Evaluation and Program Planning 71 (2018) 83–88

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Evaluation capacity building—Results and reflections across two years of a multisite empowerment evaluation in an HIV prevention context

Peter T. Lindeman[a,b,*], Emily Bettin[a,b], Lauren B. Beach[a,b], Christian N. Adames[a,b], Amy K. Johnson[c,d], Dave Kern[e], Patrick Stonehouse[e], George J. Greene[a,b], Gregory Phillips II[a,b]

[a] Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, 625 N. Michigan Ave., 14th Floor, Chicago, IL, 60611, USA
[b] Institute for Sexual and Gender Minority Health and Wellbeing, Northwestern University, 625 N. Michigan Ave., 14thFloor, Chicago, IL, 60611, USA
[c] Center for Gender, Sexuality and HIV Prevention, Division of Adolescent Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, 225 E. Chicago Ave., Chicago, IL, 60611, USA
[d] AIDS Foundation of Chicago, 200 W. Jackson Blvd. #2100, Chicago, IL, 60606, USA
[e] HIV/STI Bureau, Chicago Department of Public Health, 333 S. State Street, Chicago, IL, 60604, USA

ARTICLE INFO

Keywords:
Empowerment Evaluation
Multisite evaluation
Evaluation planning
Evaluation capacity building
Community engagement

ABSTRACT

As the need for rigorous evidence of program efficacy increases, integrating evaluation activities into program implementation is becoming crucial. As a result, external evaluators are placing increased focus on evaluation capacity building as a practice. However, empirical evidence of how to foster evaluation capacity in different contexts remains limited. This study presents findings from an evaluation capacity survey conducted within a multisite Empowerment Evaluation initiative, in which an external evaluator worked with 20 project teams at diverse community agencies implementing HIV prevention projects. Survey results revealed representatives from project teams (n = 33) reported significantly higher overall evaluation capacity after engaging with the external evaluator on planning and implementing their evaluation. Improvements differed across organization type, intervention type, staff position, and reported engagement on various activities throughout the course of the evaluation. Results indicated empowerment evaluation and other stakeholder-focused evaluation approaches are broadly applicable when evaluation capacity building is a desired outcome, particularly when able to engage project staff in the planning of the evaluation and in delivering technical assistance services. Accordingly, efforts should be made by program funders, staff, and evaluators to encourage active engagement starting in the early stages of program and evaluation planning.

1. Introduction

In a time of increasingly limited funding for public health activities, funders have heightened the level of accountability to which they hold grantee agencies, often requiring detailed reporting on outcomes rather than on scopes alone. To accomplish this, grantee organizations must gather meaningful evidence of program effectiveness through rigorous evaluation (Centers for Disease Control & Prevention, 2012). This emphasis on evaluation is particularly relevant in the field of HIV prevention, as small budget HIV prevention services agencies often lack the resources needed to foster the technical ability, staff capacity, and external funding required to conduct an in-depth evaluation of their programming (Kegeles & Rebchook, 2005). Increasingly, funders of HIV prevention interventions allocate a portion of their monies for an external evaluator to consult with funded agencies, following a model used by the Health Resources and Services Administration (HRSA) Special Projects of National Significance (SPNS) initiatives. This approach allows for evaluation, technical assistance, and capacity building to take place without drastically increasing the burden placed on agency staff and finances. However, in public health fields, particularly HIV prevention, target populations and social contexts are always changing, which can diminish the value of a point-in-time evaluation given that findings are nested within a system (Veniegas, Kao, Rosales, & Arellanes, 2009).

The field of evaluation has attempted to address these challenges by promoting evaluation capacity building (ECB) as a central topic within all community-engaged work. Successful ECB aims to foster an organizational learning culture where evaluation activities are integrated

* Corresponding author at: 625 N. Michigan Ave., #14-010, Chicago, IL, 60611, USA.
E-mail addresses: peter.lindeman@northwestern.edu (P.T. Lindeman), emily.bettin@northwestern.edu (E. Bettin), lauren.beach@northwestern.edu (L.B. Beach), cna@northwestern.edu (C.N. Adames), akjohnson@luriechildrens.org (A.K. Johnson), david.kern@cityofchicago.org (D. Kern), patrick.stonehouse@cityofchicago.org (P. Stonehouse), george-greene@northwestern.edu (G.J. Greene), glp2@northwestern.edu (G. Phillips).

https://doi.org/10.1016/j.evalprogplan.2018.09.001
Received 19 April 2018; Received in revised form 26 July 2018; Accepted 5 September 2018
Available online 07 September 2018
0149-7189/ © 2018 Elsevier Ltd. All rights reserved.

into the day-to-day operations of the organization, allowing them to self-monitor and improve program performance. Theoretical models (Preskill & Boyle, 2008) and research syntheses on ECB (Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012) suggest that the evaluation field has reached consensus around the specific metrics indicative of successful ECB: increased knowledge of evaluation, improved attitudes about evaluation, and improved incorporation of evaluation activities within organizations. Despite this consensus, limited empirical research has been conducted to assess ECB in multi-site evaluation contexts. This study seeks to quantitatively measure successful ECB within a multi-site evaluation of HIV prevention organizations in Chicago, in which ECB was a central goal of the evaluation approach employed.

1.1. Empowerment evaluation's role in ECB

There are several key components of program evaluation that help stakeholders and evaluators better understand program activities and intended outcomes. First, generation of a logic model allows the evaluator and stakeholders to map intervention activities onto intended outcomes. Second, creation of data collection tools and an evaluation plan facilitate accurate measurement, which is essential to determining success based on outcomes defined in the logic model. Third, creation and utilization of a fidelity assessment plan is necessary to ensure the activities listed in the logic model match those implemented. Without fidelity assessment, an evaluation is limited in its ability to measure the success of the program, as there may be significant deviations from the ideal implementation of the program. Finally, generation of a program manual allows stakeholders to organize all information about implementing and evaluating the program in a single document. This final component includes all evaluation tools, a summary of the staff and resources needed to implement and evaluate this program, and other relevant project materials (e.g., handouts, links to videos, flyers) needed to replicate the intervention in different settings. Only after establishing each of these components will evaluators and stakeholders be able to fully understand the program they are evaluating.

Program evaluators are accustomed to creating the aforementioned tools. However, the organization evaluands may have limited experience with creating an evaluation framework in tandem with program development, which is key to a rigorous evaluation. By completing the logic model, data collection tools, fidelity measures, and program manual collaboratively with an external evaluator, organization staff build their internal evaluation capacity (Kaufman et al., 2006; Rodriguez-Campos, 2012). This form of ECB is one of the guiding principles of the Empowerment Evaluation (EE) approach, and can occur at both the individual and organizational level (Fetterman & Wandersman, 2005). First, by granting ownership of the evaluation to program staff, the evaluator helps staff build skills at an individual level. Simultaneously, encouraging project directors and field staff to lead the completion of critical evaluation components promotes knowledge and expertise that would allow them to develop future program evaluation plans for their organization. Second, EE has an overarching goal of organizational ECB. Organizational learning and integration of evaluation activities into the day-to-day operations occurs by engaging program staff in the planning and implementation of evaluation activities (Duignan, 2003). Furthermore, the EE theory of process use states involving stakeholders will increase the likelihood that results will be used to make recommendations and inform future project activities (Fetterman, Kaftarian, & Wandersman, 2015). Together, the increased evaluation capacity and increased likelihood of use allow the evaluator to cultivate a more sustainable impact than a traditional, evaluator-led, point-in-time evaluation.

Although approaches, such as EE, place an emphasis on ECB at organizations, limited research has focused on the specific scenarios in which these approaches are most effective. The type of intervention, organizational setting, and project staff engagement with the evaluator each play significant roles in affecting the extent to which capacity building occurs. Further, there is a dearth of knowledge about ECB and the role of an external evaluator in the context of HIV prevention organizations. Addressing this gap in the literature is relevant to evaluators, service organizations, and funders. Evaluators will be better informed when deciding whether to propose an EE model, particularly when working with an HIV prevention organization. Additionally, they can increase effectiveness by tailoring future EE models with best practices and lessons learned from this study. In turn, service agencies will benefit from evaluators using evidence-informed approaches, ultimately resulting in an increased likelihood of successful ECB occurring during a rigorous, stakeholder-led evaluation. This built capacity will not only allow these agencies to implement programs that work through ongoing monitoring and refinement, but it will also enhance their ability to communicate results with current and potential funding agencies. Each of these instances assures funders that they can hold their delegate agencies accountable for implementing effective programming, which also increases their ability to assess the success of ongoing funding initiatives, as well as inform future opportunities they will announce.

This study aims to assess the impact of an external evaluator using a stakeholder-focused evaluation approach, such as EE, on the evaluation capacity of 20 HIV prevention programs in Chicago. Results will provide lessons learned about the efficacy and perceived importance of an external evaluator, or similar entity, using an EE approach in an urban non-profit landscape. While we expect to find successful capacity building taking place in all cases, we anticipate individuals who report engaging the most with the external evaluator will see the most significant gains. Furthermore, we expect organizations implementing their own, locally developed intervention will benefit more than those implementing evidence based interventions (EBIs), and that staff at community based organizations (CBOs), as opposed health centers and hospitals, will similarly report the most significant gains in ECB.

2. Methods

2.1. Evaluation environment

The Center for the Evaluation of HIV Prevention Programs ("Evaluation Center") was funded by the Chicago Department of Public Health (CDPH) to oversee the evaluation efforts of 20 HIV prevention projects at 15 community-based agencies across Chicago. Ten of these projects are Centers for Disease Control and Prevention (CDC)-endorsed EBIs, while the other ten projects are locally-developed ("homegrown") interventions. The Evaluation Center has used an EE approach to provide technical assistance, engage in ECB activities, and ensure the site-specific evaluations result in rigorous, usable findings. This EE approach ensures that community members remain engaged and in control of the development, implementation, and evaluation of each demonstration project.

2.2. Survey development

The Evaluation Center identified six vital capacity-related domains: overarching evaluation plan history, understanding of evaluation components, success in engaging in evaluation activities, organizational evaluation capacity and approach, technical assistance reflections and needs, and satisfaction with Evaluation Center performance. To develop this survey, Evaluation Center staff adapted individual evaluation capacity measures (knowledge and skills) from the Evaluation Capacity Assessment Instrument (ECAI) to reflect the key evaluation components identified by the Evaluation Center team as being most germane to this project (Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry, & Balcazar, 2013). The study team also adapted specific measures from both the ECAI and the Organizational Evaluation Capacity Self-Assessment (Bourgeois & Cousins, 2013) to measure organizational evaluation support and capacity specific to the context of this project.

2.3. Evaluation metrics

The key metrics in assessing evaluation capacity were competence in four key evaluation materials (logic models, evaluation plan, fidelity assessment plan, and data collection tools). Participants were first asked to rate statements on a scale of 1–4 (1 = Strongly Disagree; 4 = Strongly Agree) based on how they felt before working with the Evaluation Center. These retrospective responses were considered baseline measures. They were then asked to rate the same statements on their current level of agreement. These responses, which occurred at 18 months after project launch, are considered follow up. The statements read "I understand what a program logic model is" and "I am confident in my ability to develop a program logic model" for each of the first three evaluation materials. For the final material – data collection tools – individuals were only asked about their confidence developing them. Higher current agreement with each statement, relative to retrospective ratings, served as a proxy for built evaluation capacity in each of these areas. These seven items were combined to make the Evaluation Capacity Building Scale (ECBS-7), which was found to have high internal reliability and consistency for both the retrospective ($\alpha = 0.88$) and current ($\alpha = 0.93$) sequences of questioning.

2.4. Covariates

To adequately assess effective ECB, both individual-level and organizational-level variables need to be explored as potential factors. Individual-level characteristics included the survey taker's role within the organization (field staff vs. project director) as well as their reported engagement with the Evaluation Center on three activities (planning the evaluation, implementing the evaluation, receiving technical assistance). For each activity, survey takers rated their engagement on a three-point scale ('very much,' 'somewhat,' or 'not at all'). Organizational-level characteristics included intervention type (homegrown vs. EBI), and intervention location (community-based organization [CBO], community health center, or hospital).

2.5. Survey administration

The Evaluation Center requested responses from two staff members at each partnering agency – one field-level and one director-level. The survey was self-administered online using Research Electronic Data Capture (REDCap) and asked approximately 30 questions across the six domains indicated above. This survey was launched in June 2016 and was completed by 33 individuals between June 21st and September 8th, 2016.

2.6. Statistical analysis

Basic descriptive statistics (frequencies and means) were generated based on responses to each of the questions within the Evaluation Capacity Survey. Wilcoxon Rank Sum tests were conducted to compare individual's responses on their evaluation capacity at the time of the survey with their capacity before working with the Evaluation Center. Mann-Whitney U tests were conducted to investigate differences by intervention type (EBI vs. homegrown), intervention location (clinical health center vs. CBO vs. hospital), staff role (field vs. director), and level of engagement with the Evaluation Center (not at all vs. somewhat vs. very much). All statistical analyses were performed using IBM SPSS version 24 (IBM Corp, 2015).

3. Results

3.1. Participants

At least one person from each of the 20 demonstration sites responded to the Evaluation Capacity Survey. As some agencies housed multiple projects, four staff completed a single survey for their projects they were involved with, rather than completing the survey twice (Table 1). The Evaluation Center sent the survey to 37 individuals, and 33 (89.2%) completed it; there were no unique characteristics among the 4 non-responders to anticipate they would bias the results.

Table 1
Demographic Characteristics of Individuals Completing Evaluation Capacity Survey.

	N	%
Job Type		
Project Director	20	60.61
Field Staff	13	39.39
Intervention Type		
Evidence Based	16	48.48
Homegrown/Locally Developed	17	51.52
Intervention Location		
Community Based Organization	14	42.42
Community Health Center	8	24.24
Hospital	11	33.33

3.2. Change in evaluation capacity

Participants reported significantly higher agreement with each statement about their current understanding of and confidence in developing evaluation materials, compared to their retrospective self-assessment ($p < 0.05$), except for understanding program evaluation (*n.s.*). Similarly, when collapsed into the ECBS-7, individuals reported significantly higher current capacity than retrospective capacity ($p < 0.01$) (Table 2).

3.3. Differences in evaluation capacity building

Stakeholders from demonstration sites implementing a homegrown intervention showed significantly more improvement than stakeholders from sites implementing an EBI in understanding what a logic model is ($p < 0.05$) and confidence developing data collection tools ($p < 0.01$) when comparing retrospective and current capacity. Neither comparing field staff to project directors or the different intervention locations to each other resulted in significant differences of observed ECB (Table 3).

Those who reported engaging 'very much' with the Evaluation Center on evaluation planning tasks also reported larger improvements in their understanding and confidence developing the evaluation materials than those who reported engaging 'somewhat' (Fig. 1). Improvements were observed in all evaluation capacity metrics as well as

Table 2
Retrospective and Current Evaluation Capacity among HIV Prevention Staff in Chicago.

	Retrospective		Current		
	M	SD	M	SD	*p*
Understanding logic model	3.30	0.68	3.70	0.47	0.005
Confidence developing logic model	3.03	0.85	3.45	0.67	0.007
Understanding of an evaluation plan	3.39	0.75	3.63	0.55	0.142
Confidence developing an evaluation plan	2.85	0.76	3.27	0.67	0.002
Understanding of fidelity assessment plan	3.18	0.85	3.55	0.56	0.001
Confidence developing fidelity assessment plan	2.76	0.90	3.12	0.70	0.011
Confidence developing data collection tools	3.06	0.86	3.52	0.71	0.001
ECBS-7 Total Score	**3.08**	**0.62**	**3.46**	**0.52**	**< 0.001**

Note: M = mean, SD = standard deviation; Wilcoxon Rank Sum test used to determine statistical significance.

Table 3
Differences in Capacity Building by Individual and Organizational Demographics.

	EBI (n = 16)		Homegrown (n = 17)		
	M	SD	M	SD	*p*
Understanding logic model	0.13	0.34	0.65	0.86	0.031
Confidence developing logic model	0.19	0.54	0.65	0.93	0.084
Understanding of an evaluation plan	0.25	0.93	0.24	0.90	0.478
Confidence developing an evaluation plan	0.31	0.60	0.53	0.72	0.329
Understanding of fidelity assessment plan	0.31	0.60	0.41	0.51	0.418
Confidence developing fidelity assessment plan	0.44	0.73	0.29	0.77	0.711
Confidence developing data collection tools	0.13	0.34	0.76	0.75	0.005
ECBS-7 Total Score	**0.25**	**0.47**	**0.50**	**0.61**	**0.297**

	CBO (n = 14)		CHC (n = 8)		Hospital (n = 11)		
	M	SD	M	SD	M	SD	*p*
Understanding logic model	0.21	0.70	0.50	0.54	0.55	0.82	0.381
Confidence developing logic model	0.21	0.70	0.63	0.92	0.54	0.82	0.445
Understanding of an evaluation plan	0.21	1.05	0.50	0.76	0.09	0.83	0.423
Confidence developing an evaluation plan	0.29	0.61	0.63	0.52	0.45	0.82	0.263
Understanding of fidelity assessment plan	0.36	0.63	0.25	0.46	0.45	0.52	0.639
Confidence developing fidelity assessment plan	0.36	0.84	0.25	0.46	0.45	0.82	0.799
Confidence developing data collection tools	0.36	0.50	0.38	0.74	0.64	0.81	0.623
ECBS-7 Total Score	**0.29**	0.55	**0.45**	0.49	**0.45**	0.64	**0.457**

	Field Staff (n = 13)		Director (n = 20)		
	M	SD	M	SD	*p*
Understanding logic model	0.38	0.65	0.40	0.75	0.948
Confidence developing logic model	0.38	0.65	0.45	0.89	0.982
Understanding of an evaluation plan	−0.08	0.95	0.45	0.83	0.082
Confidence developing an evaluation plan	0.31	0.63	0.50	0.69	0.351
Understanding of fidelity assessment plan	0.46	0.52	0.30	0.57	0.272
Confidence developing fidelity assessment plan	0.46	0.66	0.30	0.80	0.473
Confidence developing data collection tools	0.46	0.66	0.45	0.69	0.896
ECBS-7 Total Score	**0.34**	0.55	**0.41**	0.57	**0.834**

Note: M = mean, SD = standard deviation, EBI = effective behavioral intervention, CBO = community-based organization, CHC = clinical health center; Mann-Whitney U test used to determine statistical significance.

the ECBS-7 total score ($p < 0.05$). Conversely, there were no significant differences in capacity building between those who reported engaging with the Evaluation Center 'very much' on the implementation of the evaluation and those who reported engaging 'somewhat,' looking at both individual metrics and the ECBS-7 total score (*n.s.*). Additionally, those reporting high levels of engagement for technical assistance services had marginally insignificantly more improvement in multiple individual metrics and the ECBS-7 total score ($p = 0.06$).

4. Discussion

In this multi-site evaluation of 20 HIV prevention projects in Chicago, stakeholders reported increased understanding of evaluation materials and increased confidence in planning and implementing an evaluation. Analyses showed significant increases in six out of seven evaluation metrics, as well as in the ECBS-7. These findings add to the evidence that engaging stakeholders in the evaluation, as was done through this EE approach, builds evaluation capacity by increasing stakeholder knowledge of and confidence in developing key evaluation materials (Kaufman et al., 2006). Given this increased competency, these organizations may now be better equipped to integrate evaluation activities into their day-to-day operations. Furthermore, this competency has the potential to increase program sustainability as organizations are better able to use findings to make programmatic improvements and potentially leverage this capacity to obtain additional funding.

Stakeholders at sites delivering homegrown interventions benefitted significantly more from than those implementing EBIs (Table 2). As EBIs are pre-packaged by the CDC, individuals who implement these interventions rarely develop their own evaluation materials. Therefore, the decreased engagement with an external evaluator needed to plan and develop tools for evaluation activities at these sites led to fewer opportunities to apply and develop these skills; this decreased engagement limited the extent to which ECB could occur. While an EE model is broadly appropriate and effective for use within community-based intervention delivery, our data reveal effects are most pronounced for new and tailored interventions. We strongly recommend EE as the optimal approach when working with organizations who are delivering novel homegrown interventions.

Given the limited evaluation and organizational capacity housed at CBOs (Carman & Fredericks, 2010), some might expect an EE approach to be more effective at engaging these organizations compared to the more traditional research settings of hospitals and health centers. However, we found evidence that EE is useful at building evaluation capacity among stakeholders across all organizational settings. In fact, individuals from hospitals and CHC tended to report higher levels of built capacity. Accordingly, funding opportunities should always include evaluation components that allow for an external evaluator to

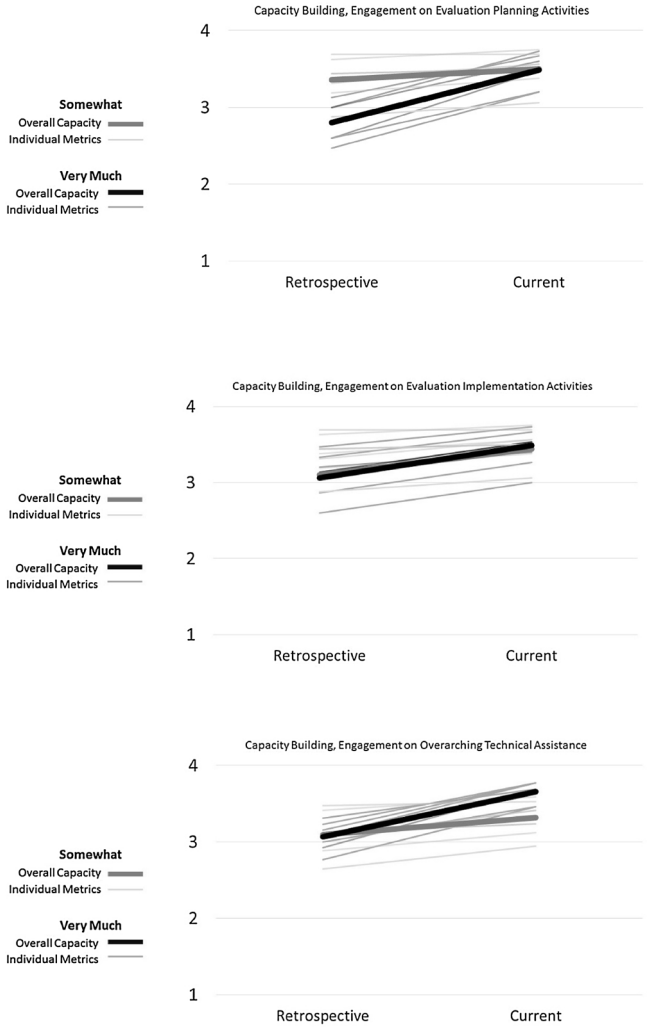

Fig. 1. Differences in Capacity Building by Level of Engagement in Evaluation and Program Planning Activities.

engage with the intervention staff, regardless of the types of agencies applying for the funding. Evaluators seeking to build their clients' evaluation capacity need to actively pursue partnerships with all types of organizations, rather than solely engaging with organizations that they perceive to have low levels of capacity.

External evaluators should attempt to engage individuals from all levels of the organization equally in the development and implementation of an evaluation plan. In this project, field staff reported significantly more engagement with the Evaluation Center than project directors, yet evaluation capacity building was similar for these two groups. Considering that project directors are in a more prominent position to foster a sustained evaluation culture at their organization, evaluators must strive to engage with them during the planning of evaluation activities. Furthermore, it is especially necessary to engage multiple individuals at different levels in this context, given the high levels of staff turnover, particularly among field staff at HIV prevention agencies (Chillag et al., 2002). Turnover was a major problem at delegate agencies in this project, so the Evaluation Center had to thoroughly engage with all members of the project team, especially director-level staff, to ensure institutional knowledge learned through this experience persisted.

The most reliable predictor of increased overall evaluation capacity was engagement with the Evaluation Center. However, this was not consistent across all types of engagement. For instance, higher levels of engagement in evaluation planning activities were significantly associated with increases in individuals ECBS-7 total scores (Fig. 1). More specifically, we observed that organizations with low baseline capacity engaged with the Evaluation Center more in evaluation planning, and saw a correspondingly greater increase in capacity. In contrast, organizations with moderate to high baseline capacity engaged less with the Evaluation Center in evaluation planning which resulted in less capacity increase. Meanwhile, higher levels of engagement in technical assistance activities were associated with increases in ECBS-7 total score regardless of capacity levels at baseline. Notably, this indicates that while engaging with an external evaluator in planning activities can close the gap between organizations with different levels of baseline capacity, engaging with the evaluators in technical assistance services may be a uniquely important component of successful capacity building. These findings have important implications for program funders, staff, and evaluators in planning future efforts.

Based on these findings, funders should expand the use of external evaluators using EE, or another evaluation approach that focuses on engaging stakeholders, to increase evaluation capacity at their delegate agencies. Funders should set aside additional time up front to allow for more collaboration during the planning phase of the evaluation, particularly for those sites with limited baseline evaluation experience, and encourage community agencies to use the external evaluator as a resource for technical assistance services, as both of these activities increase the likelihood of improved evaluation capacity. This is consistent with theories of process use highlighted in a variety of evaluation texts (King, 2007; Patton, 2002). Thus, the evaluator must leverage their expertise, partnerships, and other existing resources to engage stakeholders in the evaluation process.

4.1. Limitations

There were some key limitations to this study. First, all measures were self-reported. Future studies looking to gain empirical evidence about ECB during evaluation activities could attempt to find a better proxy for measuring built evaluation capacity, such as an "Evaluation 101" assessment or monitoring the quality of the evaluation materials developed through an EE approach compared to materials developed without this guidance for prior evaluation. To our knowledge, these approaches have not been used or disseminated within the current context. These self-reported measures included primarily Likert scales that may have resulted in ceiling effects that could make measuring differences in learning between groups much less efficient. While confidentiality of survey responses were promised, social desirability and recall bias could also affect the results of this study. Similarly, as the baseline measures for these analyses were retrospective, it is possible that organizations reported more positive impact, given our close collaboration during the project. However, as we were an external evaluator with no role in funding decisions, sites had little incentive to misrepresent their experience working with us. Given that this was a real-world evaluation scenario, the Evaluation Center only engaged with the 20 sites funded by CDPH, which resulted in a relatively small sample size. This shortcoming prevented more robust analyses of these results, particularly when comparing the evaluation capacity of different categories of survey takers. Larger projects in the future should explore evaluation capacity building across a larger and more diverse group of program stakeholders. Moreover, the sample was limited to individuals working in HIV prevention organizations in the City of Chicago, meaning these findings are not generalizable outside of this context. Future studies should measure sustained ECB, by surveying individuals a certain amount of time after engagement with the external evaluator to more accurately assess the lasting impact of engaging stakeholders in capacity building activities. Finally, by using a general descriptor, rather than a specific measure for level of engagement, there may have been differential responses based on an individual's perception of the three response options. This may be particularly true when comparing field staff and project directors.

5. Lessons learned

While prior studies have described the efficacy of EE to build evaluation capacity at organizations, this study is a unique opportunity to measure these changes across several agencies and provides insight and guidance about the contexts in which EE is most useful. Future work should build on these results by collecting more empirical evidence to identify predictors of successful ECB. Through the identification of these predictors, funding institutions will be able to more efficiently allocate funding for external evaluators to engage with delegate agencies to build evaluation capacity and create a widespread culture of learning at service organizations. Additionally, it will allow for best practices to be shared by evaluators to refine the ECB nature of EE and offer tailored evaluation support and technical assistance to client agencies.

Funding

This work was supported by the Chicago Department of Public Health (RFP # DA-41-3350-11-2014-003).

Declarations of interest

This project was funded by the Chicago Department of Public Health HIV/STI Bureau. Two authors of this article, Mr. Kern and Mr. Stonehouse, are members of the leadership team of this entity.

Acknowledgments

We would like to acknowledge the twenty project teams who implemented the programs being evaluated and participated in the planning and implementation of the evaluation. Similarly, we want to thank the staff at CDPH for helping us work with the community based agencies.

References

Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *The American Journal of Evaluation, 34*(3), 299–319. https://doi.org/10.1177/1098214013477235.

Carman, J. G., & Fredericks, K. A. (2010). evaluation capacity and nonprofit organizations is the glass half-empty or half-full? *The American Journal of Evaluation, 31*(1), 84–104. https://doi.org/10.1177/1098214009352361.

Centers for Disease Control and Prevention (2012). *Improving the use of program evaluation for maximum health impact: Guidelines and recommendations.* Retrieved fromAtlanta, GA: Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services. https://www.cdc.gov/eval/materials/finalcdcevaluationrecommendations_formatted_120412.pdf.

Chillag, K., Bartholow, K., Cordeiro, J., Swanson, S., Patterson, J., Stebbins, S., ... Sy, F. (2002). Factors affecting the delivery of HIV/AIDS prevention programs by community-based organizations. *AIDS Education and Prevention, 14*(3), 27–37. https://doi.org/10.1521/aeap.14.4.27.23886.

Duignan, P. (2003). Mainstreaming evaluation or building evaluation capability? Three key elements. *New Directions for Evaluation, 99*(2003), 7–21. https://doi.org/10.1002/ev.90.

Fetterman, D. M., & Wandersman, A. (2005). *Empowerment evaluation principles in practice.* New York, New York: Guilford Press.

Fetterman, D. M., Kaftarian, S., & Wandersman, A. (2015). *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability.* Thousand Oaks, CA: Sage.

IBM Corp (Cartographer). (Released 2015). IBM SPSS Statistics for Windows, Version 24.0.

Kaufman, J. S., Crusto, C. A., Quan, M., Ross, E., Friedman, S. R., O'Rielly, K., ... Call, S. (2006). Utilizing program evaluation as a strategy to promote community change: Evaluation of a comprehensive, community-based, family violence initiative. *American Journal of Community Psychology, 38*(3-4), 191–200. https://doi.org/10.1007/s10464-006-9086-8.

Kegeles, S. M., & Rebchook, G. M. (2005). Challenges and facilitators to building program evaluation capacity among community-based organizations. *AIDS Education and Prevention, 17*(4), 284–299. https://doi.org/10.1521/aeap.2005.17.4.284.

King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation, 116*, 45–59. https://doi.org/10.1002/ev.242.

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *The American Journal of Evaluation, 33*(3), 307–338. https://doi.org/10.1177/1098214011434608.

Patton, M. Q. (2002). *Qualitative research and evalution mentods.* Thousand Oaks, CA: Sage.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *The American Journal of Evaluation, 29*(4), 443–459. https://doi.org/10.1177/1098214008324182.

Rodriguez-Campos, L. (2012). Advances in collaborative evaluation. *Evaluation and Program Planning, 35*(4), 523–528. https://doi.org/10.1016/j.evalprogplan.2011.12.006.

Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Iriarte, E., Henry, D. B., & Balcazar, F. E. (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *The American Journal of Evaluation, 34*(2), 190–206. https://doi.org/10.1177/1098214012471421.

Veniegas, R. C., Kao, U. H., Rosales, R., & Arellanes, M. (2009). HIV prevention technology transfer: Challenges and strategies in the real world. *American Journal of Public Health, 99*(Suppl 1), S124–130. https://doi.org/10.2105/AJPH.2007.124263.

Peter T. Lindeman, MA, is a Program Coordinator in the Evaluation, Data Integration, and Technical Assistance (EDIT) Program, currently working on the Evaluation Center project discussed in this article. He received a formal training as an evaluator in Michigan State University's MA in Program Evaluation in 2016. He brings experience conducting multisite evaluation projects of HIV prevention and care interventions and has successfully submitted abstracts to and presented at local, national, and international conferences. His research interest focus on how public health intervention can address health and wellbeing disparities of marginalized populations.

Emily Bettin, BA, is a Research Study Coordinator spanning several projects within the Institute for Sexual and Gender Minority Health and Wellbeing (ISGMH) including the Evaluation Center. She has a bachelor's degree in Political Science from Kenyon College, and a graduate certificate in Pre-Clinical Psychology from Northwestern University. She brings experience working on a variety of health-focused research projects. Her clinical and research interests broadly include issues related to trauma and resilience, particularly among members of underserved and marginalized communities.

Lauren B. Beach, PhD/JD, is a postdoctoral research fellow at ISGMH in the research group of Dr. Gregory Phillips, II. Scientifically, she investigates how stigma affects chronic disease management and health outcomes in LGBTI populations. As an alumna of the Joint Degree Program in Law, Science & Technology, she received her JD specialized in Health Law & Bioethics from the University of Minnesota Law School in 2012 and her PhD in Molecular, Cellular, Developmental Biology & Genetics from the University of Minnesota, Twin Cities in 2014. She has worked on several dissemination activities from the Evaluation Center project.

Christian N. Adames, BA, is a Research Assistant with the EDIT Program who works primarily with the Evaluation Center. He earned his bachelor's degree in Psychology with a minor in Biological Sciences from the University of Chicago in 2015. He brings experience coordinating a variety of health research projects. His research interested include health care disparities among minority populations, with a particular focus on social factors contributing these disparities.

Amy K. Johnson, PhD, is a research assistant professor in the Division of Adolescent Medicine at Lurie Children's and Department of Pediatrics at Northwestern University's Feinberg School of Medicine. She earned her Ph.D. in Epidemiology at the University of Illinois at Chicago. Dr. Johnson has advanced training in program evaluation, quantitative and qualitative analysis and is interested in applied HIV and STI prevention and care intervention research. She is a co-investigator on this project.

Dave Kern, BA, is Deputy Commissioner of the HIV/STI Bureau at the Chicago Department Health. He has also served as the Manager of the Infectious Disease Prevention Section at the Washington State Department of Public Health, and as the Director of the Prevention Program and Strategic Initatitves at the National Alliance of State and Territorial AIDS Directors.

Patrick Stonehouse, AM, is Director of HIV/STI Prevention at the Chicago Department of Public Health. Formally trained in Social Work at University of Chicago, he brings experience managing HIV prevention programs at a community organizations such as Chicago House and Social Service Agency, and Heartland Alliance for Human Needs & Human Rights.

George J. Greene, PhD, is the associate director of the EDIT Program and a research assistant professor in the Department of Medical Social Sciences at Northwestern University. He earned his Ph.D. in Community and Prevention Research at the University of Illinois at Chicago. Dr. Greene's research focuses on health disparities in lesbian, gay, bisexual, and transgender (LGBT) adolescents and young adults, with particular emphasis on ethnic/racial disparities in the HIV epidemic. He employs community-based research approaches to develop effective interventions and applies mixed methods approaches to conduct rigorous program evaluation. Dr. Greene is a co-investigator of the Evaluation Center.

Gregory Phillips II, PhD, is the Director of the EDIT Program and an assistant professor in the Department of Medical Social Sciences at Northwestern University. His research focuses on understanding social-, sexual-, and network-level factors driving the HIV epidemic among men who have sex with men (MSM), particularly among young MSM and Black MSM. Trained in epidemiology at The George Washington University, he has spent his career working on HRSA-, CDC-, and NIH-funded projects to identify behaviors associated with HIV infection and effective interventions tailored to serve adversely affected population. He is principal investigator of the Evaluation Center project.