to design and implement viable, effective, and generalizable real-world programs. This knowledge is highly useful—not only to assist stakeholders in designing quality real-world programs, but also to serve as a bridge between academic sciences and community practice. On the one hand, real-world science can help researchers work with stakeholders to study and design interventions that have strong implications for real-world applications. On the other hand, real-world science passes on to academic community successful interventions for further study that can expand propositions and theories of changing human behavior. Real-world science could even become an arena for greater dialogue and collaboration between academic and practical communities.

Acknowledgments

The author is grateful to Paul Garbe, Michele Mercier, Elizabeth Herman, Wallace Sagendorph, and anonymous reviewers for their comments, which enhance the original manuscript.

References

- Aber, J. L., Brown, J. L., Chaudry, N., Jones, S. M., & Samples, F. (1996). The evaluation of the resolving conflict creatively program: An overview. *American Journal of Preventive Medicine*, 12(5 Suppl.), 82–90.
- Alkin, M. (Ed.). (2004). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.
- Altschuld, J. (1995). *Planning and conducting needs assessment: A practical guide*. Thousand Oaks, CA: Sage.
- Atkins, D., Fink, K., & Slutsky, J. (2005). Better information for better health care: The evidence-based practice center program and the agency for healthcare research and quality. *Annals of Internal Medicine*, 142(12 Pt. 2), 1035–1041.
- Bamberger, M., Rugh, J., & Mabry, L. (2006). *Realworld evaluation: Working under time, data, and political constraints*. Thousand Oaks, CA: Sage.
- Bluthenthal, R. N., Kral, A. H., Erringer, E. A., & Edlin, B. R. (1998). Use of an illegal syringe exchange and injection-related risk behaviors among street-recruited injection drug users in Oakland, California, 1992 to 1995. *Journal of Acquired Immune Deficiency Syndromes Human Retrovirology*, 18(5), 505–511.
- Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- CDC. (1999). Framework for program evaluation in public health. *MMWR*, 48(RR-11), 1–40.
- Chen, H. T. (1988). Validity in evaluation research: A critical assessment of current issues. *Policy and Politics*, 16(1), 1–16.
- Chen, H. T. (1990). *Theory-driven evaluations*. Thousand Oak, CA: Sage.
- Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oak, CA: Sage.
- Chen, H. T., Yip, F., Iqbal, S., & Garbe, P. (2009). Evaluating a Carbon Monoxide Ordinance: Using Program Theory to Enhance External Validity. *Manuscript presented at American Evaluation Association Annual Meetings, November 10–14, 2009, Orlando, Florida*.
- Chen, H. T. (2006). A Theory-Driven Evaluation Perspective on Mixed Methods Research. *Research in the Schools*, 13(1), 75–83.
- Chen, H. T., & Rossi, P. H. (1983). The theory-driven approach to validity. *Evaluation and Program Planning*, 10, 95–103.
- Chinman, M., Hannah, G., Wandersman, A., Ebener, P., Hunter, S. B., Imm, P., et al. (2005). Developing a community science research agenda for building community capacity for effective preventive interventions. *American Journal of Community Psychology*, 35(3/4), 143–157.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Cross, J. E., Saunders, C. M., & Bartelli, D. (1998). The effectiveness of educational and needle exchange programs: A meta-analysis of HIV prevention strategies for injection drug users. *Quality & Quantity*, 32, 165–180.
- Cunningham-Sabo, L., Carpenter, W. R., Peterson, J. C., Anderson, L. A., Helfrich, C. D., & Davis, S. M. (2007). Utilization of prevention research: Searching for evidence. *American Journal of Preventive Medicine*, 33(1 Suppl.), S9–S20.
- Des Jarlais, D. C., & Friedman, S. R. (1998). Fifteen years of research on preventing HIV infection among injecting drug users: What we have learned, what we have not learned, what we have done, what we have not done. *Public Health Reports*, 113(Suppl. 1), 182–188.
- Des Jarlais, D. C., Marmor, M., Paone, D., Titus, S., Shi, Q., Perlis, T., et al. (1996). HIV incidence among injecting drug users in New York City syringe-exchange programmes. *The Lancet*, 348(9033), 987–991.
- Donaldson, S. L. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York: Lawrence Erlbaum.
- Donaldson, S. L., Christie, C. A., & Mark, M. M. E. (2008). *What counts as credible evidence in applied and evaluation practice?* Newbury Park, CA: Sage.
- 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 intensive school-based social network development program. *American Journal of Health Promotion*, 8(3), 202–215.
- FDA. (1992). *Guideline for the clinical evaluation of analgesic drugs*. Rockville, MD: Food & Drug Administration.
- Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15(5), 451–474.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151–175.
- Gibson, M., Labonte, R., & Laverack, G. (2002). Evaluating community capacity. *Health & Social Care in Community*, 10(6), 485–491.
- Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health*, 93(8), 1261–1267.
- Glasgow, R. E., Green, L. W., Klesges, L. M., Abrams, D. B., Fisher, E. B., Goldstein, M. G., et al. (2006). External validity: We need to do more. *Annals of Behavioral Medicine*, 31(2), 105–108.
- 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.
- In J. Greene & V. J. Caracelli (Eds.), *Advanced in mixed-method evaluation: The challenge and benefits of integrating diverse paradigm (74)*. San Francisco: Jossey-Bass.
- Greenwald, P., & Cullen, J. W. (1984). *The scientific approach to cancer control*. CA: A Cancer Journal for Clinicians, 34(6), 328–332.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Hagan, H. C. J. D., Friedman, S. R., Purchase, D., & Alter, M. J. (1995). Reduced risk of hepatitis B and hepatitis C among injection drug users in the Tacoma syringe exchange program. *American Journal of Public Health*, 85(11), 1531–1537.
- 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.
- Haynes, B. (1999). Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *BMJ*, 319(7211), 652–653.
- Huffman, D., & Lawrenz, F. (Eds.). (2006). *Critical issues in STEM evaluation*. San Francisco: Jossey-Bass.
- Kellam, S. G., & Langevin, D. J. (2003). A framework for understanding "evidence" in prevention research and programs. *Prevention Science*, 4(3), 137–153.
- Klesges, L. M., Dzewaltowski, D. A., & Glasgow, R. E. (2008). Review of external validity reporting in childhood obesity prevention research. *American Journal of Preventive Medicine*, 34(3), 216–223.
- Kohatsu, N. D., Robinson, J. G., & Torner, J. C. (2004). Evidence-based public health: An evolving concept. *American Journal of Preventive Medicine*, 27(5), 417–421.
- Ksobiech, K. (2003). A meta-analysis of needle sharing, lending, and borrowing behaviors of needle exchange program attendees. *AIDS Education and Prevention*, 15(3), 257–268.
- Lurie, P. (1997). Invited commentary: Le mystere de Montreal. *American Journal of Epidemiology*, 146(12), 1003–1006.
- Masson, C. L., Sorensen, J. L., Perlman, D. C., Shopshire, M. S., Delucchi, K. L., Chen, T., et al. (2007). Hospital- versus community-based syringe exchange: A randomized controlled trial. *AIDS Education and Prevention*, 19(2), 97–110.
- McGinnis, J. M., & Foege, W. (2000). Guide to community preventive services: Harnessing the science. *American Journal of Preventive Medicine*, 18(1 Suppl.), 1–2.
- Miller, R. L., & Shinn, M. (2005). Learning from communities: Overcoming difficulties in dissemination of prevention and promotion efforts. *American Journal of Community Psychology*, 35(3/4), 169–183.
- Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet*, 357(9263), 1191–1194.
- Mrazek, P. J., & Haggerty, R. J. (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.
- Nutbeam, D. (1999). Evaluating health promotion. *BMJ*, 318(7180), 404A.

- Patrick, K., Scutchfield, F. D., & Woolf, S. H. (2008). External validity reporting in prevention research. *American Journal of Preventive Medicine*, 34(3), 260–262.
- Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). CA: Thousand Oaks.
- Persaud, N., & Mamdani, M. M. (2006). External validity: The neglected dimension in evidence ranking. *Journal of Evaluation in Clinical Practice*, 12(4), 450–453.
- Posavac, E. J., & Carey, R. G. (2007). *Program evaluation: Methods and case studies*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.
- Rotheram-Borus, M. J., & Duan, N. (2003). Next generation of preventive interventions. *Journal of American Academy of Child and Adolescent Psychiatry*, 42(5), 518–526.
- Sackett, D. L., Rosenber, W. M. C., Gray, J. A. M., Hayes, R., & Richardson, W. S. (1996). Evidence-based medicine: What is and what isn't. *BMJ*, 312, 71–72.
- Sanders, J., & Evaluation, J. C. o. S. f. E. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.
- Sandler, I., Balk, D., Jordan, J., Kennedy, C., Nadeau, J., & Shapiro, E. (2005). Bridging the gap between research and practice in bereavement: Report from the Center for the Advancement of Health. *Death Studies*, 29(2), 93–122.
- Shadish, W. R., & Luellen, J. K. (2004). Donald Campbell: The accidental evaluator. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 80–87). Thousand Oaks, CA: Sage.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundation of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Speller, V., Learmonth, A., & Harrison, D. (1997). The search for evidence of effective health promotion. *BMJ*, 315(7104), 361–363.
- Spoth, R. L., & Greenberg, M. T. (2005). Toward a comprehensive strategy for effective practitioner-scientist partnerships and larger-scale community health and well-being. *American Journal of Community Psychology*, 35(3/4), 107–126.
- Steckler, A., & McLeroy, K. R. (2008). The importance of external validity. *American Journal of Public Health*, 98(1), 9–10.
- Stephenson, J., & Imrie, J. (1998). Why do we need randomised controlled trials to assess behavioural interventions? *BMJ*, 316(7131), 611–613.
- Stimson, G. V. (1996). Has the United Kingdom averted an epidemic of HIV-1 infection among drug injectors? *Addiction*, 91(8), 1085–1088.
- Suchman, E. (1967). *Evaluation research*. New York: Russell Sage.
- Sussman, S., Valente, T. W., Rohrbach, L. A., Skara, S., & Pentz, M. A. (2006). Translation in the health professions: Converting science into action. *Evaluation & Health Professions*, 29(1), 7–32.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.
- Tilford, S. (2000). Evidence-based health promotion. *Health Education Research*, 15(6), 659–663.
- Van den Hoek, J. A., van Haastrecht, H. J., & Coutinho, R. A. (1989). Risk reduction among intravenous drug users in Amsterdam under the influence of AIDS. *American Journal of Public Health*, 79(10), 1355–1357.
- Vlahov, D., Junge, B., Brookmeyer, R., Cohn, S., Riley, E., Armenian, H., et al. (1997). Reductions in high-risk drug use behaviors among participants in the Baltimore needle exchange program. *Journal of Acquired Immune Deficiency Syndrome Human Retrovirology*, 16(5), 400–406.
- Wandersman, A. (2003). Community science: Bridging the gap between science and practice with community-centered models. *American Journal of Community Psychology*, 31(3/4), 227–242.
- Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., et al. (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.
- Weiner, B. J., Helfrich, C. D., Savitz, L. A., & Swiger, K. D. (2007). Adoption and implementation of strategies for diabetes management in primary care practices. *American Journal of Preventive Medicine*, 33(1 Suppl.), S35–S44 quiz S45–S39.
- Weiss, C. (1973/1993). Where politics and evaluation research meet. *American Journal of Evaluation*, 14(1), 93–106.
- Weiss, C. (1998). *Evaluation* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Wholey, J. S. (1987). Evaluability assessment: Developing program theory. In Bickman, L. (Ed.), *Using program theory in evaluation*. 33 (pp.77–92). San Francisco: Jossey-Bass.
- Zaltman, G., & Duncan, R. (1977). *Strategies for planned change*. New York: Wiley.

Huey T. Chen is senior evaluation scientist at the Centers for Disease Control and Prevention (CDC). Previously he was professor of the School of Public Health at the University of Alabama at Birmingham from 2002 to 2007. Dr. Chen has contributed to the advancement of evaluation theory and methodology, especially in areas of program theory, theory-driven evaluation, mixed methods, and validity model and approaches. His 1990 book, *Theory-Driven Evaluations*, is considered the classic text for understanding program theory and theory-driven evaluation. His 2005 book, *Practical Program Evaluation: Assessing and Improving Planning, implementation, and effectiveness*, provides a major expansion of the scope and usefulness of theory-driven evaluations. He was the 1993 recipient of the American Evaluation Association Paul Lazarsfeld Award for contributions to evaluation theory. He received the senior biomedical research award in 1998 and the 2001 Award for Dedication and Scientific Direction in the Development and Implementation of Program Evaluation Research Branch from CDC.